



Mutation landscapes

Akira Sasaki^a, Martin A. Nowak^{b,*}

^aDepartment of Biology, Faculty of Science, Kyushu University Graduate Schools, Fukuoka 812-8581, Japan

^bInstitute for Advanced Study, Olden Lane Einstein Drive, Princeton NJ 08540, USA

Abstract

A traditional picture of evolutionary dynamics with constant fitness is that of genomes living in sequence space and adapting on fitness landscapes. Mutation rates are considered to be constant or externally regulated. If, however, we take into account that genomes also encode for enzymes that perform replication and error correction, then individual genomes not only have a specific replication rate (fitness), but also a specific mutation rate. This leads to the concept of a mutation landscape. We explore evolution on mutation landscapes. Localization in pure mutation landscapes is only possible under extremely restrictive conditions. Coupling of mutation landscapes and fitness landscapes leads to localization and hence adaptation and evolution. We analyse how mutation landscapes facilitate localization in fitness landscapes and vice versa. Finally, we show that for mutation landscapes, at equilibrium, with constant environment, there is not necessarily selection for the minimum mutation rate. Instead, the target of selection is an optimum distribution of mutation rates, a ‘mutational quasispecies’.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Mutation landscape; Fitness landscape; Quasispecies; Error catastrophe; DNA repair

1. Introduction

Evolution is the interplay of mutation and selection. Mutation means that genomes are passed on from parent to offspring with some changes. Selection means that fitter individuals reproduce faster than others. Fitness can be frequency dependent or not. Frequency dependence emerges when the fitness of a phenotype depends on the frequency of itself and other phenotypes in the population. Here we study constant fitness. Genomes are given by sequences, and each sequence has a constant fitness value. Genomes are arranged in sequence space and adapt on a fitness landscape.

Wright (1932) introduced fitness landscapes, while Maynard Smith (1970) invented sequence space. Eigen and Schuster (1977) combined these two ideas. In the formalism of quasispecies theory, there is a replicating population of RNA or DNA genomes. Let us consider binary sequences of length n . There are 2^n possible sequences. Sequences replicate subject to mutation. Let us only consider point mutations. Denote by u the

mutation rate per bit per round of replication. The probability that replication of sequence i yields sequence j is given by $q_{ij} = u^{h_{ij}}(1-u)^{n-h_{ij}}$. The Hamming distance, h_{ij} , denotes the number of point mutations between sequences i and j . This distance defines the metric of sequence space. All sequences are arranged in such a way that nearest neighbors differ by a single point mutation. This gives rise to a hyper-cube of dimension n . Each sequence, i , has a certain reproduction rate, a_i , which defines its fitness. The fitness landscape is a function from sequence space into positive real numbers. Thus the fitness landscape is an n -dimensional mountain range (Fig. 1).

Denote by x_i the frequency of sequence i in an infinitely large population. Deterministic evolutionary dynamics can be written as

$$\dot{x}_i = \sum_j a_j x_j q_{ji} - \phi x_i. \quad (1)$$

This is the quasispecies equation. It describes constant selection with mutation. The average fitness of the population is given by $\phi = \sum_i a_i x_i$. Quasispecies dynamics have been studied in stochastic and deterministic frameworks (Fontana and Schuster, 1987; Swetina and Schuster, 1988; Eigen et al., 1989; Nowak and Schuster 1989).

*Corresponding author. Tel.: +1-609-734-8389; fax: +1-609-951-4438.

E-mail addresses: asasascb@mbox.nc.kyushu-u.ac.jp (A. Sasaki), nowak@ias.edu (M.A. Nowak).

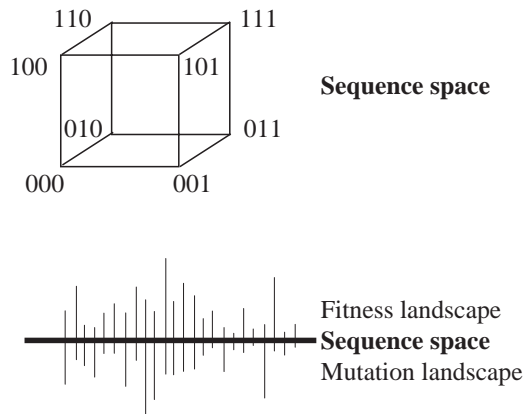


Fig. 1. In sequence space, sequences are arranged such that immediate neighbors differ by a single point mutation. All sequences have the same length, n . The resulting sequence space has n dimensions. For binary sequences, each dimension has two possibilities: 0 or 1. Fitness landscape is a function that assigns each sequence a specific fitness value (a reproductive rate), which is usually a positive real number. Genomes of organisms, however, encode for proteins that determine both reproductive and mutation rates. Hence it makes sense to consider the concept of a mutation landscape, which assigns each sequence a specific mutation rate. The figure shows sequence space for $n=3$ and the illustration of a mutation and a fitness landscape over a projection of sequence space.

The quasispecies adapts on the fitness landscape. It will move from regions of lower fitness to regions of higher fitness. It will climb up mountain slopes and edges to reach local peaks. There will be transitions from lower peaks to higher peaks. For deterministic dynamics, there is only one globally stable equilibrium, which is given by the eigenvector associated with the largest eigenvalue of the linear operator $w_{ij} = a_i q_{ij}$. The largest eigenvalue corresponds to the highest average fitness, ϕ , at equilibrium.

A crucial phenomenon of quasispecies dynamics is the error threshold. For many reasonable fitness landscapes the mutation rate, u , has to be less than $1/n$, for the equilibrium quasispecies distribution to be localized in a particular region of sequence space. Localization is required for the ability of the quasispecies to adapt to peaks in the fitness landscape. Localization and adaptation are required for evolution. In contrast, if $u > 1/n$, then for most fitness landscapes, the quasispecies will be delocalized, which means that all sequences will be roughly equally frequent. In this case, any finite population will wander endlessly in sequence space.

The new idea that we introduce in the present paper is that individual genomes are not only associated with a specific fitness value, but also with a specific mutation rate. We imagine that large parts of the genome encode for enzymes involved in DNA replication or error correction. For example, more than 30 genes encode for DNA replication and error correction in the T4 phage whose total genome contains 300 genes. In the eukaryotic genome, hundreds or thousands of genes

contribute to an overall low mutation rate: these are genes that participate in DNA replication, synthesis of materials for DNA replication, error correction, recombination mechanisms, cell cycle check points, chromosome segregation, etc. We want to understand the selection pressure that ensures the evolution and adaptation of genes that affect mutation rates. The particular base sequences of such genes determine the overall mutation rate. In a quasispecies framework, sequence j is associated with a specific mutation rate u_j . This gives rise to the concept of a mutation landscape (Fig. 1). Evolution is now described by adaptation of the quasispecies on both mutation and fitness landscapes. We want to study the conditions for adaptation (that is localization) of quasispecies on mutation landscapes and on combined mutation and fitness landscapes.

Mutation landscapes provide a general framework for studying the evolution of mutation rates. There is a large and interesting literature on this topic. Kimura (1967) suggested that mutation rates of natural populations are determined by a balance between (i) selection against deleterious mutants favoring lower mutation rates and (ii) costs incurred by any further reduction in the mutation rate. Such costs can include energetic and kinetic considerations of DNA replication, proof-reading and repair. This idea was further analysed by Kondrashov (1995) and Dawson (1999). In the absence of any costs associated with reduced mutation rates, Fisher (1918, 1930) and Leigh (1970, 1973) argued that in an asexual population the optimum mutation rate can be positive given there is a possibility of producing advantageous mutants (see also Sniegowski et al., 1997; Taddei et al., 1997). In this context, the mutation rate is determined by selection against deleterious mutants and selection of beneficial mutants. Most of the mathematical models on the evolution of mutation rates include modifier alleles that affect the mutation rate of a genome without directly affecting its fitness (Karlín and McGregor, 1974; Gillespie, 1981; Liberman and Feldman, 1986; Ishii et al., 1989; Haraguchi and Sasaki, 1996).

The term ‘mutation landscape’ was first used by Gillespie (1984, 1991), but with a different meaning. Gillespie pictured a population being stuck on a local optimum of a fitness landscape unable to reach the global optimum, because the mutation rate being too low. He referred to this phenomenon as ‘mutation landscape’. This idea is very different from ours.

There is a fascinating and growing literature dealing with evolutionary dynamics on fitness landscapes (Weinberger, 1990; Holland, 1992; Kauffman 1993; Hordijk, 1996; Stadler, 1996; Fontana and Schuster 1998; Stadler and Wagner, 1998; Newman and Engelhardt, 1998; van Nimwegen et al., 1999; Naudts and Kallel, 2000; Krakauer and Plotkin, 2002; Smith et al., 2002). One aim of this paper is to induce similar investigations into mutation landscapes.

2. Mutation landscapes

Consider a quasispecies of binary sequences with length n . Let x_0 be the frequency of the sequence with lowest mutation rate. We refer to this sequence also as the highest fidelity sequence. Let x_i denote the frequencies of sequences that are i point mutations away from the highest fidelity sequence. All sequences have the same fitness, and we ask what is the condition for the maintenance of the highest fidelity sequence.

Suppose that the mutation rate u_i of the sequences are an increasing function of the Hamming distance, i , from the highest fidelity sequence: $u_0 < u_1 < \dots < u_n$. Evolutionary dynamics are given by

$$\begin{aligned} \dot{x}_0 &= -nu_0x_0 + u_1x_1, \\ \dot{x}_i &= (n - i + 1)u_{i-1}x_{i-1} - nu_ix_i + (i + 1)u_{i+1}x_{i+1}, \\ i &= 1, \dots, n - 1, \end{aligned} \tag{2}$$

$$\dot{x}_n = u_{n-1}x_{n-1} - nu_nx_n.$$

At equilibrium, we have the detailed balance condition $(n - i + 1)u_{i-1}\hat{x}_{i-1} = iu_i\hat{x}_i$. This yields

$$\hat{x}_i = \binom{n}{i} \frac{1}{u_i} / \sum_{j=0}^n \binom{n}{j} \frac{1}{u_j}, \quad i = 0, 1, \dots, n. \tag{3}$$

Let us now consider two specific examples of simple mutation landscapes.

2.1. Symmetric mutation landscapes

If the mutation rate of sequences increase with the distance from the highest fidelity sequence as $u_i = u_0\alpha^i$ with $\alpha > 1$, then the stationary distribution is binomial

$$\hat{x}_i = \binom{n}{i} \left(\frac{1}{\alpha + 1}\right)^i \left(1 - \frac{1}{\alpha + 1}\right)^{n-i}. \tag{4}$$

The frequency of the highest fidelity sequence is $\hat{x}_0 = (1 + 1/\alpha)^{-n}$. For $\alpha \gg 1$, we have $\hat{x}_0 \approx e^{-n/\alpha}$. In this case, the population is localized at the highest fidelity sequence only if $n < \alpha$. As the genome length, n , increases, this condition implies a very steep mutation landscape.

2.2. Single-valley mutation landscapes

Let us now analyse the case where one sequence has a lower mutation rate than all other sequences, and all other sequences have the same mutation rate. The mutation rates are given by $u_0 < u_1 = u_2 = \dots = u_n$. At equilibrium, we obtain

$$\dot{x}_0 = \frac{u_1}{u_0} / \left[2^n + \left(\frac{u_1}{u_0} - 1\right) \right], \tag{5}$$

$$\dot{x}_i = \binom{n}{i} / \left[2^n + \left(\frac{u_1}{u_0} - 1\right) \right], \quad i = 1, 2, \dots, n. \tag{6}$$

For the localization at the high fidelity sequence, the mutation rate u_0 of the high fidelity sequence must be of the order of 2^{-n} . Again this condition is extremely restrictive for reasonable long genome length, n . In comparison the condition for localization under a single-peak fitness landscape is about $u_0 < 1/n$, which is the standard error threshold.

In general, if we ignore back mutation to the highest fidelity sequence, then the first equation of system (2) is replaced by $\dot{x}_0 = -nu_0x_0$. Thus, the highest fidelity sequence is never maintained without back mutation. In contrast, for single peak fitness landscapes, the error threshold specifies also the condition for maintaining the fittest sequence in the absence of back mutation.

3. Coupling mutation and fitness landscapes

3.1. A single fitness peak and a single mutation valley

Suppose that the whole genome of an organism consists of n positions that determine mutation rate and of m positions that determine fitness. We could also include k positions that determine both fitness and mutation rate, but this does not change our analysis. We assume a single-peak mutation landscape: only one particular sequence of length n ensures the low mutation rate, u_0 , while all other sequences have mutation rate u_1 with $u_0 < u_1$. We also assume a single-peak fitness landscape for the fitness-defining part of the genome of length m . Only one sequence of length m (the wild type) ensures a higher fitness $a_0 > 1$ than the rest of sequences, which have fitness $a_1 = 1$. We define the Malthusian parameter of the wild type as $s_0 = \log a_0 > 0$. The Malthusian parameter of the other sequences is 0. There are 2×2 classes of sequences: a single high-fidelity high-fitness sequence, m high-fidelity low-fitness sequences, n low-fidelity high-fitness sequences, and $2^{n+m} - n - m + 1$ low-fidelity low-fitness sequences. We denote the frequencies of these classes by x_{00} , x_{01} , x_{10} , and x_{11} , respectively.

Neglecting the back mutations to high-fidelity or high-fitness sequences, the evolutionary dynamics are given by

$$\dot{x}_{00} = -(n + m)u_0x_{00} + (s_0 - \bar{s})x_{00}, \tag{7a}$$

$$\dot{x}_{01} = mu_0x_{00} - nu_0x_{01} - \bar{s}x_{01}, \tag{7b}$$

$$\dot{x}_{10} = nu_0x_{00} - mu_1x_{10} + (s_0 - \bar{s})x_{10}, \tag{7c}$$

$$\dot{x}_{11} = nu_0x_{01} + mu_1x_{10} - \bar{s}x_{11}. \tag{7d}$$

The population average Malthusian fitness is given by $\bar{s} = s_0(x_{00} + x_{10})$. This is a quasispecies equation with separate mutation and selection terms. There are three equilibria:

(i) *High-fidelity high-fitness equilibrium*. In this case, the equilibrium frequency of the sequence with high fidelity and high fitness, \hat{x}_{00} , is positive. The equilibrium frequencies are given by

$$\hat{x}_{00} = \left(1 - \frac{(n+m)u_0}{s_0}\right) \left(1 - \frac{nu_0}{m(u_1 - u_0)}\right), \tag{8a}$$

$$\hat{x}_{01} = \frac{mu_0/s_0}{1 - mu_0/s_0} \left(1 - \frac{(n+m)u_0}{s_0}\right) \left(1 - \frac{nu_0}{m(u_1 - u_0)}\right), \tag{8b}$$

$$\hat{x}_{10} = \left(1 - \frac{(n+m)u_0}{s_0}\right) \frac{nu_0}{m(u_1 - u_0)}, \tag{8c}$$

$$\hat{x}_{11} = 1 - \hat{x}_{00} - \hat{x}_{01} - \hat{x}_{10}. \tag{8d}$$

This equilibrium exists and is stable if

$$u_0(n+m) < s_0, \quad u_0(n+m) < u_1m. \tag{9}$$

The first condition states that the mutation rate per genome has to be less than the Malthusian fitness. This is the classical error threshold. The second condition states that the mutation rate per genome has to be reduced by adding genes that reduce the mutation rate: for the high-fidelity, high-fitness organism, the mutation rate per genome is given by $u_0(n+m)$; this quantity has to be lower than u_1m which is the mutation rate per genome for the low-fidelity, high-fitness organism.

The two conditions can also be written as a limit on the total genome length, $n+m < (1/u_0) \min\{s_0, mu_1\}$. Maintenance of low mutation rate and high fitness is favored if n and u_0 are as small as possible and if s_0 is as large as possible. Interestingly, for the number of selective positions, m , we obtain both an upper and a lower bound:

$$\frac{nu_0}{u_1 - u_0} < m < \frac{s_0}{u_0} - n. \tag{10}$$

If m is too small there is not enough selection for low mutation rate, if m is too large then the sequence with high fitness cannot be maintained, because the population is beyond the error threshold.

(ii) *Low-fidelity, high-fitness equilibrium*. Here we have $\hat{x}_{00} = \hat{x}_{01} = 0$ but $\hat{x}_{10} > 0$ and $\hat{x}_{11} > 0$. From the right-hand side of (7c), we see that $\bar{s} = s_0 - mu_1$ must be satisfied at equilibrium with $\hat{x}_{10} > 0$, and then $\hat{x}_{01} = 1 - (mu_1/s_0)$ and $\hat{x}_{11} = mu_1/s_0$. This equilibrium exists and is stable if $u_1m < s_0$ and $u_1m < u_0(n+m)$.

(iii) *Low-fidelity, low fitness equilibrium*. In this case, we have $\hat{x}_{00} = \hat{x}_{01} = \hat{x}_{10} = \hat{x}_{11} = 0$ and $\hat{x}_{11} = 1$. This equilibrium exists and is stable if $u_0 = s_0 > 1/(n+m)$ and $u_1/s_0 > 1/m$.

The phase diagram is illustrated in Fig. 2.

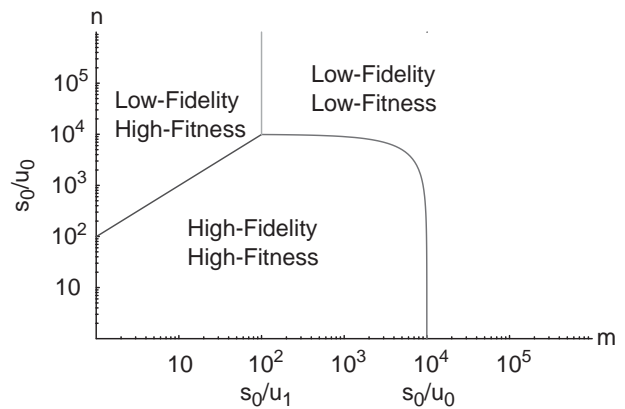


Fig. 2. Coupling of mutation and fitness landscapes can lead to selection of high fitness and high fidelity (that is low mutation rate). Suppose n positions of a binary sequence have to be maintained to ensure a low mutation rate, u_0 . Any change in these n positions increases the mutation rate to u_1 . Suppose m different positions have to be maintained to ensure a high Malthusian fitness value, s_0 . Any change in these m positions reduces the Malthusian fitness to 0. Quasispecies dynamics are given by system (7). There is always a single stable equilibrium. Depending on conditions outlined in the text and illustrated in the figure, this equilibrium will contain the high-fidelity/high-fitness sequence, the low-fidelity/high-fitness sequences or only the low-fidelity/low-fitness sequences.

3.2. A single fitness peak and a complex mutation landscape

Let us now consider a general mutation landscape that is linked to a single-peak fitness landscape. As before, we assume there are n positions that determine the mutation rate. We label the binary sequences of length n as $i = 1, \dots, 2^n$. The mutation probability from sequence i to sequence j is given by

$$q_{ij} = u_j^{h_{ij}} (1 - u_j)^{n-h_{ij}}. \tag{11}$$

As before, we assume there are m positions that determine the fitness. The fitness landscape has a single peak. Hence, there is one particular sequence of length m that has fitness, $\alpha > 1$, while all other sequences have fitness 1. Denote by x_i the frequency of the genome with a high fitness sequence and with sequence i in the mutational part. Denote by y_i the frequency of genomes with any low fitness sequence and with sequence i in the mutational part. Evolutionary dynamics are given by

$$\dot{x}_i = \sum_j ax_j q_{ji} (1 - u_j)^m - \phi x_i, \tag{12}$$

$$\dot{y}_i = \sum_j ax_j q_{ji} [1 - (1 - u_j)^m] + \sum_j y_j q_{ji} - \phi y_i. \tag{13}$$

The average fitness of the population is $\phi = \sum_j (ax_j + y_j)$.

Fig. 3 shows a computer simulation. The mutation rates, u_i , are uniformly distributed between 0.02 and 0.5. The number of positions that affect the mutation rate is

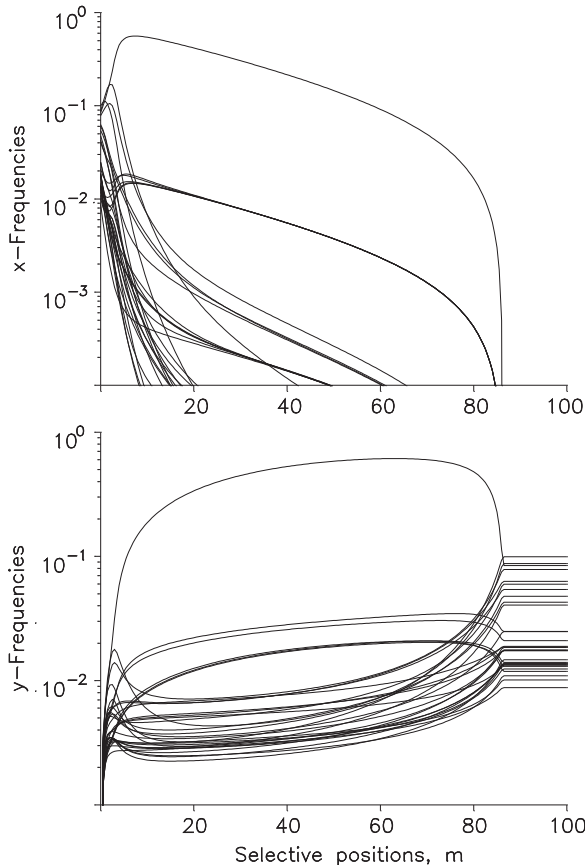


Fig. 3. Localization and error threshold in a mutation landscape. In this example we couple a single-peak fitness landscape with a complex mutation landscape. The fitness landscape is defined on a sequence space of m positions. This number varies from $m = 1$ to 100 and is presented on the x -axis. The all-0 sequence has fitness $a = 10$, while all other sequences have fitness 1. The mutation rate is determined by 5 other positions. Therefore, we have $2^5 = 32$ different mutation sequences, whose mutation rates are uniformly distributed between 0.02 and 0.5. The variables x_i and y_i denote, respectively, the relative abundances of these 32 sequences in conjunction with high and low fitness. As predicted, we observe a minimum and a maximum number m that is compatible with selection for low mutation rates. The lowest mutation rate in our random sample was $u_0 = 0.02493$, while the average mutation rate is $u_1 \approx 0.25$. Using Eq. (10), we find the $0.5 < m < 87$ is required for localization in the mutation landscape. This result is in perfect agreement with the numerical simulation. For $m < 87$, the population is centered around a high-fitness sequence with a low mutation rate. For $m > 87$, the high-fitness sequences disappear; there is a fairly that distribution of sequences with low fitness and different mutation rates.

$n = 5$. Hence there are $2^5 = 32$ different mutation sequences. The number of selective positions varies from $m = 1$ –100. The fitness peak is $a = 10$. There is a minimum and a maximum number m that is compatible with selection for low mutation rates. We can use the calculation of Section 3.1 to estimate these boundaries. The Malthusian fitness is given by $s_0 = \log a$. The lowest mutation rate in our random sample was $u_0 = 0.02493$. The average mutation rate is $u_1 \approx 0.25$. From Eq. (9), we

obtain $0.5 < m < 87$, which is in perfect agreement with the simulation.

3.3. Selection of ‘mutational quasispecies’

Let us now couple a single-peak fitness landscape with a mutation landscape that has two valleys. Valley A is deep and narrow, valley B is less deep but broader. In a constant environment, the expectation is that mutation rates are as low as possible. Thus there should always be selection for A over B. Here we show that this is not the case. Whether A or B wins, depends on the fitness landscape.

Fig. 4 shows a computer simulation. As before, the number of positions that affect the mutation rate is $n = 5$. The sequence 00000 has mutation rate 0.019. The sequence 11111 has mutation rate 0.0195. The 5 one-error mutants of the sequence 11111 have mutation rates 0.02, 0.0205, 0.021, 0.0215, 0.022. All other sequences have mutation rate 0.5. The fitness peak is $a = 10$. For m less than about 42, the dominant sequence is 11111. Hence, in this case there is no selection for the lowest mutation rate. As m exceeds 42 there is a sharp transition and suddenly the dominant sequence is 00000. From Eq. (9), we can calculate that sequences with high fidelity and high fitness will disappear at the critical value $m = [\log(10)/0.019] - 5 = 116$. This is exactly the case.

In this example, for $m < 42$, we have the interesting result that for a constant environment at equilibrium there is no selection for minimum mutation rate. This is contrary to a current dogma of evolutionary biology. The target of selection is not the sequence with minimum mutation rate, but the optimum distribution of mutation rates, the ‘mutational quasispecies’ that maximizes fitness at equilibrium.

A similar concept has been described for fitness landscapes (Swetina and Schuster, 1988). Imagine a fitness landscape with two peaks. Peak A contains a single sequence with fitness a_0 . Peak B contains several sequences with fitness $a_1 < a_0$. All other sequences have fitness 1. All sequences have the same mutation rate. There are two critical mutation rates, $u_1 < u_2$. If $u < u_1$, the quasispecies at equilibrium will be centered around peak A. If $u_1 < u < u_2$, however, the quasispecies is centered around peak B. In this case, the sequence with highest fitness, a_0 , is not present in the population. Hence, there is no ‘survival of the fittest’. Instead there is survival of the quasispecies (with highest average fitness, ϕ).

4. Adding genes to reduce mutation

We can use the model of Section 3.1 to study the conditions for adding a gene that would reduce the

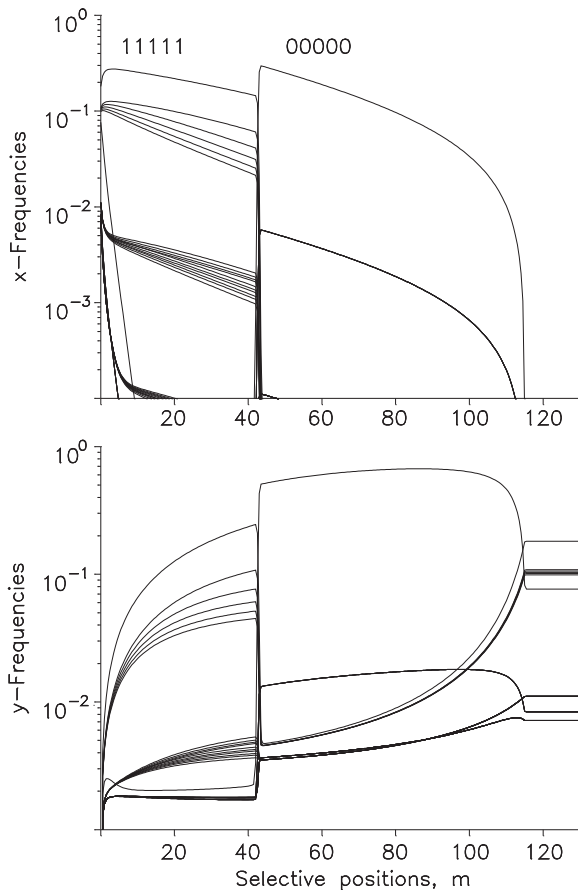


Fig. 4. In a constant environment, at equilibrium, there is selection for a mutational quasispecies rather than selection for minimum mutation rate. In this example, $n=5$ positions affect the mutation rate. The sequence 00000 has mutation rate 0.019, while the sequence 11111 has the slightly higher mutation rate 0.0195. The 5 one-error mutants of the sequence 11111 have mutation rates 0.02, 0.0205, 0.021, 0.0215, 0.022. All other sequences have mutation rate 0.5. These values define a mutation landscape with two valleys. The valley around 00000 is deep and narrow, whereas the valley around 11111 is slightly less deep but broader. For the fitness landscape we assume a single peak of fitness $a=10$; all other sequences have fitness 1. There are m positions that determine the fitness landscape. For m less than about 42, the dominant sequence in the mutation landscape is 11111. Hence, in this case there is no selection for the lowest mutation rate. As m exceeds 42 there is a sharp transition and suddenly the dominant sequence is 00000. From Eq. (10), we can calculate that sequences with high fidelity and high fitness will disappear at the critical value $m = \lceil \log(10)/0.019 \rceil - 5 = 116$. This is confirmed by our numerical analysis.

mutation rate. Consider a genome of length m with mutation rate u_1 and Malthusian fitness s_0 . Evolutionary stability in terms of the error threshold requires that $s_0 > mu_1$. Let us now add a gene of length n that reduces the mutation rate to u_0 . Thus $u_0 < u_1$. From Eq. (9), we find that this gene can be maintained in the population provided

$$\frac{m}{m+n}u_1 - u_0 > 0. \tag{14}$$

If $n \ll m$ then we simply need $u_1 - u_0 > 0$. Hence, any gene that reduces the mutation rate can be added to a large genome.

A gene that reduces the mutation rate, however, should not come for free. There could be a constant cost, c , that reduces the Malthusian fitness to $s_0 - c$. Furthermore, there could be a cost proportional to the total genome size, thereby further reducing the Malthusian fitness to $s_0 - c - \gamma(n + m)$. In this case, the gene can be maintained in the population if

$$\frac{m}{m+n}u_1 - u_0 > \gamma + \frac{c}{m+n}. \tag{15}$$

Again if $n \ll m$ then $u_1 - u_0 > \gamma + c/m$. Therefore the difference in the mutation rates has to exceed the costs. These are fairly traditional considerations linking the new concept of mutation landscapes to the existing literature on the evolution of mutation rates.

5. Discussion

Genomes encode for proteins that determine both reproductive rates (fitness) and mutation rates. Hence, evolution is adaptation on both fitness landscapes and mutation landscapes. We have studied some basic properties of mutation landscapes. If the fitness landscape is completely flat, there is no localization in a mutation landscape (other than in extremely restrictive circumstances). The interaction between fitness and mutation landscapes leads to localization in mutation landscapes. We have derived analytic results for conditions of localization (error thresholds) on simple fitness and mutation landscapes.

Suppose m positions in a genome contribute to an increased Darwinian fitness, a_0 , relative to a fitness of 1 if at least one of these positions is mutated. The Malthusian fitness is given by $s_0 = \log a_0$. If the mutation rate per base is u_1 , then the error threshold is given by $u_1 m < s_0$. This means, the mutation rate per genome has to be less than the Malthusian fitness. Since fitness advantages cannot be arbitrarily large, this condition provides a limit for the maximum genome size for a given mutation rate. Let us now add n positions to the genome that reduce the overall mutation rate from u_1 to u_0 . The evolutionary stability of the new genome of length $n + m$ requires two conditions to hold:

- (i) $u_0(n + m) < s_0$;
- (ii) $u_0(n + m) < u_1 m$

As before the total mutation rate per genome has to be less than the Malthusian fitness. This is the error threshold for a genome of length $n + m$ and mutation rate u_0 . In addition, the total mutation rate per genome, $u_0(n + m)$, has to be less than the total mutation rate per genome, $u_1 m$, for the genome of size m and mutation

rate of u_1 . This two conditions have to be fulfilled for the evolution of complex life.

Furthermore, if a mutation landscape has two valleys, one deep and narrow, the other one shallower but broader, there is not necessarily selection for lowest possible mutation rate. Hence, in a constant environment, natural selection does not necessarily favor the genome with the lowest possible mutation rate. Instead the target of selection is a ‘mutational quasispecies’.

References

- Dawson, K.J., 1999. The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theor. Popul. Biol.* 55, 1–22.
- Eigen, M., Schuster, P., 1977. The hypercycle. *Naturwissenschaften* 64, 541–565.
- Eigen, M., McCaskill, J., Schuster, P., 1989. The molecular quasispecies. *Adv. Chem. Phys.* 75, 149–263.
- Fisher, R.A., 1918. The correlation between relatives under the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* 52, 399–433.
- Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*, 1st Edition. Clarendon Press, Oxford, 2nd Edition, 1958, Dover Publications, New York.
- Fontana, W., Schuster, P., 1987. A computer model of evolutionary optimization. *Biophys. Chem.* 26, 123–147.
- Fontana, W., Schuster, P., 1998. Continuity in evolution: on the nature of transitions. *Science* 280, 1451–1455.
- Gillespie, J.H., 1981. Mutation modification in a random environment. *Evolution* 35, 468–476.
- Gillespie, J.H., 1984. Molecular evolution over the mutational landscape. *Evolution* 38, 1116–1129.
- Gillespie, J.H., 1991. *The Causes of Molecular Evolution*. Oxford University Press, Oxford.
- Haraguchi, Y., Sasaki, A., 1996. Host-parasite arms race in mutation modifications: indefinite escalation despite a heavy load? *J. Theor. Biol.* 183, 121–137.
- Holland, J., 1992. *Adaptation in Natural and Artificial Systems*, 2nd Edition. MIT Press, Cambridge, MA.
- Hordijk, W., 1996. A measure of landscapes. *Evol. Comput.* 4, 335–360.
- Ishii, K., Matsuda, H., Iwasa, Y., Sasaki, A., 1989. Evolutionarily stable mutation rate in a periodically fluctuating environment. *Genetics* 121, 163–174.
- Karlin, S., McGregor, J., 1974. Towards a theory of the evolution of modifier genes. *Theor. Popul. Biol.* 5, 59–103.
- Kauffman, S., 1993. *The Origins of Order: self-organization and selection in evolution*. Oxford University Press, Oxford.
- Kimura, M., 1967. On the evolutionary adjustment of spontaneous mutation rate. *Genet. Res.* 9, 23–34.
- Kondrashov, A.S., 1995. Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. *Genet. Res.* 66, 53–69.
- Krakauer, D.C., Plotkin, J., 2002. Redundancy, antiredundancy and the robustness of genomes. *PNAS* 99, 1405–1409.
- Leigh, E.G., 1970. Natural selection and mutability. *Am. Nat.* 104, 301–305.
- Leigh, E.G., 1973. The evolution of mutation rates. *Genetics* 73 (Suppl.), 1–18.
- Liberman, U., Feldman, M.W., 1986. Modifiers of mutation rate: a general reduction principle. *Theor. Popul. Biol.* 30, 125–142.
- Maynard Smith, J., 1970. Natural selection and the concept of a protein space. *Nature* 225, 563–564.
- Naudts, B., Kallel, L., 2000. A comparison of predictive measures of problem difficulty in evolutionary algorithms. *IEEE Trans. Evol. Comput.* 4, 1–15.
- Newman, M., Engelhardt, R., 1998. Effects of selective neutrality on the evolution of molecular species. *Proc. R. Soc. London B* 265, 1333–1338.
- Nowak, M.A., Schuster, P., 1989. Error thresholds of replication in finite populations. *J. Theor. Biol.* 137, 375–395.
- Smith, T., Husband, P., Layzell, P., O’Shea, M., 2002. Fitness landscapes and evolvability. *Evol. Comput.* 10, 1–34.
- Sniegowski, P.D., Gerrish, P.J., Lenski, R.E., 1997. Evolution of high mutation rates in experimental populations of *E. Coli*. *Nature* 387, 703–705.
- Stadler, P., 1996. Landscapes and their correlation functions. *J. Math. Chem.* 20, 1–45.
- Stadler, P., Wagner, G., 1998. Algebraic theory of recombination spaces. *Evol. Comput.* 5, 241–275.
- Swetina, J., Schuster, P., 1988. Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.* 50, 635–660.
- Taddei, F., Radman, M., Maynard-Smith, J.M., Toupance, B., Gouyon, P.H., Godelle, B., 1997. Role of mutation alleles in adaptive evolution. *Nature* 387, 700–702.
- van Nimwegen, E., Crutchfield, J., Huynen, M., 1999. Neutral evolution of mutational robustness. *Proc. Nat. Acad. Sci.* 96, 9716–9720.
- Weinberger, E., 1990. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol. Cybernet.* 63, 325–336.
- Wright, S., 1932. The role of mutation, inbreeding, crossbreeding and selection in evolution. In: Jones, D. (Ed.) *Proceedings of the Sixth International Congress on Genetics*, vol. 1, pp. 356–366. Reprinted in Ridley, M., 1997. *Evolution*, pp. 32–40, Oxford University Press, Oxford.