Evolution models with lethal mutations on symmetric or random fitness landscapes

Zara Kirakosyan, David B. Saakian, and Chin-Kun Hu

1Yerevan Physics Institute, Ali Khanian Brothers Street 2, Yerevan 375036, Armenia
2Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
3National Center for Theoretical Sciences: Physics Division, National Taiwan University, Taipei 10617, Taiwan
4Center for Nonlinear and Complex Systems and Department of Physics, Chung-Yuan Christian University, Chungli 32023, Taiwan

(Received 1 March 2010; published 7 July 2010)

We calculate the mean fitness for evolution models, when the fitness is a function of the Hamming distance from a reference sequence, and there is a probability that this fitness is nullified (Eigen model case) or tends to the negative infinity (Crow-Kimura model case). We calculate the mean fitness of these models. The mean fitness is calculated also for the random fitnesses with logarithmic-normal distribution, reasonably describing the situation with RNA viruses.

DOI: 10.1103/PhysRevE.82.011904

PACS number(s): 87.23.Kg, 87.10.Mn

I. INTRODUCTION

The investigation of evolution models [1–3] is one of the most fruitful applications of statistical mechanics [4–17]. The main activity is connected with the investigation of quasispecies model [1,2], where there is a direct connection to the simple statistical models of spin chains, with fitness being the analog of energy (with a minus sign) and mean fitness being an analog of free energy. It has considered continuous time models with connected mutation-selection scheme [2] (mutations are during the birth) and with parallel scheme (mutations are during the life) [3,11], as well as discrete time models [4]. The real fitness landscapes (distribution of fitness for different genomes) are highly complicated with both rugged (the fitness could be drastically changed after a few mutations) and neutral (the fitness could be unchanged after specific mutations) features. Usually one considers a simple version of selection, when fitness depends on a Hamming distance from some reference sequence, and there is a mutation which changes the sequence, keeping its length constant. Such a simplification of fitness landscape is popular in population genetics. We slightly aim to modify such a choice of fitness landscapes: it is originally symmetric, and then with some probability some of sequences become lethal ones. Even such a simple case has not been well investigated. While the existence of lethal mutants is well established experimentally [18,19], there are rather few theoretical papers: approximate results [20–23] and recent exact results for truncated selection case (all the mutants after some number of mutations are lethal) [24]. The role of lethal mutants was investigated especially for the problem of error threshold. In this work we will first of all carefully define what we mean by a model with lethal mutants; first consider the simplest case of Sec. II and important scaling by Eq. (8), then more involved cases of Crow-Kimura and Eigen continuous time models. We will consider the problem of lethal mutants for the case of general fitness landscape, where some sequences are randomly chosen as lethals, while other sequences correspond to the viable mutants. We will give analytical results for this case and clarify how the error threshold changes in the case of lethal mutants. We will see that the existence of lethal mutants (40% according to [18]) could substantially change the error threshold (about 50%).

For simplicity we consider the case of two-letter alphabet, where the genome is identified with the chain of N two value spins $s_i = \pm 1$, $1 \leq i \leq N$, and there are $M = 2^N$ such configurations (sequences). We denote the probability of the $i$th genotype as $P_i$.

In the case of the Eigen (connected) model, we consider

$$\frac{dP_i}{dt} = \sum_j Q_{ij} P_j - P_i \langle \sum_j r_j P_j \rangle.$$  \hspace{1cm} (1)

where the matrix $Q_{ij}$ describes the transition probabilities from the type $j$ to the type $i$. The elements of mutation matrix are $Q_{ij} = q^{N-d(i,j)}(1-q)^{d(i,j)}$; $d(i,j)$ is the Hamming distance between sequences $j$, $i$ and $q$ is the probability of errorless replication per nucleotide. The Hamming distance $d_{ij}$ between configurations $i$ and $j$ is the number of point mutations between these configurations. The diagonal terms of the mutation matrix are $Q_{ii} = q^N = Q = e^{-\gamma}$, where $\gamma = -N \ln(q) \approx N(1-q)$ is the parameter of mutation in the Eigen model. $r_i$ is identified as a fitness of the sequence $i$. As index $i$ is equivalent to the set of indices $s_1, \ldots, s_N$, we can define $r_i$ through the function $f_0$,

$$r_i = f_0(s_1, \ldots, s_N),$$  \hspace{1cm} (2)

where $s_1, \ldots, s_N$ correspond to the sequence $i$. We choose the zeroth sequence with all up spins, $s_i=1$. Here, $P_i$ satisfies normalization condition $\sum_{i=1}^{M} P_i = 1$.

How to define the lethal mutants? In the case of the Eigen model we simply take $r_i = 0$. Our choice means that the lethal mutant has no offspring. We will use this definition of fitness for lethal mutants to define the corresponding fitness in the Crow-Kimura model. In Sec. II we will define the proper fitness for the parallel model and the dilution of the sequence space by lethal mutants. The definition of the lethal mutants is rather direct. In Sec. III we will define lethal mutants, following the ideas of Sec. II, and calculate the mean fitness for the Crow-Kimura model. In Sec. IV we will use the methods of Sec. III to the Eigen models with symmetric

--

huck@phys.sinica.edu.tw

1539-3755/2010/82(1)/011904(5) 011904-1 ©2010 The American Physical Society
fitness landscapes. In Sec. V we will solve the model of uncorrelated random fitnesses with lethal mutants.

II. DISCRETE TIME EIGEN AND CONTINUOUS TIME CROW-KIMURA MODELS WITH LETHAL MUTANTS

A. Discrete time Eigen model

Consider now the discrete time Eigen model [4]. In the case of Crow-Kimura model there is mixing of generations, while in the discrete time Eigen model there is no such a mixing. It is possible to realize experimentally this scheme of evolution. Now the probabilities are defined at discrete time moments \( n \); thus, we consider \( P_{i,n} \) as a probability of type \( i \) at the \( n \)th moment of time. We have

\[
P_{i,n+1} = \frac{\sum_j Q_{ij} P_{j,n}}{\sum_j \hat{r}_j P_j(n)}.
\]

We can again define the lethal mutants as the sequences where \( \hat{r}_j = 0 \). We used notation \( \hat{r}_j \) for the fitness of this model. We will connect them with the corresponding fitness \( r_j \) in the Crow-Kimura model.

B. Crow-Kimura (parallel) model

It is well known [5,6] that the discrete time Eigen model could be mapped into the continuous time Crow-Kimura model [11], using the mapping

\[
P_j(n \tau) = P_{i,n},
\]

\[
\hat{r}_j = \exp(\tau \mu_0),
\]

\[
\gamma = \tau \mu_0,
\]

where \( \tau \) is the duration of an elementary time step and \( \mu_0 \) is a mutation rate per unit time period. We denote \( P_i = P_i(t) \) and get the equations of [11]

\[
dP_i = r_i P_i + \sum_j P_j \mu_{ij} - P_i (\sum_j r_j P_j),
\]

where \( \mu_{ij} \) describes the mutation phenomenon: \( \mu_{ii} = -\mu_0 N \), \( \mu_{ij} = 1 \) when the Hamming distance \( d_{ij} \) between two sequences is equal 1, and \( \mu_{ij} = 0 \) for the other case. The Hamming distance between two configurations is the number of different spins (alleles) between two genomes. There is a balance condition for probability,

\[
\sum_j \mu_{ij} = 0.
\]

Having the mapping by Eqs. (4), we get for the lethal mutants with \( \hat{r}_j = 0 \), therefore, \( r_j = \ln \hat{r}_j / \tau \rightarrow -\infty \). The number of offsprings should be positive. The point is that we can define the fitness in the Crow-Kimura model only with the additive constant \( c \) to all fitnesses to make all \( r_j \) non-negative. Thus, we used the discrete time version of Eigen model to explain the definition of fitness of lethal mutants in the Crow-Kimura model.

C. Dilution of sequence space by lethals

The next important point is the definition of the fraction of lethal sequences among all the \( M = 2^N \) sequences. First of all we define accurately the distribution of lethal mutants in the sequence space. The number \( M_v \) of nonlethal (viable) sequences scales as a power of the total sequence number,

\[
M_v \sim (2^N)^c,
\]

where \( c \) is a constant under the condition \( 0 < c < 1 \). We choose a reference sequence.

Consider the master sequence, having nonlethal \( N(1-\lambda) \) neighbors with single mutations, where \( 0 \leq \lambda < 1 \) is a parameter describing the fraction of lethal mutations. When the fitness is a function of Hamming distance from the reference sequence, we simplify the evolution equations using this symmetry. We assume that any sequence having at least one lethal mutation (plus some nonlethal mutations) is lethal. Therefore, at the \( l \)th Hamming class we have \( N_{i,\lambda} = N(1-\lambda)!/[N(1-\lambda)!] \) point mutants, and as a maximal \( l \) we take \( N(1-\lambda) \). For the small \( l \ll N \) there is a dilution of the sequence space via a factor \( (1-\lambda) \), while we have for the total number of viable sequences,

\[
M_v = \sum_{j=0}^{N(1-\lambda)} N_{j,\lambda} = 2^{(1-\lambda)N}.
\]

III. PARALLEL MODEL WITH LETHAL MUTANTS

In the Crow-Kimura model [3,11], one can calculate \( P_i \) from the solution of linear equations for \( x_j \),

\[
\frac{dx_i}{dt} = \sum_j A_{ij} x_j, \quad P_i = \frac{x_i}{\sum_j x_j},
\]

where the matrix \( A_{ij} \) corresponds to the linear terms in Eq. (5). The mean fitness in the steady state, \( R = \sum_j P_j r_j \), is identified with the maximal eigenvalue of the matrix \( A \). Thus, the calculation of mean fitness is equivalent to the search of the maximal eigenvalue of matrix \( A_{ij} \). This can be done using the quantum-mechanical approach [11].

The first equation in Eq. (9) could be mapped into the Schrödinger equation in the imaginary time [14] using a quantum Hamiltonian [11],

\[
H = -\gamma (1 - \sum_l \sigma_l^z / N) + f_0 (\sigma_1^x, \ldots, \sigma_N^x).
\]

where \( \sigma \) denotes the Pauli spin operator, and the mean fitness \( R \) can be expressed as [14]

\[
R = \lim_{\beta \to \infty} \frac{1}{N \beta} \ln Z(\beta), \quad Z(\beta) = \text{Tr} e^{-i\beta H},
\]

011904-2
where $\beta$ is the inverse temperature and we took the zero-
temperature limit to get the ground-state energy. To trans-
form the quantum statistical-mechanical problem into the one in classical mechanics, we use the identity
\begin{equation}
\text{Tr} e^{CH} = \text{Tr}[e^{CL} e^{HL}]^L
\end{equation}
at $L \to \infty$, where $C$ and $H$ are arbitrary operators. To calculate
the last expression, we introduce $(L+1)$ classical spins $s^i_l$, $1 \leq k \leq N$, $1 \leq l \leq L+1$, instead of quantum spins $\sigma^i_l$, $\sigma^c_l$
(this trick corresponds to the introducing the identity
$I = \sum_{a} |a\rangle\langle a|$ between any brackets $\langle \rangle$) [13,25]. We use the following transformation:
\begin{equation}
Z = \text{Tr} \left\{ \exp \left[ \frac{\mu \beta}{L} \sum_{k=1}^{N} (\sigma^i_k - 1) \right] \exp \left[ \frac{\beta}{L} f_0(\sigma^i_1, \ldots, \sigma^i_N) \right] \right\}^L.
\end{equation}

We take for the boundary configuration $s^i_1 = s^i_{L+1}$ and use a
representation of $\sigma^i_l$ in the basis of $|s\rangle$, $s = \pm 1$ [26],
\begin{equation}
\langle s_1 | e^{x \sigma^i_1} | s_2 \rangle = \cosh(x) e^{B(s_1 z_{s_2} - 1)}, \quad e^{-2B} = \tanh(x).
\end{equation}

For $\chi = \mu \beta / L \ll 1$ or $\cosh(x) \approx 1$, $\sinh(x) = x$, we have an
equality
\begin{equation}
\langle s_1 | e^{x \sigma^i_1} | s_2 \rangle = e^{B(s_1 z_{s_2} - 1)}, \quad e^{-2B} = x.
\end{equation}

Then
\begin{equation}
Z = \sum_{s_{i \in \{ \pm 1 \}}} \prod_{l} \exp \left[ - \beta N + \frac{NB}{L} \sum_{l} f_0(s^i_1, \ldots, s^i_N) \right]
\times \exp [B \sum_{n} (s^i_n + 1 - 1)].
\end{equation}

Equation (16) is valid for any fitness.

For the lethal mutant we have $r_i \to -\infty$. Let us introduce a
function $P(s_1, \ldots, s_N)$, where $P(s_1, \ldots, s_N) = 1$ for nonlethal
$2^{N(1-\lambda)}$ spin configurations, and $P(s_1, \ldots, s_N) = 0$ for the lethal
configurations.

We assume the following form of fitness function. We use
assumptions that for the nonlethal sequences the fitness is a
function of Hamming distance from the reference sequence,
\begin{equation}
f_0(s_1, \ldots, s_N) = \frac{1}{N(1-\lambda)} \sum_{l} s^i_l,
\end{equation}
and for the lethal configurations we have $f_0 \to -\infty$.

For the lethal sequences, we have $-\infty$ in the exponent.
Introducing auxiliary variables $h_l$, $m_l$, we derive [14]
\begin{equation}
Z = e^{-\beta \mu} \prod_{l} \int_{-\infty}^{\infty} dh_l \int_{-\infty}^{\infty} dm_l \exp \left[ - \frac{\beta \mu}{2} \sum_{l} h_l m_l (1-\lambda) \right]
\times \exp \left[ \frac{\beta B}{L} \sum_{l} f(m_l) \right]
\times \sum_{s_{i \in \{ \pm 1 \}}} \prod_{l} P(s^i_1, \ldots, s^i_N) \exp [B \sum_{n} (s^i_n + 1 - 1) + h_l s^i_l].
\end{equation}

Let us calculate the sum over spins. We have
\begin{equation}
\sum_{s_{i \in \{ \pm 1 \}}} \prod_{l} P(s^i_1, \ldots, s^i_N) \exp [B \sum_{n} (s^i_n + 1 - 1) + h_l s^i_l] = [Q_l]^{N(1-\lambda)},
\end{equation}
\begin{equation}
Q_l = \sum_{s_{i \in \{ \pm 1 \}}} \exp [B \sum_{n} (s^i_n + 1 - 1) + h_l s^i_l].
\end{equation}

While deriving the last result, we assumed simply that any
mutation from a group of $N\lambda$ spins is lethal. Without lethal
mutants, in Eq. (19) there was an expression $Q_l^{N}$, with $N$
being the number of spins; now we just replaced $N$ with $N(1-\lambda)$. $Q_l = z(B, \{h_l\}, L)$ is calculated in [14] as follows:
\begin{equation}
z(B, \beta, h(x)) = \text{Tr} \exp \left[ \int_{0}^{\beta} dx [h(x) \sigma^i + \sigma^c] \right].
\end{equation}

The exponent is time ordered on the right-hand side of Eq.
(20). Eventually we have
\begin{equation}
Z = \int \mathcal{D}M(\beta') \mathcal{D}H(\beta') \exp [N(1-\lambda) \ln Q_1]
+ N \int_{0}^{\beta} d\beta' f[M(\beta')] - N(1-\lambda) H(\beta') M(\beta') - N\beta] \right].
\end{equation}

In Eq. (21) we used the notation $M(\beta'), H(\beta')$ for $h_l, m_l$. At the
limit of large $\beta$ we consider the symmetric saddle-point
solution $M(\beta') = m$ and derive the following expression for
the mean fitness $R$:
\begin{equation}
R = \frac{\ln Z(\beta)}{N\beta} = \max_{m}[f(m) - \mu_0 + (1-\lambda) \mu_0 \sqrt{1-m^2}],
\end{equation}
where the magnetization parameter $m$ is defined as
\begin{equation}
m = \frac{1}{N(1-\lambda)} \langle \sum \sigma^i \rangle.
\end{equation}

Looking the maximum of Eq. (22) we define the value of $m$.

The observable variable, surplus, is defined in the steady
state as $s = \sum P_l (1-2d_0/N)$. We calculate it from the
expression
\begin{equation}
f(x) = R.
\end{equation}

Comparing the general $\lambda$ case in Eq. (22) with the $\lambda=0$ case, we have
\begin{equation}
R(\mu_0, 0) = R(\mu_0[1-\lambda], 0) - \lambda \mu_0.
\end{equation}

Thus, the mean fitness of the model with lethal mutants has
been expressed via the corresponding expression of the
standard Crow-Kimura model.

IV. EIGEN MODEL WITH LETHAL MUTANTS

The Eigen model [1,2] has been mapped into the quantum
spin model (see Eq. (6) in Ref. [13]),
\[ H = - \{ \exp[ - \gamma (1 - \sum \sigma_i^2/N) \hat{f}_0(\sigma_1, \ldots, \sigma_N)] \}. \]  (26)

In [15] we calculated the mean fitness \( R_e \), expanding \( e^{-\beta H} \) via \( \beta \) and using the Suzuki-Trotter form [22], as in the case of parallel model [13]. Let us follow an alternative simpler method.

Instead of \( \text{Tr} e^{-\beta H} \), we consider
\[ Z = \text{Tr}(\hat{H})^\beta = \text{Tr}(\exp[ - \gamma (1 - \sum \sigma_i^2/N) \hat{f}_0(\sigma_1, \ldots, \sigma_N)])^\beta. \]  (27)

Thus, at the limit of large \( \beta \),
\[ Z \rightarrow R_e^\beta, \]  (28)
where \( R_e \) is the mean fitness of Eigen model or the largest eigenvalue of the linear operator \(-H\).

Consider now the model with lethal mutants. We take the same distribution of lethal mutations as in the parallel model. For nonlethal configuration,
\[ r_i = \hat{f} \left( 1 - 2 \frac{d_{i0}}{N(1 - \lambda)} \right). \]  (29)

Equation (27) is equivalent to Eq. (13) with the mapping
\[ \beta = L, \quad \gamma = \mu_0, \quad \hat{f} = f. \]  (30)

Then the mean fitness in Eigen model, \( R_e = \sum \sigma_i P_i \), is mapped into the mean fitness \( R \) of parallel model,
\[ \ln R_e = R. \]  (31)

Using Eqs. (30) and (22), we find the following relation for the mean fitness:
\[ R_e(N[1 - q], \lambda) = Q^\lambda R([1 - \lambda][1 - q]N, 0), \]  (32)
where
\[ Q = q^N \]  (33)
is the probability of errorless replication of the whole genome, and the scale in the fitness landscapes is also changed [see Eq. (28)]. For \( R(N[1 - q], 0) \), Ref. [15] has derived an expression
\[ R_e = \max_m(\hat{f}(m)\exp[ \gamma(-1 + \sqrt{1 - m^2})]). \]  (34)

Thus, for the nonzero \( \lambda \) we get by combining Eqs. (32) and (34)
\[ R_e = \max_m(\hat{f}(m)\exp[ \gamma(-1 + (1 - \lambda)\sqrt{1 - m^2})]). \]  (35)

For the nonselective phase with \( m=0 \) we derive from Eq. (35)
\[ R_e = Q^\lambda. \]  (36)

We see that the mean fitness in the nonselective phase decreases after switching on the lethal mutations. Without the lethal mutations, the mean fitness was \( R=1 \), as after mutations the sequence was transformed into other sequence with the same fitness 1. Now after lethal mutations the mutant has zero fitness; therefore, the mean fitness decreases.

In the case of the single peak fitness \( f(1)=A \) and \( f(m) = 1 \) for \( m \leq 1 \); putting \( m=1 \) we get from Eq. (35)
\[ R_e = QA. \]  (37)

Thus, the mean fitness is unchanged due to lethal mutations, while the error threshold point is changed to
\[ AQ^{1-\lambda} = 1. \]  (38)

Thus, while the error threshold constraint becomes softer, the transition still exists. The transition disappears only at the extreme case \( \lambda \rightarrow 1 \).

\section*{V. Random Uncorrelated Fitness with Lethal Sequences}

The evolution models with uncorrelated random fitnesses are rather popular in statistical physics approach to evolution theory [4,27–31], where they are identified with the “rugged” fitness landscapes. We consider the case, when \( \ln r \) has a random uncorrelated distribution,
\[ p(r) = \frac{1}{k\sqrt{\pi}} \exp\left[ - \frac{N(\ln r)^2}{k^2} \right]. \]  (39)

Such a logarithmic-normal distribution for the fitnesses \( r \) has been confirmed by the experimental data of [18,19,32]. Thus, our aim is the solution of Eigen model with the distribution by Eq. (39).

Let us find the maximal fitness among all the \( 2^N \) sequences. This can be done either by using the extreme value statistics or by means of statistical mechanics of the random energy model (REM) by Derrida [33], considering the ground-state energy of the REM with the distribution of energies \( E_i = \ln r_i \) by Eq. (39). It has been found that [33]
\[ \ln r_o = k\ln\ln 2, \]  (40)
where \( k \) is some parameter. Equation (40) gives the maximum fitness value. The vast majority of other fitnesses are equal to 1.

As has been derived in [4], the error threshold is at
\[ r_oQ = 1, \]  (41)
where \( Q \) is the errorless replication probability. Let us make lethal some of the \( 2^N \) sequences, holding nonzero fitnesses only for \( M_e = 2^{N(1-\lambda)} \) sequences [see Eq. (8)]. Now the maximum fitness is changed. We have for the maximal fitness \( r_o \) in this case [33]
\[ \ln r_o = k\ln 2(1-\lambda). \]  (42)

The error threshold could be found by comparing \( r_o \) and \( Q^\lambda \) [4],
\[ Q^{1-\lambda} = e^{k\ln 2(1-\lambda)}. \]  (43)

Comparing the above equation with Eq. (38), we see that the introduction of lethal mutants changes the error threshold less than in case of a single peak fitness.
VI. CONCLUSION

In this paper we considered the evolution models with general symmetric fitness landscape, when sequences with some probability become lethal. While the lethal mutations are typical for viruses, there have been only few exact theoretical results. First of all one should carefully define how the sequence space of the model is diluted by lethal mutants. We defined this dilution via Eqs. (7) and (8). The next step is the definition of the lethal sequence in the case of parallel (Crow-Kimura) model. We defined the lethal mutations here using the discrete time Eigen model as an intermediate stage. We then calculated the mean fitness for the considered models. It is simply expressed via the corresponding expressions of the models without lethal mutants: one needs only to rescale the mutation rate. Equations (25) and (32) are simply generalized for the case of general mean fitness like fitness functions, when the fitness is a function of Hamming distances from the K reference sequences. We also solved the model where fitnesses are randomly distributed variables, giving the crude picture of experimental data.

In the case of RNA viruses [18] the probability of lethal mutations is about 40%. Thus, evolution characteristics may be substantially changed due to lethal mutations: for the single peak fitness the error threshold could be changed by 50% according to Eq. (38). This phenomenon could be especially important to solve the problem of the error catastrophe at the origin of life. It has been assumed to use the error catastrophe ideas for antiviral therapy [34,35]. Our accurate formulas could be used for the estimate of the error threshold in the case of lethal mutants, also for the calculation of critical mutation rate for the lethal mutagenesis [36–38].

Our another result is that the mean fitness could be seriously affected in the case of the nonselective phase. This phase is less important for biology than the selective one.

We derived our results for the scaling of the number of nonlethal mutants by Eq. (8), with additional hypothesis that any sequence is lethal if it carries at least one lethal mutation. Another (a bit abstract from the biological perspective) situation corresponds to the case of uniform distribution of lethal mutants in the whole sequence space. This case is mathematically much more involved, while it is tractable via a set of recursive equations like Eq. (21) in [15]. The profile of distributions via Hamming classes and the error threshold are modified compared with the case solved in the current paper. Another possible extension of our results is the investigation of the simplest evolution models with both neutral and lethal mutations.

ACKNOWLEDGMENTS

This work was supported by NSC Grant Nos. 96-2911-M 001-003-MY3 and 98-2811-M-001-080, Academia Sinica, and National Center for Theoretical Sciences in Taiwan.

[38] P. Chen and E. I. Shakhnovich, Genetics 183, 639 (2009).