Tradeoff between short-term and long-term adaptation in a changing environment

Robert Forster[1] and Claus O. Wilke[2,1]

[1]*Digital Life Laboratory, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Section of Integrative Biology, University of Texas, Austin, Texas 78712, USA*
(Received 16 December 2004; revised manuscript received 15 July 2005; published 20 October 2005)

We investigate the competition dynamics of two microbial or viral strains that live in an environment that switches periodically between two states. One of the strains is adapted to the long-term environment, but pays a short-term cost, while the other is adapted to the short-term environment and pays a cost in the long term. We explore the tradeoff between these alternative strategies in extensive numerical simulations and present a simple analytic model that can predict the outcome of these competitions as a function of the mutation rate and the time scale of the environmental changes. Our model is relevant for arboviruses, which alternate between different host species on a regular basis.

I. INTRODUCTION

The quasispecies model [1] is the premier model to study the evolution of asexual replicators, such as self-replicating molecules or viruses [2,3]. Originally formulated for constant environments, the quasispecies model has recently been extended to describe adaptation to a changing environment [4–8]. The various extensions of the model all work within the original deterministic framework developed by Eigen and coworkers, and thus assume that the population size is infinite. This assumption implies that a population can never lose any genetic information. However, the loss of genetic material, and the mechanisms that prevent it from occurring, are probably major forces shaping the evolutionary dynamics of finite populations in time-dependent environments: A finite population can lose to mutation pressure previously useful genetic material that has become meaningless after a change in the environment, and the population may not be able to reacquire this material when the environment changes back to its original state. Consequently, the population will be at a selective disadvantage in comparison to another population that has managed to prevent a similar loss, even if the second population had to pay some short-term cost to keep the useless genetic material. Therefore, in an environment that alternates between two (or more) states, natural selection faces two conflicting agendas—(i) specialization to the current state of the environment or (ii) adaptation to the long-term environment, which includes both environmental states.

Here, we study the evolutionary dynamics of competing finite populations of asexual replicators in an environment that alternates between two states, remaining in each state for a time interval of length $T/2$ before switching to the other. To study the tradeoff between the competing forces of selection in this environment, we consider two different strains of replicators, as previously proposed [9]. The first strain, which we refer to as the *fused strain*, has a single gene that performs equally well in both environmental states. The second strain, which we refer to as the *divided strain*, has two genes, each of which is advantageous in one environmental state and useless in the other. Clearly, if the fused strain performs as well as the divided strain in both environments, without paying any additional cost, then the divided strain cannot have a selective advantage over the fused strain, regardless of the time scale on which environmental changes happen. If, however, the fused strain does pay some small cost, then which strain is advantageous depends on the exact interplay of the cost, the time scale of environmental change, and the mutation rate. Here, we develop a method to assess and analyze this interplay and to predict which strain is advantageous in a given setting. As cost, we consider the differential mutation pressure that arises when the fused and the divided strain have genes of different length [10,11]. However, it is straightforward to extend our approach to other types of costs.

Note that while we refer to separate genes throughout this paper, our model can also apply to separate functions carried out by a single gene. In this case, the divided strain corresponds to a gene that can adapt to either function, but not to both at the same time; whereas the fused strain corresponds to a gene that can adapt to both functions at the same time. Such a situation has been observed in an artificial-life simulation with a changing environment [12], where the fusion of genetic function evolved presumably through changes in the amount of epistatic interactions among the different parts of the organisms' genomes.

II. MATERIALS AND METHODS

A. Model

Here, we model the evolutionary dynamics of a finite population in a time-dependent environment. For comparison, for an *inifite population* in a time-dependent environment, the quasispecies equation reads [5]

$$\frac{dy_i(t)}{dt} = \sum_j w_j(t)\mu_{ij}y_j(t) - y_i(t)\sum_j w_j(t)y_j(t), \qquad (1)$$

where y_i is the fraction of type i in the population, $w_i(t)$ is the replication rate (i.e., fitness) of type i at time t, and μ_{ij} is the mutation rate per unit time from type j to type i. The quadratic term corresponds to the total production of new organisms per unit time and is subtracted to keep the $y_i(t)$ normalized.

We represent all genes as binary strings. The divided strain has two genes, each of length L_{div}, and the fused strain

has a single gene of length L_{fuse}. For each gene, there exists a single functional sequence (the master sequence) that confers the selective advantage, and all alternative sequences are nonfunctional, regardless of the environment. The reproductive fitness $w_i(t)$ of an individual is determined by whether the individual has a functional gene specialized for the current state of the environment. An individual with the correct functional gene has fitness $1+s$, while an individual without such a gene has fitness 1. Mutations occur upon reproduction with a per-site probability μ, corresponding to a per-gene mutation rate of $U = \mu L$.

B. Simulation

Both the speed of environmental change and the mutation rate μ are important factors in determining the outcome of the competition between the divided and fused strains. To assess their relative importance, we simulated a population of $N = 1000$ individuals reproducing in discrete generations and with probability of reproduction proportional to their fitness (Wright-Fisher sampling). Initially the population was divided equally between fully functional members of the two strains and the simulation continued until one strain became extinct. We fixed the length of the divided strain at $L_{\text{div}} = 5$, while L_{fuse} varied from 3 to 11. We performed 10 000 replicates at each pair of period lengths and mutation rates for $T/2$ (10, 30, 100, 300, 1000, 3000, 10 000) and $U_{\text{div}} = \mu L_{\text{div}}$ (0.0001, 0.0003, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3).

III. RESULTS

A. Time scales

To understand the dynamics of competition between the divided and fused strains in a finite population, we consider two time scales—(i) the competition time scale T_{c} and (ii) the drift time scale T_{dr}. In order to calculate these time scales, we need to know the selective advantage of one strain over the other. The selective advantage s of genotype i over genotype j is a key quantity in theoretical population genetics and is defined as $s = (w_i - w_j)/w_j$ [13]. While this definition is *a priori* applicable only to individual genotypes, it turns out that to a good approximation, standard results from population genetics can be applied to separate strains (which consist of a mixture of closely related mutants) if we treat each strain as an individual genotype with fitness given by the strain average $\langle w \rangle$ [14]. Throughout this paper, we refer to the selective advantage of one strain over another as the effective selective advantage of this strain. We define the effective fitness advantage of the fused strain over the divided strain by

$$s_{\text{eff}} = \frac{\langle w_{\text{fuse}} \rangle - \langle w_{\text{div}} \rangle}{\min\{\langle w_{\text{fuse}} \rangle, \langle w_{\text{div}} \rangle\}}. \tag{2}$$

This definition guarantees that the magnitude of s_{eff} corresponds to the fitness advantage of the superior strain, while the sign indicates whether the fused strain is superior (positive s_{eff}) or inferior (negative s_{eff}). If one strain has an effective fitness advantage $|s_{\text{eff}}|$ over the other, the competition

time scale T_{c} is defined to be the typical time until extinction of the inferior strain in a constant environment. Neglecting finite population effects and applying Eq. (1) to strains rather than genotypes, we find that the population fraction $x(t)$ of the superior strain changes approximately according to the logistic equation

$$\dot{x}(t) = |s_{\text{eff}}|x(t)[1 - x(t)], \tag{3}$$

subject to our initial condition that $x(0) = 1/2$. To determine the typical extinction time of the inferior strain, we solve Eq. (3) for the time when a single member remains of the inferior strain. Thus

$$T_{\text{c}} = \frac{\ln(N-1)}{|s_{\text{eff}}|} \approx \frac{\ln N}{|s_{\text{eff}}|}. \tag{4}$$

The drift time scale is defined as the average time for a neutral mutation to go to fixation. Neutral drift becomes important when the fitness advantage between the competing strains is small compared to the fluctuations due to finite sampling effects, i.e., when $s_{\text{eff}} \lesssim 1/N$ [15]. For our initial conditions, the diffusion theory [16] predicts that

$$T_{\text{dr}} = 2N \ln 2 \approx 1390 \tag{5}$$

generations.

B. Quasispecies effects

Given sufficient time to reach equilibrium, each strain will adopt a quasispecies distribution consisting primarily of those members with a functioning gene adapted to the current environment, together with the deleterious mutants that are constantly regenerated through mutation pressure. In our dynamic fitness landscape, the fused strain will attain this equilibrium distribution after some time, while the divided strain will attempt to equilibrate to the current environment and then go through a period of transition when the environment changes. As an example, Fig. 1 illustrates the formation of the fused-strain quasispecies and the dynamics of the divided-strain quasispecies in the limit of large population size.

Before we can use the time scales derived in the preceding section to predict the competition's outcome, we must estimate the average fitness of each strain appearing in Eq. (2). In our model, the average fitness $\langle w \rangle$ of a strain is given by $\langle w \rangle = 1 + s y_0$, where y_0 is the fraction of the population that has a functional gene for the current environment. Thus, Eq. (2) becomes

$$s_{\text{eff}} = \frac{s y_{0,\text{fuse}} - s y_{0,\text{div}}}{1 + s \min\{y_{0,\text{fuse}}, y_{0,\text{div}}\}}. \tag{6}$$

To estimate y_0, we assume that the quasispecies immediately reaches its equilibrium distribution. Denoting $y_i(t)$ as the population fraction of a given strain with i errors at time t, we neglect the back mutation term (that is, the term that represents mutations from genotypes with errors to error-free genotypes) in Eq. (1) and obtain
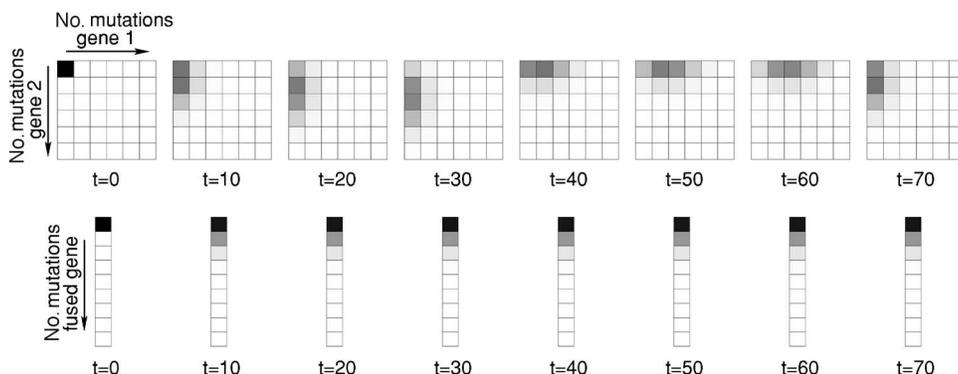
FIG. 1. Population structure of the divided and fused strains at various time points. Gray levels indicate the fraction of sequences at the given mutational distance from the respective error-free sequence. Parameters are oscillation period $T=60$, per-site mutation rate $\mu=0.02$, length of a single gene of the divided strain $L_{div}=5$, length of the fused gene $L_{fused}=8$, selective advantage of functional gene $s=1$, and infinite population size. From Ref. [9].

$$\dot{y}_0(t) = (1+s)Q_{00}(L)y_0(t) - [1+sy_0(t)]y_0(t), \qquad (7)$$

where $Q_{ij}(L)$ is the probability that a string of length L with j errors mutates into one with i errors. $Q_{ij}(L)$ has been given for example in Ref. [17]. Setting $\dot{y}_0=0$ in equilibrium, we find $y_0=[(1+s)Q_{00}(L)-1]/s$. However, when back mutations become significant and y_0 approaches zero in this expression, we reach the classical error threshold [1]. For mutation rates beyond this point, we assume that the population is randomized uniformly over all possible states, and hence we use

$$y_0 = \max\left\{\frac{(1+s)Q_{00}(L)-1}{s}, 2^{-L}\right\} \qquad (8)$$

for all mutation rates. For mutation rates below the error threshold, Eq. (8) yields a simple form for the magnitude of the fitness advantage, $|s_{eff}| = (1-\mu)^{-|L_{div}-L_{fuse}|}-1$, while the sign of s_{eff} is given by $\text{sgn}(L_{div}-L_{fuse})$. This result shows that the effect of mutational load on fitness is to favor whichever strain has the shorter length.

The result of Eq. (8) applies to a quasispecies that has reached equilibrium. While the equilibrium assumption provides a good estimate for the fused strain, the rate of environmental changes may prevent the divided strain's quasispecies from ever reaching this equilibrium. If the environment changes quickly (relative to the competition and drift time scales), the divided strain persists in an average environment that requires both genes for functionality [5], and hence, this strain has an effective gene length of $2L_{div}$. In this case, we approximate the resulting quasispecies as one with a single gene of length $2L_{div}$ and replace $Q_{00}(L_{div})$ by $Q_{00}(2L_{div})$ in Eq. (8). Even though this approximation disre-

gards the two-dimensional nature of the divided-strain quasispecies, it gives a reasonably good estimate of the true y_0 for most mutation rates. Note, however, that we do not replace $2^{-L_{div}}$ by $2^{-2L_{div}}$. At any point in time, the environment favors only one of the divided strain's two genes, and hence, beyond the error threshold the probability that a randomly chosen individual from the divided strain carries a functional gene remains $2^{-L_{div}}$. We refer to s_{eff} calculated with $Q_{00}(L_{div})$ as the short-term limit and to s_{eff} calculated with $Q_{00}(2L_{div})$ as the long-term limit.

C. Predicting the probability of fixation

We propose a simple ternary model to predict the probability of fixation p of the fused strain. In our model, p is 0 if the divided strain is favored, $1/2$ for neutral evolution, when both strains are equally likely to go to fixation, or 1 if the fused strain is favored. First, we classify the selective regime based on the drift time T_{dr} and the short-term competitive time scale $T_c=\ln N/|s_{eff}^{short}|$ in comparison to $T/2$, the length of time for which the environment remains constant (see Table I). If $T_c<T/2$, then we expect the competition between the two strains to end before the environment changes even once, and hence, the short-term limit applies. The value of T_{dr} is irrelevant in this case. If both times are longer than $T/2$, then we expect the competition to extend over several half periods, and hence, the long-term limit applies. Finally, if $T/2$ is smaller than T_c, but larger than T_{dr}, then we expect drift to be the dominant force. We call this regime the neutral limit and set $s_{eff}=0$. Having determined the appropriate limiting case (short-term, neutral, or long-

TABLE I. Selective regime, as determined by the relative magnitudes of T_{dr}, T_c, and $T/2$.

Condition	Selective Regime
$T/2>T_c$	short-term limit
$T_{dr}<T/2<T_c$	neutral limit
$T/2<T_{dr},T_c$	long-term limit

TABLE II. Model predictions, as determined by the relative magnitude of s_{eff} and $1/N$.

Fitness Advantage (Fused Strain)	Prediction	p
$s_{eff}\leqslant-1/N$	divided strain wins	0
$-1/N<s_{eff}<1/N$	neutral[a]	1/2
$s_{eff}\geqslant1/N$	fused strain wins	1

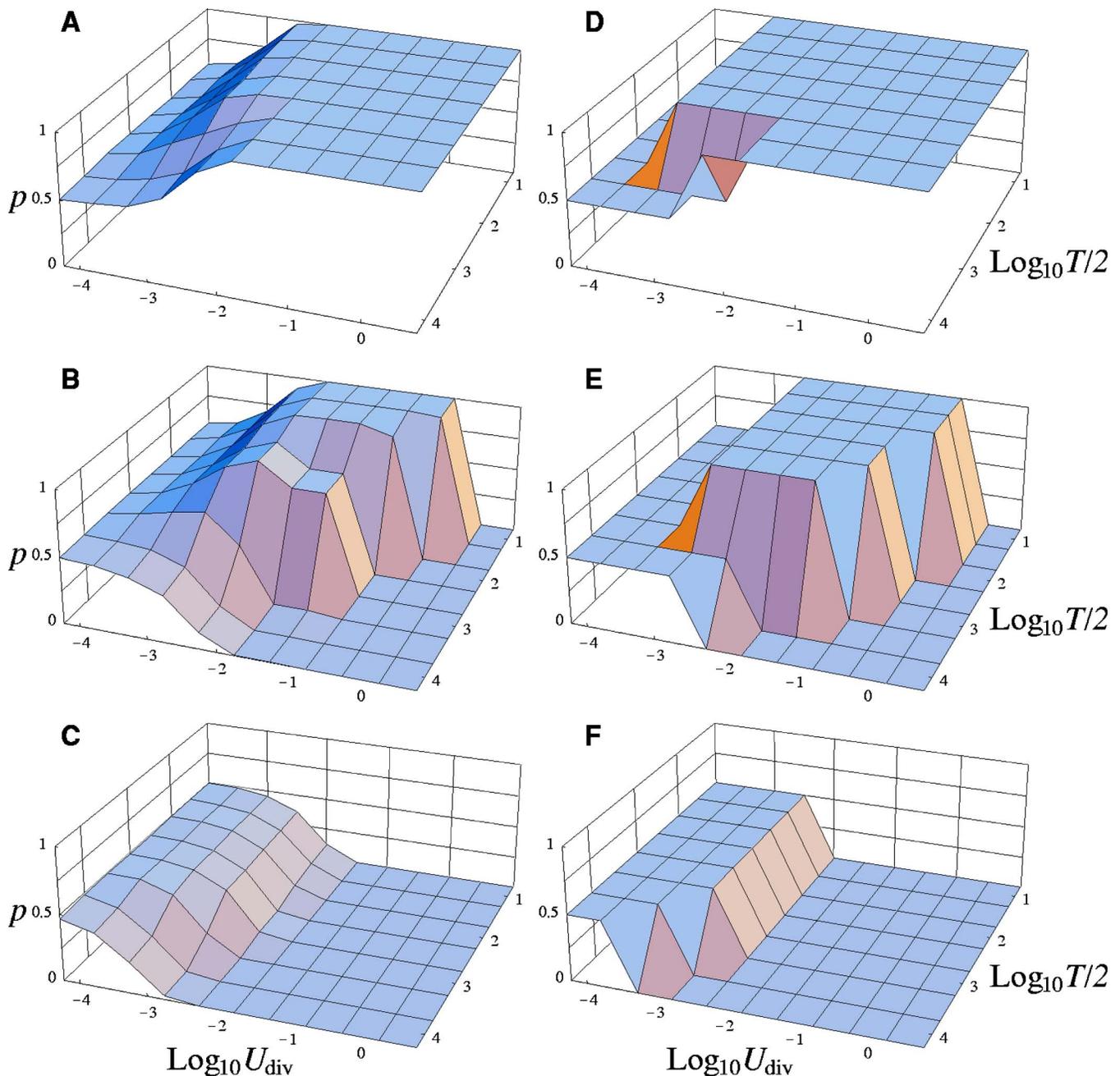[a]In the neutral case, both strains are equally likely to win.

FIG. 2. (Color online) Left (a, b, c): Simulation results for the probability of fixation of the fused strain as a function of the mutation rate $U_{div}=\mu L_{div}$ and period length T for $L_{div}=5$ and $L_{fuse}=4,6,11$ (top to bottom). Simulation results have a standard error of approximately $\pm 1\%$. Right (d, e, f): Model predictions for the same. Parameter values are $s=1$, $N=1000$.

term limit), we can use the associated fitness advantage s_{eff} to predict the probability p of fixation for the fused strain (Table II). If $|s_{eff}|<1/N$, then the two strains are effectively neutral. Hence, the outcome of the competition is determined by drift and $p=1/2$. Otherwise, $p=0$ or 1 depending on whether s_{eff} is negative or positive.

D. Comparison with simulation results

From our simulations, we estimated the probability p that the fused strain would fixate in the population as a function of $T/2$, U_{div}, and L_{fuse}. Results are shown in Figs. 2(a)–2(c),

for representative values of L_{fuse}. The probabilities obtained by simulation have a standard error of $\pm 1\%$. Figures 2(d)–2(f) show the corresponding predictions of our model. Over the full range of parameters described in Sec. II B, our model predicted that the divided strain was superior in 235 cases ($p=0$), that the strains were neutral in 240 cases ($p=1/2$), and that the fused strain was superior in 155 cases ($p=1$). The range of simulation results corresponding to each of these three predictions is shown in Fig. 3. When the model predicted $p=0$, 85% of all simulation probabilities fell in the range 0–0.1. When $p=1$ was predicted, 84% of simulation probabilities fell between 0.9 and 1; while in the neutral case
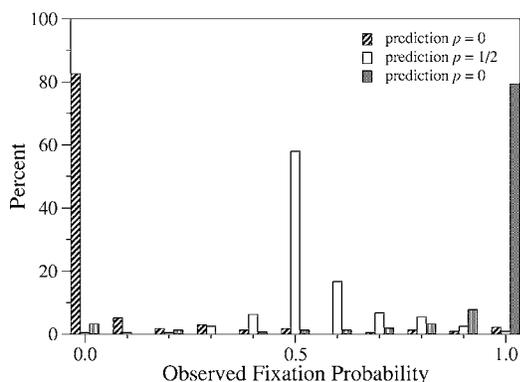
FIG. 3. Fixation probabilities for the fused strain (as determined by simulation), classified by model prediction. The model predicted $p=0$ in 235 cases, $p=1/2$ in 240 cases, and $p=1$ in 155 cases. The x axis is binned in 0.1 increments, with the bins' midpoint values shown.

($p=1/2$ predicted), 58% of simulation probabilities fell between 0.45 and 0.55. The rms error between the simulation values and the model predictions is 18.6% averaged over all cases. For comparison, the best possible rms error for any such ternary model is 7.6% on this data.

IV. DISCUSSION

The study of quasispecies dynamics in a time-dependent fitness landscape to date has primarily focused on the limit of infinite population size [4–8]. In a periodic fitness landscape, an infinite population size guarantees that competition between two strains will result in the deterministic extinction of the inferior strain or, in certain finely tuned cases, an unstable coexistence between the strains (although frequency-dependent selection may stabilize this equilibrium [18]). In contrast, the generalization of these models to a finite population presents a continuous range of possibilities from almost certain extinction to the complete randomness of neutral drift.

In this study, we present a model for predicting the outcome of competition between finite quasispecies in a periodic environment. As applied to our specific case of competition between a divided and a fused strain, our model shows good qualitative and quantitative results in comparison with simulation (Figs. 2 and 3). When $L_{fuse} < L_{div}$ or $L_{fuse} > 2L_{div}$, one of the two strains is strictly superior, and the outcome of the competition is determined by whether drift or competition are more important at the given mutation rate. For intermediate values of L_{fuse}, the competitive dynamics are more complex, as certain combinations of mutation rate and period lengths favor the fused strain while others favor the divided strain.

The only qualitative feature of the competitive dynamics that we have ignored is the time for the quasispecies to reach its equilibrium distribution. We decided not to include a quasispecies time scale in our model after preliminary efforts

showed that the added complexity failed to significantly improve model accuracy. While the fused strain quickly reaches its equilibrium independent of the environment's state, the divided strain may fare worse than predicted in some special cases when it spends significant time in transitions between the quasispecies distributions for each environment, being poorly adapted in the meanwhile. Still, these cases are relatively rare, since at low mutation rates, quasispecies effects are less important; while at high mutation rates, the error threshold masks any such effects.

Our results suggest that, under appropriate circumstances, a selective pressure exists to fuse or divide complementary genes in a periodic environment. The tendency of uncertain enviromental conditions to facilitate large scale genetic changes such as this one has recently been studied [19]. Most genomes are full of apparently useless or nonfunctional genetic material, which in our model corresponds to the excess length of the fused strain over a single gene of the divided strain. In populations as small as 1000, such as those we studied, such temporarily useless genetic material comprising 20% of the genome could be stably maintained for periods of dormancy up to 1000 generations [Fig. 2(b)]. Alternatively, a fused gene exposed to an appropriate periodic environment might undergo division in response to the selective pressure we describe. The beginnings of a genome segmentation has been observed in foot-and-mouth disease virus (FMDV) in response to conditions of high multiplicity of infection [20]. One explanation of this segmentation is that translational speed favors shorter genes [21,22], a fitness effect similar to the length-based mutational pressure of our model.

Our model applies directly to the evolution of arboviruses, which are viruses transmitted by arthropods. For example, the arbovirus West Nile virus is transmitted from birds to birds (and the occasional human) by mosquitoes, experiencing the alternating environments of the avian and insect hosts. Experimental virologists have long tried to determine whether viruses subjected to such alternating environments adapt to the short-term or the long-term environment, but have not found a conclusive answer. Experiments with vesicular stomatitis virus (VSV), eastern equine encephalitis virus (EEEV), and Dengue virus in cells of insect and mammalian origin have shown that in some cases, adaptation to one cell type leads to loss of fitness in the other cell type; while in other cases, fitness can increase in both cell types at the same time [23–29]. Which of the two cases occurs depends on the time spent in each of the two hosts and also on the details of the fitness landscape in the two hosts. It stands to reason that in future experiments, in which the time scale of environmental change is varied over a wide range, a switch in the adaptation strategy from short term to long term will be observed, and that the time scale at which the switch occurs can be predicted with the methods we have developed here.

In conclusion, we have demonstrated that the competitive dynamics of finite populations in a time-dependent environment can be quite complex, but that, nevertheless, estimates of the effective fitness advantages of the different strains

together with an understanding of the drift and competitive time scales can lead to remarkably accurate predictions of the evolutionary dynamics. We believe that similar tech-niques will prove useful to interpret and predict outcomes of virus-evolution experiments in changing environments.

[1] M. Eigen and P. Schuster, *The Hypercycle—A Principle of Natural Self-Organization* (Springer-Verlag, Berlin, 1979).

[2] M. Eigen, J. McCaskill, and P. Schuster, J. Phys. Chem. **92**, 6881 (1988).

[3] E. Domingo, C. K. Biebricher, M. Eigen, and J. J. Holland, *Quasispecies and RNA Virus Evolution: Principles and Consequences* (Landes Bioscience, Georgetown, TX, 2001).

[4] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).

[5] C. O. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).

[6] M. Nilsson and N. Snoad, Phys. Rev. E **65**, 031901 (2002).

[7] C. Kamp and S. Bornholdt, Phys. Rev. Lett. **88**, 068104 (2002).

[8] Y. Brumer and E. I. Shakhnovich, Phys. Rev. E **69**, 061909 (2004).

[9] C. O. Wilke, R. Forster, and I. S. Novella, Current Topics in Microbiology and Immunology (to be published).

[10] C. O. Wilke *et al.*, Nature (London) **412**, 331 (2001).

[11] C. O. Wilke and C. Adami, Mutat Res. **522**, 3 (2003).

[12] Y. Li and C. O. Wilke, Artif. Life **10**, 123 (2004).

[13] J. F. Crow and M. Kimura, *An Introduction to Population Ge-netics Theory* (Harper & Row, New York, 1970).

[14] C. O. Wilke, Evolution (Lawrence, Kans.) **55**, 2412 (2001).

[15] M. Kimura, J. Appl. Probab. **1**, 177 (1964).

[16] M. Kimura and T. Ohta, Genetics **61**, 763 (1969).

[17] G. Woodcock and P. G. Higgs, J. Theor. Biol. **179**, 61 (1996).

[18] C. O. Wilke, D. Reissig, and I. Novella, Evolution (Lawrence, Kans.) **58**, 900 (2004).

[19] D. J. Earl and M. W. Deem, Proc. Natl. Acad. Sci. U.S.A. **101**, 11531 (2004).

[20] J. García-Arriaza *et al.*, J. Virol. **78**, 11678 (2004).

[21] S. Nee, J. Mol. Evol. **25**, 277 (1987).

[22] E. Szathmary, J. Theor. Biol. **157**, 383 (1992).

[23] I. S. Novella *et al.*, J. Virol. **69**, 6805 (1995).

[24] I. S. Novella *et al.*, J. Mol. Biol. **287**, 459 (1999).

[25] S. C. Weaver, A. C. Brault, W. Kang, and J. J. Holland, J. Virol. **73**, 4316 (1999).

[26] P. E. Turner and S. F. Elena, Genetics **156**, 1465 (2000).

[27] L. A. Cooper and T. W. Scott, Genetics **157**, 1403 (2001).

[28] W-J. Chen, H-R. Wu, and S-S. Chiou, Intervirology **46**, 289 (2003).

[29] S. Zárate and I. S. Novella, J. Virol. **78**, 12236 (2004).