Evolutionary dynamics of a polymorphic self-replicator population with a finite population size and hyper mutation rate

Takuyo Aita [a], Tetsuya Yomo [a,b,c,*]

[a] Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Yamadaoka 1-5, Suita, Osaka, Japan
[b] Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Yamadaoka 1-5, Suita, Osaka, Japan
[c] Graduate School of Frontier Biosciences, Osaka University, Yamadaoka 1-5, Suita, Osaka, Japan

HIGHLIGHTS

- We analyzed the evolutionary dynamics of an asexual self-replicator population.
- The Kauffman's NK model was used as a model of the fitness landscape.
- The dynamics was considered based on the local fitness distribution on the landscape.
- The first three cumulants of fitnesses in the evolving population were formulated.
- We examined how the dynamics depends on several parameters.

ARTICLE INFO

ABSTRACT

Self-replicating biomolecules, subject to experimental evolution, exhibit hyper mutation rates where the genotypes of most offspring have at least a one point mutation. Thus, we formulated the evolutionary dynamics of an asexual self-replicator population with a finite population size and hyper mutation rate, based on the probability density of fitnesses (fitness distribution) for the evolving population. As a case study, we used a Kauffman's "NK fitness landscape". We deduced recurrence relations for the first three cumulants of the fitness distribution and compared them with the results of computer simulations. We found that the evolutionary dynamics is classified in terms of two modes of selection: the "radical mode" and the "gentle mode". In the radical mode, only a small number of genotypes with the highest or near highest fitness values can leave offspring. In the gentle mode, genotypes with moderate fitness values can leave offspring. We clarified how the evolutionary equilibrium and climbing rate depend on given parameters such as gradient and ruggedness of the landscape, mutation rate and population size, in terms of the two modes of selection. Roughly, the radical mode conducts the fast climbing but attains to the stationary states with low fitness, while the gentle mode conducts the slow climbing but attains to the stationary states with high fitness.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Many kinds of experimental evolution, by using bacteria (Barrick et al., 2009), viruses (Meyer et al., 2012) or self-replicating molecular systems (Kita et al., 2008), have been conducted over the world, and then it becomes more and more necessary to analyze the evolutionary dynamics and statistical properties of fitness landscapes (Betancourt and Bollback, 2006; Pitt and Ferre-D'Amare, 2010; Steinbruck and McHardy, 2011; Otwinowski and Nemenman, 2013). Meanwhile, many theorists have tackled to construct mathematical theories for experimental evolution (Eigen, 1985; Gerrish and Lenski, 1998; Wahl and Krakauer, 2000; Voigt et al., 2000; Schiffels et al., 2011). These theoretical works can be categorized into two viewpoints: models of fitness landscapes and regimes of evolution.

As for the models of fitness landscapes, there are three categories: (1) the abstract mathematical model based on the sequence space or the allele frequency space, (2) the molecular structure-based model and (3) the "stairway to heaven" model. There are so many varieties of the mathematical model based on the sequence space, such as the "flat landscape" (Derrida and Peliti, 1991), the "additive fitness landscape" (Woodcock and Higgs, 1996; Prügel-Bennett, 1997),

* Corresponding author at: Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Yamadaoka 1-5, Suita, Osaka, Japan.
E-mail address: yomo@ist.osaka-u.ac.jp (T. Yomo).

Notation			
ν	sequence length	d	number of mutated sites. Hamming distance between a parent genotype and each of its mutants
λ	number of all available letters at each site	α	parameter that governs the skewness of an unsymmetrical Gaussian
ϵ	parameter that governs the "gradient" of the NK landscape	W_m	fitness of the mth fittest among M genotypes in a population
σ	standard deviation of site-fitnesses over all available λ letters at each site. Approximately, $\sigma \approx \epsilon/\sqrt{3}$	X_m	standardized fitness of W_m
k	parameter that governs the "degree of ruggedness" of the NK landscape	$\mathcal{W}, \mathcal{U}, \mathcal{S}$	mean, standard deviation and skewness of fitnesses over M genotypes
M	population size. Number of all different (parent) genotypes in a population	$\mathcal{W}_t, \mathcal{U}_t, \mathcal{S}_t$	the $\mathcal{W}, \mathcal{U}$ and $\mathcal{S}$ at the tth generation, respectively
		$\mathcal{W}^*, \mathcal{U}^*, \mathcal{S}^*$	the $\mathcal{W}, \mathcal{U}$ and $\mathcal{S}$ in the stationary state, respectively

"random rugged landscape" (Macken and Perelson, 1989), "royal road landscape" (Nimwegen et al., 1999), "Babel-tower landscape" (Suzuki and Iwasa, 1999), "block model landscape" (Perelson and Macken, 1995; Orr, 2006a), "NK fitness landscape" (Kauffman, 1993; Altenberg et al., 1997), "rough mount Fuji-type" (Aita et al., 2000) and "neutral network landscape" (Nimwegen et al., 1999). For example of the molecular structure-based model, RNA folding landscape (Fontana et al., 1993; Huynen et al., 1996) and protein folding landscape (Govindarajan and Goldstein, 1997) were studied well. Particularly, the concept of the "neutral network" was proposed through the study of the sequence- (secondary) structure mapping for RNA molecules (Schuster et al., 1994). Then, the neutral molecular evolution is interpreted as a random drift along the neutral network (Schuster and Fontana, 1999; Schultes and Bartel, 2000). The stairway to heaven model, in which one assumes a constant distribution of selection coefficients or a constant fraction of beneficial mutations for every point on the landscape, has been used in many studies for evolution of asexual populations (Gerrish and Lenski, 1998; Wilke, 2004; Park and Krug, 2007; Hosoda et al., 2014). This model is effective when the sequence length is infinitely long and the final goal as the global optimum is far away. Therefore, this model can be applied to evolution of bacteria and is not so suitable for evolution of viruses. Several articles inferred the structural properties of real landscapes by fitting parameters of the mathematical model to experimental data (Kauffman and Weinberger, 1989; Wahl and Krakauer, 2000; Aita et al., 2007; Kryazhimskiya et al., 2009; Szendro et al., 2013; Neidhart et al., 2013).

As for regimes of evolution, population size M and mutation rate μ per site per replication are crucial parameters. A representative study with infinitely large population size was done by Eigen's group, as is known as the "quasi-species" theory (Eigen, 1985). The quasi-species is a polymorphic population of self-replicators with high mutation rate and large population size. Its concept has been applied to interpret the evolutionary dynamics of viruses (Elena et al., 2008). Lenski's group has been examining the evolutionary dynamics of bacteria experimentally and theoretically (Gerrish and Lenski, 1998; Wiser et al., 2013). Since the mutation rate of bacteria is much lower than that of viruses, it becomes easy to pursue the dynamics of clonal interference among a few genotypes and their fixation (Gerrish and Lenski, 1998; Park and Krug, 2007; Fogle et al., 2008; Schiffels et al., 2011). Recently, the synthetic biology is contributing to experimental evolution. Ichihashi et al. (2013) has conducted an experimental evolution of a translation-coupled RNA replication system. This system takes a hyper mutation rate ($\mu = 10^{-3}$) and large population size ($M = 10^8$), showing a polymorphic population and highly complex dynamics (Ichihashi et al., in preparation). We consider that Ichihashi's system is one of the most simplified biological systems constructed artificially and then it is worth to analyze it theoretically.

In this paper, based on our previous studies (e.g. Aita et al., 2007), we examined the evolutionary dynamics of finite self-replicator population by using the NK fitness landscape and the regime of evolution with hyper mutation rate and large population size such like Ichihashi's system. Our aim is (1) to formulate a dynamics of population's fitness distribution along a single fitness coordinate without considering the details of individual sequences and (2) to clarify how parameters, such as mutation rate, population size and the gradient and ruggedness of the landscape, affect evolutionary equilibrium and the rate of fitness increase. As a previous study, Campos et al. (2002) performed a numerical study of the evolutionary dynamics on an NK fitness landscape, on the aim to identify the mutation rate at which either the mean or the maximum fitness of the population is maximized and a transition point at which the population delocalizes at evolutionary equilibrium. Our study focuses on the mutation–selection–drift balance on this landscape and an evolutionary process toward it.

To focus on population's fitness distribution is effective. Kryazhimskiya et al. (2009) took a similar approach and applied to estimate the types of the fitness landscape from the observed time series of fitness. Prügel-Bennett (1997) formulated the evolutionary dynamics based on the cumulants of fitness distribution for an additive fitness landscape and showed excellent results of prediction consistent with the simulations. In our paper, we deduced explicit recurrence relations of the first three cumulants (mean, variance and skewness) of fitnesses in an evolving Sections 2 and 3.1–3.3 do not specify types of fitness landscapes and have a generality. The remaining parts are based on an NK fitness landscape defined in Appendix B.

2. Model of an evolving population through natural selection and mutation

We consider an evolving population of asexual self-replicators in a fixed environment. Each of their genotypes has the genome sequence with chain length of ν, in which λ letters are available at every site. At every generation, the population consists of M different genotypes, where M is fixed through all generations. Let W_m be the "fitness"[1] of the mth fittest among M genotypes in the population: $W_1 > W_2 > \cdots > W_M$. First, all genotypes reproduce offspring. The mth genotype reproduces cMe^{W_m}/Z offspring by replicating its sequence, where $Z \equiv \sum_{m=1}^{M} e^{W_m}$ and $c \gg 1$ is an infinitely large constant. During the reproduction process of each offspring genotype, d-fold point mutations occur randomly in each sequence, where d corresponds to the number of mutated sites and the Hamming distance between a parent genotype and each of its offspring. We assume that d is fixed to a constant value throughout

---

[1] Although, in biology, e_m^W should be called the (true) "fitness", we call W_m the "fitness" for convenience.