A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

Società Italiana di Fisica

edp sciences

IOP Institute of Physics

LETTER

A solution of the Crow-Kimura evolution model on fluctuating fitness landscape

To cite this article: Vladimir Suvorov *et al* 2023 *EPL* **142** 57003

View the article online for updates and enhancements.

A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

June 2023

EPL, **142** (2023) 57003
doi: 10.1209/0295-5075/acd65b

www.epljournal.org

A solution of the Crow-Kimura evolution model on fluctuating fitness landscape

VLADIMIR SUVOROV[1], DAVID B. SAAKIAN[2(a)] and MICHAEL LYNCH[3]

[1] *Auriga Inc. - 400 TradeCenter, Ste 5900 Woburn, MA 01801, USA*
[2] *A.I. Alikhanyan National Science Laboratory (Yerevan Physics Institute) Foundation - 2 Alikhanian Brothers St., Yerevan 375036, Armenia*
[3] *Biodesign Center for Mechanisms of Evolution, Arizona State University - Tempe, AZ, 85287, USA*

received 27 November 2022; accepted in final form 17 May 2023
published online 1 June 2023

Abstract – The article discusses the Crow-Kimura model in the context of random transitions between different fitness landscapes. The duration of epochs, during which the fitness landscape is constant over time, is modeled by an exponential distribution. To obtain an exact solution, a system of functional equations is required. However, to approximate the model, we consider the cases of slow or fast transitions and calculate the first-order corrections using either the transition rate or its inverse. Specifically, we focus on the case of slow transitions and find that the average fitness is equal to the average fitness for evolution on static fitness landscapes, but with the addition of a load term. We also investigate the model for a small number of genes and identify the exact transition points to the transient phase.

Introduction. – The study of evolutionary dynamics on fluctuating fitness landscapes has emerged as a key area of modern evolutionary theory [1–37]. Fluctuations in fitness landscapes can arise due to various factors, such as the presence of toxins, changes in pH, temperature, or other environmental changes.

According to evolutionary theory, populations adapt to a given landscape, and non-trivial evolution arises in a dynamic environment. It has been widely assumed that changing environments are a key factor contributing to the complexity of population evolution. Typically, simple gene models are used to study the dynamics of evolutionary adaptation in a dynamic environment [1,20]. However, we are interested in the case of evolution with many genotypes and epistasis, which has far-reaching implications in the real world.

In this study, we investigate the dynamics of quasispecies models in a dynamic environment with a focus on large genome lengths. The models are studied in a stochastic environment with small or large transition rates [23–28]. We consider a haploid, infinite population model with symmetric mutations, where the forward and backward mutation rates are equal, and there is no recombination. The model assumes that the genome is a chain of L genes, with each gene being one of two types: $+$ or $-$. Denoting the gene type as s_l, we define the fitness of the genome $s_1,\ldots,s_L$ as $f(s_1,\ldots,s_L)$. If the fitness can be decomposed into $f(s_1,\ldots,s_L) = f_1(s_1) + \ldots + f_L(s_L)$, there is no epistasis; otherwise, we encounter non-zero epistasis. In this study, the fitness depends on the mean number of gene types, leading to a symmetric fitness landscape.

There have been extensive studies of quasispecies models on static fitness landscapes with a constant mutation rate [23,28]. Exact solutions for the dynamics on a smooth fitness landscape at the large genome limit have been obtained [29,30]. In quasispecies models on static fitness landscapes, two phases are observed: the selective phase at small mutation rates and the non-selective phase at high mutation rates. On the other hand, interesting qualitative results have been derived for the quasispecies model on fluctuating fitness landscapes [14,15] for the case of random fluctuations in the fitness landscape. More precise results have been obtained for the periodic change of the single peak fitness positions [16–19] using the methodology of our previous work [31]. An important finding reported in [14] is that at very high levels of fluctuations, the system can lose the selective phase, even for small mutation rates.

Our research aims to investigate the impact of random transitions between two general fitness landscapes on the evolution of populations. This problem resembles

(a) E-mail: saakian@yerphi.am (corresponding author)

somehow the mutator phenomenon, where the mutation in a specific gene in the genome results in a drastic change of the mutation rates of other genes or their fitness landscape [32], so we again have an evolution with two fitness landscapes. In the case of the mutator, there is a well-formulated mathematical model with a system of $2(L+1)$ equations (where L is the genome length) [32].

However, the accurate consideration of evolution on fluctuating fitness landscapes requires a complicated system of functional equations, it is a much harder mathematical tool than the one used in the mutator model [32]. The significant difference between the two phenomena lies in the fact that we have macroscopic transition rates for the mutator model, while our model of evolution on fluctuating fitness landscapes assumes stochastic transitions. As a result, the mathematical treatment of the two problems differs significantly.

First we solve the case of slow transitions by using our previous results for the dynamics of evolution on a static fitness landscape, and then solve the fast transition case by introducing an effective Hamilton-Jacobi equation.

In the next section, we present our model by formulating a large system of ordinary differential equations. In the third section, we apply our method to solve the dynamics of the Crow-Kimura (CK) model [25,26] with a given fitness function, starting from a steady state of the model with an alternative fitness function.

In the fourth section, we apply our method to random fluctuations in the fitness landscape and calculate the mean fitness of the model with a fluctuating fitness landscape in the case of slow transitions. In the fifth section, we solve the fast transition case.

Finally, in the sixth section, we consider a model with few genotypes and directly calculate the transition point to the transient phase without using the Hamilton-Jacobi equation.

Formulation of the mathematical model. – We consider a model with L two type genes. We denote the gene type as ± 1, which we further denote as spins. Thus the genome is a sequence of L spins; we have 2^L different (genomes) sequences. We denote different sequences via an index $0 \leq i \leq 2^L - 1$. The Hamming distance $d(i,j)$ between two sequences (genomes) is the number of differences in the signs.

We denote by μ_{ij} the mutation rate from the i-th to the j-th state. For $d(i,j) = 1$, we have a mutation rate μ/L, also $\mu_{ii} = -\mu$. Thus we consider only point mutations, and μ is the mutation rate per genome. The Crow-Kimura model (or parallel mutation selection model [25]) is defined via a system of equations,

$$\begin{aligned} \frac{\mathrm{d}P(i,t)}{\mathrm{d}t} &= (r_i + \mu_{ii} - R)P(i,t) + \sum_{j \neq i} \mu_{ji} P(j,t) \\ &= (r_i - \mu - R)P(i,t) + \frac{\mu}{L} \sum_{j, d(i,j)=1} P(j,t), \end{aligned} \tag{1}$$

where the sum is over the neighbors with the Hamming distance $d(i,j) = 1$, r_i is the fitness of i-th genome, mean fitness $R(t) = \sum_j P(j,t) r_j$, $P(i,t)$ is the fraction of i-th genotype in the population, and there is a balance condition $\sum_i P(i,t) = 1$.

We define the 0th sequence as the one with only + gene type. For the symmetric fitness case, when the fitness is a function of the total number of mutations (steps) from the 0th sequence, we introduce a fitness function $f(m)$, where the variable m_j is an equivalent of the magnetization in the Ising model with L spins taking + and − spins and having j number of − spins, $m_j = 1 - 2d(j,0)/L$. We give such a definition of the variable m from the statistical physics perspective, and the model is mapped to the system of 1-dimensional Ising models in the transverse magnetic field, m is the equivalent of spin [31].

We collect all of the sequences with the Hamming distance l from the 0th sequence, to the Hamming class, and denote by $\hat{p}_l$ the total probability of the l-th Hamming class. There are $\frac{L!}{l!(L-l)!}$ possible sequences in the l-th class. Following [24,26] we get the following equation from eq. (1):

$$\begin{aligned} \frac{\mathrm{d}\hat{p}_l(t)}{\mathrm{d}t} &= (f(x_l,t) - \mu - R(t))\,\hat{p}_l \\ &\quad + \mu \left(\hat{p}_{l-1} \frac{L-l+1}{L} + \hat{p}_{l+1} \frac{l+1}{L} \right), \end{aligned} \tag{2}$$

where we introduce the fitness function $r_l = f(x_l,t)$, $R(t) = \sum_l f(x_l,t)\hat{p}_l$ is the mean fitness, $\sum_l \hat{p}_l = 1$, and we define $x_l = 1 - 2l/L$. The coefficients $\frac{l+1}{L}$, $\frac{L-l+1}{L}$ arise because we work with the Hamming class probabilities instead of single sequence probabilities [24,26]. For $l = 0$, we drop the $\hat{p}_{l-1}$ term in eq. (2) and drop the $\hat{p}_{l+1}$ term for $l = L$. Later in the article we take $\mu = 1$ for the simplicity of the formulas.

We can omit the nonlinear terms $R\hat{p}_l$ in eq. (2), solve the system of equations and obtain a solution p_l. Then we can recover the solution for $\hat{p}_l$ simply as

$$\begin{aligned} &\frac{\mathrm{d}p_l(t)}{\mathrm{d}t} = (f(x_l,t) - 1)\,p_l + p_{l-1}\frac{L-l+1}{L} + p_{l+1}\frac{l+1}{L}, \\ &\hat{p}_l(t) = \frac{p_l}{\sum_n p_n}, \end{aligned} \tag{3}$$

Now we describe the fluctuations of the landscape. We have the first landscape for the period of time T_1 and T_2 for the duration of the second fitness landscape.

We take the fitness $f(x,t) = f_1(x)$ for the first environment with a period T_1, having a distribution

$$\pi_1(T_1) = \alpha_1 \exp(-\alpha_1 T_1)$$

and we look at the ordinary CK equation, eq. (2) in the interval $0 < t < T_1$.

Then we get new duration period T_2 for the choice of fitness $f(x,t) = f_2(x)$ with the probability density

$$\pi_2(T_2) = \alpha_2 \exp(-\alpha_2 T_2)$$

and continue the solution in the interval $T_1 < t < T_1 + T_2$, using $f(x,t) = f_2(x)$. Let us denote the i-th period of epoch as $\hat{T}_i$, where either $\hat{T}_i = T_1$ or $\hat{T}_i = T_2$. We perform an averaging, obtaining the mean fitness

$$R = \frac{\sum_i R_i \hat{T}_i}{\sum_i \hat{T}_i},$$

where $R_i = \int_{t_{i-1}}^{t_i} \sum_l r_l^i \hat{p}_l(t) \mathrm{d}t / \hat{T}_i,\ t_i = \sum_{l=1}^{i} \hat{T}_i$ and r_l^i represents the fitness of the l-th sequence during the i-th epoch, with $r_l^i = f_1(1-2l/N)$ for the first fitness landscape and $f_2(1-2l/N)$ for the second fitness landscape.

The dynamics. –

Dynamics with initial δ function distribution for a static landscape. Equation (3) is a large system of ordinary differential equations (ODE). We assume that the fitness function $f(x)$ is a smooth function of x, so $f(x_l) - f(x_{l+1}) \sim 1/L$. With relative accuracy $O(1/L)$ we map the system of ODEs to the partial differential equation. To solve the dynamics of the Crow-Kimura model (eq. (3)) with the static fitness landscape $f(x_l, t) = f(x_l)$, we used the Hamilton-Jacobi equation version of the model [30]. We make an ansatz,

$$p_l(t) = \exp[LU(x,t)], \tag{4}$$

where the coordinate $x = 1 - 2l/L$ ranges over the interval $[-1, 1]$. We just introduced a continuous function $U(x,t)$ to describe the set of p_l. $U(x,t)$ is the action function if we look at a related Hamilton-Jacobi equation, or it can be interpreted as an energy in the stochastic thermodynamic approach.

Then we immediately get

$$\begin{aligned} p_{l-1}(x,t) &= \exp[LU(x + 2/L, t)] \approx \exp[LU(x,t) + 2U'],\\ p_{l+1}(x) &= \exp[LU(x + 2/L, t)] \approx \exp[LU(x,t) - 2U'], \end{aligned}$$

where $x = 1 - 2l/L, U' = \frac{\partial U}{\partial x}$. Using the last equation and $l = L(1-x)/2, (L-l) = L(1+x)/2$ from eq. (3), we obtain the following Hamilton-Jacobi equation:

$$\frac{\partial U(x,t)}{\partial t} = f[x] - 1 + \frac{1+x}{2} e^{2U'} + \frac{1-x}{2} e^{-2U'}. \tag{5}$$

Consider the relaxation after large period of time. Since eq. (3) is a linear system of differential equations, therefore, the asymptotic solution is $p_l \sim e^{Rt} g_l$, where R is the maximal eigenvalue of the matrix in eq. (3). Taking a logarithm of the last equation we get an asymptotic formula for the solutions of eq. (4),

$$U(x,t) = Rt + \hat{U}(x). \tag{6}$$

We got the following equation for the asymptotic (steady-state) distribution:

$$R = f[x] - 1 + \frac{1+x}{2} e^{2\hat{U}'(x)} + \frac{1-x}{2} e^{-2\hat{U}'(x)}. \tag{7}$$

We calculate $\hat{U}'(x)$ from eq. (7) and then integrate to obtain $\hat{U}(x)$ in the steady state. We define the surplus of the distribution as $s = \sum_l p_l(1 - 2l/L)$. If we assume that the distribution by eq. (4) has a maximum at some point x, then $s = x$ with the accuracy $O(1/L)$. Putting $\hat{U}'(s) = 0$ in eq. (7), we get for the surplus a simple equation [26],

$$f(s) = R.$$

We calculate steady state mean fitness R looking for the minimum of the r.h.s of eq. (7) via $\hat{U}'$, then the maximum via x,

$$R = \max[f(x) + \sqrt{1 - x^2} - 1]. \tag{8}$$

We need to solve the Hamilton-Jacobi equation (eq. (5)) to obtain the dynamics of the model. Having the Hamiltonian-Jacobi equation, we define the characteristics $y(t)$ with a parameter q, see [33],

$$\begin{aligned} \mathrm{d}y/\mathrm{d}t &= (1+y)e^{2v} - (1-y)e^{-2v},\\ q &= f(y) - 1 + \frac{1+y}{2} e^{2v} + \frac{1-y}{2} e^{-2v}, \end{aligned} \tag{9}$$

where $v = U'$.

For the given initial distribution $U_0(x)$, we can immediately obtain the value of $U(x,t)$ at point x using the equation

$$U(x,t) = U_0(x_0) + \int_{x_0}^{x} v(y) \mathrm{d}y + qt, \tag{10}$$

where y is the trajectory of the characteristics that starts at the point x_0 and arrives at point x at time t. Therefore, the characteristics are defined by x_0 and q.

With the solution for $U(x,t)$ obtained from eqs. (9) and (10), we can determine the dynamics of the distribution $p_l(t)$ using eq. (4). If $U(x,t)$ has a maximum at point $x = s$, then $p_l(t)$ has a maximum at $l = N(1-s)/2$.

The dynamics with a smooth initial distribution. Equation (7) describes the distribution via the number of mutations after a long period of time for a fitness function $f(x)$. Consider a population in steady state with the fitness function $f_1(x)$, which evolves according to the model with fitness function $f_2(x)$. We will investigate how the mean number of mutations changes for the population distribution.

For the initial distribution, we have

$$R_1 = f_1(x) - 1 + \frac{1+x}{2} e^{2v} + \frac{1-x}{2} e^{-2v}, \tag{11}$$

where we denote $v = \hat{U}'(x)$ and the steady state mean fitness R_1 and the surplus s_1 are defined by the equations

$$\begin{aligned} R_1 &= \max[f_1(x) - 1 + \sqrt{1 - x^2}]_x,\\ f_1(s_1) &= R_1. \end{aligned} \tag{12}$$

The initial distribution has a maximum at the point s_1.

For the dynamics of the model with fitness function $f_2(x)$, we examine the equation along the characteristics,

$$q = f_2(x) - 1 + \frac{1+x}{2}e^{2v} + \frac{1-x}{2}e^{-2v}. \tag{13}$$

We can examine the characteristics arriving at the maximum point with $q = f_2(x)$ and calculate the time.

As the characteristics are defined by the parameter q and the starting point x_0, we need to find x_0 for the trajectory arriving at point x. Equation (13) is valid for point x_0, so we set $x = x_0$. We obtain x_0 from the system of eqs. (11) and (13),

$$R_1 - f_1(x_0) = q - f_2(x_0) \equiv f_2(x) - f_2(x_0). \tag{14}$$

We verify that x_0 changes with x, whereas in [30], we used the same x_0 for different values of x.

We define two functions for x_0: $x_{02}(x)$ for the case of eq. (14) and $x_{01}(x)$ for the alternative case,

$$\begin{aligned} R_1 - f_1(x_{02}(x)) &= f_2(x) - f_2(x_{02}(x)), \\ R_2 - f_2(x_{01}(x)) &= f_1(x) - f_1(x_{01}(x)). \end{aligned} \tag{15}$$

Figures 1, 2 correspond to a scenario in which the population distribution is relaxed at fitness $f_1(x)$ before time $t = 0$, after which the fitness changes to $f_2(x)$.

By integrating the ODE between points $x_{02}(x)$ and x, we obtain

$$\begin{aligned} t &= \frac{1}{2}\int_{x_{02}(x)}^{x} \mathrm{d}y \frac{\mathrm{d}y}{F(f_2(x), y)}, \\ F(f(x), y) &= \sqrt{(f(x) + 1 - f(y))^2 - 1 + y^2}. \end{aligned} \tag{16}$$

A more advanced situation is possible with two characteristics.

For the two-characteristics solution, we have

$$t = \frac{1}{2}\int_{x_0}^{x_1} \frac{\mathrm{d}y}{F(f_2(x), y)} + \frac{1}{2}\int_{x}^{x_1} \frac{\mathrm{d}y}{F(q, y)}, \tag{17}$$

where x_1 is determined by the equation

$$(q + 1 - f_2(x_1))^2 - 1 + (x_1)^2 = 0 \tag{18}$$

and $q = f_2(x)$. Thus, we examine the dynamics first along one characteristic between the points x_0 and x_1, and then along the other characteristic between the points x_1 and x.

Our solution in eq. (17) becomes applicable after the first term in eq. (16) disappears, *i.e.*, when $x_0 = x_1$, so

$$\begin{aligned} &F(f_2(x), x_1) = 0, \\ &f_1(x_1) - R_1 = f_2(x_1) - f_2(x), \end{aligned} \tag{19}$$

Then the corresponding time $\hat{T}_2$ is defined as

$$\hat{T}_2 = \frac{1}{2}\int_{x_0}^{x_1} \frac{\mathrm{d}y}{F(f_2(x), y)}. \tag{20}$$

If x given by eq. (18) is outside the interval (s_1, s_2), then we have only the solution by eq. (16).

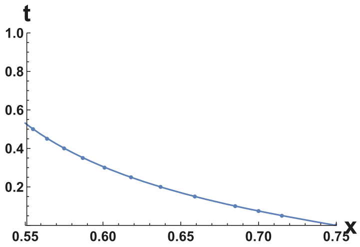

Fig. 1: The dynamics of the maximum of distribution $s = x$. The vertical axis corresponds to time, the horizontal one to x. The smooth line is our theoretical result by eq. (16), the solid dots are given by numerics. $f_2(x) = x^2, L = 1000, f_1(x) = 2x^2$. Before the time $t = 0$ the fitness function $f_1(x)$ has been chosen for the long period of time; later $f_2(x)$ starts at $t > 0$.

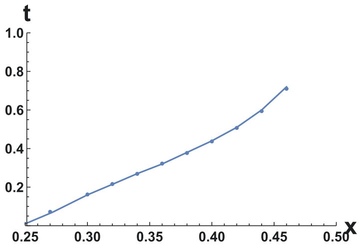

Fig. 2: The dynamics of the maximum of distribution $s = x$. The vertical axis corresponds to time, the horizontal one to x. The smooth line is our numerical result, the solid dots are given by our analytics eq. (17). $f_1(x) = (2/3)x^2, L = 1000, f_2(x) = x^2$. Before the time $t = 0$, the fitness function $f_1(x)$ has been chosen for a long period of time; later $f_2(x)$ starts at $t > 0$. We use a solution with one characteristics until the point $x < 0.433$ and two characteristics for $x > 0.433$.

For time periods $t < \hat{T}_2$, we take $X_2(t)$ to be the inverse of $t(x)$ given by eq. (16), and later use the inverse of $t(x)$ given by eq. (17). We define $X_1(t)$ in the same way replacing $f_2 \to f_1, f_1 \to f_2, R_1 \to R_2$, and we numerically derive the inverse functions of $t(x)$.

Our analytics is well confirmed by numerics, see figs. 1, 2.

The fluctuation load. – In the case of a fluctuating fitness landscape, we first look at an initial distribution given by eq. (10), or the equivalent form with the change $f_1(x) \to f_2(x), R_1 \to R_2$. We look at the dynamics for long periods of times, randomly drawn from the exponential distribution, with the parameters α_1, α_2. The first landsape has a probability $\frac{\alpha_2}{\alpha_1+\alpha_2}$, the second one $\frac{\alpha_1}{\alpha_1+\alpha_2}$. Using the exponential distribution of the epoch periods and ignoring $O(\alpha^2)$ terms, we get

$$\begin{aligned} R &= R_1\frac{\alpha_2}{\alpha_1+\alpha_2} + R_2\frac{\alpha_1}{\alpha_2} - \Delta, \\ \Delta &= \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\int_0^\infty \mathrm{d}t(f_1(X_1(t)) - f_1(s_1)) \\ &\quad + \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\int_0^\infty \mathrm{d}t(f_2(X_2(t)) - f_2(s_2))), \end{aligned} \tag{21}$$

where s_1 is the surplus in the statics with the first fitness landscape, and s_2 for the second fitness case. The derivation of Δ in eq. (21), supported by eqs. (15)–(20), is the main result of the work.

How many different analytical expressions for Δ exist? Throughout our investigations (monotonic fitness functions in the interval $[0, 1]$) we used 1 characteristic solution for the dynamics on one fitness landscape and 2 characteristics solution for the dynamics with an alternative choice of a steeper landscape.

In principle, there could be a situation with two characteristic solutions in evolution on both landscapess, so at least 3 different solutions are possible.

Let us make some speculations, looking for some formal analogies with information therodynamics [38]. There one considers the dynamics around several vacuums, with F_i defined as a free energy of the statistical physics dynamics around the i-th vacuum. If p_i is the probability of the vacuum choice, then the effective free energy is defined as

$$F = p_1 F_1 + p_2 F_2 - T \ln 2. \tag{22}$$

We can map our eq. (21) to eq. (22), identifying R_i with the local free energies around the given vacuum. Our speculation is meaningful for $\Delta > 0$, which, as we verified, is the case. Figure 3 illustrates the accuracy of our analytical result, eq. (21).

Random transitions between many landscapes. Consider the case of multiple landscapes with each period of residence in the i-th landscape given by $\alpha_i \exp(-T_i \alpha_i)$. We assume there are transition probabilities Q_{ij} between the landscapes.

We examine the Markov process with the given transition probabilities and calculate the set of probabilities of the system moving from the i-th landscape to the j-th. We denote this probability as y_{ij}.

If the steady-state distribution of the discrete Markov process with transition probabilities Q_{ij} is x_i, then we obtain $y_{ij} = x_i Q_{ij}$. The mean fitness is given by

$$\sum_i \frac{x_i R_i}{P \alpha_i} + \sum_{i,j} \frac{x_i Q_{ij}}{P} \int_0^\infty \mathrm{d}t (f_j(X_{ij}(t)) - f_j(S_j))). \tag{23}$$

Here, $P = \sum_i \frac{x_i}{\alpha_i}$ and $X_{ij}(t)$ is the mean surplus in the j-th landscape when we have the steady-state distribution of the i-th landscape at the start.

The fast transitions between landscapes. – Our aim is to derive analytical expressions for the average fitness, as explained in the introduction. Equations (21) and (23) give the first-order correction in the transition rate to the mean fitness expression, assuming weak transition rates α. We now seek an alternative expression for the opposite case of fast transitions, where the epoch mean length h is smal using the Runge-Kutta scheme for numerical integration.

We consider the HJE for a time period t_1 with the fitness function $f_1(x)$ and then t_2 with the function $f_2(x)$. We want to calculate the small time periods, so we need to find $u(x,t) = u(x,0) + At + Bt^2 + Ct^3$. To solve the differential equation $\mathrm{d}y/\mathrm{d}t = g(y)$ in the interval t, we use a standard 3rd-order Runge-Kutta scheme,

$$k_1 = g(y)t, \quad k_2 = g(y + k_1/2)t, \quad k_3 = g(-k_1 + 2k_2)t,$$
$$y(t) = y(0) + \frac{k_1 + 4k_2 + k_3}{6}.$$

Using this method, we obtain the following expression for the effective Hamiltonian for the HJE of this system:

$$\begin{aligned} & f[x] \\ & +(1 + df(x)h)\frac{1+x}{2}e^{2p} + \frac{1-x}{2}e^{-2p}(1 - df[x]h) \\ & + \left(\frac{1+x}{2}e^{2p} + \frac{1-x}{2}e^{-2p}\right)\frac{1}{\alpha} df^2[x]\frac{7}{24}, \end{aligned} \tag{24}$$

where $df(x) = f_2'(x) - f_1'(x)$, $f(x) = (f_1(x) + f_2(x))/2$, and $p = U'$. To find the steady state, we look for the minimum of eq. (23) via p, then look for the maximum via x. We treat the last term in eq. (24) as a perturbation, and use it to derive the following expression for the mean fitness:

$$R = f(x) - 1 + \sqrt{1 - x^2}\left(1 + \frac{1}{\alpha} df^2(x)\frac{7}{24}\right). \tag{25}$$

Here, x is taken as the maximum point for the function $f(x) - 1 + \sqrt{1 - x^2}$. Figure 4 illustrates the accuracy of our analytical result, eq. (25) derived in the limit of small $1/a$.

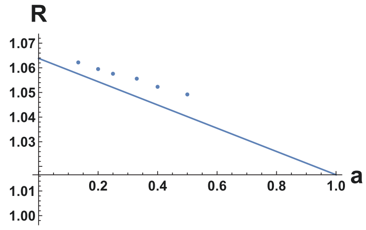

Fig. 3: The average fitness *vs.* the transition rate a, $f_1(x) = (2/3)x^2$, $L = 200$, $f_2(x) = 3x^2/2$. The smooth line is our analytical result, the solid dots correspond to our numerics. The accuracy is better for the small transition rates.

Exact expression of the scaling index for the model with several gene case. – In the previous sections we derived approximate expressions for the average fitness of population, while looking for the evolution on fluctuating fitness landscape. We have focused on the long genome length case. Here we consider several genomes. We are interested in exploring possible phase transitions in the problem. We do not need a Hamilton-Jacobi equation approach and to avoid confusion use different notations from those used in the previous section.

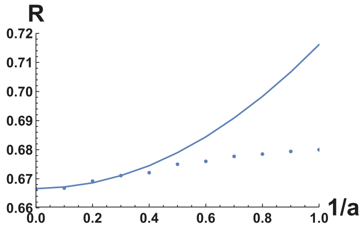

Fig. 4: $f_1(x) = 2x^2$, $L = 200, f_2(x) = 4x^2/2$. The smooth line is our analytical result, the solid dots correspond to our numerics.

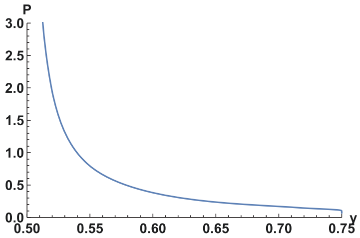

Fig. 5: The probability distribution $p_1(y)$, for the model by eq. (21) with random transitions between the landscapes and the parameters $J_1 = 2, J_2 = 4, \alpha = 0.1$.

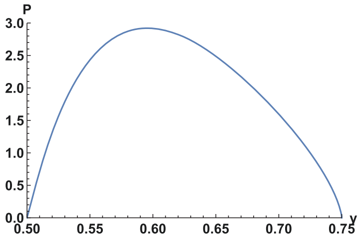

Fig. 6: The probability distribution $p_1(y)$, for the model by eq. (21) with random transitions between the landscapes and the parameters $J_1 = 2, J_2 = 4, \alpha = 2$.

A single peak fitness case. Let us solve eq. (2) for $L = 1$ and denote $r_0 = J$, so we have a zero fitness r_l for $l \geq 1$ and we take $\mu = 1$,

$$\frac{\mathrm{d}y}{\mathrm{d}t} = (J - 1 - Jx)x. \tag{26}$$

The solution for $y(t) \equiv p_0(t)$ for a static landscape with a single gene is

$$y(t) = \frac{(J-1)}{J(1 - \exp(-(J-1)(t-t_0)))}, \tag{27}$$

where y_0 is the solution at the $t = t_0$. After a long period of time we have an asymptotic expression,

$$y(t) = Y(t) \equiv y_c + c\exp[-k(t-t_0)], \tag{28}$$

$k = J - 1, y_c = 1 - 1/J$.

While considering the transitions with the rate α between 2 fitness landscapes with the peak fitnesses J_1, J_2, we have a distribution in the support $[y_{1c}, y_{2c}]$, where $y_{1c} = 1 - \frac{1}{J_1}$ and $y_{2c} = 1 - \frac{1}{J_2}$ are the steady state points in the first and second landscapes. In figs. 5, 6 we give the distribution for $p_1(y)$, the value of x when the system is in the first landscape. The $p_2(y)$ has a similar distribution, with a peak at y_{2c}. At small α we have a peak at $y = y_{1c}$, fig. 5, later the singularity disappears, fig. 6. We identify the second case with the transient phase, as the system has lost the memory about first landscape.

Let us derive the scaling behavior near the peak. By considering $p(y) = \langle \delta(y - Y(t)) \rangle$, we obtain

$$\int \mathrm{d}t\alpha\exp(-\alpha t)\delta(y - Y(t)) = \alpha\exp(-\alpha t)/Y'(t), \tag{29}$$

where $Y(t)$ is given by eq. (28). We are interested in the behavior of x near the point $y \to y_c$. We obtain

$$\begin{aligned} p(y) &= \frac{\alpha}{y_c}\exp((k-\alpha)(tt_0)) = \frac{1}{(y-y_c)^{n-1}}, \\ n &= \alpha/k. \end{aligned} \tag{30}$$

We get an exact scaling index [10].

Many genes in the genome. Let us consider an evolution model with more than three genes. The solution is described by the functions $Y_l(t) \equiv p_l$ for the fractions of genotype groups. By dropping the nonlinear terms, we can look at the right part as $\hat{A}\mathbf{p}$, where $\hat{A}$ is a quadratic matrix. Its maximal eigenvalue is just the steady state mean fitness. We denote the next eigenvalue by $R - \lambda$. Then for the general case, we assume the following asymptotics for the solution:

$$Y_l \to y_l e^{Rt} + c_l e^{(R-\lambda)t}, \tag{31}$$

where c_l is proportional to the second eigenvalue's related eigenvector. We can use the method of the previous subsection, just replacing k by $R - \lambda$. Then we get that the mean fitness distribution will change its behavior when

$$\alpha = \lambda. \tag{32}$$

Consider the model with three genomes, with the fitnesses J_0, J_1, J_2, and mutation rates μ from the first to second and from the second to third,

$$\begin{aligned} \frac{\mathrm{d}y_0}{\mathrm{d}t} &= (J_0 - \mu - R)y_0 = 0, \\ \frac{\mathrm{d}y_1}{\mathrm{d}t} &= (J_1 - \mu - R)y_0 + \mu y_1 = 0, \\ R &= J_0 y_0 + J_1 y_1 + J_2(1 - y_0 - y_1). \end{aligned} \tag{33}$$

We take $J_0 > J_1 > J_2 + \mu$. The we get simply $R = J_0 - \mu, \lambda = J_0 - J_1$. We verified numerically that the behaviour of the distribution is changed when $\alpha > \lambda$.

Conclusion. – In this article, we tackled the evolutionary dynamics of a quasispecies model on a fluctuating fitness landscape and derived an exact transition point to the transient phase for general evolution models. The first result is significant, as the quasispecies model has broad applications ranging from virus evolution to cancer and artificial intelligence [34,35]. The second result about the transient phase has a similar mathematical framework to other models, from ecology [36] to random matrix products [37]. Our focus was on a specific case of the Crow-Kimura model with random transitions between two landscapes and slow transition rates. The problem is analytically challenging, and we used our previous results for the dynamics of the Crow-Kimura model on static fitness landscapes [30]. Our goal was to calculate the mean fitness of the model. This quantity is equal to the mean fitness of the population averaged via several landscape cases, minus a quantity called the load, calculated by eq. (21) for the small transition rate α and for the high transition rate $h \equiv 1/\alpha$, see eq. (25). To solve the model, we first mapped our large system of ordinary differential equations to the Hamilton-Jacobi equation and then solved the latter using the method of characteristics. We observed an analogy with information thermodynamics, where different landscapes are the analogs of different vacuums in thermodynamics, while the load is the analog of the entropy of the ensemble of vacuums. We also explored the transient phase in the Crow-Kimura model, when the mutation-selection balance on one landscape ceases to work. Our mathematical result is intuitive: the transient phase arises when the transition rate from that landscape case is higher than the gap between two eigenvalues of the evolutionary dynamics matrix, which is the relaxation rate in the static landscape. Our eqs. (14)–(20) can be applied to derive the Berry phase in evolution.

* * *

DBS thanks for the support the SCS of Armenia, grants Nos. 20TTAT-QTa003 and 21T-1C037 and Faculty Research Funding Program 2022 implemented by the Enterprise Incubator Foundation with the support of PMI Science. ML thanks for the support National Institutes of Health, R35-GM122566-01, 2017-2022, National Science Foundation, DBI-2119963, 2021-2026, Moore and Simons Foundations, Grant 735927, 2020-2023.

Data availability statement: All data that support the findings of this study are included within the article (and any supplementary files).

REFERENCES

[1] Kussell E. and Leibler S., *Science*, **309** (2005) 2075.

[2] Parmesan C., *Annu. Rev. Ecol. Evol. Syst.*, **37** (2006) 637.

[3] He Q., Mobilia M. and Täuber U. C., *Phys. Rev. E*, **82** (2010) 051909.

[4] Assaf M. E., Roberts E. E., Luthey-Schulten Z. and Goldenfeld N., *Phys. Rev. Lett.*, **111** (2013) 058102.

[5] Mustonen V. and Lassig M., *Phys. Rev. Lett.*, **100** (2008) 108101.

[6] Mustonen V. and Lässig M., *Trends Genet.*, **25** (2009) 111.

[7] Mustonen V. and Lässig M., *Proc. Natl. Acad. Sci. U.S.A.*, **107** (2010) 4248.

[8] Dobramysl U. and Täuber U. C., *Phys. Rev. Lett.*, **110** (2013) 048105.

[9] Assaf M., Mobilia M. and Roberts E., *Phys. Rev. Lett.*, **111** (2013) 238101.

[10] Hufton P. G., Lin Y. T., Galla T. and McKane A. J., *Phys. Rev. E*, **93** (2016) 052119.

[11] Yuan R., Zhu X., Wang G., Li S. and Ao P., *Rep. Prog. Phys.*, **80** (2017) 042701.

[12] Skanata A. and Kussell E., *Phys. Rev. Lett.*, **117** (2016) 038104.

[13] Mayer A., Mora T., Rivoire O. and Walczak A. M., *Proc. Natl. Acad. Sci. U.S.A.*, **113** (2016) 8630.

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

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

[16] Ancliff M. and Park J. M., *Phys. Rev. E*, **80** (2009) 061910.

[17] Ancliff M. and Park J. M., *Phys. Rev. E*, **82** (2010) 021904.

[18] Neves A. G. M., *Phys. Rev. E*, **82** (2010) 031915.

[19] Wonpyang G., *J. Korean Phys. Soc.*, **57** (2010) 287.

[20] Rivoire O., *J. Stat. Phys.*, **162** (2016) 1324.

[21] Mayer A., Mora T., Rivoire O. and Walczak A. M., *Proc. Natl. Acad. Sci. U.S.A.*, **113** (2016) 8630.

[22] Saakian D. B., *Phys. Rev. E*, **98** (2018) 062115.

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

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

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

[26] Baake E. and Wagner H., *Genet. Res.*, **78** (2001) 93.

[27] Hermisson J., Redener O., Wagner H. and Baake E., *Theor. Popul. Biol.*, **62** (2002) 9.

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

[29] Saakian D. B., *J. Stat. Phys.*, **128** (2007) 781.

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

[31] Saakian D. B. and Hu C.-K., *Phys. Rev. E*, **69** (2004) 046121.

[32] Saakian D. B., Yakushkina T. and Hu C.-K., *Sci. Rep.*, **6** (2016) 34840.

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

[34] Schuster P. and Domingo E., *Quasispecies: From Theory to Experimental Systems* (Springer) 2016, `https://doi.org/10.1007/978-3-319-23898-2`.

[35] Burger R. and Lynch M., *Evolution*, **49** (1995) 151.

[36] Hufton P. G., Ting Lin Y., Galla T. and McKane A. J., *Phys. Rev. E*, **93** (2016) 052119.

[37] Saakian D. B., *Phys. Rev. E*, **98** (2018) 062115.

[38] Parrondo J. M. R., Horowitz J. M. and Sagawa T., *Nat. Phys.*, **11** (2015) 131.