# Effects of Spatial Competition on the Diversity of a Quasispecies

Jacobo Aguirre and Susanna C. Manrubia

*Centro de Astrobiología, CSIC-INTA. Ctra. de Ajalvir km. 4 28850 Torrejón de Ardoz, Madrid, Spain*

(Received 23 July 2007; published 25 January 2008)

The diversity harbored by populations of RNA viruses results from high mutation rates, as well as from the characteristics of the environment where they evolve. By means of a simple model for structured quasispecies, we quantify how competition for space among phenotypic types shapes their distribution at the mutation-selection equilibrium. We introduce a general framework to treat this problem and relate mutation rate and competition strength to the quasispecies composition. For diffusion limited competition, diversity typically increases and the asymptotic growth rate of the population diminishes as diffusion decreases. Limited mobility confers a relative advantage to worse competitors. The stationary state is characterized by an over-production of viral particles. Empirical data allow an estimation of mutation rates compatible with the diversity observed in viral populations infecting cellular monolayers.

DOI: 10.1103/PhysRevLett.100.038106 PACS numbers: 87.23.−n, 87.10.−e

RNA viruses count amongst the most plastic organisms on Earth. Elevated mutation rates maintain high levels of diversity in their populations, a feature that underlies their fast response to environmental changes [1]. Simple models that focus on the diversity of phenotypes in a quasispecies [2] have successfully explained a number of collective behaviors displayed by RNA viruses. These models tacitly assume that a phenotype can be represented by a large number of different genotypes, and that the relevant quantity behind adaptive properties is the diversity maintained in the quasispecies. In this framework, the ways in which viral quasispecies can escape Muller's ratchet have been quantified [3,4], as well as the kinetics of the process of fitness gain [5], or the effect that infection transmission modes [6] and interaction between phenotypic classes [7] have on virulence. Further, empirical observations indicate that quasispecies diversity is related to pathogenesis [8].

The phenotypic composition of a quasispecies results from properties intrinsic to the population, such as the mutation rate, but also from features of the environment where evolution and adaptation occur. Among the latter, population bottlenecks are known to decrease diversity and average fitness [9], a situation that can be reverted if massive population passages are applied [10]. Another example is the extremely different equilibrium distributions of phenotypes yielded by horizontal or vertical transmission [6]. Finally, the structural properties of infected tissues also condition the composition of quasispecies, as revealed for instance when comparing plant viruses (with restricted mobility and spreading mostly in two dimensions) with animal viruses [11].

The asymptotic properties of quasispecies structured into two classes (master and mutant) evolving in physical spaces of finite dimensionality have been theoretically addressed in the literature. Limited diffusion confers an advantage to the mutant class, and thus pushes the critical error threshold to lower values of the mutation rate as diffusion decreases [12,13]. The use of field theory has demonstrated as well that lower spatial dimensions enhance coexistence [14]. In this Letter, we use a quasispecies structured into $F$ different phenotypic classes to quantify how diversity becomes modified by limited diffusion. Classes above average are impaired as diffusion decreases, while those below average obtain a relative advantage that we precisely quantify.

Our model is inspired by an often applied protocol where the characteristics of infective agents are studied by means of infection spreading on cellular monolayers [15]. The process is initiated by a single particle that replicates inside a cell. Eventually, the cell is killed and new viral particles (possibly mutated) are released to the medium. A fraction of that progeny infects adjacent, fresh cells. The process repeats and the size of the plaque formed by dead cells grows. After a transient period, all activity occurs at the perimeter of the plaque. In the limit $t \to \infty$, the number of cells killed per generation attains a constant value $N \gg 1$, and the diversity distribution of viral particles produced reaches a stationary shape that depends on the mutation rate and on the transport properties of the medium.

The two main processes (mutation and competition) are implemented in our model as a phenotypic mutation rate and as a set of coefficients representing the effect of competition between fitness types, respectively. We consider $f = 1, \ldots, F$ fitness classes for viral particles, each represented by $n_f(g)$ individuals at generation $g$. The fitness $f$ corresponds to the number of offspring per generation produced by an individual in that class. Successful replication depends not only on $f$, but also on the properties of neighboring individuals: we define the phenomenological competition coefficient $\alpha_f$ of class $f$ as the fraction of actually infecting offspring of that class. By definition, these quantities fulfill $0 \leq \alpha_f \leq 1$, with a precise functional form that will be specified below. Evolution proceeds in discrete time steps, or generations, and each new population substitutes the old one [16]. Finally, offspring

 

from an individual of class $f$ mutates to class $f-1$ with probability $p$ and to class $f+1$ with probability $q$. In the limit $N \gg 1$, the model can be described through average values that evolve according to the deterministic dynamical equations

$$n_f(g+1) = \alpha_f(1-p-q)fn_f(g) + \alpha_{f+1}p(f+1)n_{f+1}(g) + \alpha_{f-1}q(f-1)n_{f-1}(g), \quad (1)$$

with reflecting boundaries at $f=1$ and $f=F$, such that $n_1(g+1) = \alpha_1(1-q)n_1(g) + 2\alpha_2 p n_2(g)$ and $n_F(g+1) = \alpha_F(1-p)Fn_F(g) + q\alpha_{F-1}(F-1)n_{F-1}(g)$.

The limit $g \to \infty$ is characterized by an asymptotic distribution of fitness classes $n_f(g\to\infty)/N \equiv u_f$ independent of the initial condition. For constant population size, the coefficients $\alpha_f$ fulfill

$$\sum_i i\alpha_i u_i = 1. \quad (2)$$

Let us start by presenting a self-consistent solution to the problem. A formal solution to the dynamical Eqs. (1) can be written in terms of the mean matrix $M_{f,f'}$, defined as the expected number of individuals with fitness $f$ in generation $g+1$ arising from an individual of fitness $f'$ at generation $g$ [4,17,18]. Its largest eigenvalue for $N$ constant is $\lambda = 1$, and the corresponding right eigenvector $u_f$ yields the stationary distribution of fitness classes,

$$Mu_f = u_f. \quad (3)$$

The competition coefficients $\alpha_f$ depend on the phenotypic mutation rates $p$ and $q$, on the transport properties of the medium, and on $\bar{f}$. Given a functional form for $\alpha_f$ consistent with Eq. (2), the stationary solution of the problem can be explicitly obtained for $F \le 4$ from Eq. (3), where $\bar{f}$ results from the condition that the maximum eigenvalue of the mean matrix equals unity. For $F > 4$, the additional condition $\sum_i iu_i = \bar{f}$ can be used in order to simultaneously solve that equality and Eq. (3): A numerical solution yielding $\bar{f}$ and $u_f$ can be iteratively obtained as the fixed-point of the dynamical system defined by the two equations.

In order to go further, we analyze a specific implementation of spatial competition that yields a functional form of the coefficients $\alpha_f$. Consider a population occupying an array of $L$ columns $\times\, d$ rows, $L \gg d$, $L \gg 1$. At each generation, the progeny of this population will move to a similar array with the condition that all individuals in domains $D$ of size $S_D = d \times (2d+1)$ compete for empty sites (see Fig. 1). For $d=1$, for example, the individual at site $r_0$ competes with its nearest neighbors at $r_0 - 1$ and $r_0 + 1$ for the site immediately in front of it. Increasing values of $d$ mimic the mixing effect of larger diffusion rates. The probability $p(f)$ that an individual with fitness $f$ infects a particular cell out of the $S_D$ available to the next generation is $p(f) = f(\sum_D f_{nn})^{-1}$, where $f_{nn}$ corresponds

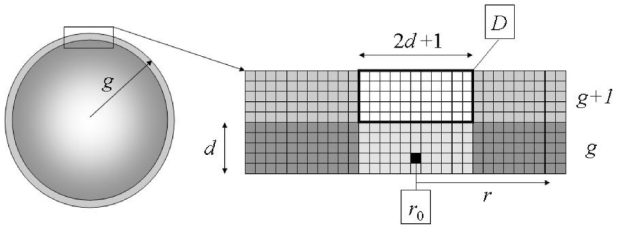


FIG. 1. Schematic representation of the geometrical model parameters. In the limit $g \to \infty$ the growth of a lytic plaque is approximated by an array of depth $d$ rows and length $L$ columns. The offspring produced at generation $g$ by a site at position $r_0$ infects at generation $(g+1)$ any of the cells in a domain $D$ of size $S_D = d \times (2d+1)$, centered at $r_0$, with a probability proportional to its fitness. It competes with all the rest of sites in a domain of the same size at generation $g$.

to the fitness of each site in the area defined by the domain $D$ at generation $g$. The fraction $\alpha_f$ of actually infecting offspring is the product of $p(f)$ times the number of trials allowed to each daughter particle (i.e., the size of the domain where competition occurs) normalized through the average number of offspring produced by class $f$, i.e.,

$$\alpha_f = f^{-1}p(f)S_D. \quad (4)$$

Accurate predictions of the average fitness in the domain $D$ require knowledge of the average fitness $\mathcal{F}(r)$ of sites at distance $r \ge 1$ conditional to having fitness $f$ at $r_0$: $\mathcal{F}(r)|_{\mathcal{F}(r_0)=f}$, such that $\sum_D f_{nn} = f + \sum_{r(D)\ge 1}\mathcal{F}(r)|_{\mathcal{F}(r_0)=f}$. The function $\mathcal{F}(r)|_{\mathcal{F}(r_0)=f}$ contains all the relevant information on the correlations developed through competition for space, and fulfills $\mathcal{F}(r\to\infty)|_{\mathcal{F}(0)=f} \to \bar{f}$, $\forall\ f$. The quantity $\sum_D f_{nn}$ is a nondecreasing function of $f$, implying that $\alpha_f$ is a nonincreasing function of $f$: $\alpha_f \ge \alpha_{f+1}$. As a general result, we thus obtain that the fraction of successful offspring for an individual of class $f$ is *larger* (or equal, in the limit $d \to \infty$) than that of an individual of class $f+1$.

Let us consider the limit of weak spatial correlations, corresponding to $q < p \to 1$. In that case mixing is very efficient and correlations decay rapidly, so we approximate $\mathcal{F}(r)|_{\mathcal{F}(r_0)=f} = \bar{f}$, $\forall\ r \ge 1$ to get

$$\alpha_f^d = \frac{S_D}{f + (S_D - 1)\bar{f}^d}. \quad (5)$$

(From now on we specify the mixing rate in the system as a superindex $d$.) This expression bounds $\alpha_f^d$ from below for $f < \bar{f}^d$ and from above for $f > \bar{f}^d$. The solution of the one-dimensional model, given now by Eqs. (3) and (5), is explicitly solvable for $F \le 4$. For those cases, we obtained a functional dependence of the form

$$\bar{f}^{\,d} = \bar{f}^{\mathrm{MF}} + \beta(p,q,F)d^{-2} + \gamma(p,q,F)d^{-3} + O(d^{-4}) \quad (6)$$

that we believe to be generic for any value of $F$. For $F=2$, for example, we get $\bar{f}^{\mathrm{MF}} = (3-2p-q+K)/2$,

$\beta(p,q,2) = \frac{p(4p+4q-2K-2)+q(q-K+1)}{4K}$, and $\gamma(p,q,2) = -\beta(p,q,2)/2$, with $K^2 = 4p^2 + (q+1)^2 + 4p(q-1)$. In general, the average fitness tends to the asymptotic mean-field value $\bar{f}^{\mathrm{MF}}$ plus a dominant term inversely proportional to the size of the domain $D$. Hence, in higher dimensions $\bar{f}^d \simeq \bar{f}^{\mathrm{MF}} + \beta(p,q,F)d^{-\mathcal{D}}$, where $\mathcal{D}$ is the physical dimension of the domain where competition occurs.

The functional form of $\bar{f}^d$ can be substituted in Eq. (5) to analyze the quantitative behavior of the set of competition coefficients $\alpha_f^d$. Their average value $\bar{\alpha}^d = F^{-1}\sum_{i=1}^{F}\alpha_i^d$, can be developed in powers of $d$,

$$\bar{\alpha}^d = \frac{1}{\bar{f}^{\mathrm{MF}}} - \frac{1-2\bar{f}^{\mathrm{MF}}+F+4\beta}{(2\bar{f}^{\mathrm{MF}})^2}\frac{1}{d^2} + \frac{1-2\bar{f}^{\mathrm{MF}}+F-8\gamma}{2(2\bar{f}^{\mathrm{MF}})^2}\frac{1}{d^3} + O(d^{-4}). \quad (7)$$

The advantage enjoyed by type $f$ in the quasispecies is

$$\alpha_f^d - \bar{\alpha}^d = \frac{1-2f+F}{(2\bar{f}^{\mathrm{MF}})^2}\left[\frac{1}{d^2} - \frac{1}{2d^3}\right] + O(d^{-4}), \quad (8)$$

demonstrating that, typically (i.e., $\forall\ d > 1/2$), limited mobility favors competitors of lower fitness [those with $f < (F+1)/2$], while phenotypes above-average [$f > (F+1)/2$] are impaired. Both effects are due to the spatial clustering of fitness classes caused by local site competition. The absolute value of the advantage or disadvantage experienced by each fitness type decreases as $d$ increases, and, as for the relative advantage at fixed $d$, tends to zero in the limit of complete mixing ($d \to \infty$):

$$\alpha_f^d - \alpha_f^{d+1} = \frac{\bar{f}^{\mathrm{MF}} - f - 2\beta}{(\bar{f}^{\mathrm{MF}})^2}\frac{1}{d^3} + O(d^{-4}). \quad (9)$$

All results actually converge to mean-field in the limit $d \to \infty$, where $\bar{f}^d\alpha_f \to 1$, $\forall\ f$, and $\lim_{d\to\infty}\bar{f}^d \to \bar{f}^{\mathrm{MF}}$. The derivation of the exact solution for the mean-field system can be found in [4].

We simulate computationally the one-dimensional model (see Fig. 1) for varying $d$ and fixed mutation parameters $p$ and $q$ to obtain a numerical estimation of the average stationary fitness $\bar{f}^d$ and the dispersion $\sigma^d$ of $u_f^d$ in the complete model. Figure 2 compares numerical results for $\bar{f}^d$ and $\sigma^d$ with analytical estimates obtained by solving Eq. (3) using the values of $\alpha_f^d$ yielded by Eq. (5). Analytical and numerical results are in qualitative agreement. For any meaningful combination of parameters, we obtain a monotonous increase in the average fitness for any value of $F$, $\bar{f}^d < \bar{f}^{d+1}$ and a monotonous decrease in the diversity of the quasispecies, $\sigma^d > \sigma^{d+1}$. Figure 3 represents the distribution $u_f^d$ and the associated set of competition coefficients $\alpha_f^d$ obtained from the simulations for $d = 1$ and $d \to \infty$. Any other value of $d$ yields diversity distributions and competition coefficients $\alpha_f^d$ bounded by those two functions.

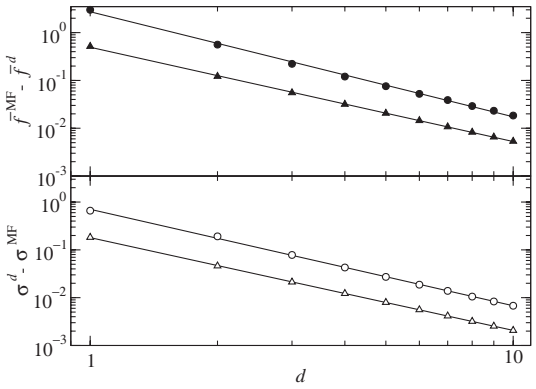


FIG. 2. Average fitness $\bar{f}^d$ and population diversity $\sigma^d$ approach mean-field values as the mixing rate increases as $d^{-2}$, in agreement with the analytical result given in Eq. (6) (here $p = 0.2$, $q = 0.03$, $F = 10$). Results are shown for the analytical approximation yielded by Eqs. (3) and (5) (triangles) and for numerical results of the one-dimensional model (circles). Solid lines are least-square fits to the data.

In our model, the average value of fitness at stationarity, $\bar{f}^d$, is well above unity for any biologically meaningful combination of parameters and finite $d$. This implies that a number $N(\bar{f}^d - 1)$ of the offspring produced at each generation is not able to infect cells, and thus causes a steady flux of individuals leaving the system. The limit of no competition corresponds to $\alpha_f^0 = 1/f$, according to the definition of the competition coefficients given in Eq. (4). This situation is exactly solvable and yields an average fitness $\bar{f}^0 = 1 + \epsilon + \epsilon^2 + \cdots + \epsilon^{F-1} + (1-F)\epsilon^F + O(\epsilon^{F+1})$, with $\epsilon = q/p$: the average fitness approaches unity as $\epsilon \to 0$ in the absence of competition. The overproduction of particles is thus an unavoidable by-product of the competition between types, and becomes negligible if competition is turned off.

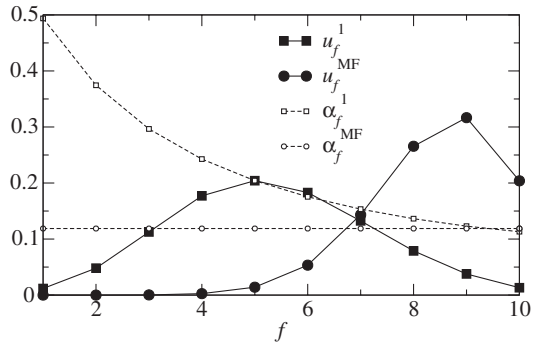


FIG. 3. Asymptotic distribution $u_f^d$ of phenotypes in the quasispecies and corresponding values of the competition coefficients $\alpha_f^d$ for $d = 1$ and $d \to \infty$, and mutation rates $p = 0.2$, $q = 0.03$. Limited mobility confers an advantage to low-fitness types and impairs the relative performance of high-fitness types. Simultaneously, the dispersion $\sigma^d$ increases and the average fitness $\bar{f}^d$ of the population decreases.

A complete solution to the problem, currently unavailable, requires the calculation of the correlation functions $\mathcal{F}(r)|_{\mathcal{F}(r_0)=f}$. Our numerical results indicate that $\mathcal{F}(r)$ decays exponentially with $r$, so spatial correlations are short ranged. This behavior is likely behind the fact that our approximation of weak correlations (which assumes short-ranged correlations), Eq. (5), does not change the qualitative properties of the system (see Fig. 2, for example). Similar problems to the one here tackled have been studied in the context of spatial multitype branching processes, for which few exact results are known [19].

Equation (3) can be used as well to estimate the coefficients $\alpha_f$ compatible with a diversity distribution $u_f$, assuming $p$ and $q$ are known. At present, the competition coefficients $\alpha_f$ cannot be obtained from experimental data. Instead, one may know with relative ease the composition $u_f$ of the population. Consider the results reported in [20], where the fitness of 98 clones of vesicular stomatitis virus was analyzed. From those data, we can infer a distribution $u_f$ with $\bar{f} \simeq 15.9$ and dispersion $\sigma \simeq 1.93$, for $F = 20$. For each mixing rate $d$, the pair $(\bar{f}^d, \sigma^d)$ maps into a pair of parameters $(p, q)$. These values run from (0.030, 0.003) for $d = 1$ to (0.248, 0.013) for $d \to \infty$, maintaining in all cases a ratio $p$:$q$ around 10:1–20:1 that is compatible with experimental knowledge, as obtained in recent experiments that have addressed the effect of mutations on phenotype [21]. Once other experiments of the kind allow to determine more precise values for $p$ and $q$, it will be possible to obtain the value of $d$ that fits the empirical distribution $u_f$. This procedure might eventually return additional knowledge on the degree of competition between classes in the quasispecies through the set of parameters $\alpha_f$, and thus on the properties of the environment where the virus has evolved.

The model here presented is able to capture the effect of different degrees of mixing in heterogeneous populations for fixed intrinsic parameters (as $p$ and $q$). In infections of cellular monolayers, the parameter $d$ has a natural counterpart in the physical properties of the agar where cells are deposited and in the characteristics of the supernatant fluid, so it could be easily modified in properly designed *in vitro* experiments. Quantitative results should be comparable to natural populations in cases where spatial competition is the main competitive interaction between fitness classes, and whenever the population is sufficiently close to the mutation-selection equilibrium. The dynamical rules used could be modified to account for other situations where those conditions do not apply. Such might be out-of-equilibrium states (e.g., shortly after a population bottleneck [18]), other forms of interaction between fitness types (as when viral interference inside the cell occurs [22]), or if the effect of beneficial and deleterious mutations is highly asymmetric, as seems to be the case [21]. Modifications of the current model, together with a deeper comparison with the ever increasing amount of experimental results appears as an interesting avenue to advance in our knowledge of the relationship between mutation, environment, and viral diversity.

The authors acknowledge discussions with C. Briones and E. Lázaro, and the support of the Spanish Ministerio de Educación y Ciencia under Project No. FIS2004-06411.

[1] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow, Genetics **148**, 1667 (1998).
[2] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[3] J. F. Fontanari, A. Colato, and R. S. Howard, Phys. Rev. Lett. **91**, 218101 (2003).
[4] S. C. Manrubia, E. Lázaro, J. Pérez-Mercader, C. Escarmís, and E. Domingo, Phys. Rev. Lett. **90**, 188102 (2003).
[5] L. S. Tsimring, H. Levine, and D. A. Kessler, Phys. Rev. Lett. **76**, 4440 (1996).
[6] C. T. Bergstrom, P. McElhany, and L. A. Real, Proc. Natl. Acad. Sci. U.S.A. **96**, 5095 (1999).
[7] A. Grande-Pérez, E. Lázaro, P. Lowenstein, E. Domingo, and S. C. Manrubia, Proc. Natl. Acad. Sci. U.S.A. **102**, 4448 (2005).
[8] M. Vignuzzi, J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino, Nature (London) **439**, 344 (2006).
[9] C. Escarmís, E. Lázaro, and S. C. Manrubia, Current Topics Microbiol. Immunol. **299**, 141 (2006).
[10] 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).
[11] F. García-Arenal, A. Fraile, and J. M. Malpica, Ann. Rev. Phytopathology **39**, 157 (2001).
[12] S. Altmeyer and J. S. McCaskill, Phys. Rev. Lett. **86**, 5819 (2001).
[13] S. Toyabe and M. Sano, Physica D (Amsterdam) **203**, 1 (2005).
[14] R. Pastor-Satorras and R. V. Solé, Phys. Rev. E **64**, 051909 (2001).
[15] S. C. Manrubia and E. Lázaro, Phys. of Life Rev. **3**, 65 (2006).
[16] This corresponds to a scenario of lytic infection where the total population $N(g) = \sum_f n_f(g)$ can be understood as the number of cells killed at generation $g$.
[17] K. B. Athreya and P. Ney, *Branching Processes* (Springer-Verlag, New York, 1972).
[18] J. Aguirre and S. C. Manrubia, Europhys. Lett. **77**, 38001 (2007).
[19] K. Fleischmann and V. A. Vatutin, Probab. Theory Relat. Fields **116**, 545 (2000).
[20] E. A. Duarte, I. S. Novella, S. Ledesma, D. K. Clarke, A. Moya, S. F. Elena, E. Domingo, and J. J. Holland, J. Virol. **68**, 4295 (1994).
[21] R. Sanjuán, A. Moya, and S. F. Elena, Proc. Natl. Acad. Sci. U.S.A. **101**, 8396 (2004).
[22] P. Whitaker-Dowling and J. S. Younger, Microbiol. Rev. **51**, 179 (1987).