



**HAL**  
open science

## Effect of lethality on the extinction and on the error threshold of quasispecies

Hector Tejero, Arturo Marín, Francisco Montero

► **To cite this version:**

Hector Tejero, Arturo Marín, Francisco Montero. Effect of lethality on the extinction and on the error threshold of quasispecies. *Journal of Theoretical Biology*, 2010, 262 (4), pp.733. 10.1016/j.jtbi.2009.10.011 . hal-00564082

**HAL Id: hal-00564082**

**<https://hal.science/hal-00564082>**

Submitted on 8 Feb 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Author's Accepted Manuscript

Effect of lethality on the extinction and on the error threshold of quasispecies

Hector Tejero, Arturo Marín, Francisco Montero

PII: S0022-5193(09)00486-X  
DOI: doi:10.1016/j.jtbi.2009.10.011  
Reference: YJTBI5739

To appear in: *Journal of Theoretical Biology*

Received date: 29 June 2009  
Revised date: 7 October 2009  
Accepted date: 8 October 2009

Cite this article as: Hector Tejero, Arturo Marín and Francisco Montero, Effect of lethality on the extinction and on the error threshold of quasispecies, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2009.10.011](https://doi.org/10.1016/j.jtbi.2009.10.011)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

**Effect of lethality on the extinction and on the error threshold of quasispecies.****Hector Tejero<sup>a</sup>, Arturo Marín<sup>a</sup>, Francisco Montero<sup>a,\*</sup>**

<sup>a</sup>*Departamento de Bioquímica y Biología Molecular I, Universidad Complutense de Madrid, Avd. Complutense s/n,  
28040 Madrid, Spain*

<sup>\*</sup>*Departamento de Bioquímica y Biología Molecular I, Universidad Complutense de Madrid, Avd. Complutense s/n,  
28040 Madrid, Spain. Tel.: +34 913944255; fax: +34 913944159. E-mail address: [framonte@quim.ucm.es](mailto:framonte@quim.ucm.es) (F. Montero).*

**Abstract**

In this paper the effect of lethality on error threshold and extinction has been studied in a population of error-prone self-replicating molecules. For given lethality and a simple fitness landscape, three dynamic regimes can be obtained: quasispecies, error catastrophe, and extinction. Using a simple model in which molecules are classified as master, lethal and non-lethal mutants, it is possible to obtain the mutation rates of the transitions between the three regimes analytically. The numerical resolution of the extended model, in which molecules are classified depending on their Hamming distance to the master sequence, confirms the results obtained in the simple model and shows how an error catastrophe regime changes when lethality is taken in account.

**Keywords:** single-peaked landscape, population dynamics, mathematical model, error catastrophe, selection-mutation balance.

**1. Introduction**

The Quasispecies model was originally developed by Eigen to study the origin of genetic information in a prebiotic scenario (Eigen, 1971). However, it has turned out to be very useful as a general model of evolutionary dynamics of populations with both

selection and high mutation rates. Although mathematically equivalent to the classical selection-mutation balance of population genetics (Biebricher and Eigen, 2006; Wilke, 2005), the explicit consideration of high mutation rates, and a great number of loci, make it especially appropriate for populations with high genetic variability such as prebiotic self-replicative molecules (Eigen, 1971), RNA viruses (Domingo et al., 1978), the immune system (Kamp et al., 2002), bacteria (Webb and Blaser, 2002), and even cancer cells (Solé and Deisboeck, 2004).

One of the most remarkable results of the model is that, in certain fitness landscapes, a quasispecies presents an error threshold, i.e. a maximum mutation rate beyond which the population suffers an error catastrophe. Error catastrophe has been classically interpreted as a loss of the genetic information contained in the quasispecies (Biebricher and Eigen, 2005; Eigen, 2002). On the basis of this result, treatment with mutagenic drugs has been postulated as a possible therapy against highly variable pathogens such as RNA viruses (Holland et al., 1990) and cancer cells (Solé and Deisboeck, 2004). After having been successfully probed with RNA viruses in cell culture, the experimental result was called lethal mutagenesis (Domingo et al., 2006; Grande-Perez et al., 2002; Loeb et al., 1999; Sierra et al., 2000).

Initially, lethal mutagenesis was explained as the result of an RNA virus crossing the error threshold due to the increase in the mutation rate. However, it has recently been proposed that lethal mutagenesis cannot be reducible to crossing the error threshold, and that in fact they are essentially different phenomena (Bull et al., 2005; Bull et al., 2007). These papers consider that error catastrophe is a genetic-evolutionary process whereby a population shifts to the most mutational robust phenotypes, due to the imposition of mutation over selection. However, no experimental evidence has been obtained of a shift to a more robust phenotype after mutagenic treatment of viral

populations (Martin et al. 2008). Lethal mutagenesis is considered to be extinction, i.e. a demographic process that can be independent of population structure.

Error catastrophe is a characteristic result of the quasispecies model in some fitness landscapes. Knowing its relation to extinction can be very useful in understanding the effect of high mutation rates in error-prone biological populations, particularly in RNA virus populations.

Taking previous results into account (Wagner and Krall, 1993), some authors have stated that error catastrophe cannot take place in the presence of lethal mutants (Summers and Litwin, 2006; Wilke, 2005). However, a recent paper (Takeuchi and Hogeweg, 2007) has demonstrated the existence of error catastrophe in a fitness landscape with lethal mutants. The paper also affirmed that the results described by Wagner and Krall were based on a very restrictive initial hypothesis: the use of a one-dimensional genotype space, and a simple mutation mechanism in which each genotype can only mutate to the adjacent genotypes. (Takeuchi and Hogeweg, 2007). On the other hand, as Wilke has pointed out (Takeuchi and Hogeweg, 2007), Takeuchi and Hogeweg considered a constant population size, which meant that extinction was not allowed, and therefore its results cannot be interpreted in the context of the relationship between error catastrophe and lethal mutagenesis. But in most cases, constant population size is a general constraint in quasispecies models. Thus, extinction by an increasing mutation rate has rarely been studied (Bonnaz, 1999) .

In this paper a simple model in which there is no restriction to population size has been developed, in order to study the relationship between extinction and error catastrophe. The main novelty of the model, with respect to previous formulations, is the introduction of lethality in such a way that some regions of the sequence space cannot self-

replicate. The second section of the paper presents a general quasispecies model in the presence of lethality. The third section proposes a simple case study of the general model. Its analysis has enabled the effect of lethality on the error threshold and on the extinction threshold to be determined analytically, i.e. the mutation rate beyond which population becomes extinct. The model has been extended in the fourth section in order to study the effect of lethality on the error catastrophe regime, and to validate the results obtained in the previous section. Finally, the results are discussed in the context of the quasispecies dynamics.

## 2. Quasispecies model with lethality

The quasispecies model describes the evolutionary dynamics of a population composed of error-prone self-replicative species subjected to selection. In this model, a quasispecies is defined as the stationary population distribution resulting from the action of the two evolutionary forces, namely mutation and selection (Eigen, 1971).

Self-replicative species are represented by binary strings of length  $v$ . Each sequence replicates with an amplification rate ( $A_i$ ) and degrades with a degradation rate ( $D_i$ ). New sequences arise by mutation during the replication process. Each position in the sequence is copied correctly with probability  $q$ , this being constant for all the positions. Consequently, the mutation rate per position is  $\mu = 1-q$ .

Following the original formulation of the model, a sharply peaked fitness landscape has been used. In this approach a sequence, called the *master*, has an amplification rate ( $A_m$ ), higher than that of any other sequence. The sequences different from the master have the same amplification rate  $A_k$ , irrespective of the number of mutations (Swetina and Schuster, 1982). The degradation rate is constant for the whole sequence space

( $D_m = D_k = D$ ). To deal with this problem, the  $2^v$  sequences are grouped in  $v+1$  error classes, depending on the number of mutations with respect to the master sequence.

The lethality scheme introduced in the fitness landscape implies that there are  $n$  fixed positions in the sequence, whose mutation with respect to the master sequence implies that the emerging sequence cannot replicate, so that its amplification factor is zero ( $A_l = 0$ ). In this scheme, the distribution of lethal mutants among the sequence space is not uniform. The sequences with greater Hamming distance from the master have a higher fraction of lethal sequences. Fig 1. shows the lethal sequence fraction as a function of the Hamming class for a different number of lethal positions,  $n$ .

### 2.1. A simple model

As a first approach, all the sequences have been grouped in three classes: master ( $A = A_m$ ), non-lethal mutants ( $A = A_k$ ) and lethal mutants ( $A = A_l = 0$ ). For the sake of simplification, back mutation from lethal mutants to the master sequence has not been taken into account. This simplification is usually used in quasispecies models (Swetina and Schuster, 1982) because the probability of obtaining a given specific sequence by mutation is extremely low.

The linear differential equation system of this model is:

$$\begin{aligned}\frac{dx_m}{dt} &= (A_m q^v - D)x_m \\ \frac{dx_k}{dt} &= A_m q^n (1 - q^{v-n})x_m + (A_k q^n - D)x_k \\ \frac{dx_l}{dt} &= A_m (1 - q^n)x_m + A_k (1 - q^n)x_k - Dx_l\end{aligned}\tag{1}$$

which refers to the time evolution of master ( $x_m$ ), non-lethal ( $x_k$ ) and lethal ( $x_l$ ) classes, respectively. In the first place, the equations show that, discarding back mutation, the production of master sequence depends only on the correct self-replication of its  $v$  positions. The class of non-lethal mutants can be produced from master sequence if one or more mutations take place in any of its  $v - n$  non-lethal positions, but not in the rest of the lethal positions. The probability of this happening is  $q^n (1 - (q^{v-n}))$ . Next, if any non-lethal mutant copies its  $n$  lethal positions correctly, with probability  $q^n$ , it will give rise to another non-lethal species in spite of the number of non-lethal positions that have been miscopied. Finally, a lethal species is produced when, during the replication of any master or mutant non-lethal species, at least one mutation is introduced in any of the  $n$  lethal positions, which happens with probability  $1 - q^n$ .

No population constraint has been introduced in these equations. It can be shown that population fractions are the same as in the constant populations quasispecies model (Schuster and Stadler, 2008).

The linear differential equations system can be expressed in matrix form as:

$$\dot{x} = Wx \quad (2)$$

where the  $W$  matrix is

$$W = \begin{pmatrix} A_m q^v - D & 0 & 0 \\ A_m q^n (1 - q^{v-n}) & A_k q^n - D & 0 \\ A_m (1 - q^n) & A_k (1 - q^n) & -D \end{pmatrix} \quad (3)$$

$W$  being a triangular matrix. Its eigenvalues are the elements of the main diagonal, each one with its associated eigenvector. The model presented is linear, so populations grow infinitely or decrease until extinction. However, during this process the population reaches an equilibrium distribution. In biological terms, eigenvalues are the growth or decrease rates of the population, whereas eigenvectors are the



population distribution associated to those rates (Caswell, 2001). Among the possible eigenvalues of  $W$ , the system reaches the final state given by the dominant eigenvalue or spectral radius, i.e. the one with greater value.

The dominant eigenvalue is the growth rate of the population. If it is greater than zero, the population will grow indefinitely. If it is lesser than zero, it will become extinct, i.e. the dominant eigenvalue refers to the demographic process. On the other hand, the normalized dominant eigenvector is the stationary population distribution expressed in the molar fraction

$$y_i = \frac{x_i}{\sum_i x_i} \quad (4)$$

resulting from the selection-mutation process, so it refers to the genetic-evolutionary component. Each eigenvalue-eigenvector pair defines an evolutionary regime. The first eigenvalue  $\lambda_1 = w_{11} = A_k q^v - D$  is associated with the eigenvector  $v_1$

$$v_1 = \begin{pmatrix} \frac{A_m q^n (1 - q^{v-n})}{A_m - A_k} \\ \frac{A_m q^v - A_k q^n}{A_m - A_k} \\ \frac{(A_m - A_k)(1 - q^n)}{A_m - A_k} \end{pmatrix} = \begin{pmatrix} \frac{A_m (q^n - q^v)}{A_m - A_k} \\ \frac{A_m q^v - A_k q^n}{A_m - A_k} \\ (1 - q^n) \end{pmatrix} \quad (5)$$

which is characterized by a given stationary distribution of master, non-lethal and lethal species. Consequently, this eigenvector-eigenvalue pair defines the quasispecies regime.

The eigenvalue  $\lambda_2 = W_{22} = A_k q^n - D$  is associated with the eigenvector  $v_2$ :

$$v_2 = \begin{pmatrix} 0 \\ q^n \\ 1 - q^n \end{pmatrix} \quad (6)$$

in which there is no master sequence, the population being composed exclusively of non-lethal and lethal mutants. This eigenvector-eigenvalue pair defines error catastrophe. In single-peaked landscapes without back mutation, error catastrophe implies the disappearance of the master sequence. Finally, the third eigenvalue,  $\lambda_3 = W_{33} = -D$ , can only be dominant in the particular case of  $A_k = A_m = 0$ . In this case all sequences are lethal, so the population just decays exponentially.

When two eigenvalues are equal, a transition from one evolutionary regime to another occurs. If  $q$  is the only parameter to be changed, the value at which this transition takes place can be obtained.

Although it is simple, this model has the great advantage of clearly separating genetic processes from demographic processes, so it is very easy to study each one separately, and then its relationship.

Thus, the error threshold ( $q_c$ ) is defined as the value of  $q$  at which the transition to error catastrophe occurs (Biebricher and Eigen, 2005). In the same way, the extinction threshold ( $q_{ex}$ ) is defined as the value of  $q$  at which the population begins to become extinct (Bull et al., 2005). It is possible to obtain explicit values for both thresholds from the  $W$  matrix.

## 2.2. The error threshold

As we have seen, the error threshold can be obtained by equalling both eigenvalues associated with the quasispecies and error catastrophe regimes:

$$W_{11} = W_{22} \quad (7)$$

$$A_m q^v - D = A_k q^n - D \quad (8)$$

and therefore

$$q^{v-n} = \frac{A_k}{A_m} \quad (9)$$

$$q_c = \sqrt[v-n]{\frac{A_k}{A_m}} = \sqrt[v-n]{\frac{1}{\sigma}} \quad (10)$$

where  $\sigma$  is the superiority of the master sequence over the error tail which, in this case, is just the ratio between amplification factors.

Equation (10) shows that, in this lethality scheme, the population behaves as if its sequence length were decreased, and equivalent to the number of non-lethal positions. Therefore, the error threshold decreases when the number of lethal positions ( $n$ ) increases. The value  $v_{ef} = v - n$  is called the effective sequence length.

Particularizing the original premise of Eigen's paper (Eigen, 1971) to the presence of lethality, it is possible to obtain a general expression for any single-peaked fitness landscape. In Eigen's original paper, the error threshold is obtained by equalling the selective value of the master sequence to the residual mean productivity, or error tail productivity:

$$A_m Q_m - D_m = \bar{E}_{i \neq m} \quad (11)$$

where  $\bar{E}_{i \neq m} = \frac{\sum_{i \neq m} E_i x_i}{\sum_{i \neq m} x_i}$

In the presence of lethality, the mean productivity of the error tail can be expressed as the sum of all the contributions of lethal and non-lethal populations:

$$\bar{E}_{i \neq m, \text{let}}(q) = (A_k - D)y_k - Dy_l = A_k y_k - D \quad (12)$$

since only non-lethal mutant fractions ( $y_{nl}$ ) replicate. By inserting (12) into equation (11) and assuming D to be constant, the following expression can be obtained:

$$q_c^v = \frac{A_k y_k(q)}{A_m} = \frac{y_k(q)}{\sigma} = \frac{1}{\sigma_{ef}} \quad (13)$$

where

$$\sigma_{ef}(q) = \frac{\sigma}{y_k(q)} \quad (14)$$

is the effective superiority of the master over the error tail.

Since  $y_{nl} < 1$ , this implies that  $\sigma_{ef} > \sigma$ , and therefore the error threshold will always be lower when lethality is introduced in the landscape, and its value will decrease whenever lethality increases in the population. So, in general, lethality decreases the error threshold due to an increase in the effective superiority of the master sequence over the error tail.

Applying this to the simple model established in section 2, the error threshold can be found from the expression (14), by simply evaluating the effective superiority in the error threshold:

$$\sigma_{ef}(q_c) = \frac{\sigma}{y_k(q_c)} = \frac{\sigma}{q_c^n} \quad (15)$$

The error threshold will therefore be:

$$q_c^v = \frac{1}{\sigma_{ef}} = \frac{q_c^n}{\sigma} \quad (16)$$

$$q_c = \sqrt[v-n]{\frac{1}{\sigma}} \quad (17)$$

which is the same value obtained before, since the original premise expressed by Eigen is formally identical to the equality of eigenvalues in the simple linear model. However, it is worth emphasizing that in the lethality scheme used here, the general premise of an increase in the effective superiority is equivalent to a lower effective sequence length.

### 2.3. The extinction threshold

The extinction threshold is defined as the value of  $q$  at which the dominant eigenvalue becomes negative. Therefore, its value will depend on the evolutionary regime of the population.

When the population is in the quasispecies regime, the extinction threshold is obtained from the eigenvalue  $\lambda_1 = W_{11}$

$$W_{11} = A_m q^v - D = 0 \quad (18)$$

$$q_{ex} = \sqrt[v]{\frac{D}{A_m}} \quad (19)$$

which is constant, and independent from lethality, since it depends only on the absolute fitness of the master sequence.

When the population has crossed the error threshold, and is in the error catastrophe regime, the extinction threshold is obtained from the eigenvalue  $\lambda_2 = W_{22}$

$$W_{22} = A_k q^n - D = 0 \quad (20)$$

$$q_{ex} = \sqrt[n]{\frac{D}{A_k}} \quad (21)$$

This value depends on both lethality ( $n$ ), and the absolute fitness of the mutant species. Beyond the error threshold, the population's growth rate is the mean residual productivity, which is the sum of the contribution of lethal and non-lethal subpopulations:

$$\bar{E}_{k \neq m} = (A_k - D)y_k - Dy_l = A_k y_k - D \quad (22)$$

It is easy to show that, beyond the error threshold, extinction happens when a lethal mutant fraction in the population exceeds a threshold value defined by:

$$\bar{E}_{k \neq m} = A_k y_k - D = 0 \quad (23)$$

$$y_k = \frac{D}{A_k} \quad (24)$$

$$y_l = 1 - \frac{D}{A_k} = \frac{A_k - D}{A_k} \quad (25)$$

#### 2.4. Critical lethality

When lethality is low, the error threshold is greater than the extinction threshold in the quasispecies regime. Increasing the mutation rate causes the population to cross the error threshold, and it subsequently becomes extinct when the mutation rate becomes greater than the extinction threshold defined for error catastrophe. As lethality increases, the error threshold decreases, until it finally falls below the extinction threshold of the quasispecies. Beyond this point, that we call *critical lethality*, the population becomes extinct before it enters error catastrophe, i.e. beyond this point the error threshold disappears. This value can be obtained by equalling both expressions of error threshold for both evolutionary regimes:

$$n = \nu \frac{\ln(A_m / D)}{\ln(A_k / D)} \quad (26)$$

The value of  $n$  depends on the length sequence, and on the ratio between the master and the mutant fitness. As the number of lethal positions in the sequence must be an integer, critical lethality  $n_c$  is the first integer above the value of  $n$  obtained from (26).

A phase diagram can be drawn, using the results obtained in previous sections, to describe the behaviour of the population as a function of lethality, and the mutation rate for the given values of  $A_m$ ,  $A_k$  and  $D$ .

Figure 2 shows this diagram for different single-peaked landscapes, each one characterized by different amplification or degradation factors for the master and mutant species. In every case, three different regimes appear: quasispecies, error catastrophe and extinction. The borders between the regimes are the error and extinction thresholds. When the amplification factor of the mutant class increases from  $A_k = 2$  (fig. 2a) to  $A_k = 5$  (fig. 2b), the error catastrophe regime expands out of both the quasispecies and the extinction regime. On the other hand, beyond the critical lethality, the extinction threshold does not change, as it only depends on master amplification ( $A_m$ ) and degradation factor ( $D$ ). The increase of degradation factor from  $D = 1$  (fig. 2a and 2b) to  $D = 4$  (fig. 2c) expands the extinction regime, increasing the extinction threshold for all  $n$ . On the other hand, the error threshold does not change as  $D$  does not affect the error threshold when  $D$  is constant for all the population. Finally, an increase in the amplification factor of the mutant and the master class that keeps constant the effective superiority  $\sigma_{ef}$ , (fig. 2a and 2d) does not change the error threshold, but decrease the extension of the extinction regime.

## 2.5. Other lethality schemes

In order to compare the effect of lethality distribution on the extinction and error thresholds of the population, other simple lethality schemes have been applied to the model proposed here.

If a uniformly distributed lethality in the whole sequence space is assumed (Bonnaz, 1999), every error class will have a constant fraction  $(1-p)$  of lethal mutants, and the  $W$  matrix will be:

$$W = \begin{pmatrix} A_m q^v - D & 0 & 0 \\ A_m(1-q^v)p & A_k p - D & 0 \\ A_m(1-q^v)(1-p) & A_k(1-p) & -D \end{pmatrix} \quad (27)$$

where the probability of producing non-lethal mutants is proportional to  $p$ , i.e. to the non-lethal mutant fraction on the sequence space. Similarly, the probability of producing a lethal mutant is proportional to  $(1-p)$ .

Proceeding in the same way as in the previous lethality scheme, the error threshold is:

$$q_c = \left( \frac{pA_k}{A_m} \right)^{\frac{1}{v}} = \left( \frac{p}{\sigma} \right)^{\frac{1}{v}} \quad (28)$$

which can be expressed in terms of an effective superiority as:

$$\sigma_{ef}(q_c) = \frac{\sigma}{y_{nl}(q_c)} = \frac{\sigma}{p} \quad (29)$$

$$q_c = \left( \frac{1}{\sigma_{ef}} \right)^{\frac{1}{v}} \quad (30)$$

Consequently, the error threshold decreases as the lethal mutant fraction on sequence space,  $p$ , decreases. The extinction threshold depends, as in the previous scheme, on the evolutionary regime. For the quasispecies regime it is the same as in the previous lethality scheme:



$$q_{ex} = \sqrt{\frac{D}{A_m}} \quad (31)$$

However, an extinction threshold for an error catastrophe regime is not defined using  $q$ , but by the lethality degree,  $p$ :

$$A_k p - D = 0 \quad (32)$$

$$p_c = \frac{D}{A_k} \quad (33)$$

$$1 - p_c = \frac{A_k - D}{A_k} \quad (34)$$

This means that if the population is beyond the error threshold, and the lethal mutant fraction  $(1-p)$  is greater than the critical value  $(1-p_c)$ , the population will become extinct. However, if  $(1-p)$  is less than  $(1-p_c)$ , the population will survive in an error catastrophe regime. This means that, in this scheme, extinction and error catastrophe cannot coexist for any given value of  $p$  (Bonnaz, 1999).

### 3. The extended model

In the previous section a simple model was studied in order to obtain explicit expressions of the error and extinction thresholds as a function of lethality. In the model, sequences were grouped in three classes: master, lethal mutants, and non-lethal mutants, which allowed an analytical treatment. However, it is necessary to classify and group the sequences in  $v+1$   $H_i$  error classes, depending on the number of mutations with respect to the master sequence, for a more detailed study of how lethality affects the delocalization of the population over the whole sequence space. It is also possible to discard the no-back-mutation assumption which, as will be shown, has no qualitative effect.

To take the lethality into account in this model, each error class was divided into two subpopulations of lethal and non-lethal mutants. Since the master sequence is always non-lethal, and the sequence with  $v$  mutations is always lethal,  $2v$   $I_k$  subpopulations are obtained where subscript  $k$  refers to the non-lethal subpopulation of Hamming class  $H_k$  when  $k \leq v$ , and refers to lethal subpopulations of Hamming class  $H_{k-(v+1)}$  when  $k > v$ .

With these subpopulations a  $2v \times 2v$   $W$  matrix is obtained. The elements of the  $W$  matrix are the production rates of the different subpopulations of each Hamming class. Since each class has been divided into two subpopulations, the subscripts of the matrix elements do not refer directly to the Hamming class, but to the different subpopulations given above.

The elements  $W_{ij}$  are the production rate of a subpopulation  $i$  from any other subpopulation  $j$ :

$$W_{ij} = Q_{ij}A_j - D \quad \forall \quad i = j$$

$$W_{ij} = Q_{ij}A_j \quad \forall \quad i \neq j$$

where  $Q_{ij}$  is the probability of obtaining a sequence of the subpopulation  $i$  from any sequence of the subpopulation  $j$  by mutation. As the amplification rate of lethal species is  $A_i = 0$ , the elements of the  $W$  matrix are:

$$W_{ij} = -D \quad \forall \quad i = j \quad j > v+1$$

$$W_{ij} = 0 \quad \forall \quad i \neq j \quad j > v+1$$

To find the probabilities of mutation from non-lethal sequences, namely the  $Q_{ij}$  with  $j \leq v+1$ , the expression of  $Q_{ij}$  in the absence of lethality (Nowak and Schuster, 1989), has been taken into account, with the following premises:

1. The probability of the master sequence replicating without error is:

$$Q_{0,0} = q^v \quad (35)$$

2. The probability of passing from any class  $j$  to a non-lethal subpopulation of any class  $i$ , plus the probability of passing from the same class  $j$  to the lethal subpopulation of the class  $i$ , equals the probability of passing from class  $j$  to class  $i$  in the absence of lethality.

$$Q_{k=i,j}(q,v) + Q_{k=i+v,j}(q,v) = Q_{i,j}(q,v) \quad (36)$$

3. To pass from any class  $j$  to the non-lethal subpopulation of class  $i$ , the  $n$  lethal positions must be correctly copied, and then the number of necessary mutations must be introduced in the  $v-n$  remaining positions. That is:

$$Q_{k=i,j}(q,v) = q^n Q_{k=i,j}(q,v-n) \quad (37)$$

4. From this, the following can be deduced by substituting (37) in (36):

$$Q_{k=i+v,j}(q,v) = Q_{i,j}(q,v) - Q_{k=i,j}(q,v) = Q_{i,j}(q,v) - q^n Q_{i,j}(q,v-n) \quad (38)$$

It is possible to obtain the  $W$  matrix for any values of  $q$ ,  $n$  and  $v$  for any given fitness landscape from these equations. As stated in the previous section, this paper is limited to sharply peaked fitness landscapes, so the amplification rates of the elements of any Hamming class, excepting the master class, are the same. Moreover, the degradation rates of all the species, including the master one, are the same.

The dimension of the resulting  $W$  matrix, even for small  $v$ , is too great to make possible the derivation of the explicit expressions of its eigenvalues and eigenvectors. However, it is possible to calculate the eigenvalues and eigenvectors of the  $W$  matrix numerically. MATLAB® software has been used to do this.

### 3.1 The Effect of lethality on the error catastrophe

Beyond the error threshold, population gets delocalized, drifting over the whole sequence space. Therefore, the deterministic approach becomes inappropriate to study error catastrophe, as this approach predicts that all the sequences will be equally populated, which is materially impossible. However, if this result is taken as the distribution, not of the population, but of the probability of the population to be in a given region of the sequence space, the effect of lethality on delocalization can be studied. This effect will change the error catastrophe distributions.

Taking this into account, the error catastrophe is, in the deterministic approach, an essentially constant evolutionary regime, despite in reality describing populations that are continuously changing. In the error catastrophe, any increase in the mutation rate hardly affects the probability distribution of the population on the sequence space. That is, if a population had been in error catastrophe its mutation rate could not be inferred from its probability distribution, even though all the other parameters were known. Figure 3 shows, for  $n = 0$ , that when  $q$  decreases, the average Hamming distance progressively increases until it reaches a maximum value of  $1/2$  at the error threshold. At this point a discontinuity takes place, beyond which the average Hamming distance remains constant. In the deterministic approach, this result comes from the population delocalization and the fact that sequence space classification in Hamming or error classes is not uniform, as each error class comprises a different number of sequences.

When lethality is considered this behavior changes. Figure 3 shows the variation of average Hamming distance versus  $(1 - q)$  for different degrees of lethality. On one hand, a decrease in the error threshold can be observed, as the simple model predicted. On the other, a decrease in the average Hamming distance at the error threshold can also be observed. Finally, the average Hamming distance is no more constant beyond error threshold, linearly increasing with mutation rate, being this increase greater as greater is the lethality degree.

#### 4. Other fitness landscapes

To assure the validity of the previous results other fitness landscapes have been studied. In the first place, the lethality scheme used throughout this paper has been applied to a multiplicative fitness landscape. In this fitness landscape each mutation has an independent effect  $(1-s)$  on the amplification factor of the self-replicating specie. Thus, the amplification factor of Hamming class  $i$  is  $A_i = A_m(1-s)^i$ , where  $A_m$  is the amplification factor of the master sequence. In this case a simple analytical solution cannot be obtained, so the eigenvalues and the eigenvectors of the system have been numerically evaluated.

As has been described in previous papers, (Schuster and Stadler, 2008; Woodcock and Higgs, 1996), this fitness landscape does not have an error threshold or an error catastrophe regime, so a population can only be in the quasispecies or the extinction regime. Figure 4 shows that the extinction threshold increases as the deleterious effect of each mutation ( $s$ ) is increased.

Moreover, a generalization of the single-peak landscape based on Bull et al., 2005 has been studied. In this fitness landscape there are three phenotypes with decreasing amplification factors ( $A_i$ ). The robustness and uniformly distributed fraction of the non-lethal mutants ( $p_i$ ) are also different. Each phenotype mutates only to the next phenotype, i.e. back-mutation is not allowed. A lethal phenotype, in which lethal mutants from any other phenotype are grouped, is considered. The resulting  $W$  matrix is:

$$\begin{pmatrix} A_1\tilde{Q}_1p_1 - D & 0 & 0 & 0 \\ A_1(1-\tilde{Q}_1)p_2 & A_2\tilde{Q}_2p_2 - D & 0 & 0 \\ 0 & A_2(1-\tilde{Q}_2)p_3 & A_3\tilde{Q}_3p_3 - D & 0 \\ A_1\tilde{Q}_1(1-p_1) + A_1(1-\tilde{Q}_1)(1-p_2) & A_2\tilde{Q}_2(1-p_2) + A_2(1-\tilde{Q}_2)(1-p_3) & A_3\tilde{Q}_3(1-p_3) + A_3(1-\tilde{Q}_3) & -D \end{pmatrix} \quad (39)$$

Where  $\tilde{Q}_i$  is a phenotypic quality factor in which the degeneration of its neutral network is taken into account by means of a parameter  $r$ .

$$\tilde{Q}_i = [1 - r_i(1 - Q_i)] = [1 - r_i(1 - q^v)] \quad (40)$$

$$1 - \tilde{Q}_i = r_i(1 - Q_i) = r_i(1 - q^v) \quad (41)$$

where  $r_i$  ranges between 0 and 1, and expresses the effect of the genotypic mutation rate on the phenotypic change. Thus, it can be taken as an inverse measure of the robustness of the phenotype. When  $r_i = 1$ ,  $\tilde{Q}_i = q^v$ , i.e. there is no robustness, the phenotype behaves as if it were composed of just one sequence. When  $r_i = 0$ ,  $\tilde{Q}_i = 1$ , i.e. the phenotype reproduction is not affected by mutation rate, it behaves as if it had infinite robustness.

As there is no back mutation, the four eigenvalues of the  $W$  matrix are the elements of the main diagonal. As stated in section 2.1, these eigenvalues define the four possible evolutionary regimes. The numerical calculation of the eigenvalues and eigenvectors shows that the greatest eigenvalue defines a quasispecies regime in which the three phenotypes and their lethal mutants coexist. In contrast, the second eigenvalue defines a first error catastrophe regime in which the first phenotype disappears, while the third eigenvalue defines a second error catastrophe regime in which both the first and the second phenotypes disappear. It is possible to find the analytical expression of error and extinction thresholds as was done in sections 2.2 and 2.3. This fitness landscape

displays two error thresholds and three extinction thresholds, whose general expressions are:

$$q_{c,i} = \sqrt[3]{1 - \frac{A_i p_i - A_{i+1} p_{i+1}}{A_i p_i r_i - A_{i+1} p_{i+1} r_{i+1}}} \quad (42)$$

$$q_{ex,i} = \sqrt[3]{\frac{A_i p_i (r_i - 1) + D}{A_i p_i r_i}} \quad (43)$$

Figure 5 shows a phase diagram for this fitness landscape. For the sake of simplicity, all the phenotypes have the same lethal fraction of mutants. Figure 6 shows the variation of phenotype fractions with the mutation rate.

## 5. Discussion

In this paper the effect of a given lethality scheme on the population dynamics of self-replicating species in a sharply peaked landscape has been studied. The most remarkable effects are the decrease of the error threshold with lethality, the changes in the error catastrophe regime, and the occurrence of extinction at a mutation rate which depends on the degree of lethality.

As lethality increases, the effective superiority of the master sequence increases and then, necessarily, the error threshold decreases, so a higher mutation rate is needed to reach error catastrophe. This has been studied in detail for a single-peaked landscape in section 2.2. In the same section, we show that in the particular lethality scheme used in this paper, the concept of effective superiority is equivalent to an effective sequence length, both being inversely related. This effective sequence length corresponds to the number of non-lethal positions.

The dependence of the error threshold on effective superiority for a uniform lethality scheme is derived in section 2.5. An important conclusion can be drawn from the concepts of effective superiority and effective sequence length: the lethal mutants do not contribute to the transition from quasispecies regime to the error catastrophe regime. Equation (14) shows this as the effective superiority depends explicitly on the fraction of non-lethal mutants in the population. Taking into account the fact that the dimension of the sequence space depends on the sequence length, the effective sequence length means that the populations behave as if the sequence space were formed exclusively by the non-lethal mutants. In fact, the extended model can be used to verify that the transition to error catastrophe of a population of sequence length  $v$  and  $n$  lethal positions can be approximated by the transition of a population of length  $v-n$  without lethality (fig. 3). This is because the lethal regions of the sequence space cannot replicate, and therefore they behave as population sinks. On the basis of these results, the effect of lethality is better understood as a withdrawal of non-lethal mutants than as a production of lethal ones.

This “sink effect” of the lethal regions is also responsible for the changes in the error catastrophe regime. As lethal mutants can only be produced by error-prone replication of other non-lethal mutants, the amount produced depends on the mutation rate. Moreover, as the lethal regions concentrate mainly far from the master sequence in the lethality scheme used in this paper, a greater mutation rate will increase accessibility to these regions. Immediately beyond the error threshold, the population delocalizes over a limited region of the sequence space. Increasing the mutation rate extends delocalization to the far lethal regions so that the mean Hamming distance, which as we have shown can be interpreted as a measure of delocalization, grows linearly with the mutation rate (fig. 3).



As has been shown in previous papers, a uniformly distributed lethality decreases the error threshold (Bonnaz, 1999), which then depends on superiority, sequence length, and lethality degree. In section 2.5 this lethality scheme has been introduced in the simple model, obtaining very similar results: the error threshold depends on the effective superiority of the master sequence (eq. 30), so the threshold decreases as lethality increases.

As has been shown in section 2.3, lethality introduces the possibility of extinction caused by an increasing mutation rate. Extinction is a demographic phenomenon, which can happen in both evolutionary regimes. When lethality is low, the error threshold is greater than the extinction threshold for both regimes. When the mutation rate increases, the population first reaches error catastrophe. If the mutation rate continues increasing, the population becomes extinct. As equation (21) shows, the extinction threshold in the error catastrophe regime depends on the lethality introduced, the former increasing as the number of lethal positions is higher. The cause of extinction is the “sink effect” as a population becomes extinct when its growth rate is below zero. As equation (25) shows, this takes place when the fraction of lethal mutants in the population exceeds a given threshold value.

Beyond a given lethality degree, that in this paper has been called *critical lethality*, the error threshold falls below the extinction threshold, so extinction takes place before the population enters error catastrophe. Beyond the extinction threshold, the master sequence does not self-replicate enough to compensate the lethal mutants it produces, so the quasispecies becomes extinct. The extinction condition in the quasispecies regime is equivalent to that obtained when all mutants are lethal (Belshaw et al., 2008), even though this is not strictly the case. Actually, in the lethality scheme used in this paper, the introduction of just five lethal positions meant that more than 95% of sequences with length 20 are lethal.

Extinction in the quasispecies and error catastrophe regimes has different characteristics. In the latter, a critical lethal mutant fraction exists. When it is exceeded, the population becomes extinct. When lethality is increased, this threshold is reached at lower mutation rates. Beyond the critical lethality, the extinction threshold remains constant, but the fraction of lethal mutants at the extinction threshold increases in the same way as lethality does. This fraction can be found by substituting the value of  $q_{ex}$  in the expression of the lethal mutant fraction:

$$q_{ex} = \left( \frac{D}{A_m} \right)^{\frac{1}{v}} \quad (44)$$

$$y_l = 1 - q^n \quad (45)$$

$$y_{l,ex} = 1 - q_{ex}^n = 1 - \left( \frac{D}{A_m} \right)^{\frac{n}{v}} \quad (46)$$

Besides the sharply-peaked landscape, two others have been briefly studied, namely, a multiplicative fitness landscape and a multiple phenotype landscape. The results obtained in the multiplicative fitness landscape, in which there is no error threshold, show again that the existence of error catastrophe critically depends on the fitness landscape, as has been previously stated (Schuster and Stadler, 2008). However, extinction is probably a general phenomenon when lethal mutations exist. On the other hand, the multiple phenotype landscape shows that error and extinction thresholds depend on the robustness, lethality, and reproductive capacity of the phenotypes that a population can adopt.

Obviously, the fitness landscapes studied in this paper are simplifications of any real situation, in which complex, rugged fitness landscapes prevail. However, even in these simple cases the results clearly show that extinction and error catastrophe are

essentially different phenomena, as has been previously discussed (Bull et al., 2005; Bull et al., 2007). Transition to error catastrophe is a genetic-evolutionary phenomenon, resulting from the displacement of the selection-mutation balance. Lethality heightens the effect of selection, so greater mutation rates are required for the population to enter error catastrophe. Extinction is, on the other hand, a demographic phenomenon which depends on the capacity of the population to self-maintain. If lethality exists, more lethal mutants are produced when the mutation rate is increased, so the mean productivity of the whole population decreases. These considerations are not affected by the complexity of the fitness landscapes, or by whether a particular landscape, such as the multiplicative landscape, displays error catastrophe regimes or not.

Furthermore, the results obtained from this paper can be conceived as an approximation to local regions of a complex fitness landscape. Sharply-peaked and multiplicative landscapes must be seen as extreme idealizations of certain real fitness peaks in a rugged landscape. If one or more mutations have a much greater effect than the rest, the peak can be approximated to a sharply-peaked landscape and it will have two or three regimes, depending on the lethality degree. If any mutation has approximately the same effect on fitness, so the fitness decrease is smooth, it can be approximated to a multiplicative landscape, and will display neither error threshold nor error catastrophe. Throughout this paper, a deterministic methodology has been applied. However, real populations are finite and often small. In these cases, population size and small perturbations can have a great effect on the behaviour of the population. A stochastic approach to this problem is currently being developed.

### **Acknowledgments**

This paper has been supported in part by Grants no, BFU2006-01951-BMC from MEC (Spain). Hector Tejero is supported by AP2006-01044, from MEC (Spain); and Arturo Marín is supported by Consejería de Educación de la Comunidad de Madrid (Spain)

and Fondo Social Europeo (FSE). The authors thanks the two anonymous reviewers for their helpful comments and interesting suggestions.

- Belshaw, R., Gardner, A., Rambaut, A., and Pybus, O., 2008. Pacing a small cage: mutation and RNA viruses. *Trends. Ecol. Evol.* 23, 188-93.
- Biebricher, C., and Eigen, M., 2005. The error threshold. *Virus. Res.* 107, 117-27.
- Biebricher, C., and Eigen, M., 2006. What is a quasispecies? *Curr. Top. Microbiol. Immunol.* 299, 1-31.
- Bonnaz, D., 1999. About a three states molecular evolution model. *J. Theor. Biol.* 197, 415-23.
- Bull, J.J., Meyers, L.A., and Lachmann, M., 2005. Quasispecies made simple. *PLoS Comput. Biol.* 1, e61.
- Bull, J.J., Sanjuan, R., and Wilke, C.O., 2007. Theory of lethal mutagenesis for viruses. *J. Virol.* 81, 2930-9.
- Caswell, H., 2001. *Matrix Population Models: Construction, Analysis, and Interpretation.* Sinauer Associates Inc.
- Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C., 1978. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13, 735-44.
- Domingo, E., Martin, V., Perales, C., Grande-Perez, A., Garcia-Arriaza, J., and Arias, A., 2006. Viruses as quasispecies: biological implications. *Curr. Top. Microbiol. Immunol.* 299, 51-82.

- Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465-523.
- Eigen, M., 2002. Error catastrophe and antiviral strategy. *Proc. Natl. Acad. Sci. U S A* 99, 13374-6.
- Grande-Perez, A., Sierra, S., Castro, M.G., Domingo, E., and Lowenstein, P.R., 2002. Molecular indetermination in the transition to error catastrophe: systematic elimination of lymphocytic choriomeningitis virus through mutagenesis does not correlate linearly with large increases in mutant spectrum complexity. *Proc. Natl. Acad. Sci. U S A* 99, 12938-43.
- Holland, J., Domingo, E., de la Torre, J., and Steinhauer, D., 1990. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J. Virol.* 64, 3960-2.
- Kamp, C., Wilke, C.O., Adami, C., and Bornholdt, S., 2002. Viral evolution under the pressure of an adaptive immune system: optimal mutation rates for viral escape. *Complexity* 8, 28-33.
- Loeb, L.A., Essigmann, J.M., Kazazi, F., Zhang, J., Rose, K.D., and Mullins, J.I., 1999. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc. Natl. Acad. Sci. U S A* 96, 1492-7.
- Nowak, M., and 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.
- Schuster, P., and Stadler, P.F., Early replicons: Origin and evolution., in: Domingo, E., et al., Eds.), *Origin and Evolution of Viruses*, Elsevier, Oxford 2008, pp. 1-42.
- Sierra, S., Davila, M., Lowenstein, P.R., and Domingo, E., 2000. Response of foot-and-mouth disease virus to increased mutagenesis: influence of viral load and fitness in loss of infectivity. *J. Virol.* 74, 8316-23.

- Solé, R.V., and Deisboeck, T.S., 2004. An error catastrophe in cancer? *J. Theor. Biol.* 228, 47-54.
- Summers, J., and Litwin, S., 2006. Examining The Theory of Error Catastrophe. *J. Virol.* 80, 20-26.
- Swetina, J., and Schuster, P., 1982. Self-replication with errors : A model for polynucleotide replication. *Biophys. Chem.* 16, 329-345.
- Takeuchi, N., and Hogeweg, P., 2007. Error-threshold exists in fitness landscapes with lethal mutants. *BMC Evol Biol* 7, 15; author reply 15.
- Wagner, G.P., and Krall, P., 1993. What is the difference between models of error thresholds and Muller's ratchet? *J. Math. Biol.* 32, 33-44.
- Webb, G.F., and Blaser, M.J., 2002. Dynamics of bacterial phenotype selection in a colonized host. *Proc. Natl. Acad. Sci. U S A* 99, 3135-3140.
- Wilke, C., 2005. Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* 5.
- Woodcock, G., and Higgs, P.G., 1996. Population Evolution on a Multiplicative Single-Peak Fitness Landscape. *Journal of Theoretical Biology* 179, 61-73.

**Figure 1.** Fraction of lethal mutants of the Hamming class  $i$  vs. Hamming class for different number of lethal positions in the sequence,  $n$ . Sequence length,  $v = 20$ . Hamming class  $i$  is composed by all the sequences at a Hamming distance  $i$  from the master.

**Figure 2.** Phase diagram for different single-peaked landscapes as a function of the number of lethal positions  $n$ , and the probability of correct replication per digit,  $q$ . The sequence length is  $v = 20$ . As can be seen in the figure, in every case three different regimes appear: quasispecies, error catastrophe and extinction (see text for details). a)  $A_m = 10$ ,  $A_k = 2$ ,  $D = 1$ . b)  $A_m = 10$ ,  $A_k = 5$ ,  $D = 1$ . c)  $A_m = 10$ ,  $A_k = 5$ ,  $D = 4$ . d)  $A_m = 15$ ,  $A_k = 3$ ,  $D = 1$ .

**Figure 3.** Average Hamming distance of the population vs. mutation rate  $(1-q)$  for single-peaked landscape ( $A_m = 10$ ,  $A_k = 2$ ,  $D = 1$ ,  $v = 20$ ) and for different number of lethal positions,  $n$ .

**Figure 4**

Extinction thresholds as a function of the number of lethal mutations for different values of  $s$  (the deleterious effect of mutations) for the multiplicative fitness landscape. There are only two evolutionary regimes: the quasispecies regime (over the extinction threshold), and extinction (below the threshold). As the deleterious effect of mutations increases, the extinction threshold increases for any number of lethal mutations. In all the cases,  $A_m = 10$  and the sequence length is  $v = 20$ .

### Figure 5

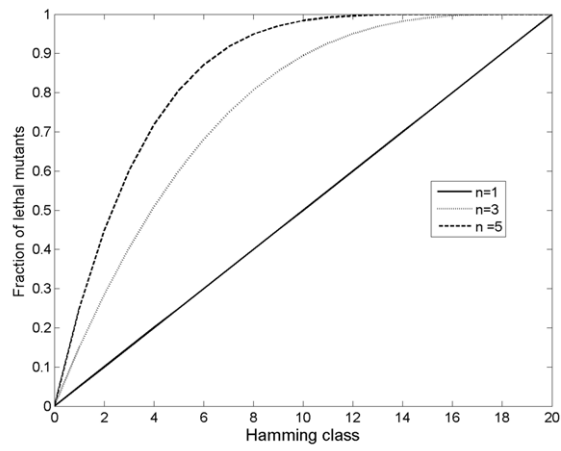
Phase diagram for the multiple phenotype landscape considered under section 4 as a function of lethal mutant fraction  $(1-p)$ , and the probability of correct replication per digit  $q$ . Four different regimes appear: the quasispecies regime, the first and the second error catastrophe regime, and the extinction regime.  $A_1 = 10$ ,  $A_2 = 5$ ,  $A_3 = 3$ ,  $D = 1$ ,  $r_1 = 1$ ,  $r_2 = 0.7$ ,  $r_3 = 0.4$ ,  $p_1 = p_2 = p_3 = p$ . The sequence length is  $v = 20$ .

### Figure 6

Molar fraction of the four phenotypes of multiple phenotype landscape as a function of mutation rate  $(1-q)$ . The arrows show the two error thresholds defined by equation 42.  $A_1 = 10$ ,  $A_2 = 5$ ,  $A_3 = 3$ ,  $D = 1$ ,  $r_1 = 1$ ,  $r_2 = 0.7$ ,  $r_3 = 0.4$ ,  $p_1 = p_2 = p_3 = 0.6$ . The sequence length is  $v = 20$ .

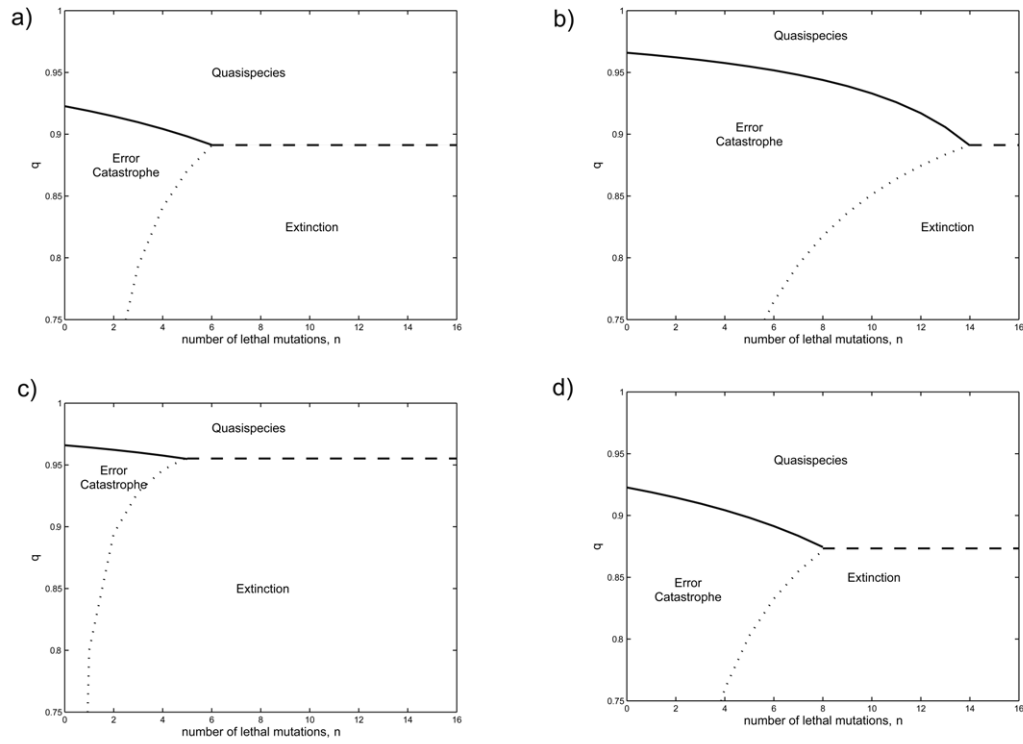


Fig 1



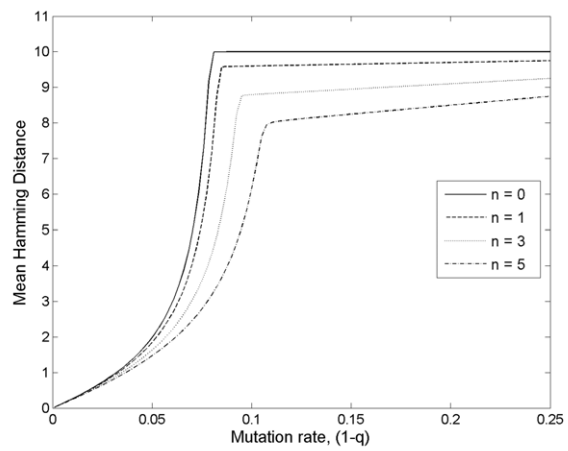
Accepted manuscript

Fig 2



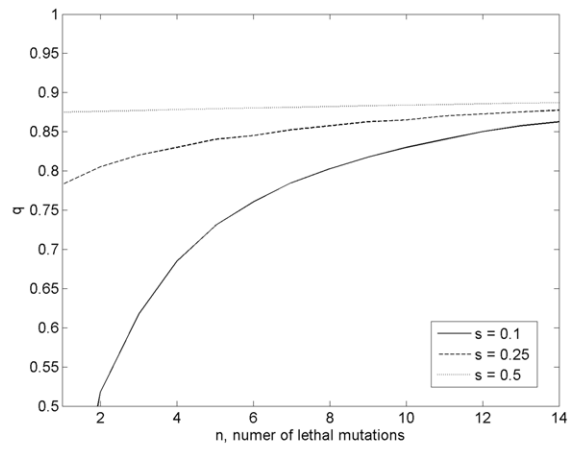
Accepted

Fig 3



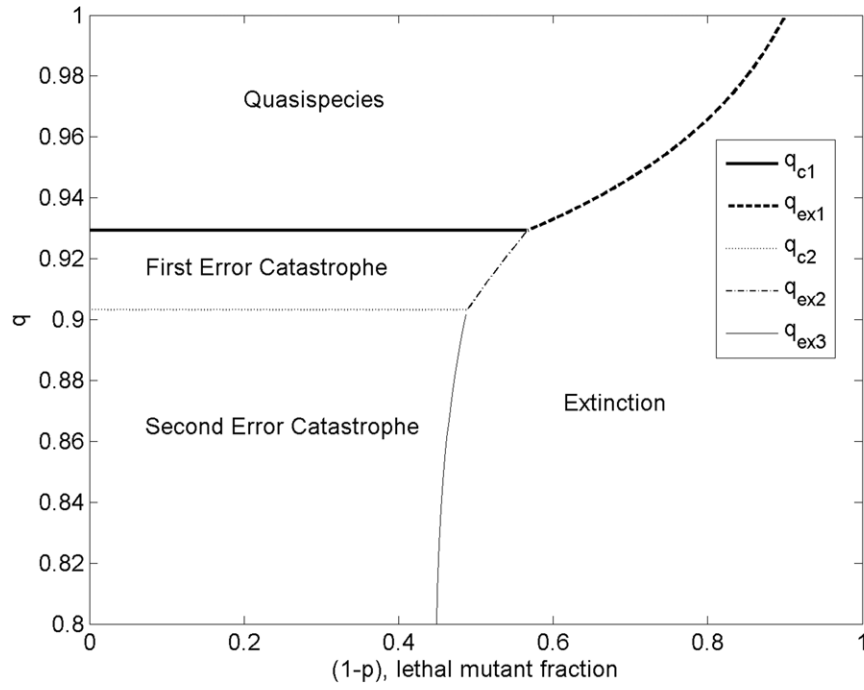
Accepted manuscript

Fig 4



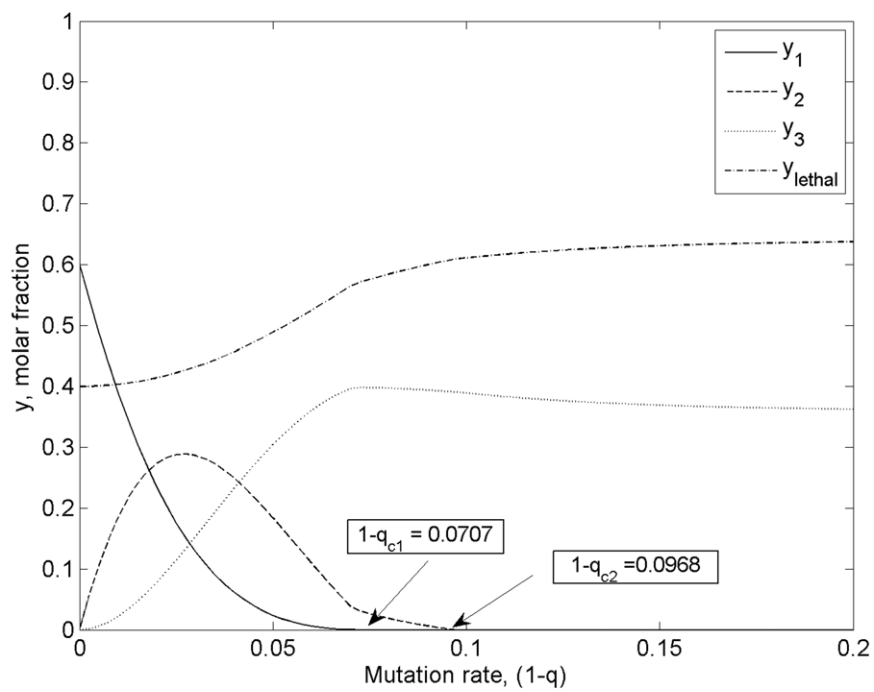
Accepted manuscript

Fig 5



Accepted

Fig 6



Accepted