Approximate perturbative solutions of quasispecies model with recombination

David B. Saakian,[1,2,*] Jin Ming Koh,[3] and Kang Hao Cheong[3,†]

[1]*Laboratory of Applied Physics, Advanced Institute of Materials Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam*

[2]*Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam*

[3]*Science and Math Cluster, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, Singapore*

(Received 9 October 2018; published 14 June 2019)

Despite the major roles played by genetic recombination in ecoevolutionary processes, limited progress has been made in analyzing realistic recombination models to date, due largely to the complexity of the associated mechanisms and the strongly nonlinear nature of the dynamical differential systems. In this paper, we consider a many-loci genomic model with fitness dependent on the Hamming distance from a reference genome, and adopt a Hamilton-Jacobi formulation to derive perturbative solutions for general linear fitness landscapes. The horizontal gene transfer model is used to describe recombination processes. Cases of weak selection and weak recombination with simultaneous mutation and selection are examined, yielding semianalytical solutions for the distribution surplus of $O(1/N)$ accuracy, where N is the number of nucleotides in the genome.

DOI: 10.1103/PhysRevE.99.062407

I. INTRODUCTION

Genetic recombination is a key factor influencing the dynamics and outcomes of biological evolution [1], the modeling of which is greatly relevant to ecology [2,3], microbiology and disease control [4,5], and statistical biophysics [6–8]. The solutions of recombination models have therefore attracted significant research attention to date—readers may refer to reviews [9,10] and Refs. [11,12] for overviews on existing mathematical and ecoevolutionary findings. Recombination is postulated to be an advantage, and therefore an evolutionary stabilizer, of sexual reproduction; especially in cases of negative epistasis, where two mutations together lead to inferior fitness than is expected from their effects individually, recombination can yield increases in the mean fitness of a population [13].

Here, we examine the recombination dynamics of a many-loci haploid genomic model, with two alleles at any locus. The genome in such a model can be considered a chain of two spin values, or equivalently, genetic letters. Such evolution models are characterized by fitness landscapes and mutation schemes dependent upon the sequence space, and have been successfully applied to the human immunodeficiency virus (HIV) [14]. While it is common in biological modeling to consider phenomenological recombination models with simplified mutation schemes, a more realistic approach can be realized by directly modeling the mutation of individual letters within the genome. An example of such a *microscopic model* had been considered in Ref. [14], in which the recombination process was taken to be a two-point crossover. Specifically, during each recombination event, two distinct points are chosen randomly along the length of each genome, and the genomes exchange their sequences between these points. The complexity of such a mechanism, whose outcome is dependent on the distance between genes along the genome, renders analytical descriptions infeasible.

On the other hand, horizontal gene transfer (HGT) is a comparatively simpler recombination process. In the HGT model [15,16] there is an exchange of a single allele between genomes, as opposed to the transfer of entire sequences. The recombination model with the exchange of a single allele is equivalent to the HGT model of Ref. [17], with a rescaling of the recombination rate. Such simplicity has enabled certain key attributes to be analytically derived through mean-field approaches, including mean fitness trends [16] and steady-state genome distributions [18]. However, the exact dynamics for the general case remains unsolved in the literature, though attempts have been made through both Hamilton-Jacobi equation (HJE) methods [19,20] and quantum statistical methods [21]. In particular, the existence of exact analytical solutions in selection-free (flat fitness landscape) special cases has been proven [22,23] and later derived [24], but exact solvability in nonzero selection is unknown, and no rigorous analytical results have been presented to date for fitness landscapes that are linear or more general in form. Addressing this lacuna is the focus of this paper.

In the current paper, we develop the Hamilton-Jacobi method to compute continuous-time surplus dynamics on general linear fitness landscapes, with the simultaneous recombination, mutation, and selection of genomes considered. The strongly nonlinear differential system describing the genetic dynamics, unfortunately, does not admit exact analytical solutions, and an approximate perturbative approach is therefore adopted. Solutions of $O(1/N)$ accuracy, where N is the genome size, typically $\sim 10^3$ for viruses and significantly greater for more complex organisms, are presented.

II. MODEL

We adopt a genome model consisting of a chain of N genes, each occurring with two possible alleles, interpretable as ± 1 spin. [15,16]. There are therefore 2^N possible genome

*david.saakian@tdtu.edu.vn

†Corresponding author: kanghao_cheong@sutd.edu.sg

sequences. The fitness of each genome can be taken to depend on the number of mutations from some reference sequence, treated here as consisting of all $+1$ spins, and we define the lth Hamming class as a collection of all sequences carrying l mutations from the reference [25]. The lth Hamming class, containing l spins of -1, is assigned fitness

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

Without loss of generality, we scale all variables such that mutation and recombination occur with unity rate and rate c, respectively. During each recombination event, one allele at a given position of a given sequence is replaced by the allele at the same position of a genome chosen randomly from the population. In this way we have defined the HGT model [24], which is in fact mathematically equivalent to a crossover recombination model with a crossover rate $c/2$. The dynamics for the lth Hamming class probability P_l can then be written

$$\frac{1}{N}\frac{dP_l}{dt} = \frac{r_l P_l}{N} + \left[\left(1 - \frac{l-1}{N}\right)P_{l-1} + \left(\frac{l+1}{N}\right)P_{l+1}\right] - P_l\left(1 + \frac{1}{N}\sum_k r_k P_k\right)$$
$$- cP_l + c\left[\left(1 - \frac{\bar{l}}{N}\right)\left(1 - \frac{l}{N}\right) + \frac{\bar{l}l}{N^2}\right]P_l + c\left[\left(1 - \frac{\bar{l}}{N}\right)\left(\frac{l+1}{N}\right)P_{l+1} + \frac{\bar{l}}{N}\left(1 - \frac{l-1}{N}\right)P_{l-1}\right], \tag{2}$$

where $\bar{l} = \sum_{l=0}^{N} lP_l$. The first line of Eq. (2) is analogous to the constant-population Crow-Kimura model for mutation and selection [26–31], with two types of alleles at any loci. Each allele can mutate to the opposite one with a probability dt during each dt interval in time. The second line in Eq. (2) describes the recombination process. During each interval in time dt, an allele in the genome is replaced with probability cdt by the allele at the same position of a randomly chosen genome (see Ref. [16] for details). This accounts for the $-cP_l$ term. The remaining terms describe the replacement of an allele with one of the same type, as well as the replacement of an allele with one of a different type. That the current model is constructed upon the Crow-Kimura genetics model necessarily implies validity in the large population limit, as opposed to the Wright-Fisher and Moran models [32] typically suitable for finite population sizes, though modifications to accurately accommodate finite populations have been shown to be possible [33].

Considering an ansatz

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

and neglecting end terms P_{l-1} and P_{l+1} when $l = 0$ and $l = N$, respectively, the following Hamiltonian of $O(1/N)$ accuracy for $u(m,t)$ can be obtained,

$$\frac{\partial u}{\partial t} + H(m, u', s) = 0,$$
$$-H(m,p,s) \equiv f(m) + \frac{1+m}{2}e^{2p} + \frac{1-m}{2}e^{-2p} + \frac{(ms-1)c}{2} + \left[\frac{1+m}{2}\frac{1-s}{2}e^{2p} + \frac{1-m}{2}\frac{1+s}{2}e^{-2p}\right]c - 1, \tag{4}$$

where $p = u'$ is the derivative of u with respect to m, and

$$s = \sum_l P_l(1 - 2l/N) \tag{5}$$

is the surplus of the distribution. Under these definitions, $u(m,t)$ has a maximum at point $m = s$ at time t. Our objective is to calculate the surplus $s(t)$ as yielded by this model. The surplus is of fundamental interest in recombination dynamics—the mean number of mutatory occurrences is given by $N(1-s)/2$, thereby providing a natural characterization for the dynamical behavior of the model. In quantum recombination models employing Ising chains, the surplus $s(t)$ is analogous to surface magnetization [16,34].

In the current work, we focus on the time dynamics of the genomic distribution, which is nontrivial to compute. The steady-state distribution in this model, on the other hand, can be found with relative ease, by substituting $\partial u/\partial t \to R$ in Eq. (4) where $R = f(s)$ denotes the steady-state mean fitness. The latter is then defined by the system of equations [18],

$$R = \max\left\{f(m) + \sqrt{(1-m^2)\left[\left(1+\frac{c}{2}\right)^2 - \left(\frac{cs}{2}\right)^2\right]} - 1 - \left(\frac{1-ms}{2}\right)c, 0\right\}, \quad f(s) = R. \tag{6}$$

III. RECOMBINATION IN THE CASE OF WEAK SELECTION

A. Exact equations of characteristic curves

We first discuss recombination in the case of weak selection (where k is small), and derive solutions for the associated characteristic curves and distribution surplus. In practical terms, this case implies that the fitness landscape imposes a small selection pressure on the population. Consider the Hamilton-Jacobi equations for the characteristic curve $x(\tau)$,

$$\frac{dx}{d\tau} = \frac{\partial H}{\partial p}, \quad \frac{dp}{d\tau} = -\frac{\partial H}{\partial x}. \tag{7}$$

With solutions for $x(\tau)$ and $p(\tau)$, the function $u(x,t)$ can be calculated as

$$u(x,t) = \int_0^t \left[p(\tau)\frac{dx(\tau)}{d\tau} - H(\tau) \right] d\tau + u(x,0), \tag{8}$$

where the integral is evaluated along the characteristic that connects the starting point (where the distribution was focused) with the point x, and $p = u'$.

Exact analytical solutions are only possible in the neutral case $f(m) = 0$ and the single-peak fitness landscape case $f(m) = \delta_{1m}$, where δ is the Kronecker delta. For general linear fitness landscapes $f(x) = kx$, a perturbative approach is necessary. We first consider the $k = 0$ neutral case, and derive the first-order perturbative solutions with respect to k. For the neutral case, we have a Hamiltonian

$$-H_0(m,p,s) \equiv \frac{1+m}{2}e^{2p} + \frac{1-m}{2}e^{-2p} + \frac{(ms-1)c}{2} - 1 + \left[\frac{1+m}{2}\frac{1-s}{2}e^{2p} + \frac{1-m}{2}\frac{1+s}{2}e^{-2p} \right] c. \tag{9}$$

The surplus $s(t)$ is exponential for the neutral case [24],

$$s_0(t) = Ae^{-2t}. \tag{10}$$

The following equations for the characteristic curves can next be derived for general k, with the Hamilton function explicitly dependent on t,

$$\begin{aligned} q &= kx + \left(1 + \frac{c}{2}\right)[\cosh(2p) - 1 + x\sinh(2p)] \\ &\quad - \frac{cs(t)}{2}\{\sinh(2p) + x[\cosh(2p) - 1]\}, \\ p &= \frac{\partial u}{\partial x}, \quad q = \frac{\partial u}{\partial t}, \quad p(s(t),t) = 0. \end{aligned} \tag{11}$$

The Hamilton-Jacobi equations then yield the following equations for the characteristic $x(\tau)$,

$$\begin{aligned} \frac{dx}{d\tau} &= \frac{\partial H}{\partial p} = -(2+c)[\sinh(2p) + x\cosh(2p)] \\ &\quad + cs(\tau)[\cosh(2p) + x\sinh(2p)], \\ \frac{dp}{d\tau} &= k - \frac{\partial H}{\partial x} = k + \left(1 + \frac{c}{2}\right)\sinh(2p) \\ &\quad - \frac{cs(\tau)}{2}[\cosh(2p) - 1]. \end{aligned} \tag{12}$$

We begin by considering the neutral-case Hamiltonian H_0. Using the surplus $s_0(t)$ in Eq. (10), we obtain $p_0(t)$,

$$\coth(p_0) = s_0(t)(1 - Be^{-ct}) = Ae^{-2t}(1 - Be^{-ct}), \tag{13}$$

where $A \equiv s(0)$ and B are parameters characterizing the solution. Henceforth, the characteristic x_0 may at times be written as $x_0(t,B)$ for clarity, in order to express its dependence on B. The following solution for $x_0(t,B)$ is then found [24],

$$\begin{aligned} x_0(t,B) &= Ae^{F_2(t)} + e^{F_2(t)}\int_0^t f_1(\tau)e^{-F_2(\tau)}\,d\tau, \\ \frac{dx_0}{dt} &= f_1 + xf_2, \end{aligned} \tag{14}$$

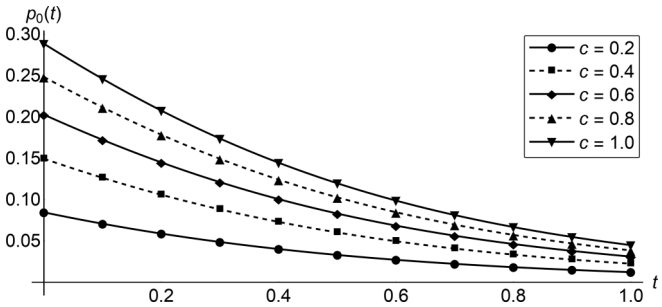

FIG. 1. Plot of $p_0(t)$ as described by Eq. (13), for different values of recombination rate c. $A \equiv s(0)$ is taken to be unity, equivalent to an initial genomic distribution concentrated at the $l = 0$ Hamming class, and B is given by Eq. (16).

where

$$\begin{aligned} -f_1(t) &= s_0(t)\cdot\frac{2(2+c)[1 - f_3(t)] - c[s_0(t)^2 f_3(t)^2 + 1]}{s_0(t)^2 f_3(t)^2 - 1}, \\ -f_2(t) &= \frac{(2+c)[s_0(t)^2 f_3(t)^2 + 1] - 2cs_0(t)^2 f_3(t)}{s_0(t)^2 f_3(t)^2 - 1}, \\ f_3(t) &= 1 - Be^{-ct}, \\ F_2(t) &= \int_0^t f_2(\tau)d\tau. \end{aligned} \tag{15}$$

The integrals in Eqs. (14) and (15) are calculated along the characteristic curves. Sample plots of $p_0(t)$ as described by Eq. (13) and $x_0(t)$ as described by Eq. (14) are presented in Figs. 1 and 2, respectively, for different values of c. It is of interest to determine the characteristic along which $u(m,t)$ is maximum at the moment of time t—this is herein termed the *main characteristic*. We first take $B = b(t)$ along this curve. From Eq. (12) the derivative $dx_0(t,B)/dt|_{B=b(t)} = -2e^{-2t}$ is obtained. We then have

$$f_1 + Ae^{-2t}f_2 = -2e^{-2t}, \tag{16}$$

which is a second-order algebraic equation for $b(t)$. The function $b(t)$ can therefore be calculated, enabling the perturbative definition $B = b(t) + b_1(t)$. Similarly, we represent the perturbative solutions for $x(\tau)$ and $p(\tau)$ as

$$\begin{aligned} x(\tau) &= x_0(\tau,B) + x_1(\tau), \\ p(\tau) &= p_0(\tau,B) + p_1(\tau), \end{aligned} \tag{17}$$

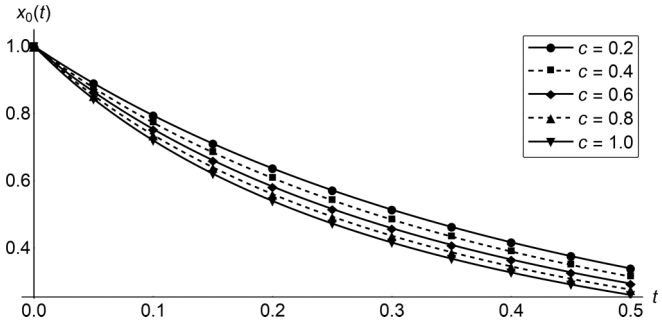

FIG. 2. Plot of $x_0(t)$ as described by Eq. (14), for different values of recombination rate c. $A \equiv s(0)$ is taken to be unity, and B is given by Eq. (16).

with the initial conditions

$$x_1(0) = p_1(0) = 0. \tag{18}$$

The surplus $s(t)$ is also expressed in the form

$$s(t) = s_0(t) + s_1(t). \tag{19}$$

We now examine the maximum points $x_0(t, b(t) + b_1(t)) + x_1(t)$, where $p = 0$, and

$$\begin{aligned} p_0(t, b(t) + b_1(t)) + p_1(t) &= \frac{dp_0}{dB} b_1(t) + p_1(t) = 0, \\ x_0(t, b(t) + b_1(t)) + x_1(t) &= x_0(t, b(t)) \\ &\quad + \frac{dx_0}{dB} b_1(t) + x_1(t) \\ &= s_0(t) + \frac{dx_0}{dB} b_1(t) + x_1(t). \end{aligned} \tag{20}$$

From Eq. (20), it can be deduced that

$$s_1(t) = \left.\frac{dx_0}{dB}\right|_{B=b(t)} b_1(t) + x_1(t). \tag{21}$$

The expression for $s(t)$ can then be derived. Henceforth, we examine the characteristic curves for the time period $0<\tau<t$. Substituting Eq. (17) into Eq. (12), the equations for $x_1(\tau)$ and $p_1(\tau)$ for different moments of time are

$$\begin{aligned} \frac{dx_1}{d\tau} &= [cs_0(\tau)\sinh(2p_0) - (2+c)\cosh(2p_0)]x_1 \\ &\quad + cs_1(\tau)[\cosh(2p_0) + x_0\sinh(2p_0)] \\ &\quad + 2p_1\{-(2+c)[\cosh(2p_0) + x_0\sinh(2p_0)] \\ &\quad + cs_0(\tau)[\sinh(2p_0) + x_0\cosh(2p_0)]\}, \\ \frac{dp_1}{d\tau} &= k + [(1+c)\cosh(2p_0) - cs_0(\tau)\sinh(2p_0)]p_1 \\ &\quad - \frac{cs_1(\tau)}{2}[\cosh(2p_0) - 1]. \end{aligned} \tag{22}$$

The solutions for x_1 and p_1 have therefore been obtained in terms of s_1. Note also that Eq. (20) yields a solution for $b_1(t)$, and Eq. (21) yields an expression for s_1. From the last equation of Eq. (22), we may write

$$\begin{aligned} \frac{dp_1}{d\tau} &= k + g'p_1 + g_1 s_1, \\ g' &\equiv [(1+c)\cosh(2p_0) - cAe^{-2\tau}\sinh(2p_0)], \\ g_1 &\equiv -c[\cosh(2p_0) - 1]/2. \end{aligned} \tag{23}$$

With $p_1(0) = 0$, we obtain

$$p_1(t) = e^{g(t)} \int_0^t [k + g_1(\tau)s_1(\tau)]e^{-g(\tau)}\, d\tau. \tag{24}$$

Combining Eqs. (22) and (24), we also obtain

$$\begin{aligned} \frac{dx_1}{d\tau} &= g_2 x_1(\tau) + g_3 s_1(\tau) \\ &\quad + g_4 e^{g(\tau)} \int_0^t [k + g_1 s_1(\tau)]e^{-g(\tau)}\, d\tau, \end{aligned}$$

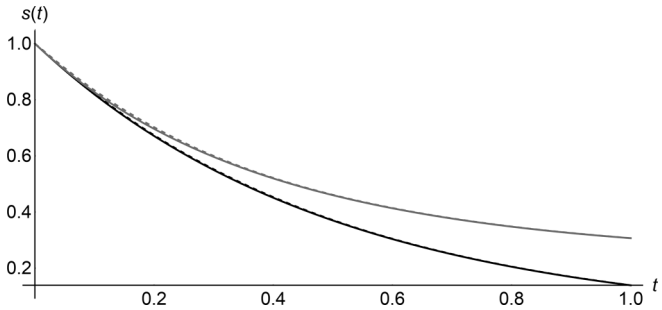

FIG. 3. Comparison of the perturbative solution of surplus $s(t)$ obtained from Eq. (28), accurate to $O(1/N)$ order, against the numerical solution of the base Eq. (2), plotted in dashed and solid lines, respectively. Curves for $k = 0.01$ (black) and $k = 0.5$ (gray) are shown, with $N = 100$.

$$\begin{aligned} g_2 &\equiv cs_0(\tau)\sinh(2p_0) - (2+c)\cosh(2p_0), \\ g_3 &\equiv c[\cosh(2p_0) + x_0\sinh(2p_0)], \\ g_4 &\equiv 2\{-(2+c)[\cosh(2p_0) + x_0\sinh(2p_0)] \\ &\quad + cs_0(\tau)[\sinh(2p_0) + x_0\cosh(2p_0)]\}. \end{aligned} \tag{25}$$

The solution is therefore

$$\begin{aligned} x_1(t) = e^{g_2(t)} \int_0^t \Bigg[& s_1(t')g_3(t') + g_4(t')e^{g(t')} \\ & \times \int_0^{t'} [k + g_1(\tau)s_1(\tau)]e^{-g(\tau)}\, d\tau \Bigg] e^{-g_2(t')}\, dt'. \end{aligned} \tag{26}$$

Substituting this result into Eqs. (20) and (21) gives

$$\begin{aligned} 0 &= \frac{dp_0}{dB} b_1(t) + e^{g(t)} \int_0^t [k + g_1(\tau)s_1(\tau)]e^{-g(\tau)}\, d\tau, \\ s_1(t) &= \frac{dx_0}{dB} b_1(t) + e^{g_2(t)} \int_0^t \Bigg[s_1(t')g_3(t') + g_4(t')e^{g(t')} \\ &\quad \times \int_0^{t'} [k + g_1(\tau)s_1(\tau)]e^{-g(\tau)}\, d\tau \Bigg] e^{-g_2(t')}\, dt'. \end{aligned} \tag{27}$$

Finally, we obtain

$$\begin{aligned} 0 = & -\frac{dx_0}{dB} e^{g(t)} \int_0^t [k + g_1(\tau)s_1(\tau)]e^{-g(\tau)}\, d\tau \\ & + \frac{dp_0}{dB} e^{g_2(t)} \Bigg\{ \int_0^t \Bigg[s_1(t')g_3(t') + g_4(t')e^{g(t')} \\ & \times \int_0^{t'} [k + g_1(\tau)s_1(\tau)]e^{-g(\tau)}\, d\tau \Bigg] e^{-g_2(t')}\, dt' - s_1(t) \Bigg\}. \end{aligned} \tag{28}$$

This represents the solution for $s_1(t)$, which in turn yields the complete perturbative solution for the surplus $s(t) = s_0(t) + s_1(t)$. Due to the truncation of end terms in Eq. (4), this solution is accurate to $O(1/N)$ order. We present in Fig. 3 a comparison between the perturbative numerical solution as obtained from Eq. (28) against that obtained from Eq. (2), with

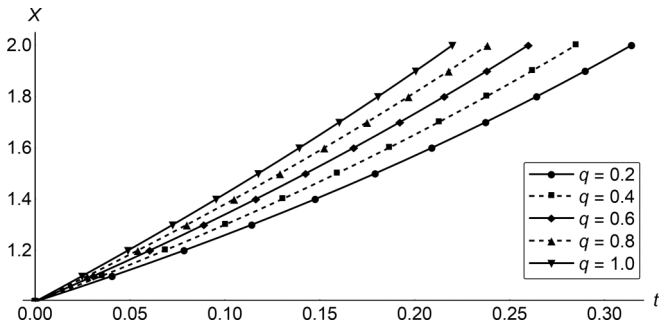

FIG. 4. Plot of $X(t, q)$ against t for different values of q, as described by Eq. (30). Parameters are taken to be $k = 0.01$ and $A = 1$ for illustration.

recombination rate $c = 1$ and selection rates $k = 0.01$ and $k = 0.5$ for illustrative purposes. The chosen recombination rate is a biologically reasonable middle-ground value for numerous species of bacteria [35], though it is notable that the presented method makes no assumptions on c and is therefore applicable to any recombination rate. Clearly, a close match between the solutions can be observed, reflecting the good accuracy of the current method. For large N, reflecting real-life conditions, utilizing Eq. (28) will be greatly more efficient than utilizing Eq. (2), the latter requiring the solving of a system of N strongly nonlinear coupled ordinary differential equations.

IV. EVOLUTIONARY DYNAMICS WITH MUTATION, SELECTION, AND WEAK RECOMBINATION

The adopted perturbative approach can be further extended to cases of evolutionary dynamics with simultaneous mutation, selection, and weak recombination. We consider neutral-case coordinate and momentum characteristics x_0 and p_0 of the form

$$\begin{aligned}\frac{dx_0}{d\tau} &= -2[\sinh(2p_0) + \cosh(2p_0)x_0],\\ \frac{dp_0}{d\tau} &= f'(x_0) + \sinh(2p_0).\end{aligned} \tag{29}$$

We characterize these characteristics by a parameter q, and define the functions $X(t, q)$ and $P(t, q)$ such that

$$\begin{aligned}t &= \int_A^X \frac{dy}{2\sqrt{(q+1-ky)^2 - 1 + y^2}},\\ P &= \frac{1}{2}\ln\left[\frac{(q+1-kx) \pm \sqrt{(q+1-kx)^2 - 1 + x^2}}{1+x}\right].\end{aligned} \tag{30}$$

Such a definition yields

$$x_0(\tau) = X(\tau, q), \quad p_0(\tau) = P(\tau, q). \tag{31}$$

The maximum of the genomic distribution without recombination occurs at $s(t) = X(t, f(s(t)))$, therefore suggesting $q = f(s(t))$. Sample plots of X for differing q are presented in Fig. 4, with $k = 0.01$ for illustrative purpose. We now consider the following ansatz,

$$\begin{aligned}x(\tau) &= X(\tau, q(\tau) + b_1(\tau)) + x_1(\tau),\\ p(\tau) &= P(\tau, q(\tau) + b_1(\tau)),\end{aligned} \tag{32}$$

which yields the correction terms of

$$\begin{aligned}\frac{dx_1}{d\tau} &= h_3 x_1 + h_4 p_1 + c h_5,\\ \frac{dp_1}{d\tau} &= h' p_1 + h_1 x_1 + h_2,\\ h' &= 2\cosh(2p_0), \quad h_1 = f''(x_0),\\ h_2 &= \frac{c}{2}\sinh(2p_0) - \frac{c s_0(t)}{2}[\cosh(2p_0) - 1],\\ h_3 &= f''(x_0) - 2\cosh(2p_0),\\ h_4 &= -4[\cosh(2p_0) + x_0\sinh(2p_0)],\\ h_5 &= s_0(\tau)[\cosh(2p_0) + x_0\sinh(2p_0)]\\ &\quad - [\sinh(2p_0) + \cosh(2p_0)x_0].\end{aligned} \tag{33}$$

Equation (33) can be solved to yield solutions for x_1 and p_1, and therefore the characteristics x and p, for general fitness functions. For linear fitness landscapes $f(x) = kx$, the solution for p_1 is especially simple,

$$\begin{aligned}\frac{dp_1}{d\tau} &= h' p_1 + h_2,\\ p_1(t) &= e^{h(\tau)}\int_0^t h_2(\tau) e^{-h(\tau)}\, d\tau.\end{aligned} \tag{34}$$

Substituting Eq. (34) into Eq. (33) then gives the solution for x_1,

$$\begin{aligned}x_1(t) = e^{h_3(t)}\int_0^t e^{-h_3(t')}\Bigg\{ & c h_5(t') + h_4(t') e^{h(t')}\\ & \times \int_0^{t'} h_2(\tau) e^{-h(\tau)}\, d\tau\Bigg\} dt'.\end{aligned} \tag{35}$$

Combining these solutions with Eq. (32), the following properties can then be obtained,

$$\frac{dP}{dq}b_1(t) + p_1 = 0, \quad \frac{dX}{dq}b_1(t) + x_1 - s_1 = 0, \tag{36}$$

or

$$-\frac{dX}{dq}p_1 + \frac{dP}{dq}(x_1 - s_1) = 0. \tag{37}$$

Thus, we have obtained a simple expression for $s_1(t)$ in cases of simultaneous mutation, selection, and weak recombination. This is of $O(1/N)$ accuracy due to the truncation of end terms in Eq. (4), in a similar fashion as the earlier weak-selection solutions.

V. CONCLUSION

Genetic recombination plays major roles in shaping evolution, but the complexity of its mechanisms has rendered analytical treatments mostly infeasible, especially when occurring simultaneously with genome mutation and selection. Exact results have been derived only for selection-free and

single-peak fitness landscape special cases to date. Our work has presented new solutions for general linear fitness landscapes, in particular for the distribution surplus.

Cases of weak selection and weak recombination had been considered, each treated with a perturbative approach of $O(1/N)$ accuracy. The solutions were only semianalytic, though numerical solving of the derived integral solutions is possible, and indeed admits significantly greater convenience and efficiency as opposed to tackling the N-dimension coupled ordinary differential system naively. These results are of great relevance to ecoevolutionary modeling [9,10,36–39], especially in epistasis where recombination processes yield significant phenotypic and genotypic effects. The current study also suggests a general method to solve for discrete-time recombination processes and for nonlinear fitness landscapes, though these will be more complex than the current solutions. Generalization of this method to diploid genomes [40] is also possible, and is an avenue for further research.

ACKNOWLEDGMENT

This work was supported by the Russian Science Foundation Grant No. 19-11-00008.

[1] J. M. Smith and J. Maynard-Smith, *The Evolution of Sex* (Cambridge University Press, Cambridge, U.K., 1978).
[2] I. Hanski, Nature (London) **396**, 41 (1998).
[3] A. A. Hoffmann and C. M. Sgrò, Nature (London) **470**, 479 (2011).
[4] M. J. Berryman, S. L. Spencer, A. G. Allison, and D. Abbott, Proc. SPIE **5471**, doi:10.1117/12.546641 (2004).
[5] M. I. Nelson and E. C. Holmes, Nat. Rev. Gen. **8**, 196 (2007).
[6] M. Perc, J. J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, and A. Szolnoki, Phys. Rep. **687**, 1 (2017).
[7] J.-M. Park and M. W. Deem, J. Stat. Phys. **125**, 971 (2006).
[8] E. Muñoz, J.-M. Park, and M. W. Deem, J. Stat. Phys. **135**, 429 (2009).
[9] S. P. Otto and T. Lenormand, Nat. Rev. Genet. **3**, 252 (2002).
[10] J. A. G. M. de Visser and S. F. Elena, Nat. Rev. Genet. **8**, 139 (2007).
[11] R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation* (Wiley, New York, 2000).
[12] W. J. Ewens, *Mathematical Population Genetics* (Springer, New York, 2004).
[13] M. W. Feldman, F. B. Christiansen, and L. D. Brooks, Proc. Natl. Acad. Sci. USA **77**, 4838 (1980).
[14] G. Bocharov, N. J. Ford, J. Edwards, T. Breinig, S. Wain-Hobson, and A. Meyerhans, J. Gen. Virol. **86**, 3109 (2005).
[15] E. Cohen, D. A. Kessler, and H. Levine, Phys. Rev. Lett. **94**, 098102 (2005).
[16] J.-M. Park and M. W. Deem, Phys. Rev. Lett. **98**, 058101 (2007).
[17] M. C. Boerlijst, S. Bonhoeffer, and M. A. Nowak, Proc. R. Soc. London, Ser. B **263**, 1577 (1996).
[18] D. B. Saakian, Z. Kirakosyan, and C.-K. Hu, Phys. Rev. E **77**, 061907 (2008).
[19] D. B. Saakian, J. Stat. Phys. **128**, 781 (2007).
[20] K. Sato and K. Kaneko, Phys. Rev. E **75**, 061909 (2007).
[21] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 046121 (2004).
[22] M. Baake and E. Baake, Can. J. Math. **55**, 3 (2003).
[23] M. Baake and E. Baake, Can. J. Math. **60**, 264 (2008).
[24] Z. Avetisyan and D. B. Saakian, Phys. Rev. E **81**, 051916 (2010).
[25] J. Swetina and P. Schuster, Biophys. Chem. **16**, 329 (1982).
[26] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper & Row, New York, 1970).
[27] D. B. Saakian, K. H. Cheong, and J. M. Koh, Phys. Rev. E **98**, 012405 (2018).
[28] C. J. Thompson and J. L. McBride, Math. Biosci. **21**, 127 (1974).
[29] B. L. Jones, R. H. Enns, and S. S. Rangnekar, Bull. Math. Biol. **38**, 15 (1976).
[30] E. M., J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[31] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[32] W. J. Ewens, Population genetics theory - the past and the future, in *Mathematical and Statistical Developments of Evolutionary Theory*, edited by S. Lessard (Springer, Dordrecht, 1990), pp. 177–227.
[33] D. B. Saakian, M. W. Deem, and C.-K. Hu, Europhys. Lett. **98**, 18001 (2012).
[34] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997).
[35] P. Fearnhead, S. Yu, P. Biggs, B. Holland, N. French *et al.*, Ann. Appl. Stat. **9**, 200 (2015).
[36] K. H. Cheong, J. M. Koh, and M. C. Jones, Proc. Natl. Acad. Sci. USA **115**, E5258 (2018).
[37] J. M. Koh, N.-g. Xie, and K. H. Cheong, Nonlinear Dyn. **94**, 1467 (2018).
[38] Z. X. Tan and K. H. Cheong, eLife **6**, e21673 (2017).
[39] K. H. Cheong, Z. X. Tan, and Y. H. Ling, Commun. Nonlinear Sci. Numer. Simul. **60**, 107 (2018).
[40] Z. Kirakosyan, D. B. Saakian, and C.-K. Hu, J. Stat. Phys. **144**, 198 (2011).