Quasispecies dynamics with network constraints

Valmir C. Barbosa, Raul Donangelo, Sergio R. Souza

A quasispecies is a set of interrelated genotypes that evolve according to the principles of selection and mutation. Quasispecies studies invariably assume that it is possible for any genotype to mutate into any other, but recent finds indicate that this assumption is not necessarily true. Here we adopt a network structure to constrain the occurrence of mutations. Our results support the theory's assertions regarding the adaptation of the quasispecies to the fitness landscape and also its possible demise.

Introduction

A quasispecies is the stationary state of a population of genotypes whose members mutate into one another while replicating without recombination. Quasispecies theory is applicable to the dynamics of RNA viruses and in cancer research, among other topics, providing interesting insight into the dynamics of any population of genotypes. Its central hypothesis is that, although each individual genotype can be ascribed a fitness that is a function of its replicative capacity, the actual fitness is a property of the population rather than of the genotype. Normally a genotype is represented as a length-L string of 0's and 1's, so the number of genotypes in the population is 2^L. Every genotype can mutate into every other, so essentially there is no structure constraining the occurrence of mutations.

Random-graph model

In contrast to standard quasispecies theory[1], which assume that it is possible for any genotype to mutate into any other, we assume[2] that the n genotypes are the nodes of a directed graph D with self-loops at all nodes. The set of in-neighbors of node i in D is denoted by I_i and its set of out-neighbors by O_i. The existence of an edge directed from node i to node j means that it is possible for genotype i to mutate into genotype j during replication. This happens with probability q_{ij}. Letting q_{ii} be the probability that genotype i remains unchanged during replication leads to $\sum_{j \in O_i} q_{ij} = 1$.

Let X_i denote the abundance of genotype i at any given time, and similarly let $x_i = X_i / \sum_{k=1}^{n} X_k$ be its relative abundance. Then:

$$\dot{x}_i = \sum_{j \in I_i} f_j q_{ji} x_j - \phi x_i, \qquad (1)$$

where $\phi = \sum_{k=1}^{n} f_k x_k$ is the average fitness of all n genotypes. Eq. (1) is the well-known quasispecies equation, now written for graph D.

We assume that both the structure of graph D and the dynamics of mutation depend on how susceptible each of the L loci in a genotype is to undergo a mutation. For $\ell = 1, 2, \ldots, L$, we let s_ℓ be a positive number that grows with the susceptibility that a genotype undergoes a mutation at locus ℓ, the same for all genotypes. Thus, an edge exists in graph D directed from genotype i to genotype j with probability p_{ij} such that

$$p_{ij} = p^{\sum_{\ell=1}^{L} h_\ell / s_\ell}, \qquad (2)$$

where p is a probability parameter and $h_\ell = 1$ if and only if the two genotypes differ at locus ℓ ($h_\ell = 0$, otherwise).

Henceforth we work on the hypothesis that, at the stationary state, x_i depends on the fitness f_i as a power law for every genotype i. That is, we assume that $x_i = b f_i^a$ for suitable $a > 0$ when $\dot{x}_i = 0$. Such functional dependency turns up in some of the cases we study and, furthermore, facilitates some of the analytical calculations that we carry out in this section. It immediately follows that the stationary-state value of the average fitness is $\phi = b \sum_{h=0}^{L} \binom{L}{h} 2^{-(a+1)h}$, yielding

$$\phi = b \left[1 + 2^{-(a+1)} \right]^L. \qquad (3)$$

From the constraint $\sum_{i=1}^{n} x_i = 1$ we obtain $b \sum_{h=0}^{L} \binom{L}{h} 2^{-ah} = 1$, and, therefore, $b = (1 + 2^{-a})^{-L}$.

Results

For fixed values of the length L and the probability parameter p, our results are based on generating 10^4 independent instances of graph D and solving Eq. (1) numerically for each instance. This is achieved by letting the initial population to be uniform over all genotypes and time-stepping the corresponding equations until attaining convergence.

We study two susceptibility scenarios. The first one, henceforth referred to as the uniform case, sets $s_\ell = 1$ for every locus ℓ. In the second scenario, which we refer to as the inverse-decay case, we have $s_\ell = 1/\ell$ for locus ℓ.

The resulting relative abundances of the quasispecies are given in Fig. 1 as a function of the genotypes' fitnesses.
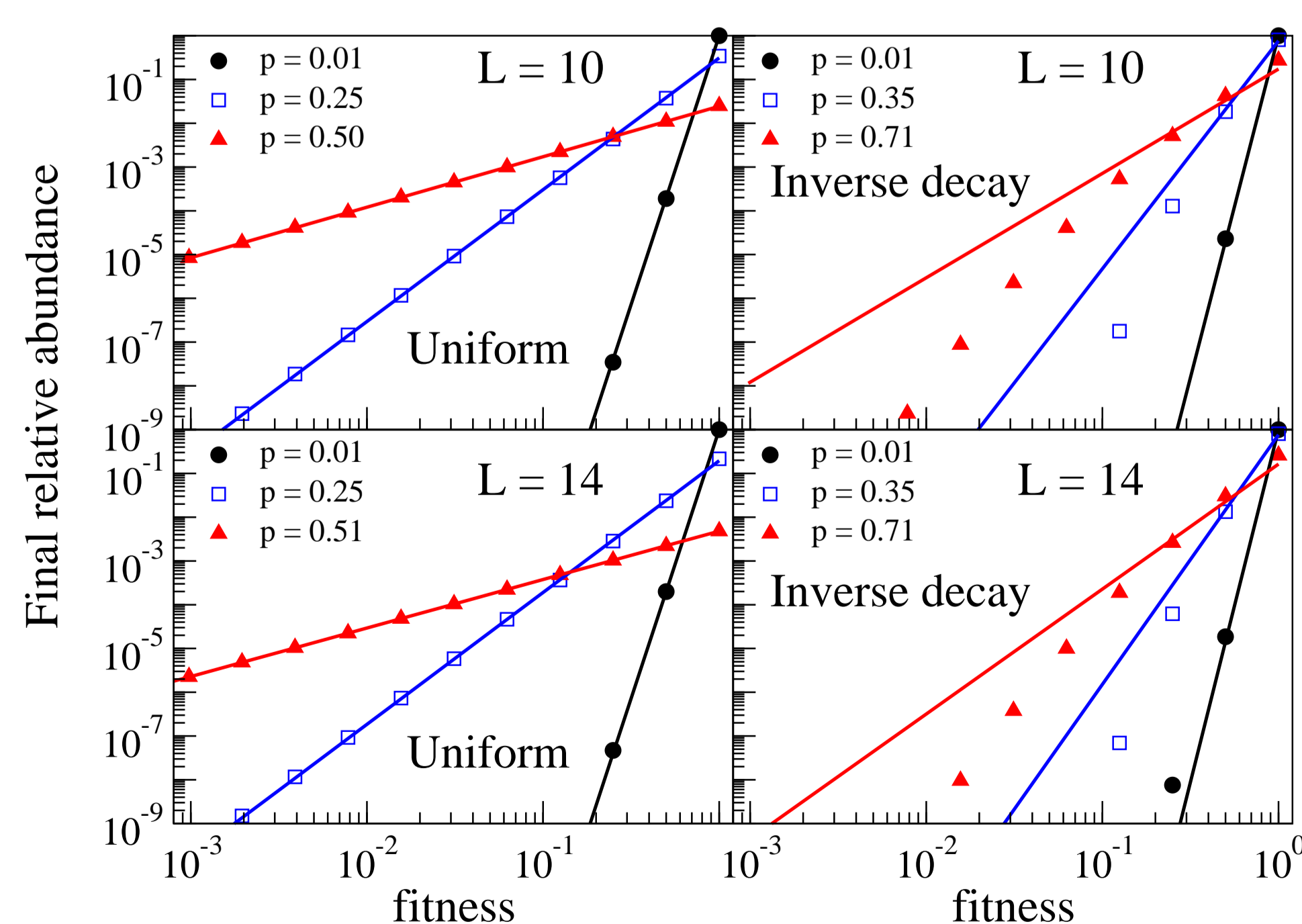

Figure 1. Relative abundances at the stationary state.

This figure also reveals how the dominance of the wild type in the population behaves as p is increased and mutations into ever more different genotypes begin to be both allowed by the structure of D and made more frequent during the dynamics.

A clearer view into this is afforded by Fig.2, where we show the relative abundance of the wild type in the quasispecies as a function of p.
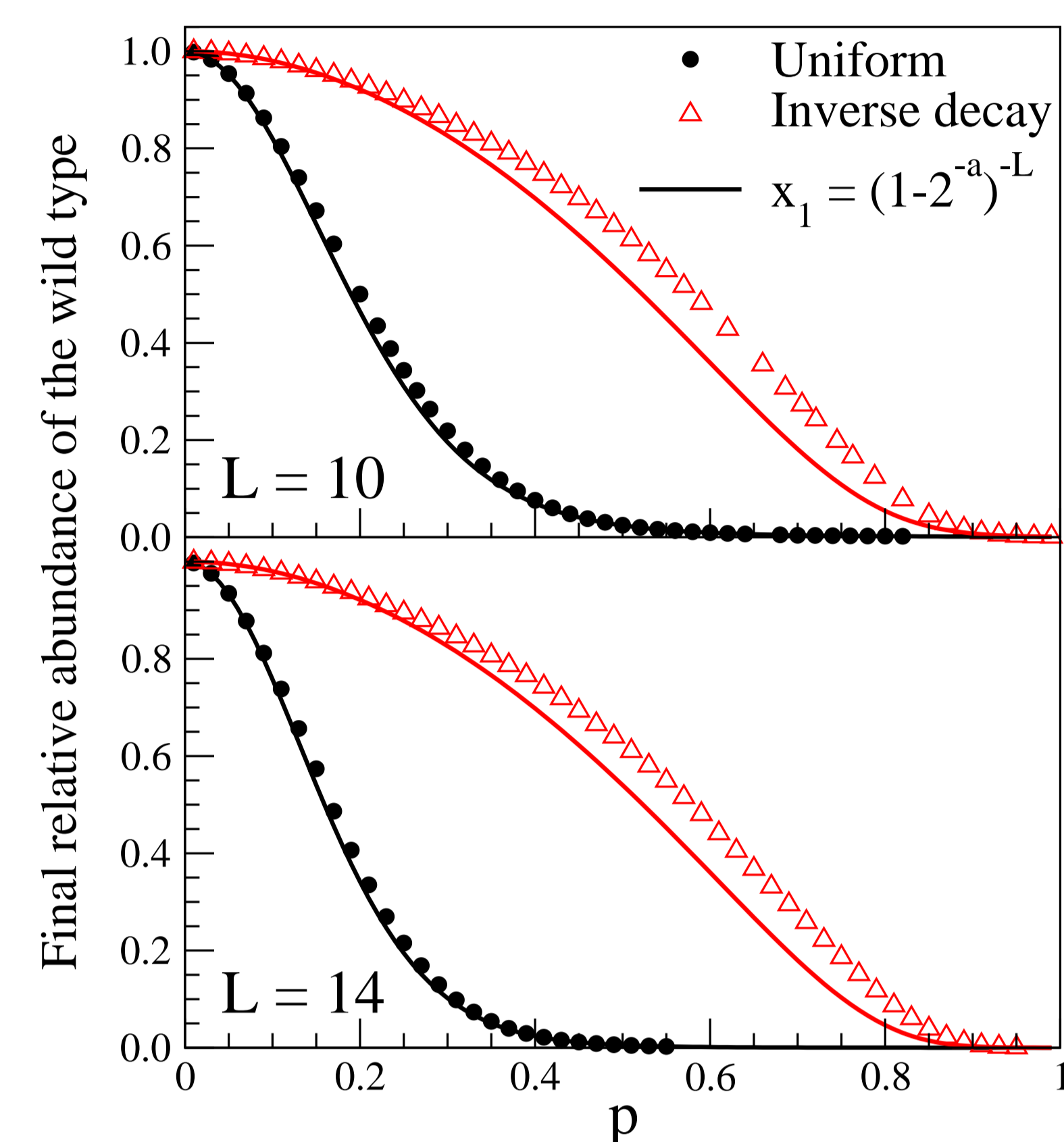

Figure 2. Relative abundance of the wild type at the stationary state.

A better glimpse into wild-type survival comes from considering the average fitness ϕ of the quasispecies.
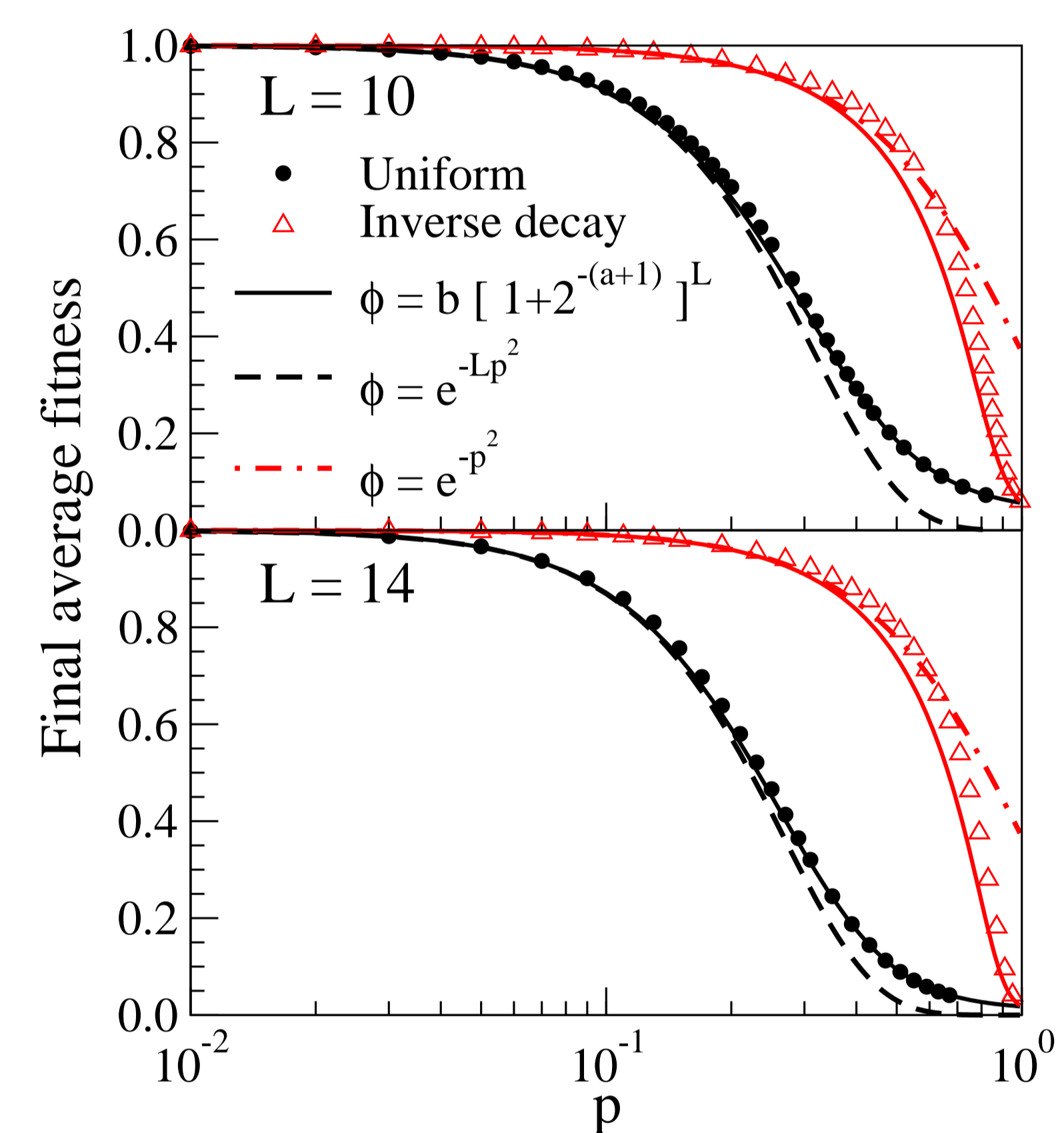

Figure 3. Average fitness at the stationary state.

Conclusions

We have revisited the quasispecies theory assumptions, in particular: that any mutation can occur and that all genotypes loci have the same susceptibility to undergo mutations. In our model the mutational interactions among genotypes are on a random graph and we adopt susceptibilities that influence both the graph's structure and the dynamics of the population. The resulting model has a probability, p, as its single parameter. Increasing p makes the graph denser and allows more mutations as the population evolves toward the quasispecies.

Our results were given for the nontrivial fitness landscape in which a genotype's fitness decays exponentially with its Hamming distance to the wild type. They have also been based on two specific susceptibility scenarios and a power-law relationship between a genotype's relative abundance in the quasispecies and its fitness.

As with other variations of the quasispecies theory, the modifications we have introduced all corroborate the theory's central idea, that selection and mutation act on the entire ensemble of genotypes. They also corroborate the crucial role of the error-related parameter (p, in our case) in separating two distinct regimes, one in which the quasispecies adapts to the fitness landscape, the other in which it becomes degenerate. It remains to be seen whether the same will continue to hold as alternative fitness landscapes and variations of the remaining assumptions are studied.

References

1. M. Eigen, Naturwissenschaften 58 (1971) 465; M. Eigen, John McCaskill and P. Schuster, Adv. Chem. Phys. 75 (1989) 149; M. A. Nowak, "Evolutionary Dynamics", Harvard University Press, Cambridge, 2006

2. Valmir C. Barbosa, Raul Donangelo, Sergio R. Souza, J. Theor. Biol. 310 (2012) 206