Different fitnesses for in vivo and in vitro evolutions due to the finite generation-time effect

David B. Saakian,[1,2,3] Araks S. Martirosyan,[4] and Chin-Kun Hu[1,5]
[1]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[2]*Yerevan Physics Institute, Alikhanian Brothers Street 2, Yerevan 375036, Armenia*
[3]*National Center for Theoretical Sciences, Physics Division, National Taiwan University, Taipei 10617, Taiwan*
[4]*Yerevan State University, Alex Manoogian 1, Yerevan 375025, Armenia*
[5]*Center for Nonlinear and Complex Systems and Department of Physics, Chung-Yuan Christian University, Chungli 320, Taiwan*
(Received 15 December 2009; revised manuscript received 7 April 2010; published 8 June 2010)

We consider the finite generation-time effect in virus evolution models, introducing differential equations with delay. The suggested approach more adequately describes the evolution in case of growing populations than the popular models of population genetics, especially for the viruses with large number of offspring during one life cycle. Now the mean fitness, as a coefficient for exponential population growth, could not be defined via instant characteristics of the model. For the constant population size the finite generation-time does not affect mean fitness in the steady state. The growing virus population is characterized by different fitness than the population with a constant size.

DOI: 10.1103/PhysRevE.81.061913 PACS number(s): 87.10.Hk, 87.10.Mn, 87.10.Rt

I. INTRODUCTION

The modeling of evolving population via mathematical models has a rich history. While from the early days of population genetics there has been investigated the evolution problem of age structured populations, see [1–5], the most of results in population genetics are derived in simpler models, considering the discrete time models with Wrightian fitness or continuum time differential equation models with Malthusian fitnesses [6,7]. The finite generation time influences, for example, the molecular clock phenomenon [8]. We will solve the simplest virus evolution model with finite generation-time effect and check that the evolution characteristics of growing virus population are substantially effected by introduction of finite generation time.

While investigating the viruses, one can investigate numerically a realistic model of virus evolution with several stages of virus reproduction cycle [9]. As an alternative approach, one considers a simplified description of the dynamics, considering different evolution processes together, via simple mapping to the discrete time Wright-Fisher [10] and Moran [11] models or considering differential equation models [6,12–16]. The popular evolution models are just those [6,12,13], written as systems of ordinary differential equations. Such differential equations predict an exponential growth of population, well observed in experiment [17–19]. The differential equations, describing the exponential growth, give the main practical tools to define the fitness from experimental data [18,19]. In this paper we will revise the virus evolution model, considering the finite generation time. As a result we get differential equations with delay as evolution models. We will derive the mean fitness using the Hamilton-Jacobi equation (HJE) method [20–22]. When the random evolution factors such as mutation or recombination change their rates it is impossible to work only with a single Malhusian fitness and miss the generation time as an irrelevant parameter. Moreover, according to our results, there are different fitnesses or selection coefficients for the growing populations and fixed size populations.

How does the exponential growth arise in simple models? Consider the virus that gives new K offsprings after any τ period of time. While this parameter fluctuates from one virus to another, it could be defined experimentally. The concentration $p(t)$ of the viruses grows according the following law:

$$p(t+\tau) = p(t)(K+1). \tag{1}$$

The latter equation is a typical form of discrete time approach to evolution processes, and $K+1$ can be identified with the Wrightian fitness.

When we observe the system for time period $t \gg \tau$, we can find an exponential growth:

$$p(t) = p(t=0)e^{rt}, \tag{2}$$

where the growth rate is defined as [16]

$$r = \ln(1+K)/\tau. \tag{3}$$

The exponential growth could be described via simple differential equation:

$$\frac{dp(t)}{dt} = rp(t). \tag{4}$$

For the experimental schemes of fitness definition see [19].

A simple generalization of Eq. (4) is the Crow-Kimura model. In case of Crow-Kimura model we consider the following system of nonlinear ordinary differential equation (ODE):

$$\frac{dP_i}{dt} = P_i r_i + \sum_j m_{ij} P_j - P_i \sum_j P_j r_j. \tag{5}$$

Here m_{ij} is the rate of mutation from configuration j to a new configuration i, and r_i is the fitness. We consider the simplified version of the model, where at any loci there are two alleles. P_i is the probability of type i, therefore $\Sigma_i P_i = 1$. The Hamming distance d_{ij} between two sequences is the number of different alleles in the same positions of genome, and $m_{ii} = -\mu$. When $d_{ij}=1$ then $m_{ij} = \mu/N$ and $m_{ij}=0$ for $d_{ij} > 1$

[7]. The nonlinear term describes the dilution of the system to hold constant population size. After relaxation we have the mean fitness at the steady state P_i:

$$R = \sum_j P_j r_j. \tag{6}$$

From the other hand we can consider the linear version of Eq. (5):

$$\frac{d\hat{p}_i}{dt} = \hat{p}_i r_i + \sum_j m_{ij}\hat{p}_j, \tag{7}$$

and define the mean fitness as

$$R = \lim_{t\to\infty} \frac{\ln\left(\sum_j \hat{p}(t)_j\right)}{t}. \tag{8}$$

$\hat{p}_i$ in Eq. (7) are relative probabilities. Two systems [Eqs. (5) and (7)] give exactly the same dynamic, after transformation [23,24]:

$$P_i = \frac{\hat{p}_i}{\sum_j \hat{p}_j}. \tag{9}$$

Two definitions of mean fitness, by Eqs. (6) and (8), are equivalent, while Eq. (7) describes the evolution of growing population. For the symmetric fitness landscape (any sequence at the Hamming distance from the reference sequence l has the same fitness r_l), we write the equation for the relative probabilities p_l of the whole Hamming class (collection of sequences with the same Hamming distance),

$$\frac{dp_l}{dt} = p_l[r_l - \mu] + \mu\frac{(N-l+1)}{N}p_{l-1} + \mu\frac{(l+1)}{N}p_{l+1},$$

$$r_l = f\left(1 - \frac{2l}{N}\right). \tag{10}$$

Equation (10) is the equivalent form of Eq. (7) for the considered case of symmetric fitness landscape.

II. EVOLUTION MODEL WITH DELAY

While the fitness is defined just as r, actually it depends on two processes: the number of offsprings K and the period τ.

We can get the same exponential growth as in Eq. (4) considering the differential equation with delay:

$$dp(t)/dt = \hat{r}p(t-\tau), \tag{11}$$

where

$$\hat{r} = re^{r\tau} \tag{12}$$

or

$$\hat{r} = (1+K)\frac{\ln(1+K)}{\tau}. \tag{13}$$

In Eq. (11) we hold the memory about the period τ of the virus, while in Eq. (4) it has been missed. We can derive Eq. (11) considering different L groups which start to divide at different moments of time $t=n\tau/L$ [4]. In our case $p(t)$ is a piecewise constant function that have jumps at any moment $t=\tau n$. Different L groups start to divide at different moments of time, starting at $t=\tau*n/L$, n is an integers, $n<L$. Actually these subsequences $p(t)$ [$t=\tau(n/L+k)$, k is any positive integer] are independent series for different n in case of free growth. For our case of free growth, we can construct a continuous exponential growth after interval τ/L, taking proper initial conditions. For the free growth there is no difference between two schemes by Eqs. (4) and (11), the difference could arose if we take into account other evolution factors: the nonlinear interaction due to dilution to hold constant population size and random processes such as mutation or recombination.

We can express r via $\hat{r}$:

$$r = F[\hat{r}\tau]/\tau, \tag{14}$$

where we used an implicit function $F(y)$:

$$ye^y = F^{-1}[y]. \tag{15}$$

We derive Eq. (3) for a fixed number of branching $K+1$. In principle could be a distribution of branching number. When can the effect of delay be ignored? Equations (3) and (13) give the following criteria:

$$\ln(1+\langle K\rangle) \ll 1. \tag{16}$$

A. Crow-Kimura model with cdelay

Let us consider a system of equations like to the one by Eq. (10), only with the time delay in the term with fitness:

$$\frac{dp_l(t)}{dt} = p_l(t-\tau)f\left(1 - \frac{2l}{N}\right) - \mu p_l(t) + \mu\frac{(N-l+1)}{N}p_{l-1}(t) + \mu\frac{(l+1)}{N}p_{l+1}(t). \tag{17}$$

Without mutation we have Eq. (11) with $\hat{r}=r_l$ which describes the exponential growth $p_l=c\exp[Nkt]$, where $k=F(r_l\tau)/\tau$ is smaller than r_l. In Eq. (17) it is assumed that the mutation process is independent of selection and acts in parallel.

There is another important issue as well. When $\tau=0$, the equations in Crow-Kimura model are invariant under the transformation $f(m)\to f(m)+c$. Now this invariance is broken, and we need therefore in absolute values of fitness.

B. Solution of the delay model with HJE method

Let us consider a solution of Eq. (17) in the form

$$p_l(t) = \exp\left[Nu\left(1 - \frac{2l}{N}, t\right)\right]. \tag{18}$$

We assume that $u(x,t)$ is a smooth function of t. Perhaps the latter property is possible for special initial conditions for the system: Eq. (17). Now Eq. (17) transforms into the following Hamilton-Jacobi equation for $u(x,t)$:

$$N\frac{\partial u}{\partial t} = f(x)e^{-N\tau\partial u/\partial t} + \mu\left(\frac{1+x}{2}e^{2\partial u/\partial x} + \frac{1-x}{2}e^{-2\partial u/\partial x} - 1\right), \tag{19}$$

where $x=1-\frac{2l}{N}$.Let us denote $q\equiv N\frac{\partial u}{\partial t}$, $p\equiv\frac{\partial u}{\partial x}$ and consider the following function:

$$q = H(x,q,p),$$

$$H(x,q,p) = f(x)\exp[-\tau q] + h_0(x,p),$$

$$h_0(x,p) = \mu\frac{1+x}{2}e^{2p} + \mu\frac{1-x}{2}e^{-2p} - \mu, \tag{20}$$

and the implicit function $H_0(x,p)$ defined from the following equation:

$$H_0(x,p) = H(x,H_0(x,p),p). \tag{21}$$

Having H_0, we can rewrite Eq. (20) in an equivalent form

$$q = H_0(x,p). \tag{22}$$

Thus we can use the methods of [20], while H_0 is defined implicitly via Eq. (21).

III. MEAN FITNESS

First of all let us calculate the mean fitness. We assume the following asymptotic for $p_l(t)$:

$$p_l(t) = \exp[Rt + Nu_0(x)], \tag{23}$$

where the coefficient of exponential growth is just identified as a mean fitness. The existence of an asymptotic regime is nontrivial issue. For example, in case of free growth its existence depends on initial conditions. For the model with both selection and mutation the asymptotic exists, as has been checked by our numerics.

Following to [20,22], we define first potential $U(x)$ as

$$U(x) = \min[H_0(x,p)]|_p, \tag{24}$$

and the mean fitness as

$$R = \max[U(x)]|_{-1\le x\le 1}. \tag{25}$$

From the other hand $H_0(x,p)$ is a monotonic function of p at fixed value of x. Equation (22) is equivalent to the equation

$$q - e^{-\tau q}f(x) = h_0(x,p), \tag{26}$$

where $q=H_0(x,p)$. The left-hand side of Eq. (26) is a monotonic function of q, thus the minimum of $H_0(x,p)$ is at the minimum of $h_0(x,p)$ or

$$U(x) - e^{-\tau U(x)}f(x) = \mu\sqrt{1-x^2} - \mu. \tag{27}$$

Then we get the mean fitness as the maximum of $U(x)$, according to Eq. (25). Equations (25) and (27) are the main results of our work.

A. Mean fitness for the single-peak fitness case

Consider the fitness $r_0=(J+c)N$ and $r_i=N$, $i\ge 1$ (Table I). For this case we have the following equation for the mean fitness in the selective phase:

TABLE I. Mean fitness for the single peak model with delay by Eq. (17), $c=0$, $\mu=1$.

N	100	100	100	100	100
τ	0	0.3	1	1	1
J	2	2	2	3	4
R_{theor}	1.005	0.647	0.375	0.617	0.799
R_{num}	1.005	0.649	0.377	0.619	0.800

$$R = (J+c)e^{-\tau R} - \mu. \tag{28}$$

For a large J R decreases exponentially with the τ. In the nonselective phase we have

$$R - ce^{-R} = 0. \tag{29}$$

The error threshold is the value of J where two fitnesses by Eqs. (28) and (29) coincide. While in case of $\tau=0$ the error threshold was c independent, now it depends on absolute value of fitness.

We should be accurate while defining the mean fitness in our case. It is the coefficient of exponential growth of population, see Eq. (8). Such choice of mean fitness is already realized in age-structured evolution models [25,26]. In the evolution models without time delay R could be calculated as $\frac{\Sigma_l r_l p_l}{\Sigma_l p_l}$. Now the mean fitness is defined via the distribution in the past:

$$R(t) = \frac{\sum_l r_l p_l(t-\tau)}{\sum_l p_l(t)}. \tag{30}$$

The surplus at the moment of time t is defined as

$$R(t) = f[s(t-\tau)]. \tag{31}$$

B. Finite generation-time models for the constant population size

Consider now the modification of Eq. (5):

$$\frac{dp_l}{dt} = p_l(t-\tau)r_l - \mu p_l(t) + \mu\frac{N-l+1}{N}p_{l-1}(t) + \mu\frac{(l+1)}{N}p_{l+1}(t) - p_l(t)\sum_n r_n p_n(t-\tau). \tag{32}$$

Choosing $\Sigma_l p_l(t\le 0)=1$, we have t $\Sigma_l p_l(t)=1$ for any $t>1$. While Eq. (32) gives the same steady state and, therefore, the same mean fitness as the Crow-Kimura model with $\tau=0$, it has completely different dynamics, see Fig. 1. The condition by Eq. (9) is broken. Thus for the nonzero generation time, the growing population and fixed population describe different systems with different evolution characteristics.

C. Different generation times for different strains

In previous sections we considered the model with fixed generation time. Consider now the case when the generation

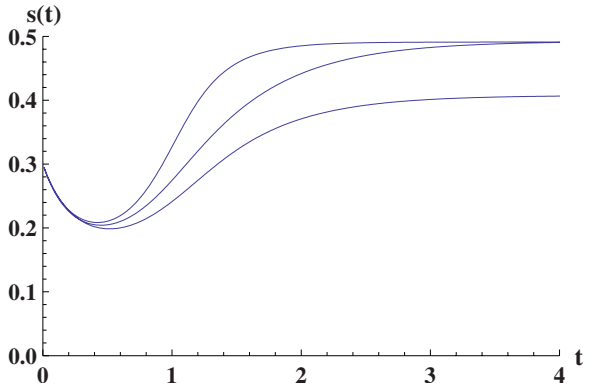

FIG. 1. (Color online) The dynamics of $s(t)=[\Sigma_l(1-2l/L)p_l]/[\Sigma_l p_l]$ for $x_0=0.3$, $f(x)=x^2$. The upper line is for the standard Crow-Kimura model by the Eq. (10), the middle line is for the nonzero generation time model with constant population size by Eq. (32) with $\tau=1$, the lower line is for the growing population model by Eq. (17) with $\tau=1$.

time depends on the sequence (Hamming class). There is a distribution of τ via a function $\rho(x,\tau)$, therefore there is a constraint $\int \rho(x,\tau)d\tau=1$. Equation (17) is modified,

$$\frac{dp_l(t)}{dt} = \int d\tau \rho\left(1-\frac{2l}{N},\tau\right)p_l(t-\tau)f\left(1-\frac{2l}{N}\right) - \mu p_l(t) + \mu\frac{(N-l+1)}{N}p_{l-1}(t) + \mu\frac{(l+1)}{N}p_{l+1}(t), \tag{33}$$

and we derive instead of Eq. (20)

$$q = H(x,q,p),$$

$$H(x,q,p) = f(x)\int d\tau \rho(x,\tau)\exp[-\tau q] + h_0(x,p),$$

$$h_0(x,p) = \mu\frac{1+x}{2}e^{2p} + \mu\frac{1-x}{2}e^{-2p} - \mu. \tag{34}$$

Let us consider the asymptotic solution, using Eqs. (23)–(25). Now Eq. (27) is modified to

$$U(x) - \int \rho(x,\tau)d\tau e^{-\tau U(x)}f(x) = \mu\sqrt{1-x^2} - \mu. \tag{35}$$

We can use the last equation to define the mean fitness. Actually the evolution potential defines also the qualitative features of the dynamics. If at any Hamming class there is a fixed generation time, i.e., $\tau=\tau(x)$, we have the following equation for the evolution potential

$$U(x) - e^{-\tau U(x)}f(x) = \mu\sqrt{1-x^2} - \mu, \tag{36}$$

then we can analyze the situation analytically, considering the maximum of Eq. (36) as a function of x.

The models, similar to Eq. (33) without mutation terms, have been considered in evolution research [5,27], mainly with host-parasite interactions. We can solve the case exactly because large genome length limit.

We derived equations for the mean fitness. The HJE approach allows to construct the steady-state solution as well [20]. We should just put $R=q$ in Eq. (20) and solve it as ODE:

$$R = H(x,R,du/dx). \tag{37}$$

For the steady-state analysis in structured population models, see [28].

IV. CONCLUSION

The evolution of structured populations [4,5] is a rather large and active topics in evolution research with the main focus on a demographic structure of (sexual) population and the matrix method [3] as the main tool. The case of many mutations accompanied with selection was not investigated yet, while the mutation-selection with two alleles has been considered in [4] for diploid evolution case.

In this paper we focused on viruses and investigated the generation time effect in virus evolution, considering a minimal generalization of existing virus evolution models to incorporate the generation time effect. For us it is important to define the mean fitness and distribution of population via genotypes. Because the phenomenon is a rather complicated, it is better to work with exact solutions. The fitness, one of central subjects of evolution research, is a complex parameter, with many traits contributing to the final value. Some of them are the replicative ability, the capacity of the virus to get out of the cell, the genomic robustness, etc. All these traits can be affected by different types of delays. For example, there can be a time interval between the production of the offspring at a generation and the infection of new cells to produce the next generation viruses. Is it possible to use in evolution models an effective fitness as have been assumed in the vast majority of papers about evolution, or we have to use several parameters to describe adequately the evolution process? In this paper we suggested delay differential equations as evolution models with both replication parameters and finite generation time and solved these models exactly (the delay differential equation models has been considered in evolution research before in [27]). When there are many offsprings during one life cycle of virus, we cannot describe the evolution solely by fitness, we have to introduce the generation time as an additional parameter as well. We should use model with finite generation time for nonstationary evolution phenomena as well time, and consequently we should use two different definitions of fitness: $\hat{r}$ in Eq. (12) for growing virus populations and r in Eq. (12) for the constant population size. Thus viruses with different sequences compete in growing populations via $\hat{r}$ and in constant size populations (similar to in vivo case) via r. While the finite generation time has no influence on evolution characteristics for the constant population size and steady-state distribution, see Fig. 1, (the upper and middle lines coincide after long period of relaxation) the finite generation time drastically changes the evolution characteristics in case of growing populations or out of equilibrium dynamics. The delay model holds an

information about finite generation-time period, and more adequately describes the real virus evolution of growing populations than the simple evolution models [6,12,13]. We checked that the effect of delay is proportional to the number of offsprings in one life cycle of virus. We calculated the mean fitness Eqs. (25), (27), and (28). There have been assumed two different mechanisms regarding to the virus extinction: via error threshold [29] and via decrease in a mean fitness to some minimal value [30]. We see that for the growing populations, prolonging the virus generation time period could suppress the viral growth even more than the pushing of the virus population to the error threshold point due to increase in mutation rate [29].

Another interesting extension of our approach could be an application of delay models to phylogeny, where the role of finite generation-time effect is recognized [8].

ACKNOWLEDGMENTS

D.B.S. thanks J. Felenstein and E. Lazaro for discussion. This work was supported by Volkswagenstiftung grant "Quantum Thermodynamics," NSC Grant Nos. 96-2911-M 001-003-MY3 and 98-2811-M-001-080, Academia Sinica and National Center for Theoretical Sciences at National Taiwan University.

[1] J. B. S. Haldane, Proc. Cambridge Philos. Soc. **23**, 838 (1927).
[2] H. T. J. Norton, Proc. London Math. Soc. **28**, 1 (1928).
[3] P. H. Leslie, Biometrika **33**, 183 (1945) on the use of matrices in certain population mathematics.
[4] B. Charlesworth, *Evolution in Age-Structured Populations*, Cambridge Studies in Mathematical Biology (Cambridge University Press, Cambridge, 1994).
[5] H. Caswell, *Matrix Population Models* (Sinauer Associates, Sunderland, 2001).
[6] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper Row, New York, 1970).
[7] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997).
[8] J. H. Gillespie, *The Causes of Molecular Evolution* (Oxford University Press, Oxford, 1991).
[9] G. Bocharov, N. J. Ford, J. Edwards, T. Breinig, S. Wain-Hobson, and A. Meyerhans, J. Gen. Virol. **86**, 3109 (2005).
[10] S. Wright, in *Proceedings of the Sixth International Congress on Genetics*, edited by D. F. Jones (Genetics Society of America, Austin, 1932), Vol. 1, p. 356.
[11] P. A. P. Moran, *The Statistical Processes of Evolutionary Theory* (Clarendon, Oxford, 1962).
[12] M. Eigen, Naturwiss. **58**, 465 (1971).
[13] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[14] D. B. Saakian and C.-K. Hu, Proc. Natl. Acad. Sci. U.S.A. **103**, 4935 (2006); D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 021913 (2004); D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 046121 (2004).
[15] D. B. Saakian, E. Munoz, C.-K. Hu, and M. W. Deem, Phys. Rev. E **73**, 041913 (2006).
[16] C. K. Biebricher and M. Eigen, Curr. Top. Microbiol. Immunol. **299**, 1 (2006).
[17] A. Wettich and C. K. Biebricher, Biochemistry **40**, 3308 (2001).
[18] A. F. M. Marée, W. Keulen, C. A. B. Boucher, and R. J. De Boer, J. Virol. **74**, 11067 (2000).
[19] S. Bonhoeffer, A. D. Barbour, and R. J. De Boer, Proc. R. Soc. London, Ser. B **269**, 1887 (2002).
[20] D. B. Saakian, J. Stat. Phys. **128**, 781 (2007).
[21] K. Sato and K. Kaneko, Phys. Rev. E **75**, 061909 (2007).
[22] D. B. Saakian, O. Rozanova, and A. Akmetzhanov, Phys. Rev. E **78**, 041908 (2008); D. B. Saakian, Z. Kirakosyan, and C.-K. Hu, *ibid.* **77**, 061907 (2008).
[23] C. J. Thompson and J. L. McBride, Math. Biosci. **21**, 127 (1974).
[24] B. L. Jones, R. H. Enns, and S. S. Rangnekar, Bull. Math. Biol. **38**, 15 (1976).
[25] R. Lande, Ecology **63**, 607 (1982).
[26] J. A. J. Metz, R. M. Nisbet, and S. A. H. Geritz, Trends Ecol. Evol. **7**, 198 (1992).
[27] W. S. C. Gurney, R. M. Nisbet, and J. H. Lawton, J. Anim. Ecol. **52**, 479 (1983).
[28] O. Diekmann, M. Gyllenberg, and J. A. J. Metz, Theor. Popul. Biol. **63**, 309 (2003).
[29] M. Eigen, Proc. Natl. Acad. Sci. U.S.A. **99**, 13374 (2002).
[30] J. J. Bull, R. Sanjuan, and C. O. Wilke, J. Virol. **81**, 2930 (2007).