Non-Markovian jumping times and evolutionary irreversibility in a computational genotype-phenotype map

Pablo Catalán[1,2], Susanna Manrubia[1,3], and José A. Cuesta[1,2,4,5]
[1]Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain
[2]Dept. de Matemáticas, Universidad Carlos III de Madrid, Leganés, Madrid, Spain
[3]Programa de Biología de Sistemas, Centro Nacional de Biotecnología, CSIC, Madrid, Spain
[4]Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Spain
[5]UC3M-BS Institute of Financial Big Data (IFiBiD), Universidad Carlos III de Madrid, Getafe, Madrid, Spain

Biological evolution is a highly complex dynamical process in which organisms reproduce and mutate through time. The information needed to build organisms (the genotype) is transmitted from parent to offspring, sometimes with mutations. The organism's features (the phenotype) are the target of natural selection, and determine its fitness or reproduction rate. At the level of genotypes, evolutionary dynamics can be modelled as a Markov process. We will assume that natural selection is strong and that mutations are rare, i.e., $\mu N \ll 1$, where μ is the mutation rate and N is the population size. This means mutations arise rarely enough so that they disappear from the population or go to fixation before another new mutation appears. The fixation rate of a new mutation in a haploid asexual population is given by

$$\phi(f,N) = \mu N \frac{f-1}{f^N-1}, \qquad (1)$$

where f is the fitness of the current phenotype relative to that of the mutant. In this scenario, evolution is a random walk in genotype space: when a new mutation appears in the population, it can either become extinct (the random walk stays in the same genotype) or dominate the population (the random walk jumps to a new state), with transition rate $\phi(f,N)$.

The relevant process, however, is evolution at the level of phenotypes, which is what we can measure. Recent work mapping genotypes to phenotypes has determined that many genotypes map to the same phenotype, forming vast neutral networks [1]. Moreover, the number of genotypes that map to a given phenotype is not homogeneous: most phenotypes are rare, but some of them are extremely abundant. Here we show that these properties imply that evolution at the phenotype level is non-Markovian and irreversible.

In order to explore these phenomena, we will make use of toyLIFE, a multilevel computational model inspired by cellular biology [2]. Our toyLIFE genotypes are formed by two binary genes that codify the expression of two proteins in time and space, forming one-dimensional patterns, which will constitute the phenotype. Because of the discrete character of the model, toyLIFE genotypes are equivalent to cellular automata.

In this work, we focus on two particular phenotypes with the same fitness, but with very different abundances: phenotype 1 is codified by 1.652×10^9genotypes, while phenotype 2 is mapped by 3×10^6 genotypes, a difference of three orders of magnitude [Fig. 1 (a)]. A naïve, coarse-grained approach to studying the dynamics between these two phenotypes (disregarding transitions to other phenotypes) is to model this system as a two-state Markov chain. This simple model captures the distribution of transition times from phenotype 2 to phenotype 1, but fails to accurately characterize jumps in the opposite direction. While the two-state Markov chain predicts long, exponential transition times from phenotype 1 to phenotype 2, our results show that two different outcomes can occur, depending on the initial condition

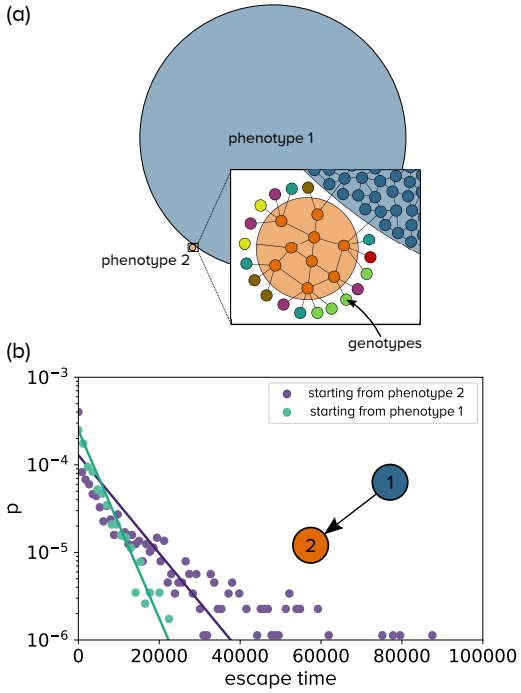

Fig. 1. Phenotypic bias. (a) Phenotype 1 is encoded by more than 10^9 genotypes, while phenotype 2 is three orders of magnitude less abundant. These two phenotypes are connected by mutations. (b) Starting from phenotype 2, the process quickly jumps to phenotype 1 and then takes a non-exponential time to return (blue). Starting from phenotype 1, the process jumps to phenotype 2 faster than the two-state Markov model predicts.

[Fig. 1 (b)]. Starting from phenotype 2, the process quickly jumps to phenotype 1 and then becomes *trapped* there [3], taking a non-exponential time to return. On the other hand, starting from phenotype 1, the population jumps (almost) exponentially to phenotype 2, but faster than predicted by the two-state model. We present a five-state Markov chain that successfully predicts these jump times. This work suggests strategies to model phenotypic evolution, taking into account the underlying structure of genotype space.

[1] J. Aguirre, P. Catalán, J. A. Cuesta, and S. Manrubia, On the networked architecture of genotype spaces and its critical effects on molecular evolution, arXiv:1804.06835.

[2] P. Catalán, A. Wagner, S. Manrubia, and J. A. Cuesta, Adding levels of complexity enhances robustness and evolvability in a multilevel genotype-phenotype map, J. R. Soc. Interface **15**, 20170516 (2018).

[3] S. Manrubia and J. A. Cuesta, Evolution on neutral networks accelerates the ticking rate of the molecular clock, J. R. Soc. Interface **12**, 20141010 (2015).