Clonal Interference and the Evolution of RNA Viruses

Rosario Miralles,[1] Philip J. Gerrish,[2]* Andrés Moya,[1] Santiago F. Elena[1]†

In asexual populations, beneficial mutations that occur in different lineages compete with one another. This phenomenon, known as clonal interference, ensures that those beneficial mutations that do achieve fixation are of large effect. Clonal interference also increases the time between fixations, thereby slowing the adaptation of asexual populations. The effects of clonal interference were measured in the asexual RNA virus vesicular stomatitis virus; rates and average effects of beneficial mutations were quantified.

Populations adapt through the appearance and subsequent fixation of beneficial mutations. In a large population, beneficial mutations may arise frequently enough that two or more are simultaneously present in independent lineages. Once beneficial mutations have arisen, there is a certain probability of losing them by drift while their frequency is low. However, after this period dominated by drift, they reach a substantial frequency in the population. For a sexual system, these beneficial mutations will eventually recombine, ensuring their fixation together. If the system is asexual, the lineages created by these beneficial mutations will compete; only the mutation with largest effect will be fixed. Thus, asexual populations must fix beneficial mutations sequentially (*1*, *2*). The possibility of simultaneous fixation of beneficial mutations in sexual populations is often contrasted with the sequential fixation in asexual populations as an argument for the evolutionary advantage of sex (*2*).

The idea that beneficial mutations must compete in asexual populations was originally proposed by Muller (*3*), and it has been developed theoretically (*4*), as well as experimentally demonstrated to be important in determining the rate of adaptation of the bacterium *Escherichia coli* (*5*). Gerrish and Lenski (*4*) modeled the fate of beneficial mutations by considering clonal interference among them as a major factor. The main conclusions of their model were as follows: (i) The probability of fixation of a given beneficial mutation decreases with both population size and mutation rate. (ii) As population size or mutation rate increase, adaptive substitutions result in larger fitness increases. (iii) The rate of adaptation is an increasing, but decelerating, function of both population size and mutation rate. (iv) Beneficial mutations that become transiently common but do not achieve fixation because of interfering beneficial mutations are relatively abundant. (v) Transient polymorphisms may give rise to a "leapfrog" effect, where the most common genotype at a given moment might be less closely related to the immediately preceding one than with an earlier genotype.

RNA viruses show the highest mutation rates in nature (*6*). This, together with their potentially large effective population sizes and the fact that their reproduction is not obligately sexual, suggests that clonal interference may play an important role in their adaptive evolution. Our goal here is to infer the presence of clonal interference acting on viral populations. Following (*4*), for increasing population sizes, we predicted that (i) the fitness effect associated with fixed beneficial mutation will tend to be larger and (ii) the rate of adaptation will tend toward a limit.

To detect experimentally the fixation of a beneficial mutation in a viral population, we mixed, at equal proportions, two variants of vesicular stomatitis virus (VSV) that differ only in their ability to grow in the presence of a monoclonal antibody (*7*). The two variants were selectively equivalent in the absence of monoclonal antibody (*7*), implying that they should stably coexist until a successful beneficial mutation appears in one of them. Seven different evolutionary regimes were designed, each one differing from the others in effective population size (N_e). As shown in Table 1, N_e ranged in these seven regimes between ~100 and ~10^8 viral particles (*8*). Each regime was independently replicated five times, for a total of 35 experimental lines. Each mixture was kept under the appropriate batch transfer conditions (*9*) until one of the two variants became fixed. Then, the winner variant that carried a beneficial mutation was isolated. This variant was then placed in head-to-head competition with its nonevolved counterpart (*10*) to estimate the fitness effect (W) associated with the beneficial mutation that drove it to fixation.

For the smallest population sizes, one can expect genetic drift to play a role in fixing neutral, or even deleterious, mutations. However, previous results have shown that for MARM C clone, the smallest N_e used here did not have a considerable deleterious effect (*11*). Thus, we can safely assume that the fixation of deleterious mutations during our experiment will be minimized by purifying selection.

The estimates obtained for W, under the seven different population-size regimes (*12*), are shown in Fig. 1. The first prediction we made on the basis of the clonal interference model is completely fulfilled: A significant correlation exists between log N_e and the magnitude of the fitness effect ($\rho_S = 0.8929$, $n = 7$, one-tailed $P = 0.0034$). The larger the population size is (that is, the stronger the clonal interference), the larger the magnitude of the beneficial effect needed to fix a mutation is. As population size increases, there is a shorter waiting time between two consecutive events of beneficial mutation, and thereby more beneficial mutations coexist at a given time.

Each one of the W values used to generate Fig. 1 was transformed into rates of evolution by subtracting from them the fitness of the initial MARM C clone (*7*) and dividing by the approximate time it took each mutation to become fixed in the population (Table 1). Following (*5*), we then regressed these rates against N_e using (i) a linear model, which implies that the rate of adaptation is proportional to the effective population size, and (ii) a hyperbolic model, which implies that clonal interference will impose a deceleration on the rate of adaptation. These data, as well as the fitting of both models, are shown in Fig. 2.

Table 1. Parameters describing the fixation of beneficial mutations under each N_e. The number of lines that increased fitness is reported in the second column. The third column shows the number of generations elapsed until fixation of a beneficial mutation (*8*). A significant correlation between log N_e and the time to fixation has been observed ($\rho_S = 0.75$, $n = 5$, one-tail $P = 0.0261$). The last column shows the probability of fixation by random genetic drift (*15*).

Log N_e	Lines that increased fitness	Time to fixation	P
2.0238	1	76.045 ± 17.168	0.0440
3.4247	4	234.003 ± 48.354	0.0113
4.3644	3	185.099 ± 32.356	0.0001
5.2940	5	288.693 ± 18.558	<0.0001
6.2103	5	328.922 ± 31.558	<0.0001
7.1065	5	347.502 ± 13.839	<0.0001
7.9695	5	283.386 ± 9.943	<0.0001

[1]Institut Cavanilles de Biodiversitat i Biología Evolutiva and Departament de Genètica, Universitat de València, Apartado 22085, 46071 València, Spain. [2]Center for Disease Control and Prevention, Atlanta, GA 30333, USA.

*Present address: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

†To whom correspondence should be addressed. E-mail: santiago.elena@uv.es

The linear model gave a significant fit to the data [$R^2 = 0.6092$, $F(1,6) = 24.7942$, $P = 0.0025$], indicating that the rate of adaptation increases with N_e. However, the hyperbolic model, despite losing a degree of freedom, provides a much better fit to the data than the linear model [partial F test: $F(1,5) = 10.1493$, $P = 0.0244$], showing a limit to the rate of adaptation of viral populations imposed by clonal interference. This finding confirms the second prediction we made on the basis of the clonal interference model (*4*).

The model of clonal interference (*4*) has the additional advantage of allowing us to estimate the beneficial mutation rate (μ_b) as well as the mean selective advantage, $E(s)$, of all beneficial mutations produced in the population (not just those that are fixed). The probability density for the selective coefficients of successful beneficial mutations is given as a part of eq. 6 in (*4*). From this density, we computed the maximum likelihood estimates and associated variance for μ_b and $E(s) = 1/\alpha$ (where α is the parameter of an exponential distribution) from the data shown in Fig. 1 (*13*). The solid line drawn in Fig. 1 represents the maximum likelihood fit of the model to the data. The estimated value for the beneficial mutation rate was $\mu_b = 6.387 \times 10^{-8}$ beneficial mutations per genome and generation, with 95% confidence interval in the range $2.57 \times 10^{-8} \leq \mu_b \leq 1.58 \times 10^{-7}$. This value of μ_b almost warrants that in all seven N_e explored, by chance a beneficial mutation will arise, because all lines reached population size of $\sim 7 \times 10^9$ viruses at the end of the daily growth ($N_f \mu_b > 1$). Also, this value supports our assumption that a single beneficial mutation fixed in each lineage was responsible for the fitness increase. [As an illustration, the probability of generating two beneficial mutations will be $N_f \mu^2_b \ll 1$.] Even at the smallest N_e, a significant beneficial effect was detected in one of the five replicates (Table 1; $W = 2.3692 \pm 0.3520$, $t_2 = 3.8903$, one-tail $P = 0.0301$). In contrast, only at the smallest N_e did a line show a significant decline in fitness ($W = 0.5286 \pm 0.0312$, $t_2 = -15.1057$, one-tail $P = 0.0022$). Drake *et al.* (*6*) estimated the total genomic mutation rate for VSV to be about 3.5 substitutions per genome and generation. Comparing this figure with our estimate of μ_b, we can infer that about one in 2×10^8 mutations produced in VSV can be considered beneficial. This number is two orders of magnitude smaller than that estimated for *E. coli* (*4*). This difference could result from the simpler genome of VSV when compared with *E. coli*: The more complex a genome is, the more room it has for improvement. Another possible explanation could be the difference in the degree of adaptation of each organisms to its experimental environment: The VSV clones used here have a history in the cell system, which can condition the number of possible beneficial mutations available (*5*).

The maximum likelihood estimate for $E(s)$ was 0.3062 per day, with 95% confidence interval in the range $0.2371 \leq E(s) \leq 0.4309$. In other words, the average fitness effect associated with the beneficial mutations produced (not necessarily fixed) is around 31% per day.

The clonal interference model used to estimate μ_b and $E(s)$ makes two important assumptions. The first is that selection coefficients of beneficial mutations are exponentially distributed. The general shape of this distribution was proposed by Fisher (*1*), and a statistical argument supporting the use of the exponential was given in (*14*). The second assumption of the model is to ignore the effect of deleterious mutations. In a small population, deleterious mutations may accumulate through fixation by drift and Muller's ratchet, thus reducing the population's fitness. As stated above, evidence of a reduction in fitness was seen in a line at the smallest N_e. To eliminate any possible effect in the computation of μ_b and $E(s)$, we corrected our data as follows. If the lowest winning fitness of the five replicates was less than one, we assumed that this population had fixed a beneficial mutation of arbitrarily small effect such that the fitness of that population was determined solely by the accumulation of deleterious mutations. The fitness loss through the accumulation of deleterious mutations was assumed to be similar in all five replicates. [This assumption is reasonable because the deleterious mutation rate is quite high (*6*) such that fitness loss is relatively deterministic.] Thus, the higher fitnesses of the other five replicates (although these may also be less than one) resulted from the fixation of beneficial mutations of significant effect. Although one may question the assumption that the smallest beneficial mutation fixed was of insignificant effect, we assert that the net effect of the assumption is negligible because (i) this correction (and hence this assumption) was only necessary in the smallest N_e and (ii) in small N_e, clonal interference does not insure the fixation of large-effect mutations; thus, mutations of very small effect may be fixed in these populations just by pure genetic drift. To check this last possibility, we computed the probability of random fixation of an allele in populations with N_e equal to those used in our experiment (*15*). These computations showed that even for our smallest N_e, the probability of chance fixation for a neutral allele was <5%. Thus, we can disregard the possibility that most of the mutations fixed during our experiment resulted from genetic drift.

Our results provide evidence that clonal interference occurs in viral populations. This evidence, along with models of clonal interference (*1–4*), allows certain properties of the adaptive evolution of RNA viruses to be inferred:

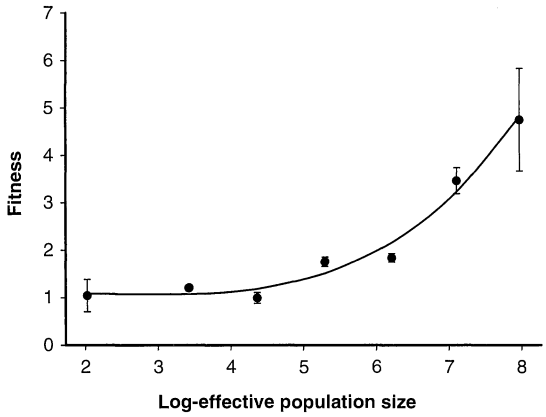

Fig. 1 (left). Influence of effective population size on the magnitude of the fixed fitness effect. Error bars represent standard errors ($n = 5$). The solid line represents the fit to the theoretical model described in brief in (*13*) and in more depth in (*4*).

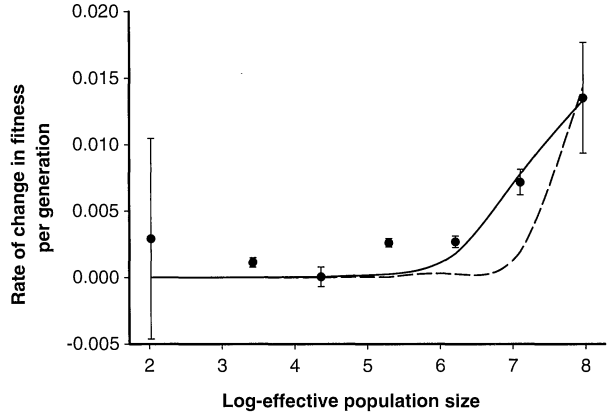

Fig. 2 (right). Influence of effective population size on the rate of evolution. Error bars represent standard errors ($n = 5$). Because the rate of adaptation in an asexual population of size zero must be null, the y intercept has been fixed at zero. The dashed line represents the fit of the experimental data to a linear model. The solid line represents the fit to a hyperbolic model [$R^2 = 0.8710$, $F(2,5) = 36.3758$, $P = 0.0010$]. Both curves appear to be exponential because of the common logarithmic scale in the x axis.

1) Adaptive substitutions appear as discrete, rare events, regardless of mutation rate or population size. They often do not occur simply as the result of a single mutation but instead represent the best of several competing mutations. This fact has consequences for the dynamics of drug resistance and the search for resistance mutations.

2) In medium to large populations, the rate of fitness increase is hardly affected by changes in either mutation rate or population size. Some have speculated that the high mutation rates of RNA viruses are maintained evolutionarily because of the great adaptive capacity they bestow (*16*). When clonal interference is present, however, this argument becomes questionable. When mutation rates are already high, changes in those rates have little effect on the adaptive capacity. Thus, a decrease in mutation rate could come about with little or no decrease in adaptive capacity. Furthermore, a decrease in mutation rate would benefit the population by slowing the accumulation of deleterious mutations. In this light, how is one to explain the high mutation rates of RNA viruses? When clonal interference is considered, it becomes more reasonable to speculate that lower mutation rates would be mechanistically costly or impossible (in terms of complex enzymatic systems required for error detection and correction) for RNA viruses than to suppose that their high mutation rates are maintained for the increased adaptive capacity they confer. If an increased production of beneficial mutations is of little or no advantage to a viral population, then mutator alleles (alleles that confer an elevated mutation rate) must be strictly deleterious because of their increased production of deleterious mutations.

3) Resident populations are protected from invaders simply because of their numerical advantage. A high-fitness VSV clone seeded at low frequency into a resident population of low-fitness variants was displaced by the low-fitness competitors (*17*). When its initial frequency was above a certain threshold, however, the high-fitness clone always outcompeted the low-fitness variants in the resident population (*17*). This observation can be easily explained in light of the clonal interference model: If the high-fitness clone is initially present at very low frequencies, it is probable that beneficial mutations arise in the most frequent genotypes, improving their fitness and interfering with the intruder, with the final result of eliminating it from the population. In contrast, when the initial frequency of the high-fitness clone is high, it increases in frequency in the population before the low-fitness variants have a chance to find the right beneficial mutation, resulting in fixation of the high-fitness invader. Such population dynamics in nature might prevent the emergence of dangerous viral pathogens: The existence of a frequency threshold for dominance imposes an element of uncertainty in virus sampling during outbreaks.

References and Notes

1. R. A. Fisher, *The Genetical Theory of Natural Selection* (Oxford Univ. Press, Oxford, UK, 1930).
2. J. F. Crow and M. Kimura, *Am. Nat.* **99**, 439 (1965); H. A. Orr, *Evolution* **52**, 935 (1998); S. P. Otto and N. H. Barton, *Genetics* **147**, 879 (1997).
3. H. J. Muller, *Am. Nat.* **8**, 118 (1932); *Mutat. Res.* **1**, 2 (1964).
4. P. J. Gerrish and R. E. Lenski, *Genetica* **102/103**, 127 (1998).
5. J. A. G. M. de Visser, C. W. Zeyl, P. J. Gerrish, J. L. Blanchard, R. E. Lenski, *Science* **283**, 404 (1999).
6. W. J. Drake, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4171 (1993); J. W. Drake, B. Charlesworth, D. Charlesworth, J. F. Crow, *Genetics* **148**, 1667 (1998).
7. One of the VSV variants used was MARM *C*, which had an Asp → Ala substitution at position 259 of the G surface protein. This substitution allows the mutant to replicate under I_1 monoclonal antibody (mAb) concentrations that completely neutralize the wild-type clone [S. B. VandePol, L. Lefrancois, J. J. Holland, *Virology* **148**, 312 (1986)]. The other viral clone used was an mAb I_1–sensitive, wild-type derivative. Both clones have been replicating under identical conditions for many years. Large volumes of each clone (~10^{10} plaque-forming units per milliliter) were produced and kept at −80°C. The relative fitness of these two clones has been determined many times, as well as up to 15 times during this experiment (mean ± SE = 1.000 ± 0.036, $t_{14} = 0.0005$, $P = 0.9996$), always yielding a relative value of one. Thus, the mutation conferring the MARM phenotype must be considered selectively neutral.
8. Each day, flasks containing an approximately constant number of host cells (*9*) were infected with different dilutions of the previous day's evolving viral mixture. The dilutions ranged from 10^{-2} to 10^{-7}, plus an extra case in which, instead of a dilution, virus was plated on solid media (*9*) and five single plaques were randomly picked, mixed, and used to infect (without further dilution) the flask. After 1 day of growth, the final viral titer in all flasks reached a value of about 7×10^9 viruses. This final value depends only on the number of cells in the flask, which is constant in all cases (*9*). Hence, we can estimate the number of generations that occurred under each regime as $\theta = \ln(N_f/N_o)$, where N_f and N_o are population sizes before and after a transfer, respectively. Let N_e denote effective population size; under a serial transfer regime, $N_e \approx \theta N_o$ [R. E. Lenski *et al.*, *Am. Nat.* **138**, 1315 (1991)].
9. Baby hamster kidney cells were grown as monolayers under Dulbecco's modified Eagle's minimum essential medium containing 5% newborn calf serum and 0.06% proteose peptone 3. Cells were grown in 25-cm^2 plastic flasks for infections or in 100-cm^2 plates for routine maintenance at a density of about 3×10^5 cells per square centimeter. Cell lines were maintained in incubators at 37°C with a 7% CO_2 atmosphere. For the infections, daily, a sample from each flask was taken and diluted, up to the convenient factor. To infect the next day's flask, we added 200 μl of diluted virus to the monolayer. In parallel, each even time-point sample was also plated (in triplicate), with and without mAb I_1, to estimate the relative abundance of each viral variant at this day's passage. We continued with this regime of transfer and plating until one of the two competing variants in the mixture became fixed. Our criterion for considering that one allele had fixed was to obtain, during three consecutive days, no evidence for the presence of the opposite genetic marker. In these platings, we always counted a minimum of 300 plaques per plate × 3 replica plates × 3 days ≥ 2700 plaques counted in total, which gives us a best estimate for the rare variant frequency of less than 0.04%.
10. At the end of each experiment, the resulting populations were assayed for relative fitness in triplicate. The competition protocol has been described elsewhere [(*16*, *17*); R. Miralles, A. Moya, S. F. Elena, *J. Gen. Virol.* **80**, 2051 (1999)].
11. I. S. Novella *et al.*, *J. Virol.* **69**, 2869 (1995); *Mol. Gen. Genet.* **252**, 733 (1996).
12. Tab-delimited ASCII files containing the raw data can be downloaded by anonymous ftp from ftp://serbio.uv.es/pub/incomming/elena/clonint1.dat and ftp://serbio.uv.es/pub/incomming/elena/clonint2.dat.
13. Define selective advantage, *s*, as per-generation excess growth in a continuously growing culture such that $(1 + s)^{\theta t} = W^t$, where θ is the per-day number of generations of growth (*8*). Then, $s = W^{1/\theta} - 1$. Assume the distribution of the selective advantage, *s*, of beneficial mutations produced to be exponential with parameter α. Let μ_b denote beneficial mutation rate. Then the selective advantage conferred by an adaptive substitution (that is, a beneficial mutation that is both produced and fixed) has probability density $p(s;\ \alpha,\ \mu_b,\ N_e) = Ks e^{-\lambda - \alpha s}$, where $\lambda = 2\mu_b N_e e^{-\alpha s}(1 + 1/\alpha s)\ln N_e$ is the expected number of interfering mutations and *K* is a constant such that $\int_0^\infty p(s;\ \alpha,\ \mu_b,\ N_e)\, ds = 1$. [See (4) for the derivation of this model and further discussion of assumptions.] This model was fit to the experimental data shown in Fig. 1 by maximizing the likelihood function, $L(\alpha, \mu_b) = \sum_{i=1}^{7} \sum_{j=1}^{5} \ln\{p[s(j); \alpha, \mu_b, N_e(i)]\}$, thereby obtaining estimates of $\hat{\mu}_b$ and $\hat{\alpha}$ (*i* is the number of population sizes used and *j* is the number of replicates per each population). The variance-covariance matrix associated with these estimates was approximated by
$$\begin{pmatrix} \mathrm{Var}(\alpha) & \mathrm{Cov}(\alpha, \mu_b) \\ \mathrm{Cov}(\alpha, \mu_b) & \mathrm{Var}(\mu_b) \end{pmatrix} \approx \begin{pmatrix} \partial L(\hat{\alpha}, \hat{\mu}_b)_{\alpha,\alpha} & \partial L(\hat{\alpha}, \hat{\mu}_b)_{\alpha,\mu_b} \\ \partial L(\hat{\alpha}, \hat{\mu}_b)_{\alpha,\mu_b} & \partial L(\hat{\alpha}, \hat{\mu}_b)_{\mu_b,\mu_b} \end{pmatrix}^{-1}$$
(*18*), and confidence intervals were determined from the assumption of normality [supported by asymptotic analysis (*18*)].
14. J. H. Gillespie, *The Causes of Molecular Evolution* (Oxford Univ. Press, Oxford, UK, 1991).
15. Let $u(p,t)$ denote the probability that one of two genotypes (either MARM *C* or wild-type) becomes fixed by the *t*th generation, given that its frequency is *p* at $t = 0$. For our case, in which the two genotypes are selectively neutral, it can be shown that $u(p,t)$ satisfies the diffusion equation $\partial u(p,t)/\partial t = p(1 - p)/2N_e \partial^2 u(p,t)/\partial p^2$ [M. Kimura, *Genetics* **47**, 713 (1962)]. Numerical solution of this diffusion equation yields $u(p,t)$, given $p = 0.5$ (the initial ratio was 1:1) and observed time to fixation, *t* (Table 1). The values of $2u(p,t)$ obtained are reported in the last column of Table 1 as *P* values for rejection of the null hypothesis that the observed fixation (of either genotype, hence the 2) was due to drift alone. Obviously, the larger the effective population size is, the smaller the probability of fixation by chance will be.
16. I. S. Novella *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5841 (1995); S. F. Elena *et al.*, *Evolution* **52**, 309 (1998); C. Escarmís, M. Dávila, E. Domingo, *J. Mol. Biol.* **285**, 495 (1999); I. S. Novella, J. Quer, E. Domingo, J. J. Holland, *J. Virol.* **73**, 1668 (1999); C. L. Burch and L. Chao, *Genetics* **151**, 921 (1999).
17. J. C. de la Torre and J. J. Holland, *J. Virol.* **64**, 6278 (1990); J. J. Holland, J. C. de la Torre, D. K. Clarke, E. A. Duarte, *ibid.* **65**, 2960 (1991).
18. R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

19. We thank J. A. G. M. de Visser, F. González-Candelas, R. E. Lenski, and P. E. Turner for constructive discussions and critical reading of the manuscript. Supported by a grant from the Spanish Dirección General de Enseñanza Superior. R.M. acknowledges a fellowship from the Spanish Ministerio de Educación y Cultura.

4 May 1999; accepted 28 July 1999