Spontaneous genetic clustering in populations of competing organisms

Pei-Wen Tsai

1 INTRODUCTION

One of the fundamental question in evolution is why species are dramatically different from each other? Through analyzing the simple mathematical models, the authors identify the sufficient and necessary conditions for this phenomenon. The species defined here is the cluster formed by similar organisms and organism is characterized by its genome sequence. Different from the previous work which treat the entire population as a whole in evolution, the authors introduce stochastics to each species. They analyze the stability of the fixed points in the limit of infinite populations, which correspond to a deterministic result. The effect of individual stochastics to the formation of genome clusters becomes clear when the system is large but finite.

2 MODEL

In this study, species are characterized by genomes which is a sequence of binary with length N. Each sequence correspond to a species. The difference between two species with genome I and J, is defined as the number of entries in difference. Take two 4-bits genomes as example. Suppose genome I is $\{0,1,0,1\}$, genome J is $\{1,0,1,1\}$, the difference between them is $n = |I - J| = 3$.

The dynamics of the population is controlled by the birth rate and the death rate. In the multiplication process, one organism might reproduce itself or mutate into other species. In their model, each bit in the genome has the probability μ to flip from 1 to 0 or vise versa. Thus, the probability of genome I grows into genome J is

$$R_{I,J} = \mu^{|I-J|}(1-\mu)^{N-|I-J|}. \tag{2.1}$$

The death rate of an organism at time t is the total competition competitions it experiences multiplied by a constant κ. The competition, $G_{I,J}$, between organism I and J depends only on their difference in genome sequence, *i.e.* ,

$$G_{I,J} = g(|I-J|). \tag{2.2}$$

The more the similar of two organisms, the competition between them is more intensive. It is applicable to the real world for similar species exploit similar resources. κ is the carrying capacity which can be interpreted as proportional to the inverse of the available resources. The larger the κ, the higher the death rate. In the beginning of the simulation, the genomes distributed uniformly. Afterwards, the distribution is shaped by birth and death rate. Looking at the birth rate, one might expect a continuous spectrum of species, i.e, we may have a intermediate species between human and apes which have human face but apes body. However, the competition between similar species would result in gaps between genomes of survivals.

3 MATHEMATICAL FORMULATION

The state of the system at time t is described by a vector $\mathbf{x}$ whose entry x_I is the number of organism with genome I. The probability of finding the system in state $\mathbf{x}$ at time t is $P(\mathbf{x}, t)$. To analyze the time evolution of the genome distribution, they write down time derivative of $P(\mathbf{x}, t)$ which combines the effects of $R_{I,J}$ and $G_{I,J}$. To investigate the system in the small κ region, the authors expand $P(\mathbf{x}, t)$ in terms of κ and achieve the time derivative of x_I. It can be classified into a deterministic term and a stochastic term which is proportional to the square root of κ and the Gaussian white noise.

In their simulation, the genome sequence is take to be 32-bit long. As a result, it is crucial to simplify the calculation by diagonalizing the R and G matrices. For this purpose, the matrix Π is introduced and formulated such that for a matrix F with entry $F_{IJ} = f(|I-J|)$, $\Pi F \Pi$ is diagonal. The transformation of x_I is denoted by y_I. The resulting time derivative of y_I can also be classified into a deterministic term plus a stochastic term with the square root of κ multiplied by Gaussian white noise.

4 ANALYSIS

Take the population size of the system into infinity reduces the time derivative of y_I into a deterministic form. The fixed point turns out to be a homogeneous state, *i.e.* , $x_I = 2^{-N}$. To analyze the stability of the fixed point, they look at the Jacobian matrix of y_I. Mutation and competition act oppositely to determine the stability of the fixed point. It can be shown by considering a particular competition scheme,

$$g(n) = \begin{cases} 1/g_w & \text{if } n \leq w \\ 0 & \text{elsewhere,} \end{cases} \tag{4.1}$$

where g_w is a constant. Plot out the $\mu - w$ phase diagram of stability, it is found out that not all choices of w may result in instability. Starting from an unstable state, x_I grows either exponential enhanced or decay which implies the formation of genome clusters.

For a finite system, the effect of stochastic term becomes obvious. The probability of randomly selected two organisms with genome difference n is deterministic term plus stochastic correction in the a order of κ. Use the competition scheme defined above and start from a stable fixed point, it is more likely to have two genomes in common than the result calculated only by the deterministic term.

5 Reference

Tim Rogers, Alan J. McKane, Axel G. Rossberg. arXiv:1207.1615 [q-bio.PE]