Characteristic time in quasispecies evolution

Arturo Marín [a], Héctor Tejero [a], Juan Carlos Nuño [b], Francisco Montero [a,*]

[a] Department of Biochemistry and Molecular Biology I, Universidad Complutense de Madrid, Avd. Complutense s/n, 28040 Madrid, Spain
[b] Department of Mathematics Applied to Natural Resources, Escuela Técnica Superior de Ingenieros de Montes, Universidad Politécnica de Madrid, 28040 Madrid, Spain

ABSTRACT

The time a phenotype takes to achieve a stationary state from an initial condition depends on multiple factors. In particular, it is a function of both its fitness and its mutation rate. We evaluate the average time, referred to as the characteristic time, T_c, that the system takes to reach a final steady state of simple models of populations formed by self-replicative sequences. The dependence of T_c on the mutation rate and on the fitness landscape is also studied. For simple fitness landscapes, e.g. single peak, the characteristic time can be analytically obtained as a function of the system parameters. In this case, T_c for obtaining the quasispecies distribution presents a maximum at a Q-value that depends on the initial conditions and decreases monotonously as the mutation rate tends to zero. For most of the complex landscapes handled in this paper, the characteristic time to achieve the quasispecies distribution picked around the fittest phenotype attains a local minimum for a given mutation rate between 0 and the Q-value at which T_c reaches its local maximum. Thus, in these cases, an optimum value for the mutation rate exists that corresponds to the lowest value of the characteristic time for quasispecies evolution.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The quasispecies model developed by Eigen (1971) is a useful general evolutionary model for error-prone self-replicative systems that has been applied in many different fields such as, for example, prebiotic self-replicating molecules (Schuster and Stadler, 2008), RNA viruses (Domingo et al., 2006), cancer cells (Solé and Deisboeck, 2004) and the immune system (Kamp et al., 2003). In all these problems, the question of how long the system takes to evolve is of great relevance. In particular, it is important to find out how this time depends on two of the main factors that govern evolution, namely the mutation rate and the fitness landscape.

In principle, as the mutation rate increases the time the system takes to evolve must decrease because the exploration through the sequences space is faster. However, the greater the value of the mutation rate, the higher the fraction of the less fitted mutant genotypes and, in consequence, the lower the average fitness of the population. In other words, the turnover of the whole population decreases, and therefore the evolution time increases. This trade-off between the generation of diversity and the fixation ability of advantageous mutants has been already described (Stich et al., 2007; Stich and Manrubia, 2011). For particular fitness landscapes, this balance is reflected in the appearance of a value for the

mutation rate at which the evolution time of the system presents a minimum. However, an explicit relationship between the optimum value of the mutation rate and the fundamental properties of the system, such as the characteristics of the fitness landscape, has not yet been exhaustively explored. Most of the models in which this problem has been examined are stochastic in nature. In this context different times have been defined, for instance, the searching time and fixation time (Traulsen et al., 2007; Gokhale et al., 2009; Stich et al., 2007; Stich and Manrubia, 2011), the evolution time (Krug and Karl, 2003) the adaptation time (Stich et al., 2007), the crossover time, the jump time, the residence time and the tunneling time (Jain and Krug, 2007; Krug, 2002).

Evolution is intrinsically a stochastic process. Phenotype existence depends on multiple factors that make it virtually impossible to be considered in its entirety. The birth and death of species occur according to probabilistic rules that, in the end, determine the population dynamics. Moreover, species interaction is also affected by randomness. Strictly speaking, mathematical models should include these stochastic factors to ensure a correct description of the population. However, it turns out that under particular conditions average variables can provide useful insights into the system's behavior. In his seminal paper (Eigen, 1971) Manfred Eigen described the evolution of self-replicative molecules in terms of ordinary differential equations. Explicitly, he was assuming that stochastic fluctuations, whether internal or external, were, at a first approximation, not relevant and therefore could be neglected. However, in the next section of his paper,

he immediately went on to establish the limitation of this phenomenological description. In any case, the deterministic formulation brought about one of the most fruitful lines of research in evolution.

The deterministic modeling of populations formed by error-prone self-replicative sequences assumes that both the elementary process of mutation and the growth dynamics are free of stochasticity. Unfortunately, even under this approximation, finding the exact solutions for the phenotypes concentration is a very hard problem. Most of the mathematical analysis carried out on these dynamical systems is qualitative, i.e. it seeks to understand the asymptotic behavior of the system. It is implicitly assumed that (stable) equilibrium states are the unique observable outcome of actual systems. Nonetheless, this is not always true. During evolution we only observe transient regimes. The system is always evolving, although transient states can appear as metastables. In these models, time takes on a new meaning, not only as the continuous independent variable, but also as an observable of the system that deserves further attention. In particular, the time a dynamical system remains in a quasistationary state or the time taken to reach another quasistationary state are of special relevance. The problem arises when we want to characterize this time within a deterministic framework since, strictly speaking, this time is infinite because the trajectories approach the equilibrium points asymptotically. To measure this time, some authors have used the inverse of the largest eigenvalue or the inverse of the difference between the two largest eigenvalues, see for instance (Kamp et al., 2003; Nowak and Schuster, 1989). However, the application of these definitions even to simple continuous dynamical systems shows serious discrepancies.

The characteristic time, hereafter referred to as T_c (Llorens et al., 1999), was previously defined to capture the global evolutionary properties of dynamical systems. Based on a geometrical interpretation, the characteristic time provides the average time the system takes to change from one state to another under the action of a perturbation or a persistent variation. For particular linear dynamical systems, whose solution can be expressed as a linear combination of exponentials, the characteristic time corresponds to the average of the inverse of the exponents weighted by the pre-exponentials, real-valued constants that depend on the initial conditions and the matrix entries.

In general, the characteristic time depends on system parameters in a more complex way. In the case studied here, it also depends on the mutation rate. To study this functional dependence this paper is organized from simple to complex models. Perhaps the simplest, though still interesting, system is the error-prone replication in a single peak fitness landscape, where a unique fittest phenotype and its indistinguishable mutants exist. Fortunately, as we will see in Section 2, this model allows a complete analytic treatment and thus, the obtention of an expression for the characteristic times for any initial condition as a function of the mutation rate and the amplification factors. For more complex landscapes, the characteristic times will be numerically computed for different parameter setups in Section 3. Finally, these results are discussed in the last section.

2. Characteristic time for a simple replicator system

We start by studying the characteristic time of a population formed by two self-replicative sequences. Fortunately, in this case the characteristic time can be analytically calculated, allowing a complete study of the dependencies with both the system parameters and initial conditions.

Let us assume a dynamic system formed by two species I_1 and I_2, and let $x_1(t)$ and $x_2(t)$ be their population at time t, respectively.

Without loss of generality we can assume a null death rate for both species. In the absence of any kind of constraint the time evolution of each of the population is given by the following system of ordinary differential equations (ODE):

$$\dot{x_1} = Q_1 A_1 x_1 + (1-Q_2)A_2 x_2$$

$$\dot{x_2} = Q_2 A_2 x_2 + (1-Q_1)A_1 x_1 \tag{1}$$

where A_1 and A_2 are the amplification factors of species I_1 and I_2, respectively, and $0 \le Q_1 \le 1$ and $0 \le Q_2 \le 1$ are their respective quality factors (i.e. $1-Q_1$ and $1-Q_2$ are the mutation rates for I_1 and I_2, respectively). Along this section, we will assume $A_1 \ge A_2$, so I_1 could be considered as the master copy, and I_2 as the error tail.

A complete description of the time evolution for each variable can be performed in terms of the molar fractions of each sequence. Let us define the molar fraction of the master sequence by

$$y_1 = \frac{x_1}{x_1 + x_2} \tag{2}$$

Note that $y_1 \in [0, 1]$, and that the corresponding molar fraction of the sequence I_2 is given by $y_2 = 1-y_1$. It is straightforward to find the differential equation that describes the time evolution of y_1 from the previous Eq. (1)

$$\dot{y_1} = (1-Q_2)A_2 + y_1(Q_1 A_1 - A_2 - (1-Q_2)A_2) - y_1^2(A_1 - A_2) \tag{3}$$

To complete the initial value problem we define the initial condition $y_1(t=0) = y_1^0$. Although non-linear, this equation is still solvable, which will allow an analytical evaluation of the characteristic time according with the following definition. The characteristic time to achieve the equilibrium point from a given initial condition $0 < y_1^0 < 1$ can be computed straightforwardly using the equation

$$T_c = \frac{\int_0^\infty t|\dot{y_1}| \, dt}{\int_0^\infty |\dot{y_1}| \, dt} \tag{4}$$

For a complete explanation of the meaning of T_c see Llorens et al. (1999). According with this definition the characteristic time provides an average time to reach the equilibrium state from any initial condition along the trajectory. Let

$$g(t) = \frac{|\dot{y_1}(t)|}{\int_0^\infty |\dot{y_1}(t)| \, dt} \tag{5}$$

be the density function obtained from the trajectory of variable y from y_0 at time $t=0$ to the equilibrium point at $t = \infty$. Then, the characteristic time is the first moment of g, i.e.

$$T_c = \langle t \rangle = \int_0^\infty tg(t) \, dt \tag{6}$$

Contrary to the time obtained from the inverse of the eigenvalues of the corresponding linearized system that only considers the dynamics on the neighborhood of the equilibrium point, the characteristic time takes into account the complete path from the initial point to the equilibrium point. In this sense, it provides a more accurate measure of the time the system takes to change its current state. However, because its dependence on the initial conditions, in general, it does not define a unique time scale of the system, as given by the inverse of the eigenvalues. Finally, note that the analytic determination of the characteristic time requires the explicit expression of the time evolution of the molar fraction that can only be obtained in simple cases. Nonetheless, as it will be done later, it can be also obtained numerically (or either from empirical data).

The initial value problem associated with Eq. (3) can be solved analytically. If $y_1(t=0) = y_0$, the general solution reads

$$y(t) = \frac{\overline{y}_p - \overline{y}_n \left(\frac{y_0 - \overline{y}_p}{y_0 - \overline{y}_n}\right) e^{(\overline{y}_n - \overline{y}_p)t}}{1 - \left(\frac{y_0 - \overline{y}_p}{y_0 - \overline{y}_n}\right) e^{(\overline{y}_n - \overline{y}_p)t}} \tag{7}$$

where $\overline{y}_p$ and $\overline{y}_n$ are the equilibrium points of Eq. (3) (for simplicity the subindex 1 of the molar fraction is removed)

$$\overline{y}_{p,n} = \frac{1}{2} \frac{Q_1 A_1 - (1+\epsilon)A_2 \pm \sqrt{(Q_1 A_1 - (1+\epsilon)A_2)^2 + 4\epsilon A_2 (A_1 - A_2)}}{A_1 - A_2} \tag{8}$$

Here, $\epsilon = 1 - Q_2$ is the mutation rate back from species I_2 to the master copy, i.e. the probability of the production of master copies from its error tail. As it can be easily checked, only the equilibrium point with the positive root $\overline{y}_p$ is non-negative (assuming $A_1 - A_2 > 0$) and then, has physical meaning. The characteristic time can be analytically evaluated using Eq. (4) for the molar fraction of the master copy to reach the equilibrium point $\overline{y}_p$

$$T_c = \frac{1}{A_1 - A_2} \frac{1}{\overline{y}_p - y_0} \ln\left(\frac{\overline{y}_p - \overline{y}_n}{y_0 - \overline{y}_n}\right) \tag{9}$$

Fig. 1A shows the characteristic time as a function of the quality factor of the master copy Q_1 and the initial conditions y_0 when the back mutation rate is $\epsilon = 10^{-2}$ and the amplification factors are $A_1 = 10$ and $A_2 = 2$. As it can be seen, T_c exhibits a maximum for all values of the initial conditions y_0 at a Q-value that will be referred as to Q_{1m} (see also Fig. 1B). The maximum value of T_c, T_c^{max} and Q_{1m} depend on the initial conditions, as well as the rest of the parameters. Besides, T_c^{max} depends on the value of ϵ. In the limit of negligible back mutation rate, i.e. when ϵ tends to 0, then T_c^{max} tends to infinity at $Q_{1m} = A_2/A_1$, the Q-value that corresponds with the error threshold.

It is worthy to compare the characteristic time T_c with the time scale given by the inverse of the eigenvalue of Eq. (3) (solid line in Fig. 1B)

$$\tau = \frac{1}{\lambda} = \frac{1}{(A_1 - A_2)(y_p - y_n)} \tag{10}$$

Obviously, τ is independent of the initial conditions and attains its maximum at a Q-value, named as Q_{1c}, that is given by

$$Q_{1c} = \frac{A_2}{A_1}(1+\epsilon) \tag{11}$$

Note that $Q_{1c} \leq Q_{1m}$ for all initial conditions y_0. As matter of fact, $Q_{1m} \to Q_{1c}$ when $y_0 \to 1$. Moreover, contrary to what occurs with the more complex landscapes we are going to investigate in the next sections, neither T_c nor τ exhibit a minimum for Q-values larger than Q_{1c} (i.e. both are monotonous decreasing function of Q_1 in this range). The value of T_c at $Q_1 = 1$ is given by

$$T_c(Q_1 = 1) = \frac{1}{A_1 - A_2} \frac{\ln\left(\frac{A_1 - A_2(1-\epsilon)}{A_1 y_0 - A_2(y_0 - \epsilon)}\right)}{1 - y_0} \tag{12}$$

that tends to the same value as τ when $y_0 \to 1$ (see Fig. 1B)

$$T_c(Q_1 = 1, y_0 = 1) = \tau(Q_1 = 1) = \frac{1}{A_1 - A_2(1-\epsilon)} \tag{13}$$

The replicator model we have qualitatively analyzed in this section can be used to measure the characteristic time for several interesting processes that appear in quasispecies theory. In particular, (A) the selection of one of the species in a double peak landscape without mutation (when $Q_1 = Q_2 = 1$), (B) the dynamic of two neutral species with different mutation rates ($A_1 = A_2$), (C) the quasispecies formation from an error-prone replicator
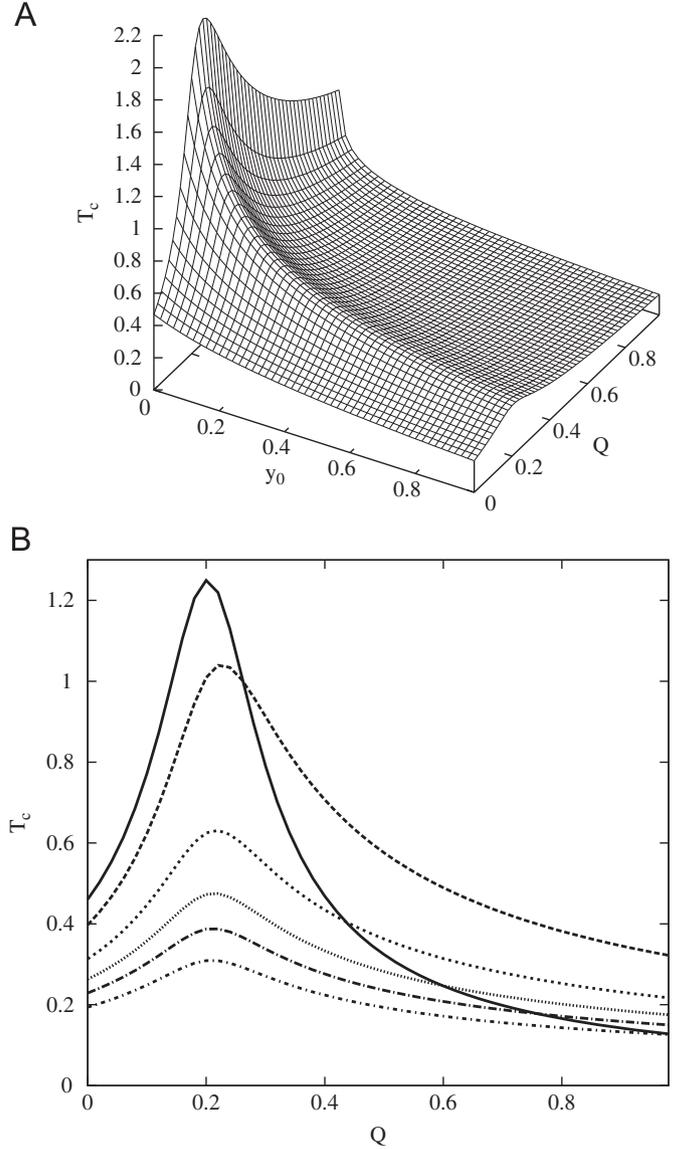

Fig. 1. Characteristic time for the fittest species in a model of two replicators. (A) Three dimensional plot of the characteristic time T_c as a function of the quality factor Q_1 and the initial condition $y(0) = y_0$. As can be seen, the surface present a maximum at a point Q_{1m} for each initial condition y_0. Both, the maximum value of T_c and the Q-value at which it is achieved depend on y_0. The rest of the parameter values are: $\epsilon = 0.01$, $A_1 = 10$ and $A_2 = 2$. (B) From top to bottom, the curve of the inverse of the eigenvalue, τ (solid line), and the characteristic time T_c as a function of the quality factor Q_1 for the initial condition: $y_0 = 0.1$, $y_0 = 0.3$, $y_0 = 0.5$, $y_0 = 0.7$, and $y_0 = 1$. The rest of the parameter values are as in (A).

without back mutations in a single peak landscape (when $Q_2 = 1$ and $Q_1 > Q_{1c}$), and (D) the displacement of the master copy by the error tail beyond the error threshold ($Q_2 = 1$ and $Q_1 < Q_{1c}$). The explicit expressions of the characteristic time for all these cases as well as for the general problem (E) are summarized in Table 1.

3. Characteristic time for the evolution in more complex fitness landscape

In the previous section we have treated a case that is analytically solvable, but real cases respond to more complex fitness landscapes, whose study requires numerical solutions. This section examines the characteristic time (T_c) of a population of replicators that evolve in different fitness landscapes of increasing complexity.

Table 1
Characteristic time (T_c) for a quasispecies model of two replicators with different assumptions in relation with the value of both the mutation and the amplification factor matrix (A, B, C and D), as described in the text (Section 2). (A) T_c depends not only on the difference between the amplification factors but also on the initial conditions. (B) T_c corresponds to the inverse of the eigenvalue of the linear Eq. (1). It is a decreasing function of both Q_1 and Q_2. (C) T_c tends to infinity when Q_1 tends to Q_c from above. It is a decreasing function of Q_1 and tends to $1/(A_1-A_2)$ as Q_1 approaches 1. (D) T_c tends to infinity when Q_1 tends to Q_c from below. It is an increasing function of Q_1. (E) T_c reaches a finite maximum value at Q_{1m} that depends on y_0. Besides, it is a decreasing function of Q_1 for $Q > Q_{1m}$.

Case	Assumptions	Initial value problem for molar fraction y	Characteristic time (T_c)
(A)	$Q_1 = Q_2 = 1$ $A_1 > A_2$	$\dot{y} = (A_1 - A_2)y(1-y)$ $y(0) = y_0$	$T_c = \dfrac{1}{A_1 - A_2} \ln \dfrac{y_0}{y_0 - 1}$
(B)	$A_1 = A_2 = A$	$\dot{y} = (1-Q_2) + y(Q_1 + Q_2 - 2)$ $y(0) = y_0$	$T_c = \dfrac{1}{A} \dfrac{1}{2 - Q_1 - Q_2}$
(C)	$Q_2 = 1$ $A_1 > A_2$ $Q_1 > (A_2/A_1) = Q_c$	$\dot{y} = (A_1 - A_2)y\left(\dfrac{Q_1 A_1 - A_2}{A_1 - A_2} - y\right)$ $y(0) = 1$	$T_c = \dfrac{\ln\left(\dfrac{Q_1 A_1 - A_2}{A_1 - A_2}\right)}{A_1(Q_1 - 1)}$
(D)	$Q_2 = 1$ $A_1 > A_2$ $Q_1 < (A_2/A_1) = Q_c$	$\dot{y} = (A_1 - A_2)y\left(\dfrac{Q_1 A_1 - A_2}{A_1 - A_2} - y\right)$ $y(0) = 1$	$T_c = -\dfrac{1}{A_1 - A_2} \ln\left(\dfrac{Q_1 A_1 - A_2}{A_1(Q_1 - 1)}\right)$
(E)	$A_1 > A_2$	$\dot{y} = (1-Q_1)A_2 + y(Q_1 A_1 - A_2 - (1-Q_2)A_2) - y^2(A_1 - A_2)$	$T_c = \dfrac{1}{A_1 - A_2} \dfrac{1}{y_p - y_0} \ln\left(\dfrac{y_p - y_n}{y_0 - y_n}\right)$

3.1. A two peak separated by a neutral valley fitness landscape

Let us assume a mutation matrix $Q = [Q_{ij}]$ that results from considering a classification of the sequence space according to the Hamming distance to the master copy I_1. We consider a landscape formed by two peaks separated by a degenerate valley. The two master copies I_1 and I_2 are the two complementary bit string, i.e. those formed by all 0 and all 1 digits, respectively. Their amplification factors are A_1 and A_2, respectively. The rest of the sequences have the same amplification factors, A_e. It is assumed that $A_2 > A_1 > A_e$.

The probability that a sequence of the Hamming class l, I_l, gives rise to a sequence of the Hamming class k, I_k, is given by Nowak and Schuster (1989)

$$Q_{k,l} = \sum_{i = l-v+k}^{min(k,l)} \binom{k}{i}\binom{v-k}{l-i} q^v \left(\frac{1-q}{q}\right)^{k+l-2i} \tag{14}$$

where v is the length of the sequences and q is the quality factor per digit. Since we are considering binary sequences, the total number of different sequences in the population is 2^v and they are grouped into $v+1$ Hamming classes.

The time evolution of the concentration of the Hamming class I_j, x_j, is described by

$$\frac{dx_j}{dt} = A_j Q_{jj} x_j + \sum_{k \neq j} A_k Q_{jk} x_k \tag{15}$$

The corresponding differential equation for the molar fraction

$$y_j = \frac{x_j}{\sum_i x_i}$$

is

$$\frac{dy_j}{dt} = y_j\left(A_j Q_{jj} - \sum_i A_i y_i\right) + \sum_{k \neq j} A_k Q_{jk} y_k \tag{16}$$

If the quality factor q is large enough, this system has a unique steady state formed by a distribution of sequences that surrounds the fittest sequence I_2 (Nowak and Schuster, 1989). Our intention is to compute the characteristic time for the molar fraction of the copy I_2 to achieve the steady state from an initial population formed exclusively by copies of species I_1. That is, we assume that $y_1(t=0) = 1$.

The numerical integration of the ODE system was performed by a Runge–Kutta method provided by the MATLAB platform. The stop control was set to keep the difference between successive steps of variable I_2 at less than 10^{-4} during 200 steps. The characteristic time was then calculated, as described in Llorens et al. (1999), by computing the quotient between the area, computed by the classical trapezoid rule, over the trajectory of I_2 and below the horizontal asymptote that corresponds to its stationary value and the height of this asymptote. Fig. 2 shows the dependence of the T_c on the Q value for different v values (10, 20, and 50). This characteristic time T_c varies with the quality factor per sequence $Q = q^v$ in a different way as described in the previous section. When q tends to 1, then T_c rises to infinity. In addition, T_c exhibits a unique maximum at a Q-value that, as before, will be referred to as Q_m. Between these two values, $Q = Q_m$ and $Q = 1$, can be found a value of Q, referred to as Q_{op}, where the characteristic time reaches a local minimum (characterized by a null first derivative). The dependence of both Q_{op} and T_c at Q_{op} on the value A_2 and A_1 is shown in Fig. 3 for $v = 10$. In all cases, $A_1 = 4$ and $A_e = 1$, whereas A_2 takes the values: 10, 20, 40, 80, 100, 200, 300, and 400. As can be observed, as A_2 is increased $T_c(Q_{op})$ decreases. However, the functional dependence of Q_{op} on A_2 is more complex, exhibiting a maximum value at an intermediate value of A_2.

3.2. Multiplicative fitness with two peaks

Multiplicative fitness landscapes have been largely used in the framework of quasispecies theory (Woodcock and Higgs, 1996). In this section, we compute the characteristic time for the evolution of the fittest sequence in the multiplicative fitness defined as follows. The amplification factor of each sequence at a Hamming distance i from the sequence I_0 whose digits are all 0 is given by

$$A_i = \left(\frac{i(1-s)^{(v-i)}}{v}\right)9 + 1 \quad i = 1, \ldots, v$$
$$A_0 = 2 \tag{17}$$

where v is the chain length and $s \in [0,1]$ allows to tune the effect that each mutation has on the fitness. In this way, the sequence whose digits are all 1, referred as I_v, has an amplification factor $A_v = 10$. Note that, for the special case $s = 1$, this fitness landscape
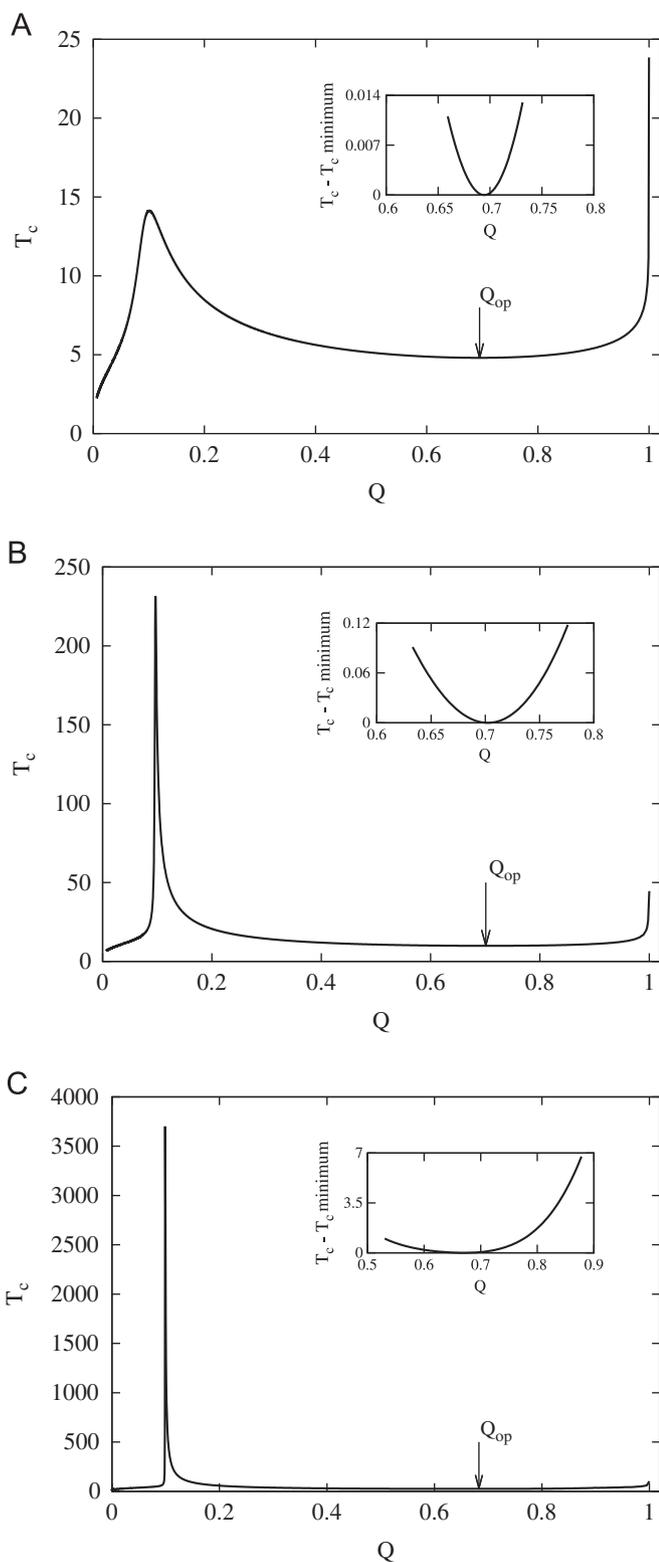
Fig. 3. Q_{op} (dashed line) and the characteristic time of I_2 at Q_{op} (dashed-dotted line) versus A_2 for a double peak with a neutral valley. In all cases, $A_1 = 4$ and $A_e = 1$.

Fig. 4 shows the dependence of the T_c on Q for different multiplicative fitness landscapes and $v = 20$. Using Eqs. (14) and (16), we determine the ODE systems that govern the time evolution of each species. These systems are numerically solved and then, the characteristic time T_c for I_v is estimated (all calculations are performed in MATLAB) (Fig. 4). As it can be seen, for the values of s larger than 0.3, T_c exhibits two maxima, one for $Q = 1$ and the other for a low Q-value (approx. 0.1), and, as a consequence, a minimum at Q_{op}.

The insets B, C and D depict the classical figure of equilibrium population of the different Hamming classes (Schuster and Swetina, 1988) for the values $s = 0.3, 0.6$ and 1, respectively. Note that only for large values of s, e.g. $s = 0.6$ and $s = 1$, the steady populations change abruptly for a Q-value around 0.1, below which all sequences are equally populated. This Q-value corresponds to the error threshold, Q_c. It is also worthy to remark that the value of Q_c estimated in these cases $s = 0.6$ and $s = 1$ approximately coincides with the Q-value for which T_c reaches its largest value (Q_m).

3.3. Binary rugged fitness landscape

Real fitness landscapes are rugged (Kauffman and Levin, 1987; Kauffman and Johnsen, 1991; Schuster, 1997). In order to rise the ruggedness of the fitness landscape, let us now consider a population of binary sequences I_k, $k = 1, \ldots, 2^v$ whose fitness is related to the natural number codified in the sequence. Specifically, let us assign to sequence I_k the amplification factor

$$A_k = \left(\frac{N_k + 1}{2^v} \right)^p \tag{18}$$

Here, N_k represents the natural number codified by I_k and p is an arbitrary positive natural number. This landscape is similar to that described in Nuño et al. to study adaptive evolution of replicators (Nuño et al., 2010). Note that with this definition $1/2^v \le A_k \le 1$ for all $k = 1, \ldots, 2^v$. The largest amplification factor $A = 1$ corresponds to the sequence whose digits are all 1. The parameter p determines the steepness of the landscape, i.e. as the

Fig. 2. Characteristic time of I_2 in a double peak with a neutral valley as a function of the quality factors per sequence Q ($Q = q^v$) for several values of sequence lengths v: (A) $v = 10$, (B) $v = 20$, and (C) $v = 50$. In all cases, $A_1 = 2$, $A_e = 1$, and $A_2 = 10$. The insets clearly show the existence of the local minimum at Q_{op}.

reduces to the degenerate fitness landscape assumed in the previous section (where all the sequences between the I_0 and I_v are assigned the same amplification factor, 1). As the value of the parameter s decreases the change of the amplification factor tends to be linear with the Hamming distance i (Fig. 4, inset A).
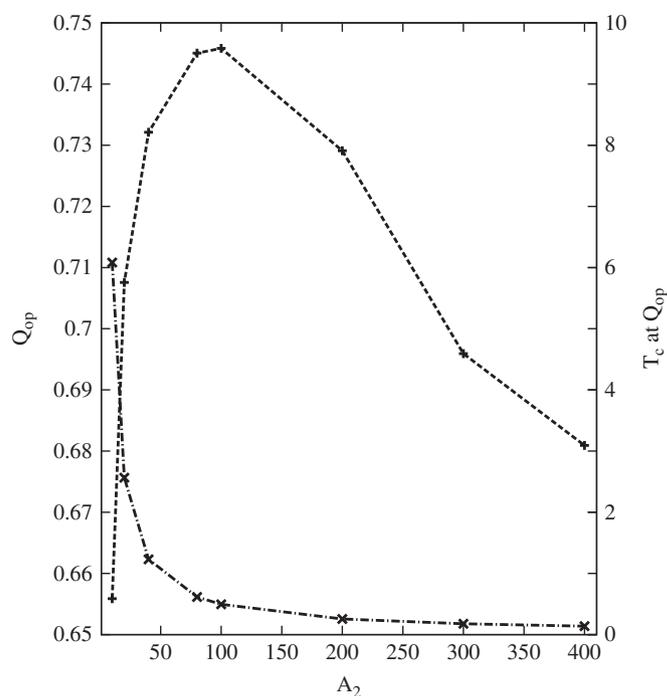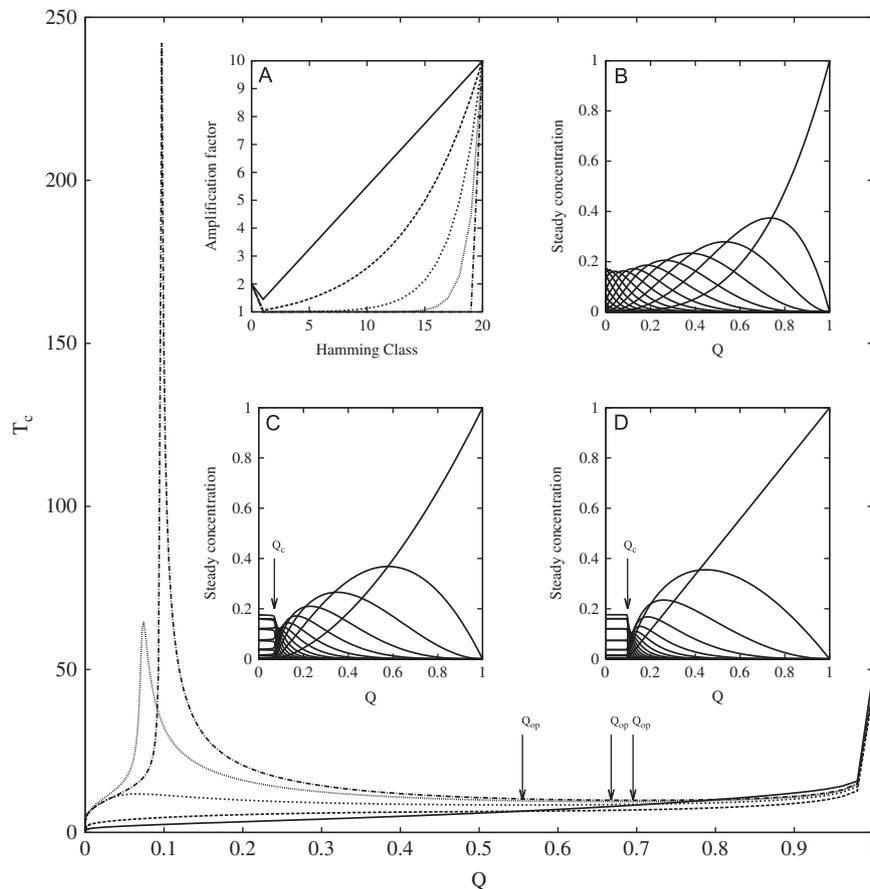
Fig. 4. Characteristic time of the fittest sequence in a multiplicative fitness landscape as a function of the quality factor per sequence Q for different values of s (from top to bottom $s = 1, 0.6, 0.3, 0.1$ and 0, respectively). Inset (A) fitness landscapes for different values of s (from top to bottom $s = 0, 0.1, 0.3, 0.6$ and 1, respectively). In insets (B), (C) and (D) steady concentrations as a function of Q for s equal to 0.3, 0.6 and 1, respectively. In all cases $v = 20$.

p value increases, the fitness landscape approaches a single peak landscape. In contrast to the fitness landscape studied in the previous section, different amplification factors now exist within the same Hamming class sequences. This fact prevents a simplification of the mutation matrix similar to that used in Eq. (14). The probability of obtaining the sequence I_i during the error-prone self-replication of the sequence I_j is now calculated from the general formula

$$Q_{ij} = (1-q)^n q^{v-n} \tag{19}$$

where n is the number of digits in which both sequences differ. This fitness landscape is rugged since sequences that differ in, for instance, only one digit can possess very different amplification factors, that is to say, they depend critically on the position of the digits.

The dynamics of the molar fraction of each sequence of the population is also governed by Eq. (16). The simulations were performed using the same algorithm in a MATLAB platform as described in Section 3.1. In all cases, a sequence length $v = 10$ was used. Thus, in this case, the dynamical system represented by Eq. (16) has 1024 ordinary differential equations and, each time, the variables satisfy $\sum_k y_k(t) = 1$. As in the previous section, initially all the population is formed by sequences of 0, i.e. $y_0 = 1$, and whose amplification factor is $A_0 = (1/2^{10})^p$. The system (16) was numerically solved for several values of q, raging from 0.5 to 1, and for each value we computed the characteristic time associated with the molar fraction of the fittest sequence.

The results are shown in Fig. 5 for several values of the exponent p. As can be seen, the more relevant facts that appear in double peak landscape are reproduced. In particular, a value of Q at which the value of T_c attains its minimum is observed. As the
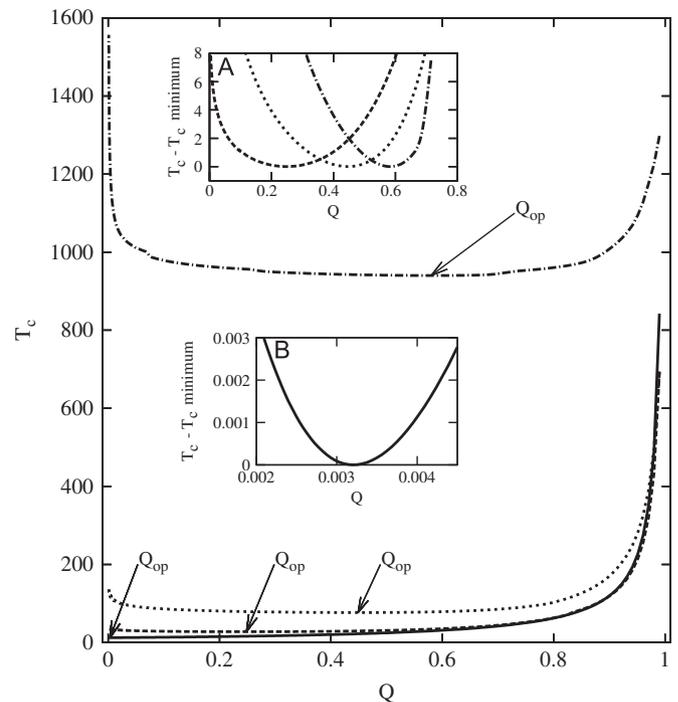

Fig. 5. The characteristic time for the evolution of I_{1024} as a function of the quality factor per sequence Q in a binary rugged fitness landscape. The parameter p is: 10 (dashed-dotted line), 4 (dotted line), 2 (dashed line), and 1 (solid line). Inset (A) shows the existence of the local minimum at Q_{op} for the values of p 10, 4 and 2. Inset (B) shows the local minimum at Q_{op} for $p=1$.
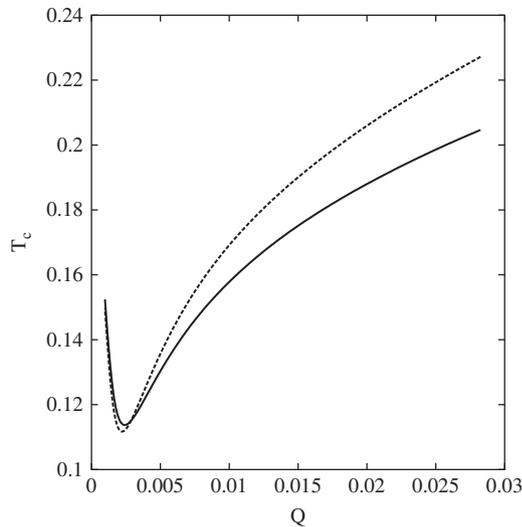
Fig. 6. The characteristic time for the evolution of fittest sequence as a function of the quality factor per sequence Q in the NK fitness landscape. for $N=10$ and $K=2$ (solid line) and $K=3$ (dashed line). Note the scale use in the Q-axis: $1/2^v < Q < 0.03$.

value of p increases, the T_c value increases. This is a straightforward consequence of the fact that the average productivity of the populations decreases with p at any q value. Moreover, the Q_{op}-value also increases with p. This fact does not have an evident explanation since as p increases the average productivity of the population decreases, while the apparent superiority of the fittest sequence increases. So, the compromise between the searching time and fixation time must vary.

3.4. NK fitness landscape

The NK model was introduced by Stuart Kauffman in the beginning of the nineties to study coevolution in simple organisms (Kauffman and Weinberger, 1989; Kauffman and Johnsen, 1991). Here, N corresponds to the chain length v, i.e. the number of positions of each sequence, whereas K ($0 < K < N$), represents the number of positions that influence on the contribution of a given position to the sequences fitness. This model generates fitness landscapes whose ruggedness increases with K.

In the classical version, the K positions that are related to each position are randomly assigned. Therefore, each position can provide 2^{K+1} values of fitness, that are computed at random, depending on its actual state and the state of the K positions that interact with it. The fitness of each species (sequence) is obtained from the average of the fitness of its N digits. As a consequence of these "long distance" interactions among the positions, sequences that belong to the same Hamming class could have very different fitness values, giving rise to rugged fitness landscapes.

Fig. 6 depicts the dependence of the characteristic time T_c on the quality factor Q in two NK fitness landscapes for $N=10$ and $K=2$ and $K=3$. The population dynamics (as in the previous section, there are 1024 sequences) is again driven by Eq. (16), with the entries of the mutation matrix given by Eq. (19). As can be seen, in both cases the characteristic time T_c exhibits a local minimum at Q-values close to the natural error threshold (located at $Q = 1/2^v$), respectively, $Q_{op} = 0.0024$ for $K=2$ and $Q_{op} = 0.0022$ for $K=3$.

4. Discussion

In this paper, we have studied the dependence of the characteristic time on the mutation rate for different fitness

landscapes. It is expected that, as a consequence of the trade-off between the searching process inherent to any error-prone system and the rate of fixation of new fitter mutants a value of Q at which the evolutionary time is minimum must exist. This opens interesting questions in evolutionary theory (Stich et al., 2007, 2010a,b; Stich and Manrubia, 2011).

The problem of evaluating this time is enormous. On the one hand, controlling transient events in complex systems is extremely difficult. On the other hand, ensuring that the system achieves a new stationary state is also risky. The time scale in continuous dynamical systems has been defined as the inverse of the module of the largest eigenvalue or, in some cases, as the inverse of the module of the difference between the two largest eigenvalues (Nowak and Schuster, 1989). However, its validity must be reconsidered since these definitions provide bad estimations even in simple linear dynamical systems since they only yield information about the approximation to a final state from its vicinity, and they do not take into account the initial conditions of the system. The characteristic time, on the contrary, takes into account the complete trajectory of the system, from an initial condition to the final stationary state. Therefore, the characteristic time is an average time and so, like any other average, can hide information relevant to the study of the system, such as the variation of the higher order moments. We are currently investigating the possibility of decomposing the characteristic time in different contributions (for example, searching time and fixation time) to avoid this limitation.

In Section 2, we have studied a population formed by two replicators with non-null back mutation rate and an explicit solution for T_c has been obtained Eq. (9). This analysis has provided the functional dependence of the characteristic time on the mutation rate and amplification factors for different initial conditions. T_c depends inversely on the selection coefficient of species I_1 over species I_2, $s = A_1 - A_2$, a result similar to that obtained by Johnson (1999) using a completely different approach. Contrary to the time obtained from the inverse of the eigenvalue, T_c also depends on the initial condition y_0. Naturally, as the initial condition is closer to the equilibrium point, T_c approaches the time provided by the inverse of the eigenvalue (Fig. 1B). When T_c is plotted as a function of the quality factor Q it exhibits a global maximum at a Q-value, Q_{1m}, that depends on y_0 and that tends to the error threshold Q_c as y_0 tends to 1.

Unfortunately, analytical solutions can only be found for simple systems. Notwithstanding this, although analytical expressions cannot be obtained in more complex systems, a numerical approach can be used. The characteristic time can be numerically obtained as its computation only requires knowledge of the trajectory, without an explicit definition of the dynamical system. In other words, the characteristic time can be applied in a semiempirical approach to study evolutionary processes described by arbitrary complex models, if a numerical solution is provided. Using this approach we have studied the dynamics of adaptation in some more complex systems in Section 3.

It is well known that, independently of the fitness landscape, an equal distribution of the sequences of the population in the equilibrium of selection is obtained for $Q = 1/2^v$. For some fitness landscapes this equidistribution of sequences occurs for a $Q > 1/2^v$. This Q-value is usually called error threshold and denoted as Q_c. As can be deduced from the results obtained in this paper, in the situations when an error threshold exists, e.g. the single-peak landscape and the multiplicative landscape for high enough values of s, the characteristic time T_c always exhibits a local maximum at a Q-value Q_m close to Q_c. In those cases where there is not evidence of the existence of an error catastrophe, T_c can present or not a largest value at $Q = 1/2^v$. As a matter of fact, while in the multiplicative fitness studied in Section 3.2 for low

values of s, T_c does not exhibit a maximum, this local maximum of T_c appears in the rugged fitness landscapes used in Sections 3.3 and 3.4.

When a second maximum of T_c appears as Q tends to 1, e.g. in fitness landscapes with more than one peak, a lowest value of the characteristic time is attained at Q_{op}, in between Q_m and $Q=1$. This lowest value depends on the fitness landscape and, in particular, on the size of the sequence and on the superiority of the master phenotype over the rest of the phenotypes (see Fig. 3). In the NK fitness landscape this minimum occurs very near the natural limit $Q = 1/2^v$ and then its relevance for evolution is questioned since the concentration of the fittest copy is very low. In the other cases, however, the minimum is achieved at larger values of Q.

Note that, despite the flat appearance of the curve of T_c as a function of Q in the neighborhood of Q_{op} (Fig. 2), small differences in the value of T_c could have strong consequences on the competitive dynamics of populations (as occurs with the tiny difference in the amplification factors). The question of how natural selection magnifies these small differences in finite size populations is currently under study.

Acknowledgments

References

Domingo, E., Martín, V., Perales, C., Grande-Perez, A., García-Arriaza, J., Arias, A., 2006. Viruses as quasispecies: biological implications. Curr. Top. Microbiol. Immunol. 299, 51–82.

Eigen, M., 1971. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften 58, 465–523.

Gokhale, C.S., Iwasa, Y., Nowak, M.A., Traulsen, A., 2009. The pace of evolution across fitness valleys. J. Theor. Biol. 259, 613–620.

Jain, K., Krug, J., 2007. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. Genetics 175, 1275–1288.

Johnson, T., 1999. The approach to mutation-selection balance in a infinite asexual population, and the evolution of mutation rates. Proc. R. Soc. Lond. B 266, 2389–2397.

Kamp, C., Wilke, C.O., Adami, C., Bornholdt, S., 2003. Viral evolution under the pressure of an adaptive immune system: optimal mutation rates for viral escape. Complexity 8 (2), 28–33.

Kauffman, S.A., Levin, S., 1987. Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Biol. 128, 11–45.

Kauffman, S.A., Weinberger, E.D., 1989. The NK model of rugged fitness landscapes and its application to maturation of the immune response. J. Theor. Biol. 141, 211–245.

Kauffman, S.A., Johnsen, S., 1991. Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and coevolutionary avalanches. J. Theor. Biol. 149, 467–505.

Krug, J., 2002. Tempo and mode in quasispecies evolution. In: Lässig, M., Valleriani, A. (Eds.), Biological Evolution and Statistical Physics. Springer, Berlin/Heidelberg, pp. 205–216.

Krug, J., Karl, C., 2003. Punctuated evolution for the quasispecies model. Physica A 318, 37–143.

Llorens, M., Nuño, J.C., Rodriguez, Y., Melendez-Hevia, E., Montero, F., 1999. Generalization of the theory of transition times in metabolic pathways: a geometrical approach. Biophys. J. 77, 23–36.

Nowak, M., Schuster, P., 1989. Error thresholds of replication in finite populations mutation frequencies and the onset of Muller's ratchet. J. Theor. Biol. 137, 375–395.

Nuño, J.C., de Vicente, J., Olarrea, J., López, P., Lahoz-Beltrá, R., 2010. Evolutionary daisyworld models: a new approach to studying complex adaptive systems. Ecol. Inf. 5, 231–240.

Schuster, P., 1997. Landscapes and molecular evolution. Phys. D: Nonlinear Phenom. 107 (2–4), 351–365.

Schuster, P., Swetina, J., 1988. Stationary mutant distribution and evolutionary optimization. Bull. Math. Biol. 50 (6), 635–660.

Schuster, P., Stadler, P.F., 2008. Early replicons: origin and evolution. In: Domingo, E., Parrish, C., Holland, J.J. (Eds.), Origin and Evolution of Viruses. Elsevier, Oxford, pp. 1–42.

Solé, R.V., Deisboeck, T.S., 2004. An error catastrophe in cancer? J. Theor. Biol. 228, 47–54.

Stich, M., Briones, C., Manrubia, S.C., 2007. Collective properties of evolving molecular quasispecies. BMC Evol. Biol. 7, 110.

Stich, M., Manrubia, S.C., 2011. Motif frequency and evolutionary search times in RNA populations. J. Theor. Biol. 280, 117–126.

Stich, M., Manrubia, S.C., Lázaro, E., 2010a. Variable mutation rates as an adaptive strategy in replicator populations. PLoS One 5, 6.

Stich, M., Lázaro, E., Manrubia, C.S., 2010b. Phenotypic effect of mutations in evolving populations of RNA molecules. BMC Evol. Biol. 10, 46.

Traulsen, A., Pacheco, J.M., Nowak, M.A., 2007. Pairwise comparison and selection temperature in evolutionary game dynamics. J. Theor. Biol. 246, 522–529.

Woodcock, G., Higgs, P.G., 1996. Population evolution on a multiplicative single-peak fitness landscape. J. Theor. Biol. 179, 61–73.