Genet. Res. Camb. (1997), **69**, *pp.* 127–136. *With 2 figures.* Printed in the United Kingdom © 1997 Cambridge University Press

Model dependency of error thresholds: the role of fitness functions and contrasts between the finite and infinite sites models

THOMAS WIEHE
Department of Integrative Biology, University of California, Berkeley, CA 94720, USA

(*Received 31 January 1996 and in revised form 26 August 1996 and 8 November 1996*)

Summary

Based on a deterministic mutation–selection model the concept of error thresholds is critically examined. It has often been argued that genetic information – for instance, an advantageous allele – can be selectively maintained in a population only if the mutation rate is below a certain limit, the error threshold, which is inversely related to the genome size. Here, I will show that such an inverse relationship strongly depends on the fitness model. To produce the error threshold, as given by Eigen (1971), requires that the fitness model is an extreme form of diminishing epistasis. The error threshold, in a strict sense, vanishes as epistasis changes from diminishing to synergistic. In the latter case even the usual definition of error thresholds becomes ambiguous. Initially, a finite sites model has been used to describe error thresholds. However, they can also be defined within the framework of the infinite sites model. I study both models in parallel and compare their properties as far as error thresholds are concerned. It is concluded that error thresholds possibly play a much less important role in molecular evolution than has often been assumed in the past.

1. Introduction

A discussion about the evolutionary role of what was later called error threshold was initiated in the early 1970s. Underlying this discussion is a model of biochemical replicator systems, devised by Eigen (1971) for a description of prebiotic evolutionary processes. A somewhat simplified version of this model is equivalent to the deterministic population genetical mutation–selection equation (see Crow & Kimura, 1970, and references therein). The concept of error thresholds translates to the question of how large the mutation rate can be in order not to completely eliminate an advantageous allele from the population, when mutation and selection are balanced. For a finite sites model the answer may – among other quantities – depend on the genome size, measured by the number of sites. Error thresholds are usually associated with an *inverse* relationship between per nucleotide mutation probability (p) and genome size (ν). The formula given by Eigen (1971) is

$$\nu_{\max} = \frac{\log \sigma}{p}, \tag{1}$$

where σ is a so-called superiority parameter which depends on the fitness model. The interpretation of this equation is the following. For a given mutation rate there is a maximal evolutionarily permissible genome size if genetic information (in the form of an advantageous allele) is to be maintained in the population. As a corollary, a given sequence length defines a maximal permissible mutation rate, the *error threshold*. Based on these relations, it was argued (Eigen & Schuster, 1979) that evolution at the early stages of life was confronted with a catch-22 situation (no enzymes without a large genome and no large genome without enzymes, which enhance the replication accuracy), also named 'information crisis'.

In this article I revisit the deterministic mutation–selection model and investigate the dependency of error thresholds on assumptions such as the underlying fitness function or the finite (cf. Wright, 1949) versus infinite sites model (Kimura, 1969). Both turn out to be crucial for the existence and magnitude of error thresholds.

Based on an infinite sites model and using discrete time difference equations Wagner & Krall (1993) reported a condition under which no error threshold exists. However, their model is somewhat more restrictive than the one used here. It allows only for single-step mutations and considers merely fitness functions which decrease monotonically as the number of mutations increases. For the continuous time model, I found results which are analogous to those of

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

the former authors. Further, I treated several generalizations of fitness functions for the finite and infinite sites models in parallel. In particular, explicit threshold formulae are derived. Since one of the important characteristics of error thresholds is the limit they set on genome size (Nee & Maynard Smith, 1990), an essential point is missed if the study of error thresholds and their dependence on fitness functions is restricted to the infinite sites model.

Here, I concentrate on the haploid version of the coupled (Hadeler, 1981) mutation–selection model. It has been used previously in this context (e.g. Wiehe *et al.*, 1995). It describes the dynamics of a wild-type allele and its variants, derived by mutation, for an infinitely large population. In this model alleles are – for simplicity – binary nucleotide sequences with a common length $\nu \leqslant \infty$, but which may differ by point mutations. Furthermore, it is assumed that reverse mutations can be neglected, and that alleles which differ from a particular type, called *wild-type* in the following, by the same number of point mutations have identical fitness.

A fitness model which is often adopted in discussions of error thresholds (Swetina & Schuster, 1982; Nowak & Schuster, 1989; Wiehe *et al.*, 1995) is the so-called *single-peaked* fitness landscape: only the wild-type is distinguished by some fitness advantage. Any other type, differing in as little as a single mutated site, is at a disadvantage, expressed in its fitness $1-s$. Such a two-class model was found to be adequate, for instance, to describe evolution of the coliphage Qβ (Domingo *et al.*, 1978). Other examples discussed (Eigen & Biebricher, 1988) in this context are the highly error-prone replication of viruses (Ortin *et al.*, 1980; Martinez-Salas *et al.*, 1985) or the serial transfer experiments of *in vitro* replication of RNA (Spiegelman *et al.*, 1965). However, concerns about the biological adequacy of a single-peaked model and its more general applicability, for instance to evolution of higher organisms, have been raised (Maynard Smith, 1983; Charlesworth, 1990; A. S. Kondrashov, personal communication).

Although some consideration has been given by Eigen & Biebricher (1988) to the 'fitness topography of sequence space', the dependency of error thresholds on this topography has not been investigated. In particular, one is left with the impression that any fitness topography with a 'superior master' allele (Eigen & Biebricher, 1988, p. 222) might produce an error threshold which is inversely related to genome size.

It therefore still appears worthwhile to ask how general error thresholds are and, if possible, to quantify them in terms of the usual parameters. Despite numerous studies, there is no unambiguous or generally accepted definition of the term 'error threshold'. I will adhere to two characterizations which are commonly used. One is to determine the mutation parameter such that the equilibrium frequency of the wild-type becomes zero. Alternatively, overall population statistics may be used, such as the average distance E of an allele from the wild-type (in terms of number of mutations) or the index of dispersion D, both defined below.

Mutation–selection balance had been an extensively discussed topic long before it attracted the attention of biochemically motivated research. Kimura & Maruyama (1966) investigated the effect of epistasis on the mutation load. Their interest has been in quantifying the genetic load for different models of selection and contrasting sexual and asexual replication. They found that diminishing epistasis produces a much higher mutational load than synergistic epistasis, if reproduction is diploid. However, they did not directly compare the magnitudes of mutation rates which are permissible for the different fitness models so as not to stall natural selection. This is done below and it is found that the kind of epistasis also strongly influences the (haploid) error threshold.

The model

The fundamental mutation–selection equation, in the form of coupled mutation and selection terms, is

$$\dot{y}_k = \sum_{i=0}^{\nu} y_i v_i m_{ki} - y_k \bar{v}, \quad 0 \leqslant k \leqslant v, \quad \nu \leqslant \infty. \tag{2}$$

Here, y_k denotes the frequency of the class of alleles which differ from the wild-type (class 0) by exactly k point mutations. v_k are their associated fitness values. The mean fitness of the population is $\bar{v} = \sum_{i=o}^{\nu} y_i v_i$. m_{ki} are entries of a mutation matrix M. They describe transitions from type i to type k. For the finite and the infinite sites models the mutation rates have to be defined separately. In the former case, single sites mutate with probability p; in the latter, p has to be substituted by a genome mutation rate, denoted λ. The probabilities m_{ij} are from a binomial distribution in the former case (see (3)). They are replaced by Poisson mutation rates $\hat{m}_{ij}$ in the second case. To first-order accuracy, p and λ are related b $\lambda = \nu p$. M has size $(\nu+1)\times(\nu+1)$ and entries (cf. Higgs, 1994)

$$m_{ij} = \begin{cases} \binom{\nu-j}{i-j} p^{i-j}(1-p)^{\nu-i}, & \text{if } i \geqslant j \\ 0, & \text{if } i < j. \end{cases} \tag{3}$$

The Poisson approximation of $\hat{M}$ of M has entries

$$m_{ij} = \begin{cases} e^{-\lambda\frac{\lambda^{i-j}}{(i-j)!}}, & \text{if } i \geqslant j \\ 0, & \text{if } i < j. \end{cases} \tag{4}$$

Matrices M and $\hat{M}$ are asymptotically equivalent (i.e. entries of M converge to the respective entries of $\hat{M}$) as ν becomes large and p small. When working without reverse mutation (see m_{ij} in (3)), this follows immediately from the approximation of a binomial distribution by a Poisson distribution. However, with

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

little more effort it can be shown that it is still true even if this condition is dropped (proof not shown). Therefore, if ν is large, model assumptions about reverse mutation do not have a bearing on the existence or magnitude of error thresholds nor is it the suppression of back mutations which distinguishes models of Muller's ratchet (Muller, 1964; Felsenstein, 1974) from those of error thresholds, as was previously suggested (Nowak & Schuster, 1989).

Further general definitions are the *average distance* from the wild-type, given by $E(p) = \sum_{i=0}^{\nu} i y_i(p)$, and the *variance in distance* from the wild-type $V(p) = \sum_{i=0}^{\nu}(i-E(p))^2 y_i(p)$ (respectively for λ instead of p). Derived from these, the *index of dispersion* $D(p) = V(p)/E(p)$ is used to measure the concentration of the population around the wild-type allele.

3. Results

(i) *Single-peaked and multiplicative fitness functions*

The single-peaked function, F_{SP}, is defined by two fitness levels: one for the wild-type, $v_0 = 1$, and one for the mutants $v_i = 1-s$, $i \geqslant 1$, $s > 0$. Under the multiplicative function, F_{M}, the fitness values are $v_i = (1-s)^i$, $i \geqslant 0$. This fitness function is often associated (Haigh, 1978; Stephan *et al.*, 1993) with the process known as Muller's ratchet: the accumulation of deleterious mutations, together with random loss of rare alleles, leads to a ratchet-like decrease of mean fitness of a population. Analytical expressions for the equilibrium frequencies $\bar{y}_i$ and the average distance, $\bar{E}(p)$, can readily be obtained for F_{M} (cf. Kimura & Maruyama, 1966, for the infinite sites case). For the finite sites model, it is a straightforward calculation to show that $\bar{y}_i$ (see Table 1) satisfy $\sum_{j=0}^{i} y_j v_j m_{ij} = y_i \bar{v}$. For F_{SP} the frequencies $\bar{y}_i$, $i \geqslant 1$ may be obtained recursively from the relation $\sum_{j=0}^{i} y_j v_j m_{ij} = y_i((1-s)+sy_0)$; an easy representation in closed form is not available. However, analytical expressions for $\bar{y}_0$ and $\bar{E}(p)$ can be derived (see Appendix). For both fitness functions the condition

$$\bar{y}_0 = 0$$

is equivalent to

$$\bar{y}_i = 0,\ i < \nu \text{ and } \bar{y}_\nu = 1\ (\nu \text{ finite})$$

$$\bar{y}_i = 0,\ i \geqslant 0\ (\nu \text{ infinite}),$$

indicating that the equilibrium is unique in these cases (note that the theorem which states uniqueness and global stability of the equilibrium distribution for the haploid mutation–selection equation (Moran, 1976) does not a priori hold if reverse mutation is suppressed, as in the model above). The error threshold may therefore either be identified as the smallest $p = p_{\max}$ ($\lambda = \lambda_{\max}$) which yields $\bar{y}_0 = 0$ (property of a single type) or, equivalently, via a property of the entire population, namely

$$\bar{E}(p_{\max}) = \nu,$$
$$\bar{E}(\lambda_{\max}) = \infty, \qquad (5)$$

for finite and infinite ν, respectively. The analytical expressions are given in Table 1. The threshold for finite ν and the single-peaked function F_{SP} is to a first-order approximation $p_{\max} \approx s/\nu$ (valid if $|s| \ll 1$). This coincides with the formula (1), given by Eigen (1971), as his superiority parameter σ equals $1/(1-s)$ for the single-peaked fitness function. For F_{M}, a threshold can be found if ν is finite. Interestingly, it does not depend on ν. Furthermore, for infinite ν, both criteria fail to detect any threshold at all. Obviously, under multplicativity no limiting relationship exists between genome size and mutation rate. The index of dispersion (depicted in Fig. 1) captures the qualitative difference between the two fitness models and its bearing on the

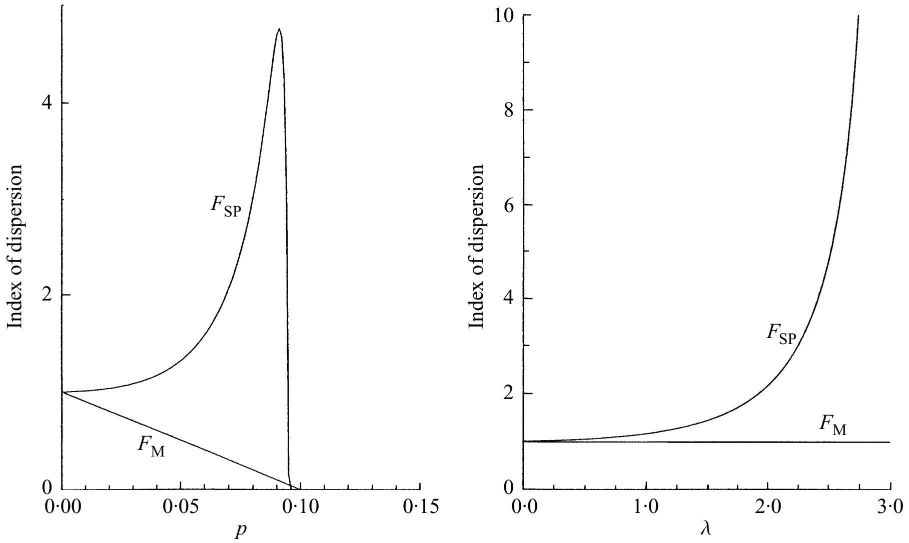

Fig. 1. Left: Comparison of single-peaked (F_{SP}) and multiplicative (F_{M}) fitness functions for finite ν. Plot of the index of dispersion $V(p)/E(p)$. Parameters are $s = 0.1$ (F_{M}), $s = 0{\cdot}95$ (F_{SP}), $\nu = 30$. Right: Index of dispersion $V(\lambda)/E(\lambda)$ for infinite ν. Parameters are $s = 0{\cdot}1$ (F_{M}), $s = 0{\cdot}95$ (F_{SP}). The error threshold is located where the dispersion function becomes singular (case F_{SP}). Obviously, mutation has no influence on dispersion in case F_{M}.

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

Table 1. *Analytical expressions for equilibrium frequencies $\bar{y}_i$, average distance from wild type $\bar{E}$, index of dispersion $\bar{D}$ and error threshold for fitness functions F_{SP} and F_M*

Model	F_{SP}	F_M
ν finite		
$\bar{y}_i(p)$	$\frac{(1-p)^{\nu}-(1-s)}{s}$	$\binom{\nu}{i}(p/s)^i(1-p/s)^{\nu-i}$
$\bar{E}(p)$	$\frac{\nu p(1-p)^{\nu-1}}{(1-p({}^{\nu-1}-(1-s)}$	$\frac{\nu p}{s}$
$\bar{D}(p)$	(see Appendix)	$\left(1-\frac{p}{s}\right)$
$p_{\max}$	$1-(1-s)^{(1/\nu)}$	s
ν infinite		
$\bar{y}_i(\lambda)$	$\frac{e^{-\lambda}-(1-s)}{s}$	$e^{-(\lambda/s)\frac{(\lambda/s)^i}{i!}}$
$\bar{E}(\lambda)$	$\frac{\lambda}{1-(1-s)e^{\lambda}}$	$\frac{\lambda}{s}$
$\bar{D}(\lambda)$	$\frac{1-(1-\lambda)(1-s)e^{\lambda}}{1-(1-s)e^{\lambda}}$	1
$\lambda_{\max}$	$-\log(1-s)$	∞

In case F_{SP} entries $\bar{y}_i(p)$ and $\bar{y}_i(\lambda)$ refer only to $i=0$.

error threshold. Under the multiplicative model the index of dispersion remains bounded. Under the single-peaked model, the index of dispersion increases around $p_{\max}$ ($\lambda_{\max}$) dramatically and exhibits a singularity, if ν is infinite. The singularity reflects the sudden loss of localization of the allele distribution.

(ii) *A more general fitness function*

An obvious generalization is to combine functions F_{SP} and F_M. The superposition, called F, sheds some more light on the dependencies of the error threshold. Let

$$v_i = t(1-s)^i+(1-t), \quad (0 \leqslant s, t \leqslant 1). \tag{6}$$

Letting $t=1$ yields the multiplicative and $s=1$ the single-peaked fitness functions. $s=0$ or $t=0$ produce a neutral model. For fitness functions with parameters on the boundaries of the s–t unit square, error thresholds can be detected using either one of the two criteria. Assuming this is true also for parameters in the interior, an analytical threshold formula can be obtained as follows. At equilibrium, the wild-type frequency satisfies

$$\begin{aligned}(1-p)^{\nu} &= \sum_k y_k v_k = t\sum_k y_k(1-s)^k+(1-t)\\ &= t(1-s)^{\nu}+(1-t)\end{aligned}$$

since the distribution will be concentrated in class ν once the threshold is surpassed (saturation of $E(p)$). Thus, for the finite sites model one has

$$p_{\max} = p_{\max}(s,t) = 1-(t(1-s)^{\nu}+(1-t))^{1/\nu}. \tag{7}$$

The analogue for the infinite sites model is obtained when one multiplies the latter equation by ν and takes the limit. The result is

$$\lambda_{\max} = \lambda_{\max}(s,t) = -\log(1-t). \tag{8}$$

As expected, the limiting threshold is independent of the multiplicative part described by the decay terms $(1-s)^i$.

To test the assumption that both threshold criteria can be used equivalently, I compared the analytical formula (7) with a numerical evaluation of the *ODE* system for various parameter choices, and found that both criteria, saturation of E and vanishing of the wild-type, yield identical error threshold values (results not shown).

In the finite sites case either parameter t or parameter s predominantly characterizes the error threshold. This depends on whether $1-t \geqslant (1-s)^{\nu}$ or $1-t < (1-s)^{\nu}$, i.e. the structure of the component v_{ν}, the minimum of the fitness function, is the decisive factor. Furthermore, the negative correlation between error threshold and genome size is weak if the slope of the fitness landscape around its 'adaptive peak' is moderate (s small).

If $\nu \to \infty$, and in agreement with (8), only parameter t plays a role. The error threshold does not depend on whether the decay in fitness around the adaptive peak is smooth or abrupt (s small versus s large). The error threshold may not exist at all ($t=1$). This observation, as shown in the next section, can be generalized to non-Fisherian fitness functions, i.e. to cases without an unique adaptive peak.

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

(iii) *Strictly positive fitness and truncation selection*

Since for large ν the quantity y_0 may not be robust enough to detect thresholds numerically, in this section the population statistics $E(\lambda)$ and $V(\lambda)$ are used to identify possible error thresholds. The following two examples represent possible generic cases. Let F_δ^Δ be defined by

$$0 < \delta \leqslant v_i \leqslant \Delta < \infty \quad (0 \leqslant i \leqslant \infty) \tag{9}$$

and $F_{\hat{\nu}}$ by

$$v_i = \begin{cases} \text{arbitrary} & \text{for } i \leqslant \hat{\nu} \\ = 0 & \text{for } i > \hat{\nu}. \end{cases} \tag{10}$$

F_δ^Δ contains F_{SP} as a special case. $F_{\hat{\nu}}$ is a model of truncation selection: carrying more than $\hat{\nu}$ mutations is lethal for an individual. From the equilibrium condition

$$\sum_{i=0}^{k} y_i v_i \hat{m}_{ki} = y_k \bar{v}, \quad (k \geqslant 0) \tag{11}$$

follows

$$\bar{v}E(\lambda) = \bar{v}\sum_k k y_k = \sum_k \sum_{i=0}^{k} i y_i v_i e^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} + \sum_k \sum_{i=0}^{k} (k-i)\, y_i v_i e^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!}. \tag{12}$$

The two series on the right side in (12) are each products of series. The first one simplifies to

$$e^{-\lambda} \sum_k k y_k v_k \cdot \sum_k \frac{\lambda^k}{k!} = \sum_k k y_k v_k,$$

the second one to

$$e^{-\lambda} \sum_k y_k v_k \cdot \sum_k k \frac{\lambda^k}{k!} = \lambda \bar{v},$$

which together gives

$$\bar{v}E(\lambda) = \lambda \bar{v} + \sum_k k y_k v_k. \tag{13}$$

For F_δ^Δ, the last series is trivially bounded from below by

$$\sum_k k y_k v_k \geqslant \delta E(\lambda).$$

This yields

$$\bar{E}(\lambda) \geqslant \frac{\lambda \bar{v}}{\bar{v} - \delta} \geqslant \frac{\lambda \delta e^{-\lambda}}{\Delta e^{-\lambda} - \delta}. \tag{14}$$

The last inequality can be seen as follows. Let $\hat{k}$ be (for a fixed λ) the lowest index such that $y_{\hat{k}} \neq 0$. Then $y_{\hat{k}} v_{\hat{k}} \hat{m}_{\hat{k}\hat{k}} = y_{\hat{k}} \bar{v}$. Thus, $\bar{v} = e^{-\lambda} v_{\hat{k}} \geqslant e^{-\lambda}\delta$ and $\bar{v} \leqslant e^{-\lambda}\Delta$, independent of $\hat{k}$. Clearly, the last expression shows a singularity for

$$\lambda = -\log\left(\frac{\delta}{\Delta}\right). \tag{15}$$

Thus, an upper bound for the error threshold for any fitness function of type F_δ^Δ is given by

$$\lambda_{\max} \leqslant -\log\left(\frac{\delta}{\Delta}\right). \tag{16}$$

To calculate the index of dispersion more must be known about the individual fitness assignments. The general expression is

$$\bar{D}(\lambda) = \frac{\lambda + \Sigma_k k^2 y_k v_k/\bar{v} - (\Sigma_k k y_k v_k/\bar{v})^2}{\lambda + \Sigma_k k y_k v_k/\bar{v}}.$$

On the other hand, for $F_{\hat{\nu}}$ any finite λ produces $E(\lambda) < \infty$. No error threshold exists. This can easily be seen by studying the worst case scenario $v_i = 0$ for all $i \neq \hat{\nu}$. Then $\bar{v} = v_{\hat{\nu}} y_{\hat{\nu}}$ and from (11) follows for $k \geqslant \hat{\nu}$

$$y_k = e^{-\lambda} \frac{\lambda^{k-\hat{\nu}}}{(k-\hat{\nu})!}. \tag{17}$$

The latter is independent of the particular choice for $v_{\hat{\nu}}$. Since $y_k = 0$ for $k < \hat{\nu}$, by summing over k one derives

$$\bar{E}(\lambda) = \sum_k k y_k = \lambda + \hat{\nu}. \tag{18}$$

$\bar{E}(\lambda)$ is therefore limited from above by an affine linear function and cannot exhibit a singularity for finite λ. The variance in this case is $\bar{V}(\lambda) = \lambda$. Therefore, the index of dispersion in the worst case scenario has an upper bound of 1, supporting the claim that there is no error threshold. These results agree with the findings by Wagner & Krall (1993). For a slightly different model, these authors proved that error thresholds may be produced by monotonically decreasing fitness functions which are bounded from below by a strictly positive value. As shown above, monotony is not even necessary; what matters is only the ratio of lowest and highest fitness values.

Analogous arguments apply to the finite sites case. An upper bound to the error threshold for F_δ^Δ is given by

$$p_{\max} \leqslant 1 - \left(\frac{\delta}{\Delta}\right)^{1/\nu}. \tag{19}$$

For truncation selection a new property is encountered: the two threshold criteria may yield differing results. This fact is dealt with more thoroughly in the following section.

(iv) *Epistasis*

Error thresholds depend strongly on the amount and kind of epistasis. Let the epistatic fitness function $F_{\text{E}(\alpha)}$ be defined by

$$v_i = (1-s)^{i^\alpha}, \tag{20}$$

with α a positive parameter. One distinguishes

(a) diminishing epistasis ($0 < \alpha < 1$),
(b) multiplicativity or absence of epistasis ($\alpha = 1$),
(c) synergistic epistasis ($\alpha > 1$).

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

F_{M} and F_{SP} are again special cases of the epistatic functions: F_{SP} is an extreme case of diminishing epistasis and recovered when $\alpha \to 0$; F_{M} corresponds to the case when $\alpha = 1$. The extreme form of synergistic epistasis is truncation selection ($\alpha = \infty$). Unfortunately, for general α, no analytical solution for the stationary frequencies $\bar{y}_i$ is known. However, analytical threshold formulae can still be derived. The two threshold conditions 'loss of wild-type' and 'saturation of $E(p)$' are in general not equivalent. The conditions coincide only if $\alpha \leqslant 1$. There is a simple heuristic explanation for this. The fraction v_i/v_{i-1} is $1-s[i^{\alpha}-(i-1)^{\alpha}]+O(s^2)$. For diminishing epistasis the expression in brackets tends to 0 as i increases. That means additional mutations cause a smaller decline of relative fitness than previous mutations. Thus, maximal mutation pressure is needed to remove the wild-type from the population. As soon as this is accomplished the stationary distribution is concentrated in the most distant mutation class ν and $E(p)$ has saturated. For synergistic epistasis ($\alpha > 1$) the situation is reversed. Each additional mutation causes a larger decline of relative fitness. Thus, to remove class i from the population requires stronger mutation pressure than is needed for removal of class $i-1$. In particular, the wild-type is lost long before $E(p)$ saturates. Therefore, one has to distinguish between $p_{\max}$ (loss of the wild-type) and the higher mutation probability $p^{\max}$ (saturation of $E(p)$). Below are analytical formulae for both.

To derive $p^{\max}$ let p be sufficiently large such that the second-to-last mutation class is just lost by mutation from a stationary population (any other class, but the last, is already lost for smaller p). The following three algebraic equations have to be satisfied in this case:

$$\begin{aligned} y_{\nu-1} v_{\nu-1} p + y_\nu v_\nu &= y_\nu (y_{\nu-1} v_{\nu-1} + y_\nu v_\nu), \\ y_{\nu-1} v_{\nu-1}(1-p) &= y_{\nu-1}(y_{\nu-1} v_{\nu-1} + y_\nu v_\nu), \\ y_{\nu-1} + y_\nu &= 1. \end{aligned}$$

The solution is

$$p = \frac{((1-s)^{(\nu-1)^{\alpha}} - (1-\mathrm{s})^{\nu^{\alpha}})(1-y_{\nu-1})}{(1-s)^{(\nu-1)^{\alpha}}}, \qquad (21)$$

which, for $y_{\nu-1} = 0$, yields

$$p^{\max} = 1-(1-s)^{\nu^{\alpha}-(\nu-1)^{\alpha}}, \quad (\alpha \geqslant 1). \qquad (22)$$

On the other hand, the wild-type is already lost if

$$p_{\max} = s, \quad (\alpha \geqslant 1). \qquad (23)$$

Eq. (23) holds for any $\alpha \geqslant 1$ and independently of ν. To justify this, note that it holds for $\alpha = 1$ (see above). Letting α go to ∞ one obtains fitnesses $v_0 = 1$, $v_1 = 1-s$ and $v_i = 0$ ($i \neq 0, 1$). At equilibrium y_0 satisfies $(1-p)^{\nu} = y_0 + y_1(1-s)$. Putting $y_0 = 0$ the latter implies $p = 1-(y_1(1-s))^{1/\nu}$. Furthermore, y_1 has to satisfy (according to (2)) $y_1(1-s)\,(1-p)^{\nu-1} = y_1^2(1-s)$. Thus, $y_1 = (1-p)^{\nu-1}$. Inserting this into the equation for p yields $(1-p)^{\nu} = (1-p)^{\nu-1}(1-s)$, or $p = s$. For fitness functions of type $F_{\mathrm{E}(\alpha)}$ the relation $\alpha_1 < \alpha_2$ implies $y_{0_{(\alpha_1)}} \leqslant y_{0_{(\alpha_2)}}$ (as functions of p). Now, by continuity, it is concluded that

$$0 = y_{0_{(\alpha=1)}}(p = s) \leqslant y_{0_{(\alpha)}}(p = s) \leqslant y_{0_{(\alpha=\infty)}}(p = s) = 0.$$

This means $y_{0_{(\alpha)}}(p = s) = 0$ for all $1 \leqslant \alpha \leqslant \infty$.

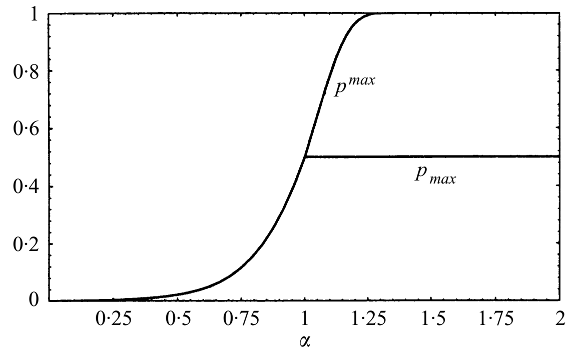

Fig. 2. Error thresholds $p_{\max}$ (see (23) and (25)) and $p^{\max}$ (see (22)) versus epistasis parameter α. A 'bifurcation' of thresholds occurs at $\alpha = 1$ (absence of epistasis). Parameters are $\nu = 10^3$ and $s = 0{\cdot}5$.

In agreement with the results before, (22) and (23) coincide for $\alpha = 1$ (the two threshold criteria are equivalent for F_{M}).

To treat the case of diminishing epistasis note that the two detection criteria are equivalent (since F_{SP} and F_{M} both have this property, a similar continuity argument can be invoked to prove the property for $0 < \alpha < 1$). When the threshold is surpassed then $\bar{v} = (1-s)^{\nu^{\alpha}}$, since the distribution is then concentrated in mutation class ν. This, together with the condition for the wild-type frequency to satisfy $y_0 v_0 m_{00} = y_0 \bar{v}$, implies

$$(1-p)^{\nu} = (1-s)^{\nu^{\alpha}}. \qquad (24)$$

Solving for p leads to

$$p_{\max} = 1-(1-s)^{\nu^{\alpha-1}}, \quad (\alpha \leqslant 1). \qquad (25)$$

These results have been validated numerically (results not shown). Results for the extrapolation to $\nu = \infty$ are given in Table 2.

Most important to note about these formulae is that the error threshold is uniquely defined (in the sense used here) and inversely related to sequence length ν, only if $\alpha < 1$. A 'bifurcation' occurs for $\alpha = 1$ and two threshold 'branches' split off (see Fig. 2). One of them is independent of the sequence length, the other is even positively correlated with ν. This is in sharp contrast to the negative correlation of ν and p observed for F_{SP} and which originally stimulated the debate about the information crisis. Furthermore, note that the error threshold ($p^{\max}$) is an increasing function of the parameter (α) of epistasis. The lower

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

Table 2. *Comparison of thresholds for different fitness functions*

		Error threshold	
Fitness function		ν finite	ν infinite
F_{SP} (single-peaked)		$1-(1-s)^{1/\nu}$	$-\log(1-s)$
F_{M} (multiplicative)		s	∞
$F_{\mathrm{E}(\alpha)}$ (epistatic)	$0<\alpha\leqslant 1$:	$1-(1-s)^{\nu^{\alpha-1}}$	∞
	$\alpha>1$:	$\begin{cases} p_{\max}: \mathrm{s} \\ p^{\max}: 1-(1-s)^{\nu^{\alpha}-(\nu-1)^{\alpha}} \end{cases}$	∞
F_{δ}^{Δ} (arbitrary, bounded above and below)		$\leqslant 1-\left(\frac{\delta}{\Delta}\right)^{1/\nu}$	$\leqslant -\log\left(\frac{\delta}{\Delta}\right)$
$F_{\hat{\nu}}$ (truncation selection[a])		$\begin{cases} p_{\max}: s \\ p^{\max}: 1 \end{cases}$	∞
F (superimposed)		$1-(t(1-s)^{\nu}+(1-t))^{1/\nu}$	$-\log(1-t)$

For definition of the fitness functions see the text. Function F_{δ}^{Δ} is not unambiguously defined. In this case, only an upper bound to a threshold can be given. Note that thresholds in the cases of finite ν refer to p (nucleotide mutation probability), whereas in the case of infinite ν they refer to λ (genome mutation rate).

[a] For finite ν truncation at $\hat{\nu}=1$, for infinite ν the truncation point is arbitrary.

bound is given by the threshold of the single-peaked fitness function ($\alpha=0$), the upper bound by the one for truncation selection ($\alpha=\infty$); in this latter case one has $p^{\max}=1$.

The effect of synergistic epistasis on Muller's ratchet has been studied by Kondrashov (1994). He showed that synergistic epistasis may effectively halt Muller's ratchet. This parallels the effect on the threshold. Synergism can – for any mutation rate – protect the fittest allele from extinction.

Summarizing, the results of this section show that

(a) error thresholds do not generally exist (dependency on the fitness function),

(b) error thresholds are ambiguous (dependency on the definition in terms of a population property or that of an individual allele),

(c) error thresholds need not be negatively correlated with the size of the genome (epistatic effects).

4. Discussion

Error thresholds have originally been described for a finite sites model, termed the quasispecies model (Eigen & Schuster, 1979). They have been interpreted as the minimal required replication accuracy to ensure that heredity of self-replicating molecules at the early stages of life does not break down. It has been suggested that the replication process of viruses and phages is adequately described within the quasispecies concept and that these organisms replicate under conditions close to an error threshold (e.g. Domingo & Holland, 1988; Eigen & Biebricher, 1988). However, even if one or a few types of the viral quasispecies may be distinguished by a fitness advantage due to better adaptation to the host environment, a full characterization of the fitness function remains largely speculative. Insight into the shape of fitness functions associated with short sequences has been gained by examining cases with a simple relationship between genotype and phenotype. For instance, for tRNAs the phenotype may be associated with the molecular secondary structure, and folding properties of the primary sequence into its secondary structure are viewed as a principal determinant for the performance (fitness) (Fontana & Schuster, 1987; Schuster *et al.*, 1994). Such models, reflecting sequence–structure relations (Forst *et al.*, 1995), show a threshold phenomenon, closely related to that observed under a single-peaked fitness function. These results, however, rely mainly on theoretical models and computer simulations, and remain to be confirmed experimentally. There seems to be little biological reason to assume that a single-peaked fitness function provides a generally applicable model underlying the evolutionary dynamics across a wide spectrum of genes and species. Whether multiplicative or epistatic models are closer to reality may surely be questioned as well. The latter have initially been studied with emphasis on theoretical aspects. The view that synergistic epistasis must be more common in nature than diminishing epistasis has been prevalent since the 1960s (Kimura & Maruyama, 1966, and references therein). The discussion of the evolution of sex has often been linked to that of the operation of Muller's ratchet in natural populations. Fitness models used in this context are the multiplicative and epistatic ones (e.g. Felsenstein, 1974; Kondrashov, 1988; Charlesworth, 1990). Experimental evidence for the operation of Muller's ratchet in populations of various species has been compiled (Bell, 1988; Chao, 1990; Duarte *et al.*, 1992; Clarke *et al.*, 1993). A recent study by Lynch (1996) detects Muller's ratchet and a gradual loss of fitness in mitochondrial tRNAs. For many nuclear eukaryotic genes synergistic epistasis appears to be a

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

more appropriate model than a single-peaked fitness function. For instance, some genetical disorders are associated with an excess in the number of trimer repeats in certain genes (e.g. the number of CTG repeats involved in penetrance of myotonic dystrophy: Harley *et al.*, 1992, 1993). A suitable fitness function, drawn over the number of trimer repeats, may be of the synergistic epistatic or truncation selection type. More generally, truncation selection is believed to play an important role in the evolutionary dynamics of repetitive sequences in eukaryotes (for a review see Charlesworth *et al.*, 1994).

With the conceptual dichotomy between models of error thresholds and Muller's ratchet (cf. the title of Wagner & Krall, 1993) which currently exists, it appears natural to concentrate the analysis on two special fitness functions, which are representative of the two models: the single-peaked (F_{SP}) and the multiplicative function (F_{M}). I investigated several extensions; first, the superposition (function F) of both. The single-peaked, multiplicative and the neutral models arise as special cases of F. Furthermore, special attention has been paid to truncation selection and a general function with the only restriction that it is bounded (a special case of which, again, is F_{SP}). In agreement with the finding by Wagner & Krall (1993), they harbour the minimum requirements to distinguish fitness functions which are error threshold free from those which are not. Finally, the set of functions $F_{\mathrm{E}(\alpha)}$ provides the possibility for a unified treatment of both models within the concept of epistasis. Two generic cases emerge: synergistic and diminishing epistasis. For the finite sites case, I showed that error thresholds exist in a strict sense (inverse relationship with genome size, uniqueness) only if epistasis is diminishing ($\alpha < 1$). Any form of synergistic epistasis implies absence of a (strict) error threshold. For the infinite sites model, the existence of error thresholds reduces even to a non-generic special case ($\alpha = 0$) (see Table 2).

Despite the study by Wagner & Krall (1993) and their proof that certain fitness functions do not produce error thresholds, the perception that the latter ubiquitously and independently of the shape of fitness functions set a limit to the evolutionary potential of a species continues to persist (Schuster, 1995, pp. 45f). Obviously, there is still a need to raise more caution about such a viewpoint. Even recent textbook accounts of error thresholds (Maynard Smith & Szathmáry, 1995) seem to overlook the fact that the superiority parameter σ (see (1)) need not be a constant for general fitness functions (but may itself depend on the mutation rate), and therefore the error threshold need not be a general phenomenon.

Furthermore, other forces, recombination for instance, have a high impact on shaping the genetic material as it is passed on from generation to generation. Boerlijst *et al.* (1996) studied recombination in a viral quasi-species and its influence on the error threshold. Nee & Maynard Smith (1990) point out that, depending on the kind of epistasis, recombination can alter the error threshold. Based on simulations, they observe that recombination together with synergistic or diminishing epistasis increases or decreases permissible mutation rates and genome sizes. They suggested that the presence of recombination might allow viruses to have larger genomes and yet avoid the pitfalls of the information crisis.

Another problem with error thresholds arises from the lack of a generally accepted definition. Two approaches have been taken in the past to identify error thresholds. One is via a property of the entire population, the other is via a property of an individual allele. The above analysis shows that the two definitions need not be congruent (i.e. produce the same threshold value). Rather, the presence of epistasis may make any definition in these terms obsolete.

The dynamics of the haploid model, treated here, is equivalent to that of a diploid model as long as dominance defects are absent (i.e. fitness of the heterozygotes is intermediate between that of the homozygotes). Qualitative new features, however, emerge if dominance plays a role. Its impact on error thresholds has been treated for the diploid analogue of the single-peaked fitness function (Wiehe *et al.*, 1995) and, recently, for a more general fitness model as well (Baake & Wiehe, 1996).

In a strict sense, the above deterministic analysis is valid for an infinitely large population only. Stochastic versions – accounting for random drift – have been studied as well (Nowak & Schuster, 1989; Stephan *et al.*, 1993; Wiehe *et al.*, 1995). The qualitative features of the deterministic model are recovered. The presence of random drift and – in its wake – random loss of rare alleles may only emphasize that it is an imminent ratchet mechanism, possibly not an information crisis, evolution has primarily to cope with.

After all, in the light of the different fitness models discussed above, it appears that the importance of error thresholds as a limiting factor to molecular evolution has been greatly overrated.

Appendix

For function F_{SP} and the mutation matrix as in (3) *hold*

$$\bar{y}_0 = \max\left(0, \frac{(1-p)^{\nu}-(1-s)}{s}\right),$$

and

$$\bar{E}(p) = \frac{\nu p(1-p)^{\nu-1}}{(1-p)^{\nu-1}-(1-s)}.$$

Proof. The first part follows immediately from the equation for the wild-type

$$y_0 v_0(1-p)^{\nu} = y_0 \bar{v}$$

and the fact that for the single-peaked fitness function

$$\bar{v} = 1-s+sy_0.$$

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press

As long as $y_0 \neq 0$ one has $\bar{v} = (1-p)^{\nu}$. To calculate $\bar{E}(p)$, one multiplies the equilibrium equation on both sides by k, then sums over k to obtain

$$\sum_{k=0}^{\nu}\sum_{i=0}^{k} k y_i (1-s)\binom{\nu-i}{k-i} p^{k-i}(1-p)^{\nu-k} + s y_0 \nu p = E(p)\cdot(1-p)^{\nu}.$$

On changing the order of summation and readjusting summation indices the latter equation becomes

$$\begin{aligned}(1-s)&\left(\sum_{i=0}^{\nu} y_i \sum_{k=0}^{\nu-i} k \binom{\nu-i}{k} p^{k}(1-p)^{\nu-i-k} + \sum_{i=0}^{\nu} i y_i \sum_{k=0}^{\nu-i} \binom{\nu-i}{k} p^{k}(1-p)^{\nu-i-k}\right) + s y_0 \nu p \\ &= (1-s)\left(\sum_{i=0}^{\nu} y_i(\nu-i)p + \sum_{i=0}^{\nu} i y_i\right) + ((1-p)^{\nu}-(1-s))\,\nu p \\ &= (1-s)\,(\nu p + E(p)\cdot(1-p)) + ((1-p)^{\nu}-(1-s))\,\nu p \\ &= E(p)\cdot(1-p)^{\nu}.\end{aligned}$$

The last equation may be solved for $E(p)$. □

In a similar manner one can find an analytic expression for the variance $\bar{V}(p)$. The ratio $\bar{V}(p)/\bar{E}(p)$, the index of dispersion, is

$$\bar{D}(p) = \frac{(1-s-(1-p)^{\nu})\,((1-p)^{\nu-1}-(1-s)\,(1-\nu p))}{(1-s-(1-p)^{\nu-1})\,((1-p)^{\nu-1}-(1-s)\,(1-p))},$$

$$p < p_{\max}.$$

Partial support of this work by a grant from the German Academic Exchange Service (DAAD) E. Boake, is gratefully acknowledged. I would like to thank A. Kondrashov, J. Mountain, R. Nielsen and an anonymous reviewer for constructive and helpful comments.

References

Baake, E. & Wiehe, T. (1997). Bifurcations in diploid models on sequence space. *Journal of Mathematical Biology*, **35**, 321–343.

Bell, G. (1988). *Sex and Death in Protozoa: The History of an Obsession*. Cambridge: Cambridge University Press.

Boerlijst, M. C., Bonhoeffer, S. & Nowak, M. A. (1996). Viral Quasi-Species and Recombination, Proceedings of the Royal Society of London, Series B **263**, 1577–1584.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455.

Charlesworth, B. (1990). Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* (*Cambridge*) **55**, 199–221.

Charlesworth, B., Sniegowski, P. & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220.

Clarke, D., Duarte, E., Moya, A., Elena, S., Domingo, E. & Holland, J. (1993). Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *Journal of Virology* **67**, 222–228.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

Domingo, E. & Holland, J. J. (1988). High error rates, population equilibrium, and evolution of RNA replication systems. In *RNA Genetics: Variability of RNA Genomes* (ed. E. Domingo, J. J. Holland & P. Ahlquist), vol. III, p. 3–36. Boca Raton: CRC Press.

Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**, 735–744.

Duarte, E., Clarke, D., Moya, A., Domingo, E. & Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proceedings of the National Academy of Sciences*, *USA* **89**, 6015–6019.

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523.

Eigen, M. & Biebricher, C. K. (1988). Sequence space and quasispecies distribution. In *RNA Genetics: Variability of RNA Genomes* (ed. E. Domingo, J. J. Holland & P. Ahlquist), vol. III, p. 211–245. Boca Raton: CRC Press.

Eigen, M. & Schuster, P. (1979). *The Hypercycle*. Berlin: Springer.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* **78**, 737–756.

Fontana, W. & Schuster, P. (1987). A computer model of evolutionary optimization. *Biophysical Chemistry* **26**, 123–147.

Forst, C. V., Reidys, C. & Weber, J. (1995). Evolutionary Dynamics and Optimization. In *Lecture Notes in Artificial Intelligence*. Vol. 929: *Advances in Artificial Life* (ed. F. Moran, A. Moreno, J. Marelo & P. Chacon) Berlin: Springer.

Hadeler, K. P. (1981). Stable polymorphisms in a selection model with mutation. *SIAM Journal of Applied Mathematics* **41**, 1–7.

Haigh, J. (1978). The accumulation of deleterious genes in a population: Muller's ratchet. *Theoretical Population Biology* **14**, 251–267.

Harley, H. G., Rundle, S. A., Reardon, W., Myring, J., Crow, S., Brook, J. D., Harper, P. S., *et al.* (1992). Unstable DNA sequence in myotonic dystrophy. *Lancet* **339**, 1125–1128.

Harley, H. G., Rundle, S. A., MacMillan, J. C., Myring, J., Brook, J. D., Crow, S., Reardon, W., Fenton, I., Shaw, D. J. & Harper, P. S. (1993). Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *American Journal of Human Genetics* **52**, 1164–1174.

Higgs, P. G. (1994). Error thresholds and stationary mutant distributions in multi-locus diploid genetics models. *Genetical Research* (*Cambridge*) **63**, 63–78.

Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.

Kimura, M. & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Kondrashov, A. S. (1994). Muller's ratchet under epistatic selection. *Genetics* **136**, 1469–1473.

Lynch, M. (1996). Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Molecular Biology and Evolution* **13**, 209–220.

Martinez-Salas, E., Ortin, J. & Domingo, E. (1985). Sequence of the viral replicase gene from foot and mouth disease virus C_1-Santa Pau (C-S8). *Gene* **35**, 55–61.

Maynard Smith, J. (1983). Models of evolution. *Proceedings of the Royal Society of London*, *Series B* **219**, 315–325.

Maynard Smith, J. & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford: W. H. Freeman.

Moran, P. A. P. (1976). Global stability of genetic systems governed by mutation and selection. *Mathematical Proceedings of the Cambridge Philosophical Society* **80**, 331–336.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutational Research* **1**, 2–9.

Nee, S. & Maynard Smith, J. (1990). The evolutionary biology of molecular parasites. *Parasitology* **100** (Suppl.), S5–S18.

Nowak, M. & Schuster, P. (1989). Error thresholds of replication in finite populations: mutation frequencies and the onset of Muller's ratchet. *Journal of Theoretical Biology* **137**, 375–395.

Ortin, J., Nájera, R., Lopez, C., Davila, M. & Domingo, E. (1980). Genetic variability of Hong Kong (H3N2) influenza viruses: spontaneous mutations and their location in the viral genome. *Gene* **11**, 319–331.

Schuster, P. (1995). Extended molecular evolutionary biology: artificial life bridging the gap between chemistry and biology. In *Artificial Life: An Overview* (ed. C. G. Langton). Cambridge, Mass.: MIT Press.

Schuster, P., Fontana, W., Stadler, P. F. & Hofacker, I. L. (1994). From sequences to shapes and back: a case study in RNA secondary structures. *Proceedings of the Royal Society of London, Series B* **255**, 279–284.

Spiegelman, S., Haruna, I., Holland, I. B., Beaudreau, G. & Mills, D. R. (1965). The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme. *Proceedings of the National Academy of Sciences, USA* **54**, 919–927.

Stephan, W., Chao, L. & Smale, J. G. (1993). The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genetical Research* (Cambridge) **61**, 225–231.

Swetina, J. & Schuster, P. (1982). Self-replication with errors: a model for polynucleotide replication. *Biophysical Chemistry* **16**, 329–345.

Wagner, G. P. & Krall, P. (1993). What is the difference between models of error thresholds and Muller's ratchet? *Journal of Mathematical Biology* **32**, 33–44.

Wiehe, T., Baake, E. & Schuster, P. (1995). Error propagation in reproduction of diploid organisms: a case study on single peaked landscapes. *Journal of Theoretical Biology* **177**, 1–15.

Wright, S. (1949). Adaptation and selection. In *Genetics, Paleontology and Evolution* (ed. G. G. Simpson, G. L. Jepson & E. Mayr), pp. 365–389. Princeton: Princeton University Press.

https://doi.org/10.1017/S0016672397002619 Published online by Cambridge University Press