Journal of Theoretical Biology 380 (2015) 53–59

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Statistical properties and error threshold of quasispecies on single-peak Gaussian-distributed fitness landscapes

Duo-Fang Li [a], Tian-Guang Cao [a], Jin-Peng Geng [a], Jian-Zhong Gu [b], Hai-Long An [a], Yong Zhan [a,*]

[a] Institute of Biophysics, School of Sciences, Hebei University of Technology, Tianjin 300401, China
[b] China Institute of Atomic Energy, P. O. Box 275 (10), Beijing 102413, China

HIGHLIGHTS

- On random fitness, the equilibrium distributions of sequence population are investigated.
- A linear relation of the error threshold width with the fitness fluctuation strength.
- Statistical properties of a given mutant class are studies.
- Study bimodal distributions around the error threshold in a random Eigen model.

ARTICLE INFO

Article history:
Received 17 December 2014
Received in revised form
29 March 2015
Accepted 8 May 2015
Available online 18 May 2015

Keywords:
Quasispecies
Error threshold
Single peak fitness
Gaussian distribution
Randomization

ABSTRACT

The stochastic Eigen model proposed by Feng et al. (2007) (Journal of Theoretical Biology, 246, 28) showed that error threshold is no longer a phase transition point but a crossover region whose width depends on the strength of the random fluctuation in an environment. The underlying cause of this phenomenon has not yet been well examined. In this article, we adopt a single peak Gaussian distributed fitness landscape instead of a constant one to investigate and analyze the change of the error threshold and the statistical property of the quasi-species population. We find a roughly linear relation between the width of the error threshold and the fitness fluctuation strength. For a given quasi-species, the fluctuation of the relative concentration has a minimum with a normal distribution of the relative concentration at the maximum of the averaged relative concentration, it has however a largest value with a bimodal distribution of the relative concentration near the error threshold. The above results deepen our understanding of the quasispecies and error threshold and are heuristic for exploring practicable antiviral strategies.

1. Introduction

The researches on the species evolution have been a focus for a long time. In the early 1970s, Eigen and Crow et al., based on Darwin's principle of nature selection, established respectively the Eigen model and the Crow–Kimura model to describe the evolution of the macromolecules (Eigen, 1971; Crow and Kimura, 1970). The Eigen model is a coupled mutation-selection model which describes the evolution of error-prone replicating molecules (Eigen, 1971). The Crow–Kimura model is known as a parallel mutation-selection model in which selection and mutation are two independent processes (Saakian et al., 2008). The Eigen model and Crow–Kimura model have appeared to be useful in understanding the origin and evolution of life. Their dynamics and equilibrium properties have been extensively studied in the past 40 years (Eigen et al., 1988; 1989; Saakian et al., 2004; Tarazona, 1992; Nilsson and Snoad, 2000). The Eigen model has two important predictions: the quasispecies (Eigen and Schuster, 1977, 1978) and error threshold (Eigen, 1971). The former means the distribution of some mutant sequences centered on the master sequence in equilibrium. The latter is a critical mutation rate above which all the macromolecule sequences lose their genetic information and are randomly distributed in the sequence space. The quasispecies and error threshold have been also confirmed in the evolution experiments of some viruses (Fishman and Branch, 2009; Wain, 1992; Domingo et al., 1992; Steinhauer et al., 1989).

In fact, the classic Eigen model was initially developed on the condition of infinite asexual population. To be closer to the reality,

* Corresponding author. Tel.: +86 02260438276.
E-mail address: yongz2013@163.com (Y. Zhan).

http://dx.doi.org/10.1016/j.jtbi.2015.05.014
0022-5193/

the theory was extended to the case of finite populations and diploid organism (Nowak and Schuster, 1989; Saakian and Hu, 2006; Wiehe et al. 1995). More realistic models were established by considering real biological effects, including genome recombination and repair effect, host immune response, complementation during coinfection and so on (Boerlijst et al., 1996; Tannenbaum and Shakhnovich, 2004; Kamp, 2003; Sardanyés and Elena, 2010). Furthermore, an empirical fitness landscape was introduced into the evolution model, which provided a broad framework of evolutionary dynamics for microbial population and cancer cells (Sardanyes et al., 2014; Rozen et al., 2002). Recently, the randomized study on biological processes both in theory and experiment has been quite active (Guo and Mei, 2014; Domingo et al., 1978). The stochastic versions of the Eigen model were built by physical methods, such as the Langevin equation and Markov process. The stochastic dynamics of species evolution were elucidated in these random models (Galstyan and Saakian, 2012; Inagaki, 1982). Nevertheless, the stochastic characteristics on the Eigen model have not been well understood. The species evolution for a real population, especially at the molecular level, is inevitably affected by various random factors, such as genetic mutations and environmental fluctuations (Neher and Shraiman, 2012). Therefore, the important physical parameters appearing in species evolution models are also subject to various random factors and should be stochastic. The foot-and-mouth disease virus (FMDV) evolution experiments found that the relative fitness of the virus appears a fluctuating pattern around a constant average fitness (Lazaro and Escarmis, 2003).

Recently, the random Eigen models in which the deterministic physical parameters were treated as Gaussian distributed random variables were proposed (Feng et al., 2007; Qiao et al., 2014). But only qualitative conclusions were given in these randomization studies. The reason for the change of the error threshold has been not yet clear. The randomization effects of the Eigen model are still worth studying. Firstly, how to measure the extension of the error threshold and what is the relationship between the extension and the fluctuation strength? Secondly, when a deterministic concentration is replaced by an ensemble of concentrations, how does the ensemble pass through the error threshold? The above questions actually motivate the present work.

In this article, the change of the error threshold and the statistical properties of the relative concentrations of the quasispecies are studied on a single peak Gaussian distributed fitness landscape. In Section 2, the classic Eigen model and the random Eigen model are introduced, respectively. In Section 3, the Gaussian distributions of random fitness of the mutant sequences and master sequence are firstly tested, then a quantitative analysis for the extension of the error threshold is performed, and finally the distribution characteristics of the quasispecies, especially those around the error threshold are examined based on the random Eigen model. In Section 4, the main conclusions of this work are drawn.

2. Models

The Eigen model: In the Eigen model, the individuals are specified by the sequences of N digits of basis κ. If one only considers purine and pyrimidine, $\kappa=2$. The total number of possible sequences is 2^N. The difference between arbitrary two sequences is represented by the Hamming distance. Those sequences with the same Hamming distance from the master sequence are combined into a class. As a result, there are $N+1$ classes of the sequences, $I_0, I_1, \cdots, I_N$. I_0 is the master class, $I_i (i>0)$ means a mutant class. The relative concentration of each class in the population, x_i, satisfies the following equation:

$$\frac{dx_i}{dt}=\sum_{j=0}^{N} f_j q_{ij} x_j - \phi(\vec{x})x_i. \tag{1}$$

Here, f_j is the fitness representing replication rate of class I_j and q_{ij} is the transition probability from class I_j to class I_i. ϕ is dilution flux, which satisfies the condition of $\phi(\vec{x})=\sum_{i=0}^{N} f_i x_i$. Eq. (1) can be converted from the nonlinear differential form into a linear differential one by a variable transformation (Thompson and McBride, 1974). The equilibrium distribution of the population could be obtained by the eigenvalue of the coefficient matrix W ($W_{ij}=f_j \cdot q_{ij}$) (Jones et al., 1976). The largest eigenvalue and its corresponding right eigenvector of the matrix W respectively give the production rate and the absolute concentration of each class (X_i) in the equilibrium state. The normalized eigenvector, $x_i=X_i/X$ represents the relative concentration of each class in population and X is the sum of all X_i.

In the deterministic Eigen model, the fitness function commonly adopted a single peak fitness landscape which assumes all sequences in a given class have exactly the same properties. Its mathematical form is written as:

$$f_0=A_0,\ f_i=A_i=A_1<A_0 (i>0) \tag{2}$$

Here, A_0 and A_1 are constants. And the uniform mutation rate is assumed, which means that only point mutation is considered and the mutation rates at different sites are the same and independent of each other (Swetina and Schuster, 1982; Nowak and Schuster, 1989). We may take

$$q_{ij}=\sum_{l=l_{\min}}^{l_{\max}} \binom{j}{l}\binom{N-j}{i-l} q^{N-j-i+2l}(1-q)^{j+i-2l}. \tag{3}$$

Here q represents copying fidelity of each site of a macromolecule sequence, and mutation rate is then $\mu=1-q$. $l_{\min}=\max\{0, j+i-N\}$, and $l_{\max}=\min\{J, i\}$. In Eq. (2) and Eq. (3), all parameters are deterministic.

The random Eigen model: Based on the assumption of single peak fitness, the deterministic fitnesses are replaced by Gaussian distributed random variables in the present work, and the probability density distributions of the random fitnesses y_i are given as follows:

$$\mathrm{P}(y_i)=\frac{1}{\sqrt{2\pi\omega_i^2}} e^{(-(y_i-\overline{y_i})^2/2\omega_i^2)} \quad i=0,1,2,3\ldots\ldots N. \tag{4}$$

Here $\overline{y_i}$ and ω_i^2 denote the average values and variances of the random fitnesses. Thus, the fitness given by the deterministic Eigen model is replaced by the above single peak Gaussian distributed fitness landscape (spGDFL). To facilitate the comparison with the deterministic model, the average values of the master class fitness and mutant class fitnesses in the numerical simulations are taken to be A_0 and A_1 respectively. The fluctuation strength of each random variable is measured by $d_i=\omega_i/\overline{y_i}$. Without loss of generality, we take $d_i=d$. It is worth noting that there exists an upper limit for the fluctuation strength of the random variables ($d=0.25$ in the present work) in numerical simulation. Beyond this limit, the system is unstable. Generally speaking, the fitness fluctuation in practice is not so large.

3. Results

Throughout the calculation, the length of the macromolecules sequence is $N=20$. The single peak fitness landscape is set to $A_0=10$, $A_1=1$. The number of realizations of random sampling is 10000. The fluctuation strength d is taken to be 0.05, 0.10, 0.15, 0.20 and 0.25 respectively. To ensure the reliability of the random fitnesses, their probability density distributions are calculated.