PHYSICAL REVIEW E **81**, 051916 (2010)

Recombination in one- and two-dimensional fitness landscapes

Zh. Avetisyan[1,*] and David B. Saakian[2,3,4]
[1]*Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, Leipzig 04103, Germany*
[2]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[3]*Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia*
[4]*Physics Division, National Center for Theoretical Sciences, National Taiwan University, Taipei 10617, Taiwan*

(Received 23 November 2009; published 17 May 2010)

We consider many-site mutation-recombination models of molecular evolution, where fitness is a function of a Hamming distance from one (one-dimensional case) or two (two-dimensional case) sequences. For the one-dimensional case, we calculate the population distribution dynamics for a model with zero fitness and an arbitrary symmetric initial distribution and find an error threshold transition point in the single-peak fitness model for a given initial symmetric distribution. We calculate the recombination period in the case of a single-peak fitness function, when the original population is located at one sequence, at some Hamming distance from the peak configuration. Steady-state fitness is calculated with finite genome length corrections. We derive analytical equations for the two-dimensional mutation-recombination model.

DOI: 10.1103/PhysRevE.81.051916 PACS number(s): 87.10.−e, 87.15.A−, 87.23.Kg

I. INTRODUCTION

Recombination is one of the key factors in evolution and is assumed to be one of the possible advantages of sex [1]. Therefore, the investigation of recombination is one of the central directions of evolution research (see recent reviews [2,3]). It has been found that recombination increases mean fitness, providing advantage in the case of negative epistasis [i.e., fitness function $f(m)$ in Eq. (4) is convex near $m=1$] [4]. Mathematical aspects of these problems have been analyzed in [5,6]. In [7], a theorem of existence of a solution to the continuous-time recombination problem has been proven and a general construction has been suggested for obtaining the solution. In [8], the existence of exact recombination dynamics in discrete time models has been questioned. In this paper, we derive several exact results for continuous-time recombination and among them the first analytical explicit solution of selection-free recombination dynamics.

During the recombination process, an exchange of parts between two genomes occurs and many different recombining allele may be involved at several loci. Models may consider both discrete time and continuous time. What is common to all recombination models in the case of continuous time is that they are described using strongly nonlinear ordinary differential equations (ODEs). It is known that related models of molecular evolution by Crow and Kimura [9,10] and Eigen [11,12] may be mapped to a system of linear ODE [13,14]. There, it is possible to find exact dynamics for an infinite population within a large-genome limit [15]. However, such simplification is not possible in the case of recombination. For this reason, the solution of recombination dynamics is a highly involved problem. While several books and many articles with thorough mathematical investigation have been devoted to recombination in infinite population case, there are no known explicit analytical results for many-allele dynamics; neither in the selection-free model nor in the simple cases of selection. In [7], it has been proven that recombination models with zero selection and linear-fitness function are exactly solvable. Here, we provide a solution to the evolution dynamics for a large-genome length, haploid, many-loci recombination model in a selection-free case for a single-peak fitness landscape. Such a fitness landscape is rather popular in quasispecies model research and is sometimes applied in population genetics. It could be considered as a first approximation of a fitness landscape with a high peak at one sequence. In order to verify our analytical results, we construct the recombination model in a two-dimensional case and complete the investigation of the single-peak fitness model by calculating the mean fitness.

We are interested in the recombination dynamics of many-loci, haploid model, with two alleles at any locus. Such models are, for example, considered to be realistic for human immunodeficiency virus (HIV) [16]. Evolution models of this kind are characterized by fitness landscapes and mutation schemes. Usually, biologists consider phenomenological models of recombination, assuming simplified mutation schemes between different genome configurations. More interesting and realistic are microscopic evolution models, where the genome can be considered as a chain of letters (alleles) and mutation is a process involving only a change of letters. Such a microscopic model with discrete dynamics and also with processes of mutation and recombination has been considered in Ref. [16]. There, during each recombination event, two genomes undergo a crossover. In other words, two points are chosen along the length of each genome and the two genomes exchange their contents between these pairs of points. Thus, as a rule, large pieces of genomes are exchanged. Such a recombination process distinguishes distance along the genome (i.e., if two points are neighbors or not), making such models analytically intractable.

Recently, multiallele models have been introduced and mean fitness has been calculated in a simpler (as compared to [16]) model of recombination. In the horizontal gene transfer (HGT) [17,18] model, there is an exchange by a single allele and, therefore, recombination is a mean-field-like process

---

[*] A part of bachelor project performed at Yerevan State University, Alex Manoogian 1, Yerevan 375025, Armenia.

(i.e., there is no concept of distance between points along the genome). Thereby, the model of [17,18] can be solved analytically: mean fitness [18] and a steady-state distribution [19] have been obtained. Unfortunately, it is impossible to find exact dynamics (in the large-genome limit) for a general symmetric fitness case (indeed, fitness is a function of the Hamming distance from one reference sequence), unlike what has been done in [15] for the Crow-Kimura and Eigen models. Here, we use the Hamilton-Jacobi equation (HJE) approach to evolution models [20,21] to solve exactly the dynamics of the recombination model with single-peak fitness. We follow the methodology of our previous paper [22].

In the model [17,18], the genome is a chain of N alleles (i.e., spins taking values ± 1). Thus, there are 2^N sequences, distributed among the Hamming classes $0 \le k \le N$ [23]. 0th class corresponds to the reference sequence and the number of -1 spins in the lth class is equal to l. We consider a symmetric distribution of population among the Hamming classes. The fitness depends on the Hamming class and all the sequences of the lth Hamming class have the same fitness

$$r_l = Nf(m), \quad m_l \equiv 1 - 2l/N, \tag{1}$$

where m_l is the mean magnetization of spins in the lth class. There is a mutation process with a rate 1 and recombination with a rate c. During the recombination event, one allele of the given sequence is replaced (with the rate c) by the allele at the same position of the genome chosen randomly from the population. This is the HGT model. The HGT model is equivalent to the recombination model, where two sequences exchange one allele with the rate $c/2$. One has the following system for P_l (total probabilities of sequences in the lth Hamming class):

$$\begin{aligned}\frac{dP_l}{dt} &= P_l r_l + [(N-l+1)P_{l-1} + (l+1)P_{l+1}] \\ &\quad - P_l N\left(1 + \frac{1}{N}\sum_k r_k P_k\right) - cNP_l \\ &\quad + cN\left[\left(1 - \frac{\bar{l}}{N}\right)\left(1 - \frac{l}{N}\right) + \frac{\bar{l}}{N}\frac{l}{N}\right)\right]P_l \\ &\quad + cN\left[\left(1 - \frac{\bar{l}}{N}\right)\frac{l+1}{N}P_{l+1} + \frac{\bar{l}}{N}\left(1 - \frac{l-1}{N}\right)P_{l-1}\right],\end{aligned} \tag{2}$$

where $\bar{l} = \Sigma_{l=0}^{N} l P_l$. The first two lines of Eq. (2) correspond to the Crow-Kimura model with two types of alleles at any loci (see Ref. [15] for the detailed description). At the instance of time dt, any allele could change to the opposite one with a probability dt. The term $-P_l \Sigma_k r_k P_k$ supports a balance condition: following the population growth, there is a uniform dilution retaining a constant population size. The following lines in Eq. (2), describe the recombination process. During the period of time dt, for a given sequence, we replace any allele in the genome with the probability cdt by the allele at the same position of other sequences, chosen randomly from the population (see [18]). Therefore, there is a term $-cP_l$ in the third line of Eq. (2). The remaining terms in the third line correspond to the replacement of an allele with the allele of the same type. The last term corresponds to the replacement of an allele with the allele of another type. The model defined with Eq. (2) describes the case of general mutation rate μ per allele as well, after mapping $c \to c/\mu$, $r_l \to r_l/\mu$, and $\mu t \to t$.

It is important to note that mutation and recombination events are independent random processes; i.e., there is no correlation between mutation-recombination events in different parts of genome or between neighbor points. Equation (2) is a nonlinear system of ODE. In the Crow-Kimura model, it is possible to avoid the balance condition, find the solution of a linear system for the relative probabilities P_l, and then rescale them to have a constraint $\Sigma_l P_l = 1$. In our case of the model with recombination, this trick is impossible.

Considering an ansatz

$$P_l = \exp[Nu(m,t)], m \equiv 1 - 2l/N, \tag{3}$$

we obtain with $1/N$ accuracy an equation for the bulk expression of $u(m,t)$,

$$\frac{\partial u}{\partial t} + H(m,u',s) = 0,$$

$$\begin{aligned}-H(m,p,s) &\equiv f(m) + \frac{1+m}{2}e^{2p} + \frac{1-m}{2}e^{-2p} - 1 + \frac{cms}{2} - \frac{c}{2} \\ &\quad + c\left[e^{2p}\frac{1+m}{2}\frac{1-s}{2} + e^{-2p}\frac{1-m}{2}\frac{1+s}{2}\right],\end{aligned} \tag{4}$$

where $p = u'$ is the derivative via m, $s = \Sigma_l P_l (1 - 2l/N)$ is the surplus (surface magnetization [10,18]) of the distribution, and $Nf(m) \equiv r_k$, $m = 1 - 2k/N$. $u(m,t)$ has a maximum at point $m = s$ at the moment of time t. Differentiating Eq. (4) via m at point $m = s(t)$, we derive

$$u''_{mt}(s,t) = f'(s(t)) + 2su''_{mm}(s(t),t), \tag{5}$$

where $u_{mm} \equiv \frac{\partial^2 u}{\partial m^2}$, $u_{mt} \equiv \frac{\partial^2 u}{\partial m \partial t}$ and in Eq. (5) we consider the derivatives at point $m = s(t)$. We use the fact that $u'_m[s(t),t] = 0$. On the other hand, near $m = s(t)$, one can use $u(m,t) \approx u''_{mm}[s(t),t][m - s(t)]^2/2$; therefore, $u''_{mt}[s(t),t] = -u''_{mm}[s(t),t]\frac{ds(t)}{dt}$. Hence, we obtain

$$\frac{\partial s(t)}{\partial t} = -2s(t) - \frac{f'[s(t)]}{u''_{mm}[s(t),t]}. \tag{6}$$

The plan of the paper is as follows. In Sec. II, we solve the case without selection, i.e. $f(m) = 0$. We calculate the probability of having $N(1-m)$ mutations, when originally the population is grouped at some reference (master) sequence. Different sequences are specified according to their Hamming distance from the reference sequence. In Sec. III, we calculate the error threshold in the case when the population originally is located at some sequence with some Hamming distance from the peak sequence (all sequences besides the peak one have the same, smaller fitness). In Sec. IV, we solve the static case of the model. We calculate the mean fitness including the finite genome length corrections. In Sec. V, we derive the system of equations for the two-dimensional recombination model. In Appendixes A and B, we derive the population dynamics for the case without recombination and compare it to the result derived by an alternative method.

II. SELECTION-FREE EVOLUTION WITH MUTATIONS AND RECOMBINATION

Equation (6) gives in the case $f(m)=0$,

$$\frac{ds}{dt} = -2s(t). \tag{7}$$

Thus the overlap of the maximum point decreases exponentially

$$s(t) = Ae^{-2t}. \tag{8}$$

Now, $s(t)$ is just a given function and we get the Hamilton-Jacobi equation [27,28] with the Hamilton function explicitly dependent of t,

$$q = -H \equiv \left(1+\frac{c}{2}\right)[\cosh(2p) - 1 + x\sinh(2p)] - \frac{s(t)c}{2}\{\sinh(2p) + x[\cosh(2p)-1]\},$$

$$p[s(t),t] = 0, p = \frac{\partial u}{\partial x}, q = \frac{\partial u}{\partial t}. \tag{9}$$

We solve the equation using the standard method of characteristics for the initial distribution $u(x,0)=u_0(x)$, considering the corresponding Hamilton equations

$$\frac{dx}{dt} = \frac{\partial H}{\partial p} = -(2+c)[\sinh(2p) + \cosh(2p)x] + s(t)c[\cosh(2p) + \sinh(2p)x],$$

$$\frac{dp}{dt} = -\frac{\partial H}{\partial x} = \left(1+\frac{c}{2}\right)\sinh(2p) - \frac{s(t)c}{2}[\cosh(2p)-1]. \tag{10}$$

By solving this system, we can find the solutions $p(t;x^0,p^0)$ and $x(t;x^0,p^0)$, where x^0 and p^0 are, respectively, the initial coordinate and the initial momentum of the characteristic. Given the pair (x,t), we find the corresponding characteristic parameters x^0 and p^0 from the equation $x=x(t;x^0,p^0)$. Thereafter, we can obtain $u(x,t)$ by integrating the formula $du = pdx - Hdt$. The solution $u(x,t)$ is given in the form of an implicit functional system (see Appendix A).

First, we solve Eq. (10) which contains only p and t and can be solved independently

$$\coth(p) = s(t)(1-Be^{-ct}) = Ae^{-2t}(1-Be^{-ct}), \tag{11}$$

where $A \equiv s(0)$ and B are parameters characterizing the solution.

Then we find the solution for $x(t)$,

$$x(t) = x^0 e^{F_2(t)} + e^{F_2(t)}\int_0^t f_1(\tau)e^{-F_2(\tau)}d\tau, \tag{12}$$

where

$$-f_1(t) = Ae^{-2t}\frac{2(2+c)(1-Be^{-ct}) - c[A^2e^{-4t}(1-Be^{-ct})^2+1]}{A^2e^{-4t}(1-Be^{-ct})^2-1},$$

$$-f_2(t) = \frac{(2+c)[A^2e^{-4t}(1-Be^{-ct})^2+1] - 2cA^2e^{-4t}(1-Be^{-ct})}{A^2e^{-4t}(1-Be^{-ct})^2-1},$$

$$F_2(t) = \int_0^t f_2(\tau)d\tau. \tag{13}$$

Finally, we get $u(t,x)$,

$$u(t,x) - u_0(x^0) = \int_0^t \left[p(\tau)\frac{dx(\tau)}{dt} - H(\tau)\right]d\tau = \int_0^t \left\{\operatorname{arccoth}[Ae^{-2\tau}(1-Be^{-c\tau})][f_1(\tau) + x(\tau)f_2(\tau)] - \left(1+\frac{c}{2}\right) + \frac{c * s(\tau)x(\tau)}{2} - \frac{1}{2}[f_2(\tau) + x(\tau)f_1(\tau)]\right\}d\tau. \tag{14}$$

The integral is calculated along the characteristics and x_0 is calculated from Eq. (11). In Fig. 1, we give the comparison of the direct numerics of Eq. (2) to our analytical results.

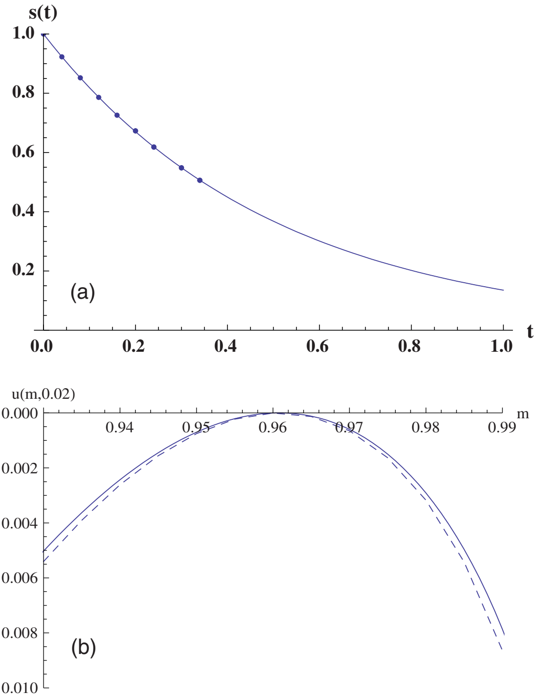

FIG. 1. (Color online) Numerical results for the free model with zero fitness $r_i=0$. (a) Numerical results (points) for $s(t)$ vs the $s(t) = \exp[-2t]$ (smooth line) for infinite population HGT model at $N = 100$. (b) Comparison of $u(m,t)$ [see Eq. (3)] from the numerical evaluation of Eq. (2) to the analytical results by HJE (dashed line) at $c=1$, $t=0.02$ for the region $0.93 \leq m \leq 0.99$

In biological applications, an important case corresponds with the population initially located at the sequence $m=1$ and thereby $A=1$. In this case, $\exp[Nu(m,t)]$ gives the probability having $N(1-m)/2$ mutations. When $A=1$, $c=0$, a solution has been derived via alternative methods [22]

$$u(x,t)=\frac{1+x}{2}\ln\cosh(t)+\frac{1-x}{2}\ln\sinh(t)-t-\frac{1+x}{2}\ln\frac{1+x}{2}-\frac{1-x}{2}\ln\frac{1-x}{2}. \tag{15}$$

In Appendixes A and B, we derive this solution using the HJE approach to verify our HJE method.

III. EVOLUTION IN THE CASE OF A SINGLE-PEAK FITNESS

A. Original population located at one sequence

In the previous section, we calculated the dynamics for a given symmetric distribution: originally, the distribution is symmetric around some reference sequence and the symmetry is conserved during the dynamics. If the fitness is a function of the Hamming distance from some other sequence, then it corresponds to the dynamics in two-dimensional (overlap) space. Such a dynamical problem cannot be solved analytically in the case of a general fitness landscape. However, in the case when the fitness landscape has a single-peak form at some other reference sequence, our analytical results of the previous section are sufficient to solve this two-dimensional problem exactly (Sec. V). This point will be proven via direct numerical evaluation of the two-dimensional evolution model.

First, we consider the evolution dynamics, when at $t=0$ the population is located at some configuration with overlap m with the master sequence. Such a situation is not a theoretical abstraction, but occurs in virological experiments [25]. An alternative version of the initial distribution could also be realized in experiment: a homogenous distribution around some reference sequence with the fixed overlap m [26]. Then, we will calculate the relaxation period of the population to move to the master sequence. The key idea of the solution is borrowed from Ref. [22]. We consider a free diffusion from the original sequence to the master sequence. Then we are looking at the growth of the population at the master sequence. Such an approach is correct with $1/N$ accuracy.

Consider the dynamics for this case. At $t=0$, we have $x(0)=A=1$. At the point $x=m$, there is a high-peak sequence with a fitness J; all other sequences have zero fitness. The probability of a given sequence is

$$\exp\left[Nu(m,t)+N\frac{1+m}{2}\ln\frac{1+m}{2}+N\frac{1-m}{2}\ln\frac{1-m}{2}\right]. \tag{16}$$

Equation (16) provides the valid dynamics for $t<t_0$. We calculate the probability of the master peak using the approach of Ref. [22]. After the point $t>t_0$, we consider the growth of the master sequence. Ignoring the nondiagonal and nonlinear terms (i.e., the time period when P_0 grows from 0.01 to 1 is $\sim 1/N$), for P_0 in Eq. (2) we have

$$\frac{dP_0(t)}{dt}=P_0(t)[(J-1)-c[1-s(t)]/2\}]. \tag{17}$$

Thus,

$$P_0=\exp\left[N\left\{\frac{1+m}{2}\ln\frac{1+m}{2}+\frac{1-m}{2}\ln\frac{1-m}{2}+u(m,t_0)+(J-1)(T-t_0)-c\int_{t_0}^{T}[1-s(\tau)]\frac{d\tau}{2}\right\}\right]. \tag{18}$$

Then, we calculate t_0 from the condition of maximum in Eq. (18). The relaxation period T is defined as the a point where $P_0(T)\sim 1$. Thus, we have a system of equations

$$\frac{\partial u(m,t_0)}{\partial t}-J+1+c[1-s(t_0)]/2=0,$$

$$\frac{1+m}{2}\ln\frac{1+m}{2}+\frac{1-m}{2}\ln\frac{1-m}{2}+u(m,t_0)+(J-1)(T-t_0)-c\int_{t_0}^{T}[1-s(\tau)]d\tau/2=0. \tag{19}$$

For $m\approx 0.797$, $J=5.196$, $c=1$, and $t_0=0.02$, we derive $T=0.123$ and $B=-0.5$ (see Appendix A).

The next task is to numerically check these results for the relaxation process, when the maximal fitness is in one sequence, and originally the population is located at another sequence. This corresponds with a two-dimensional evolution process (one dimension is the Hamming distance from the peak configuration; the second dimension is the Hamming distance from the original sequence). In the case of zero recombination, the situation is simpler. The selection-free dynamics is described by linear equations. If we are interested in calculating just the probability of $P_0(t)$ in an selection-free case, then, instead of putting the whole original population at one sequence at the Hamming distance $N(1-m)/2$, we can uniformly distribute it (at zero time) among all sequences at the Hamming distance $N(1-m)/2$. The key point is that such a symmetric distribution could be described by a one-dimensional evolution model, when originally only the probability of the lth Hamming class P_l, $l=N(1-m)/2$ is nonzero. In the case of nonzero recombination, the selection-free dynamics is nonlinear and we cannot use this trick [the population is uniformly distributed over all the sequences at the Hamming distance $N(1-m)/2$ instead of being located at one sequence with Hamming distance $N(1-m)/2$]. Therefore, we need in a two-dimensional evolution model to check our analytical results, Eq. (19). We will formulate the two-dimensional evolution model with recombination below, in Sec. V. The direct numerical evaluation of the two-dimensional model gives $T\approx 0.122$ for $N=1000$, which is consistent with the result found by the analytical formula Eq. (19).

B. Threshold values of J for the given $s(0)$

The multistability phenomenon is well established in recombination models [24]. In Ref. [18], an interesting fact has been identified: the error threshold depends on the initial distribution. Consider the fitness landscape $r_0=J$, $r_i=0$, $i>0$. According to [18], when $s(0)=1$, the error threshold is $J>1$, as was found first reported in Ref. [22]. For $s(0)=0$, new error threshold $J>1+c/2$ has been provided in Ref. [18].

Let us find the threshold value of J, when the population originally is located at some sequence at any Hamming distance from the peak sequence. We should look for the maximum of the left-hand side of second equation in Eq. (19) via T,

$$\frac{\partial u(m,t_0)}{\partial t} - J + 1 + c[1 - s(t_0)]/2 = 0,$$

$$u(m,t_0) + \frac{1+m}{2}\ln\frac{1+m}{2} + \frac{1-m}{2}\ln\frac{1-m}{2} + (J-1)(T-t_0) - c\int_{t_0}^{T}[1-s(\tau)]d\tau/2 = 0,$$

$$J - 1 - c/2 + cs(T)/2 = 0. \tag{20}$$

IV. STEADY-STATE SOLUTION FOR SINGLE-PEAK FITNESS

To complete the investigation of the single-peak fitness model, let us calculate mean fitness with finite genome corrections. Without horizontal gene transfer, the steady-state distribution is [22] $P_0=(J-1)/J$, $P_l=P_0(1-P_0)^l=P_0/J^l$. Equation (2) gives for the steady state

$$P_l\left[JP_0 + 1 - \frac{r_l}{n}\right] = -cP_l\left[\left(1-\frac{\bar{l}}{N}\right)\frac{l}{N} + \left(1-\frac{l}{N}\right)\frac{\bar{l}}{N}\right] + \frac{N-l+1}{N}P_{l-1}\left[1 + c\frac{\bar{l}}{N}\right] + \frac{l+1}{N}P_{l+1}\left[1 + c\left(1-\frac{\bar{l}}{N}\right)\right]. \tag{21}$$

For the P_0, we have

$$P_0(Jp_0 + \mu - A) = \frac{1}{N}P_1\left[1 + c\left(1 - \frac{\bar{l}}{N}\right)\right] - cP_0\frac{\bar{l}}{N}. \tag{22}$$

Replacing on the right-hand side and before the parenthesis on the left-hand side, P_0 and P_1 with their bulk expressions, $P_0 \approx (1-1/J)$ and $P_1 \approx (J-1)/J^2$, and taking

$$\bar{l} \approx 1/(J-1), \tag{23}$$

we find a second-order accuracy expression for the P_0 in the presence of horizontal gene transfer

$$P_0 = \frac{(J-1)}{J} + \frac{1}{NJ^2}\left(1 - \frac{c}{J-1}\right) + O(1/N^2). \tag{24}$$

Therefore, in the infinite-genome limit, the horizontal gene transfer does not change the mean fitness for the single-peak fitness landscape, as has been first found in Ref. [18]. We see that the recombination does not change the master type probability in the infinite-genome limit, while there are $O(1/N)$ corrections. This is a rather large correction for the short variable parts of genome [16,26], where biologists consider rather small effects of epistasis [29]. The finite-size corrections are negative; thus, the mean fitness decreases due to recombination, as has been first derived numerically in [31].

V. TWO-DIMENSIONAL RECOMBINATION MODEL

In [30], a two-dimensional model has been solved with mutation and selection. There are two reference sequences s_l^1, s_l^2 with the overlap $L \equiv (1+m)/2$ of them. Any sequence is characterized by overlap $L_1 \equiv N\frac{1+M_1}{2}$ with the first sequence and the overlap $L_2 \equiv N\frac{1+M_2}{2}$ with the second sequence. Let us first give a slightly modified derivation of the results of [30] for the two-peak case and then use it for the recombination case.

There are $N(1+m)/2$ identical alleles $(s_l^1=s_l^2)$ of two reference sequences and $N(1-m)/2$ different ones. For simplicity, in the sequence, we put first the identical alleles, then different ones. Let us take arbitrary sequence with spins s_l. We define magnetizations for the first group as m_1 and for the second group as m_2,

$$\sum_{l=1}^{L} s_l s_l^1 = N\frac{1+m}{2}m_1 = L - 2l_1,$$

$$\sum_{L+1}^{N} s_l s_l^2 = N\frac{1-m}{2}m_2 = N - L - 2l_2. \tag{25}$$

A simple calculation gives the following connection between these local magnetizations m_1, m_2, and the overlaps M_1, M_2 with the reference sequences:

$$M_1 = \frac{1+m}{2}m_1 + \frac{1-m}{2}m_2,$$

$$M_2 = \frac{1+m}{2}m_1 - \frac{1-m}{2}m_2. \tag{26}$$

We consider an original distribution, where the sequence probability depends on l_1, l_2. If the fitness depends also on l_1, l_2, then all the sequences with the same l_1, l_2 have identical probabilities during the dynamics. There are constraints $0 \le l_1 \le L$, $0 \le l_2 \le (N-L)$. To find the solution of two-peak evolution model, one should first calculate the fractions $(x_{\alpha,i}$, $\alpha=\pm 1$, $1 \le i \le 2)$ of N alleles with the given alignment along the first and second sequences (see [30])

$$x_{\alpha,1}=\frac{(1+\alpha m_1)(1+m)}{4};x_{\alpha,2}=\frac{(1+\alpha m_2)(1-m)}{4}. \tag{27}$$

Now we have the following system of equations for the probabilities of sequence classes with the same l_1, l_2:

$$\begin{aligned}\frac{dp(l_1,l_2)}{Ndt}&=\left(\frac{1}{N}r_{l_1,l_2}-R\right)p(l_1,l_2)-p(l_1,l_2)\\&\quad+\sum_{\alpha}x_{\alpha,1}(l_1+\alpha,l_2)p(l_1+\alpha,l_2)\\&\quad+x_{\alpha,2}(l_1,l_2+\alpha)p(l_1,l_2+\alpha)+c\sum_{\alpha}(x_{\alpha 1}\hat{x}_{\alpha,1}\\&\quad+x_{\alpha,2}\hat{x}_{\alpha,2}-1)p_{l_1,l_2}+c\Bigg[\sum_{\alpha}x_{\alpha,1}(l_1+\alpha,l_2)p(l_1\\&\quad+\alpha,l_2)\hat{x}_{-\alpha,1}+x_{\alpha,2}(l_1,l_2+\alpha)\hat{x}_{-\alpha,2}p(l_1,l_2+\alpha)\Bigg],\end{aligned} \tag{28}$$

where $r(l_1,l_2)$ is the fitness of configuration and NR is the mean fitness. The term $-p(l_l,l_2)$ is due to mutation with a rate 1 per allele, the third and fourth terms describe the mutation, when one allele is changed either in a first group of alleles (third term) or in the second group of alleles (fourth term). They are similar to the first line in Eq. (2): in Eq. (2) the relative number of positive alleles is l/N and for the negative alleles is $1-l/N$. The terms proportional to c correspond to the recombination. The second term in the third line corresponds to the third line in Eq. (2): the recombination terms which do not change the allele types. The fourth line is similar to the fourth line in Eq. (2): the change of allele types. We denote

$$\hat{x}_{\alpha,i}=\sum_{l_1,l_2}p(l_1,l_2)x_{\alpha,i}(l_1,l_2). \tag{29}$$

We used Eq. (28) to calculate relaxation dynamics for the single-peak fitness case and the numerical verification confirmed our analytical result of Eq. (19).

It is possible to write (nonlocal) Hamilton-Jacobi equation, similar to Eq. (4). In [19], we derived the mean fitness from the Hamilton-Jacobi equation, having a system of two equations for m, s: one is the maximum condition for the mean fitness $R=\{min[-H(m,p,s)]|_p\}|_m$ and the second one is a condition $R=f(s)$. Now, following to [19], we get only three equations for the four variables m_1, m_2, s_1, s_2. Therefore, we cannot define the mean fitness following only to the ideas of [19]. One needs a more advanced method to calculate the mean fitness.

VI. CONCLUSION

The solution of recombination dynamics in the case of many-allele models is a very hard task, and even the selection-free case has not been solved analytically to date. We solved the dynamics of the recombination model for the selection-free case and for the single-peak fitness landscape. Our analytical results are confirmed by direct numerical calculation. While the dynamics of the maximum of the distribution has a simple analytical expression, we need implicit functions to calculate the population profile. The results are exact at the large-genome limit. Our results Eqs. (12)–(14) could be applied in population genetics to calculate the probability of having $N(1-m)/2$ mutations after the period of time t. We also found how the error threshold depends on the initial distribution, generalizing the corresponding result of [18], and calculated finite-genome corrections to the mean fitness. These corrections are negative, as has been found before numerically in [31]. These finite-size corrections could play an important role in the case of short-genome length, even suppressing the contribution of small epistasis [29].

We derived the system of equations for the recombination in the case of a two-dimensional fitness landscape. The calculation of the mean fitness is a harder task than in the case of the model without recombination and one needs new ideas to calculate it.

We do not see any chances to solve the models for other fitness functions besides the three cases (i.e., selection-free case and single-peak fitness function). We considered only the models in continuous time. In [8], recombination model in discrete time for a selection-free case has been considered; however, no solution has been presented. It would be interesting to apply our HJE approach to this case.

ACKNOWLEDGMENTS

D.B.S. thanks the Volkswagenstiftung grant "Quantum Thermodynamics" and the National Science Council in Taiwan (Grant No. NSC 98-2811-M-001-080) and Academia Sinica.

APPENDIX A: THE DERIVATION OF FREE DYNAMICS

In this appendix, our task is to derive the dynamics for the selection-free case, i.e., the case $f(x)=0$. We refer now to the second line of Eq. (10) which, taking into account Eq. (7), turns into an independent ordinary first-order differential equation. Solving it, we obtain the function $p(t;B)$ along a characteristic. Then similarly, after inserting $p(t;B)$ into the first line of Eq. (10), we find the function $x(t;B)$ along the characteristic. Having this function, one can find the characteristic corresponding to any given point (x,t) using the equation $x(t;B)=x$ and thus obtain the desired B. Further, this B can be used in combination with $du=p*dx-H*dt$ to calculate the value of $u(x,t)$ by integrating along the characteristic. We denote by x^0 and p^0, respectively, the start point at $t=0$ of the characteristic and the derivative $u_x'(x^0,0)$ of the initial distribution $u(x,0)$ at that point.

We start with Eq. (10)

$$\begin{aligned}\frac{dp}{dt}&=\left(1+\frac{c}{2}\right)\sinh(2p)-\frac{s(t)c}{2}[\cosh(2p)-1]\\&=\left(1+\frac{c}{2}\right)\sinh(2p)-\frac{Ace^{-2t}}{2}[\cosh(2p)-1]\\&=\alpha\sinh(2p)-\beta e^{-2t}[\cosh(2p)-1],\end{aligned} \tag{A1}$$

where $\alpha=1+c/2$ and $\beta=Ac/2$ are used for convenience

$$\frac{dp}{dt} = \sinh(2p)\left[\alpha - \beta e^{-2t}\frac{\cosh(2p)-1}{\sinh(2p)}\right] = \sinh(2p)[\alpha - \beta e^{-2t}\tanh(p)]. \tag{A2}$$

Let us do the following substitutions: $v \doteq \tanh(p)$ and $z \doteq \exp(2t)$. We derive

$$\frac{dv}{dz} = \frac{v}{z}\left(\alpha - \beta\frac{v}{z}\right).$$

Now we do the next substitution, $y \doteq \frac{v}{z}$,

$$z = \frac{v}{y}, dz = \frac{dvy - vdy}{y^2},$$

$$\frac{dvy^2}{dvy - vdy} = y(\alpha - \beta y),$$

$$\frac{dv}{v} = \frac{dy(\alpha - \beta y)}{y(\alpha - 1 - \beta y)},$$

$$d\ \ln(v) = d\ \ln\frac{\left(y + \frac{1}{\beta} - \frac{\alpha}{\beta}\right)^{1/1-\alpha}}{y^{\alpha/1-\alpha}},$$

$$\frac{v\left(y - \frac{1}{A}\right)^{2/c}}{y^{1+2/c}} = \text{const},$$

$$z\left(1 - \frac{z}{Av}\right)^{2/c} = \text{const}, e^{ct}\left(1 - \frac{\coth(p)}{s(t)}\right) = \text{const} \doteq B.$$

$$\coth(p) = s(t)(1 - Be^{-ct}) = Ae^{-2t}(1 - Be^{-ct}), \tag{A3}$$

where B is a parameter of the characteristic and depends on $p^0: \coth(p^0) = A(1-B)$. Now we modify Eq. (10)

$$\begin{aligned}\frac{dx}{dt} &= -(2+c)[\sinh(2p) + x\cosh(2p)] + s(t)c[\cosh(2p) \\ &\quad + x\sinh(2p)] \\ &= -(2+c)\sinh(2p) + s(t)c * \cosh(2p) \\ &\quad + x[-(2+c)\cosh(2p) + s(t)c * \sinh(2p)]\frac{dx}{dt} \\ &= f_1(t) + xf_2(t), \\ f_1(t) &\doteq -(2+c)\sinh[2p(t)] + s(t)c * \cosh[2p(t)], \\ f_2(t) &\doteq -(2+c)\cosh[2p(t)] + s(t)c * \sinh[2p(t)].\end{aligned} \tag{A4}$$

Equation (A4) is easily solved. Let us denote

$$F_2(t) = \int_0^t f_2(\tau)d\tau.$$

Then the solution can be presented as

$$x(t) = x^0 e^{F_2(t)} + e^{F_2(t)}\int_0^t f_1(\tau)e^{-F_2(\tau)}d\tau. \tag{A5}$$

Now let us return to the explicit forms of functions $f_1(t)$ and $f_2(t)$,

$$\begin{aligned}-f_1(t) &= (2+c)\sinh[2p(t)] + s(t)c * \cosh[2p(t)] \\ &= \frac{2(2+c)\coth(p) - s(t)c[\coth^2(p)+1]}{\coth^2(p)-1} \\ &= Ae^{-2t}\frac{2(2+c)(1-Be^{-ct}) - c[A^2e^{-4t}(1-Be^{-ct})^2+1]}{A^2e^{-4t}(1-Be^{-ct})^2-1} \\ -f_2(t) &= (2+c)\cosh[2p(t)] + c * s(t)\sinh[2p(t)] \\ &= \frac{(2+c)[\coth^2(p)+1] - 2c * s(t)\coth(p)}{\coth^2(p)-1} \\ &= \frac{(2+c)[A^2e^{-4t}(1-Be^{-ct})^2+1] - 2cA^2e^{-4t}(1-Be^{-ct})}{A^2e^{-4t}(1-Be^{-ct})^2-1}.\end{aligned} \tag{A6}$$

One should keep in mind that the functions f_1 and f_2 depend on the selected characteristic: $f_1(t;B)$ and $f_2(t;B)$.

Now we insert $p(t)$ and $x(t)$ into Eq. (8) to obtain $H(t)$. Let us now simplify the expression for H,

$$\begin{aligned}-H &= \left(1 + \frac{c}{2}\right)[\cosh(2p) - 1 + x(t)\sinh(2p)] - \frac{s(t)c}{2}\{\sinh(2p) \\ &\quad + x(t)[\cosh(2p) - 1]\} \\ &= -\left(1 + \frac{c}{2}\right) + \frac{c * s(t)x(t)}{2} + \frac{1}{2}\{(2+c)\cosh(2p) \\ &\quad - s(t)c * \sinh(2p) + x(t)[(2+c)\sinh(2p) \\ &\quad - c * s(t)\cosh(2p)]\} \\ &= -\left(1 + \frac{c}{2}\right) + \frac{c * s(t)x(t)}{2} - \frac{1}{2}[f_2(t) + x(t)f_1(t)],\end{aligned} \tag{A7}$$

where in the last line, the definition of functions f_1 and f_2 in Eq. (A4) was used. Finally, we shall use the formula $du = pdx - Hdt$. The last line of the Eq. (A3) can be rewritten in this form

$$p(t) = \text{arccoth}[Ae^{-2t}(1 - Be^{-ct})] = \frac{1}{2}\ln\left[\frac{Ae^{-2t}(1-Be^{-ct})+1}{Ae^{-2t}(1-Be^{-ct})-1}\right], \tag{A8}$$

where $Ae^{-2t}(Be^{-ct}-1) > 1$. Equation (A4) now gives

$$\frac{dx(t)}{dt} = f_1(t) + x(t)f_2(t).$$

Finally, we have $u(x,t)$,

$$\begin{aligned} u(x,t) - u(x^0,0) &= \int_0^t \left[p(\tau)\frac{dx(\tau)}{dt} - H(\tau) \right] d\tau \\ &= \int_0^t \Bigg\{ \operatorname{arccoth}[Ae^{-2t}(1 - Be^{-ct})] * [f_1(\tau) \\ &\quad + x(\tau)f_2(\tau)] - \left(1 + \frac{c}{2}\right) + \frac{c * s(\tau)x(\tau)}{2} \\ &\quad - \frac{1}{2}[f_2(\tau) + x(\tau)f_1(\tau)] \Bigg\} d\tau. \end{aligned} \tag{A9}$$

Now one can calculate numerically the value of $u(x,t)$ for any (x,t) given the initial distribution $u(x,0)$, which relates B with x^0 through $u'_x(x^0,0)=p(0,B)$ and provides $u(x^0,0)$ and $A=argmax\{u(x,0)\}$.

APPENDIX B: THE DERIVATION OF FREE DYNAMICS WITHOUT RECOMBINATION

In this appendix, we shall use the results of Appendix A to derive the exact solution of $u(x,t)$ in the case without selection and recombination, i.e., $f(x)=0$ and $c=0$. More precisely, we shall compare this solution, obtained for the single-peak case, i.e. $A=1$ and $u(x,0)=\{-\infty, \text{when } x<1; 0, \text{when } x=1\}$, with the one derived via alternative methods, Eq. (14). Differentiating Eq. (14) by x, one gets

$$p(x,t) = \frac{1}{2}\ln\left(\coth(t)\frac{1-x}{1+x}\right). \tag{B1}$$

We see in Eq. (11) that

$$u(t;x^0) - u_0(x^0) = \int_0^t \left\{ p(\tau)\frac{dx(\tau)}{dt} - H[x(\tau),p(\tau),\tau] \right\} d\tau.$$

Having $p(x,t)$ in the whole region, $u(x,t)$ is determined uniquely up to a constant summand in the initial distribution. Therefore, let us compare $p(x,t)$ in Eq. (14) to that obtained by solving Eqs. (11)–(13) putting $c=0$, $A=1$ and $u(x,0)=-a(x-1)^2$, $a\to\infty$. [Any other $u(x,0)=\lim_{a\to\infty}F(x;a)$ with a smooth monotonous $F(x;a)$ satisfying $\lim_{a\to\infty}F(x;a)=\{0, x=1; -\infty, x<1\}$ might be used instead.] The equation $\coth(p^0)=A(1-B)$ already mentioned immediately after Eq. (A3) implies $x^0(B)=A=1$ for all B besides $B\in[0;2]$. Indeed, here we understand $p^0(B)$ as $\lim_{t\to 0}p(t;B)$ and p^0 will be certainly unbounded if $x^0=\lim_{t\to 0}x(t;B)\neq 1$ while $u'_x(x,0)=+\infty$ for $x<1$. An unbounded p^0 implies $\coth(p^0)=\pm 1$. As a result, $B=0$ or $B=2$ when $x^0\neq 1$. Furthermore, $B=0$ and $B=2$ lead to a singularity in Eq. (A6). $B\in(0,2)$ do not allow a real solution for $p^0=\operatorname{arccoth}[A(1-B)]$. Therefore, all the characteristics which are nonsingular in this sense start from the point $x^0=1$.

Let us refer to Eq. (12). Putting $c=0$, we get

$$f_1(t) = \frac{-4A(1-B)e^{-2t}}{A^2(1-B)^2e^{-4t}-1},$$

$$f_2(t) = \frac{-2[A^2(1-B)^2e^{-4t}+1]}{A^2(1-B)^2e^{-4t}-1},$$

$$F_2(t) = \int_0^t f_2(\tau)d\tau = -2t + \ln\left(\frac{A^2(1-B)^2 - e^{4t}}{A^2(1-B)^2 - 1}\right). \tag{B2}$$

Equation (12) gives after integrating and substitution

$$\begin{aligned} x(t) &= x^0 e^{-2t}\frac{A^2(1-B)^2 - e^{4t}}{A^2(1-B)^2 - 1} \\ &\quad + e^{-2t}\frac{A^2(1-B)^2 - e^{4t}}{A(1-B)} * \left[\frac{1}{A^2(1-B)^2 - 1} \right. \\ &\quad \left. - \frac{e^{4t}}{A^2(1-B)^2 - e^{4t}} \right]. \end{aligned} \tag{B3}$$

Now let us put $A=1$ into Eq. (11)

$$\coth(p) = (1-B)e^{-2t}. \tag{B4}$$

We substitute $(1-B)$ from Eq. (B4) into Eq. (B3) and get $p(x,t)$ implicitly

$$\begin{aligned} x &= x^0 e^{-2t}\frac{\coth(p)^2 - 1}{\coth(p)^2 - e^{-4t}} + e^{-4t}\frac{\coth(p)^2 - 1}{\coth(p)[\coth(p)^2 - e^{-4t}]} \\ &\quad - \tanh(p). \end{aligned} \tag{B5}$$

As it was mentioned above, $x^0=1$,

$$\begin{aligned} x &= e^{-2t}\frac{\coth(p)^2 - 1}{\coth(p)^2 - e^{-4t}} + e^{-4t}\frac{\coth(p)^2 - 1}{\coth(p)[\coth(p)^2 - e^{-4t}]} \\ &\quad - \tanh(p). \end{aligned} \tag{B6}$$

Let us do some transformations to find an explicit expression for $p(x,t)$,

$$x = \frac{\coth(p)^2 - 1}{\coth(p)^2 e^{4t} - 1}\left(e^{2t} + \frac{1}{\coth(p)}\right) - \frac{1}{\coth(p)}, \tag{B7}$$

then

$$x = \frac{\coth(p) - e^{2t}}{\coth(p)e^{2t} - 1}, \tag{B8}$$

and finally

$$p(x,t) = \frac{1}{2}\ln\left(\coth(t)\frac{1-x}{1+x}\right), \tag{B9}$$

which fully corresponds to Eq. (B1). Thus we have obtained the same solution by HJE method.

[1] M. Smith, *The Evolution of Sex* (Cambridge University Press, Cambridge, England, 1978).

[2] S. P. Otto and Th. Lenormand, Nat. Rev. Genet. **3**, 252 (2002).

[3] J. Arjan, G. M. de Visser, and S. F. Elena, Nat. Rev. Genet. **8**, 139 (2007).

[4] M. W. Feldman, F. B. Christiansen, and L. D. Brooks, Proc. Natl. Acad. Sci. U.S.A. **77**, 4838 (1980).

[5] W. J. Ewens, *Mathematical Population Genetics* (Springer-Verlag, New York, 2004).

[6] R. Burger, *The Mathematical Theory of Selection, Recombination, and Mutation* (Wiley, New York, 2000).

[7] M. Baake and E. Baake, Can. J. Math. **55**, 3 (2003); **60**, 264 (2008).

[8] U. von Wangenheim, E. Baake, and M. Baake, e-print arXiv:0906.1678, J. Math. Biol. (to be published).

[9] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper Row, New York, 1970).

[10] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997).

[11] M. Eigen, Naturwiss. **58**, 465 (1971).

[12] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).

[13] C. J. Thompson and J. L. McBride, Math. Biosci. **21**, 127 (1974).

[14] B. L. Jones, R. H. Enns, and R. S. Rangnekar, Bull. Math. Biol. **38**, 15 (1975).

[15] D. B. Saakian, O. Rozanova, and A. Akmetzhanov, Phys. Rev. E **78**, 041908 (2008).

[16] G. Bocharov, N. J. Ford, J. Edwards, T. Breinig, S. Wain-Hobson, and A. Meyerhans, J. Gen. Virol. **86**, 3109 (2005).

[17] E. Cohen, D. A. Kessler, and H. Levine, Phys. Rev. Lett. **94**, 098102 (2005).

[18] J.-M. Park and M. W. Deem, Phys. Rev. Lett. **98**, 058101 (2007).

[19] D. B. Saakian, Z. Kirakosyan, and C.-K. Hu, Phys. Rev. E **77**, 061907 (2008).

[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 and C.-K. Hu, Phys. Rev. E **69**, 021913 (2004); **69**, 046121 (2004).

[23] J. Swetina and P. Schuster, Biophys. Chem. **16**, 329 (1982).

[24] E. Baake and T. Wiehe, J. Math. Biol. **35**, 321 (1997).

[25] R. Sanjuan, J. M. Cuevas, V. Furio, E. C. Holmes, and A. Moya, PLos Gen. **3**, e93 (2007).

[26] N. N. V. Vijay, V. R. Ajmani, A. S. Perelson, and N. M. Dixit, J. Gen. Virol. **89**, 1467 (2008).

[27] L. C. Evans, *Partial Differential Equations* (Westview, Boulder, 2002).

[28] A. Melikyan, *Generalized Characteristics of First Order PDEs* (Birkhäuser, Boston, 1998).

[29] P. D. Keightley and S. P. Otto, Nature (London) **443**, 89 (2006).

[30] D. B. Saakian, E. Munoz, C. K. Hu, and M. W. Deem, Phys. Rev. E **73**, 041913 (2006).

[31] M. S. Boerlijst and S. Bonhoeffer, Proc. R. Soc. London, Ser. B **263**, 1577 (1996).