

On the molecular quasispecies model and the dominance of the fittest genotype

Principal Investigators and Correspondents

Tanya Kostova, tvassile@nsf.gov
National Science Foundation, Arlington, VA, 22230, USA

Carol Zhou, zhou4@llnl.gov
Adam Zemla, zemla1@llnl.gov
Lawrence Livermore National Laboratory, Livermore, CA, 94551, USA



This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

On the molecular quasispecies model and the dominance of the fittest genotype

Tanya Kostova *, Carol Zhou , Adam Zemla †

Abstract

We discuss quasispecies theory and the limitations of Eigen’s quasispecies model with regard to natural viral quasispecies. Analysis of Eigen’s model is performed focusing on the question: what is the effect of increasing mutation rate on the population density of the dominant (most prevalent) genotype? Rigorous mathematical analysis is possible at the two extremes: when the mutation rate is low (in the vicinity of zero) or when it is very large (close to one). At the first extreme, Eigen’s model predicts that the fittest genotype (i.e., the one that has the highest fitness) is the dominant one. At the other extreme, we prove that any genotype could gain dominance, depending on whether certain relations among the replication and degradation rates are fulfilled. We derive formulas that enable exact prediction as to which genotype will emerge as dominant at high mutation rates, and we demonstrate by means of numerical examples the validity of the predictions. We construct examples, based on the theory where we show, by computations, that the fittest (wild) type can remain dominant for all values of the mutation rate. Importantly, we do not restrict the analysis to the (typically considered) single-peaked fitness landscape, nor to the case in which all degradation rates are equal (as done in previous simulations by other authors).

Keywords: quasispecies, viral evolution, error threshold, genotype dominance, spectral abscissa, quasi-positive matrix, asymptotic stability

1 Introduction

Evidence shows that RNA viruses exist as heterogeneous populations of diverse genotypes in cell culture experiments [40, 41], within individual hosts [1, 24, 32, 18, 27] and in populations of hosts [22, 33]. This diversity is believed to result from the high frequency of mutation that occurs during RNA virus replication.

Viral mutation rate, defined as the number of misincorporations per nucleotide copied [10] during virus replication within the cell, has been measured for various viral species. It has been estimated that RNA viruses mutate on average 200-300 times faster than DNA viruses [11, 12] and that the mutation rate of RNA viruses ranges between 10^{-5} and 10^{-3} [10]. The high mutation rate of RNA viruses is explained by the absence of a proof-reading function of the RNA viral polymerase. Because viral genomes vary in length between $3 \cdot 10^3 - 3 \cdot 10^4$, the estimates of RNA viral mutation rate lead to the conclusion that during RNA genome replication very few progeny genomes escape mutation [12].

Many of the mutant progeny contain deleterious mutations that produce defective viral genomes or non-functional proteins—yet others are capable of repeating the replication and diversification rounds in millions of host cells. A fraction of these avoid immune responses, emerge from the host, and continue the cycle in other hosts. This complex process leads to the formation of a heterogeneous population of viral genotypes, called quasispecies.

A quasispecies is defined as "a dynamic distribution of nonidentical but closely related mutant and recombinant viral genomes subjected to genetic variation, competition and selection and which acts as a unit of selection" [10] and the concept is currently relatively widely adopted within the virology community ([8, 40, 34, 35, 23, 6, 29, 26], among many others) to qualify the variability of genetic sequences corresponding to the same virus species.

*National Science Foundation, 4201 Wilson Boulevard, suite 1025, Arlington, VA 22230; E-mail: tvassile@nsf.gov

†Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore CA 94550

The central tenet of the quasispecies theory, first introduced in a mathematical model by Eigen [13] and further analyzed by Eigen and Schuster [14] is that evolution acts on the population as a whole rather than on individual genotypes. Environmental pressures and bottlenecks lead to the formation of specific *equilibrium* distributions of genotypes (and phenotypes), in which a specific genotype, referred to as the master sequence, is present at dominant frequency, while the rest (the minor genotypes) occur at lower frequencies, many of which may be undetectable using any currently available sequencing technologies.

The central tenet offers an explanation of many important phenomena observed in nature and in the laboratory. For example, if environmental conditions change, the relative fitnesses of the genetic variants would presumably also change. In general, as a result of environmental changes, a new master sequence emerges surrounded by a different distribution spectrum of minor types. As an illustration, if the quasispecies is forced to replicate in a different cell type, many variants (including the master type) may be unable to replicate at all, while other, minor types may be highly successful in replication and eventually emerge in dominant proportions within the population. As another example, crossing the species barrier is readily explained by quasispecies theory. The presence of a wide distribution of genotypes generated by replication in one host species increases the chances of generating a minor genotype capable of reproducing in a different host species. Combined with opportunity to infect that species, the minor genotype may be able to replicate thereby forming the seed of a new master sequence that is well adapted to the new host. Third, the emergence of a new master sequence offers explanation of the emergence of resistance to antiviral drugs. A drug that has been developed to target a certain sequence (based on a particular isolate that may represent the current dominant genotype) may not work against some of the minor sequences, which in the presence of the drug eventually may become dominant.

Despite the popularity of quasispecies theory, it is also accompanied by skepticism, controversy and confusion. The reasons for this are two-fold.

First, the mathematical model on which it is based, was developed to represent the dynamics of self-replicating molecular species and was not meant to take into account the specifics and complexity of virus replication, including, among many other features: a) phenotypic constraints, exemplified by secondary and tertiary RNA and protein structures; b) the dependence of the replication rate on RNA-derived polymerase and virus-specific enzymes; c) complementation. The model also makes some assumptions (discussed below) that are not realistic for viruses. As such, some of the virologists are skeptical about one of the predictions of Eigen's model, namely, the existence of an "error threshold" of the mutation rate.

The error threshold was demonstrated in computational experiments done first by Swetina and Schuster [38] and later repeated by many authors [44, 15, 36]. The implication of the error threshold is that above a certain value of the mutation rate the master sequence loses dominance, becoming a very small fraction of the population, and the quasispecies consists of a large number of genotypes in low concentrations. This was interpreted as information loss [16], and the conjecture that virus quasispecies exist near the error threshold was put forward.

Some authors express skepticism regarding the existence of the error threshold in reality [37, 21, 5]. Iranzo and Manrubia [21] believe that "natural quasispecies" are not close to the error threshold, as the quasispecies model does not take into account the large number of genotypes expressing the same phenotype; thus, much of the genotypic variability does not manifest as phenotypic variability. Summers and Litwin [37] argued that the error threshold in Eigen's model exists only because of the unrealistic assumption that all mutant genotypes are able to replicate no matter how far (in terms of number of mutations) from the wild type (i.e., the fittest genotype) they have diverged. They formulated a simple model that approximates Eigen's model dynamics, and showed that without this assumption, their model does not predict error catastrophe. Cases-Gonzales et al. [5] note that "with real viruses, a large expansion through sequence space cannot occur, and . . . the increase in error rate results in a decrease of specific infectivity which can lead to the extinction of the population with modest expansion in sequence space".

Experiments with ribavirin that were meant to demonstrate the existence of the error threshold in poliovirus [6] showed that the increase in mutation rate caused by the drug was accompanied by a steep decrease in infectivity. This was interpreted as the virus having passed the error threshold. However it was later shown that poliovirus develops resistance to ribavirin expressed by the emergence of a genotype coding for a polymerase with increased fidelity which narrows back the quasispecies diversity [31, 42]. The propensity for poliovirus to adapt and avoid error catastrophe cannot be predicted by the quasispecies model because it does not take into account the complexity of the virus replication mechanism which involves virus genome - encoded production of virus species-specific polymerase and enzymes.

Second, the error threshold as a concept is not well understood. It is frequently mistaken for the extinction threshold, both in theory [43, 36, 21, 37] and experiment [9, 42]. The conceptual difference between error threshold and extinction threshold was made clear in [4] and recently reiterated by Holmes in [20]. Below we add clarification on the subject.

Extinction threshold is a term used to explain an abrupt change in density or number of a population brought on by changes in important parameters, such as habitat loss or increased death rate due to various factors (e.g., human factors, disease). The extinction threshold is a central concept in theoretical ecology and conservation biology. Mathematical models in ecology explain the extinction threshold as the value of a parameter at which the zero equilibrium (i.e., the state of the model when the population number is zero) becomes an attractor - that is to say all populations eventually decrease and go extinct with time. In epidemiology the same concept (called the eradication threshold) is used to study the conditions under which a disease can be eradicated; here the threshold again separates parameter values for which the zero (infection-free) equilibrium is stable or unstable. The ecology and epidemiology literature abounds in publications exploring extinction (eradication) thresholds under various conditions ([28, 2, 17, 25], just to mention a few).

In regards to the quasispecies model, if the parameter of interest is the mutation rate q , the extinction threshold would be a value of q beyond which the quasispecies would eventually reach extinction. Whereas Eigen's model demonstrates the error threshold, it does not support the existence of an extinction threshold. The mathematical structure of Eigen's model is such that for each set of parameters it can have only one equilibrium, which is strictly positive and globally stable. This means that independent of the initial genotype frequencies, the quasispecies population converges with time to an equilibrium in which each of the genotypes is present. Therefore, in Eigen's model the quasispecies does not have an extinction threshold. However, as simulations have demonstrated, an error threshold does exist at least for parameter values of Eigen's model for which the simulations were made.

Further, Eigen's model has some peculiarities that were mostly overlooked in the literature. If one takes a careful look at the simulations made by Swetina and Schuster and other authors, one notes (see for example [38], p.340, [16], p.118) that for very high mutation values (probability of mutation close to 1) two genotypes dominate: the original master sequence and its "mirror sequence" (i.e., the one in which all nucleotides have mutated). This peculiarity was only marginally mentioned in the original model papers and is easily overlooked in simulation plots that do not represent a range of values of q spanning from 0 to 1. Thus, according to Eigen's model, at very high mutation values, not only is it possible that the original master sequence still exists, but it may exist at a high frequency in the population, second only to its mirror sequence. In addition, according to the simulations (cited above), the mirror sequence becomes the dominant genotype at very high mutation rates. We note that for real viral genomes the mirror sequences would certainly be functionally defective and, thus, the conclusion derived from simulations of Eigen's model at very high mutation rates is unrealistic.

It is important to note that all simulations for the error threshold were done for the special case in which the degradation rates of all genotypes were equal [38, 44, 15]. In some cases the degradation rates were set to zero, and in others they were set to a chosen constant. However, it is reasonable to assume that well adapted virus genotypes would be more effective in disabling cellular mechanisms of antiviral defense, including mechanisms for degrading RNA (i.e., miRNAs reviewed in [30], p.195, [19], [39]). It would be instructive to explore how differing degradation rates would affect the outcomes of simulations using these models. Apart from this, simulations were also only performed with a limited choice of replication rates: while the wild type had one, higher replication rate, the rest of the mutant genotypes had the same, lower rate of replication (the so called "one-peak landscape").

This paper provides analysis of the above mentioned features of the quasispecies model that have been traditionally overlooked in the literature. We use methods of stability analysis and properties of quasipositive matrices to understand what is happening with the behavior of the quasispecies model at very high mutation rates. We do not limit our analysis to restricted values of the degradation and the replication rates. We prove that for very high mutation rates the wild type can be again the dominant type and we derive the analytical conditions under which this can be the case. Interestingly, we prove that if all degradation rates were equal (which was the case for the computational experiments in [38]), the wild type could not be dominant for very high mutation rates. Our analysis not only explains the computational observations in [38] but also derives the conditions under which, according to Eigen's model, a mutant genotype may emerge as dominant at high mutation rates.

Further, we examine the problem of the preservation of the master sequence. We ask the question

whether the wild type can remain dominant (i.e., is found in highest concentration compared to all other types of the quasispecies) for all values of the mutation rate. When the mutation rate is small, the wild type is dominant. When the mutation rate increases, many other mutants appear and some may become dominant. It is, however, conceivable that if the wild type is dominant for both low and high mutation rates, it might, under certain conditions, remain dominant for all values of the mutation rate. We show by means of computational experiments that this indeed could be a possible case, at least as a prediction of Eigen's model.

2 Background

2.1 Eigen's quasispecies model

The model considers the concentrations, or frequencies x_i of a given number N of "information carriers", or "sequences", S_1, S_2, \dots, S_N each of which can replicate (produce copies of itself) with a rate A_i , which in the model is a constant (i.e., independent of x_1, \dots, x_N). The replication of each S_i is imperfect; there is a non-zero probability Q_{ji} (the probability of mutation) that sequence S_i will produce by mistake a copy of any sequence S_j , while Q_{ii} is the probability that sequence i will faithfully reproduce itself. All sequences can be degraded with rates, correspondingly, D_i (the degradation rates).

The quantity $A_i - D_i$ is called the *excess productivity*, while the quantity $A_i Q_{ii} - D_i$ is the "*fitness*" of S_i [38]. Obviously, when replication has perfect fidelity ($Q_{ii} = 1$), the fitness is equal to the excess productivity. When the fidelity of replication Q_{ii} is not perfect ($Q_{ii} \neq 1$), the fitness of S_i may differ from the excess productivity and change when Q_{ii} changes. *Noting this, in the text that follows, we shall call $A_i - D_i$ also the "nascent fitness" of S_i .*

Eigen's quasispecies model is designed in such a way that the sum of all concentrations is equal to 1:

$$\sum_i x_i = 1. \quad (2.1)$$

In the language of dynamical systems, the hyperplane (2.1) is an invariant set of the model; thus, if at time 0 the initial concentrations are on the hyperplane, they will remain on it for all t. To ensure this property of the system, Eigen introduced a flux function $f(\vec{x}) = \sum (A_j - D_j)x_j$ and defined the model as follows

$$\dot{x}_i = [A_i Q_{ii} - D_i - f(\vec{x})]x_i + \sum_{k \neq i} A_k Q_{ik} x_k, i = 1, \dots, N \quad (2.2)$$

The model and its consequences were further investigated [38, 15] for binary sequences (i.e., such that each position on the sequence can take two values: 0 and 1) of a fixed length n , and under the assumption that each position can change its value (i.e., mutate) during one sequence replication event, with probability q . Thus, the probability that a sequence of length n will mutate to another sequence with the same length, but differing from the first in m positions (i.e., when the Hamming distance $d(i, j)$ between the two sequences is m) is $(1 - q)^{n-m} q^m$.

In general the probability of mutation of sequence S_i to sequence S_j with Hamming distance between the two equal to $d(i, j)$ is

$$Q_{ij} = (1 - q)^{n-d(i,j)} q^{d(i,j)}. \quad (2.3)$$

This formula is based on the assumption that each position mutates independently (i.e., there is no mutational dependency among positions).

2.2 Swetina and Schuster's simulations and the error threshold

Swetina and Schuster used the mutation rate in the form (2.3) to perform numerical simulations of the model (2.2) with increasing sequence length n [38]. For each n they varied q from 0 to 1 and for each q and n they calculated the equilibrium distribution. The simulation showed that at low mutation rates ($q \approx 0$), the dominant genotype is the one with the highest nascent fitness, calculated as $A_1 - D_1$. This genotype is called the *wild type*.

When q was increased beyond a certain "error threshold" value, the wild type lost its dominance abruptly and became a very small fraction of the population. The other genotypes existed at low equilibrium levels as

well [38]. To illustrate the simulation results, the frequencies of all species with a given Hamming distance from the wild type were summed up (let us denote these by Σ_d , where d is the Hamming distance from the wild type). Then, for values of the mutation rate above the error threshold q_{min} and below another, second threshold value q_{max} all Σ_d appeared to have the same constant value. Actually, above this second threshold, the simulations showed the resurrection of the wild type, which was dominated by its "mirror" sequence.

The reappearance of the wild type at very high values of the mutation rate above the second threshold is contrary to intuition. The wild type reappears only because the quasispecies model allows sequences with arbitrary Hamming distance from the wild type—including the mirror sequence—to remain viable and replicate. As a consequence, for very high values (close to 1) of the mutation rate, the wild type mutates primarily into its mirror sequence, and because the wild type has the highest replication rate, it and its mirror eventually become the dominant types. This assumption (that arbitrary sequences are viable and replicate), as pointed by some authors, including [37], is not valid for viral RNA. Actually, sometimes a single mutation is deleterious, and there is no reason to believe that a mirror sequence would encode anything other than a non-viable genome.

We note that the calculations in [38] were done with all values of the degradation rates equal to each other: $D_i = D, i = 1, \dots, N$. Also, all similar simulations in the quasispecies literature, such as [15, 44] were done under the same assumption. Another restriction on the model parameters used in all simulations is the so-called single peak fitness landscape, in which the wild type has the highest nascent fitness and all other genotypes have equal or lower nascent fitnesses: $A_1 - D_1 > A_2 - D_2 = \dots = A_N - D_N$. Having in mind that $D_i = D$, the latter boils down (in Swetina and Schuster's and other simulations) to $A_1 > A_2 = \dots = A_N$.

We conclude this section by noting that even though the quasispecies model makes many unrealistic assumptions with regard to viral replication, it may have fundamental properties that might be inherited by more realistic models. In addition, Swetina and Schuster's simulations were done with a restrictive set of parameters, so it is of interest to see what is gained if these restrictions are relaxed or removed.

3 Model analysis

3.1 Notations and definitions

The following represent a comprehensive list of notations and definitions used throughout the text.

I_k - the identity matrix in R^k ;

If A is a matrix (or a vector), $A \geq 0$ if all its entries are nonnegative, $A > 0$ if $A \geq 0$ and at least one of its entries is positive, and $A \gg 0$ if all its entries are positive;

A is called quasipositive if it has nonnegative off-diagonal elements;

$\lambda(A)$ - eigenvalue of matrix A ;

$s(A)$ - spectral abscissa of the matrix A , the largest real part of any of the eigenvalues of A : $s(A) = \max_i \lambda_i(A)$;

S_j - genotype sequence indexed j ;

n - number of positions ("nucleotides") in a sequence;

N - number of all possible sequences of length n ($N = 2^n$);

q - mutation rate, equal to the probability of one point mutation per one sequence replication;

Q_{ij} - probability of mutation of S_j into S_i during one sequence replication;

Q_{ii} - fidelity of replication of S_i ; $Q_{ii} + \sum_{j \neq i} Q_{ij} = 1$;

$Q = (Q_{ij})$ - the mutation probability matrix;

$diag(A_i)$ - a diagonal matrix with elements A_i ;

D_i - degradation rate of S_i ;

A_i - replication rate of S_i ;

$D = diag(D_i)$; $A = diag(A_i)$

$A_i Q_{ii} - D_i$ - fitness of S_i for a given fidelity of replication Q_{ii} ;

$F_i = A_i - D_i$ - excess productivity of type i , also called nascent fitness;

$f(x) = \sum (A_j - D_j)x_j$ - average excess production of sequences, also equal to the average fitness in the absence of mutation;

x_j^∞ - the frequency (concentration) of S_j at equilibrium;

$\vec{x}^\infty(q) = (x_1^\infty(q), \dots, x_n^\infty(q))$ - the unique positive equilibrium of sequence frequencies for a given mutation rate $q \in (0, 1)$;

Mirror sequence - given a sequence with index l and a sequence with index m , such that $d(l, m) = n$ (i.e. they differ in all positions), S_l is the *mirror* of S_m and vice versa.

3.2 The unique positive equilibrium and its meaning

If all sequences are capable of replication, i.e. $A_i > 0$, for all i , and if the probabilities of mutation are positive, Eigen's model (2.2) has a unique positive equilibrium, solving $\dot{x}_i = [A_i Q_{ii} - D_i - f(\vec{x})]x_i + \sum_{k \neq i} A_k Q_{ik} x_k = 0, i = 1, \dots, N$, denoted here as \vec{x}^∞ . In Appendix B we prove the existence of a unique positive equilibrium under slightly more relaxed conditions. Namely, if the mutation probability matrix Q is irreducible and $A \gg 0$, there is a unique \vec{x}^∞ . The mutation probability matrix is irreducible if for any pair of sequences S_i and S_j there is a non-zero probability that S_j will mutate into S_i after a certain number of replications. For example, if $Q_{ij} \neq 0$, there is a non-zero probability of mutation from S_j to S_i in just one sequence replication. If $Q_{ij} = 0$, but there is a "path" $\{i i_1, i_1 i_2, \dots, i_p j\}$ so that $Q_{i i_1} Q_{i_1 i_2} \dots Q_{i_p j} > 0$, there is a positive probability of mutation from S_j to S_i in $p + 1$ consecutive replications.

Eigen and coworkers showed in [15] that all solutions of the model, independently of the initial conditions converge to \vec{x}^∞ ; that is the equilibrium \vec{x}^∞ is *globally asymptotically stable*. Note that because the equilibrium is positive, none of the genotype frequencies can become 0. Thus, Eigen's model predicts that, for given replication and degradation rates and mutation probabilities, for all possible initial distributions of the genotype sequences, after a sufficiently long period of time the quasispecies will self-organize and approach the same equilibrium state.

For the evolutionary biologist this result translates into the following implication (i.e.,- the central tenet of quasispecies theory).

Selection acts upon the quasispecies population as a whole. It leads to the establishment of a dynamical equilibrium distribution of genotypes and not just to the "survival" of a most fit genotype. Assuming there exists only one genome with highest frequency, that one is called the master sequence. All other genotypes coexist, even if in minuscule proportions. These genotypes continually mutate into each other while maintaining the equilibrium distribution.

Although the result above is true for any form of the mutation probabilities satisfying the only restriction $Q_{ii} + \sum_{j \neq i} Q_{ij} = 1$, further we shall focus on Eigen's model with a mutation matrix defined as in (2.3). We assume without loss of generality that the nascent fitness of genotype S_1 is the highest:

$$A_1 - D_1 > A_i - D_i, i \neq 1. \quad (3.1)$$

We will explore the following questions. What happens when the mutation rate increases? Which is the master sequence when q is small and which is it when q approaches 1? What happens when q increases from 0 to 1?

3.3 Analysis for small mutation rates

When the mutation rate is very low, i.e. $q \approx 0$, and if (3.1) holds, S_1 is the dominant genotype. We see this by first considering the case with absolute fidelity of the replication, i.e. when $q = 0$.

(a) Case $q=0$.

If $q = 0$, $Q_{ij} = 0, i \neq j$. Denoting the nascent fitnesses as $F_i = A_i - D_i > 0$, the system can be written as

$$\dot{x}_i = [F_i - f(\vec{x})]x_i, i = 1, \dots, N. \quad (3.2)$$

We show in Appendix C that this system does not have *positive* equilibria. If all F_i are different, there are exactly N equilibria: the unit vectors $(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)$. Otherwise the system can have other equilibria, apart from these. We show in Appendix C that $(1, 0, \dots, 0)$ is globally stable while all other equilibria are unstable. This result has the following biological interpretation.

In the absence of mutation, independently of the initial distribution of different genotypes, the quasispecies approaches a state in which all genotypes with nascent fitness less than the maximum would disappear, while the genotype with the highest nascent fitness would persist.

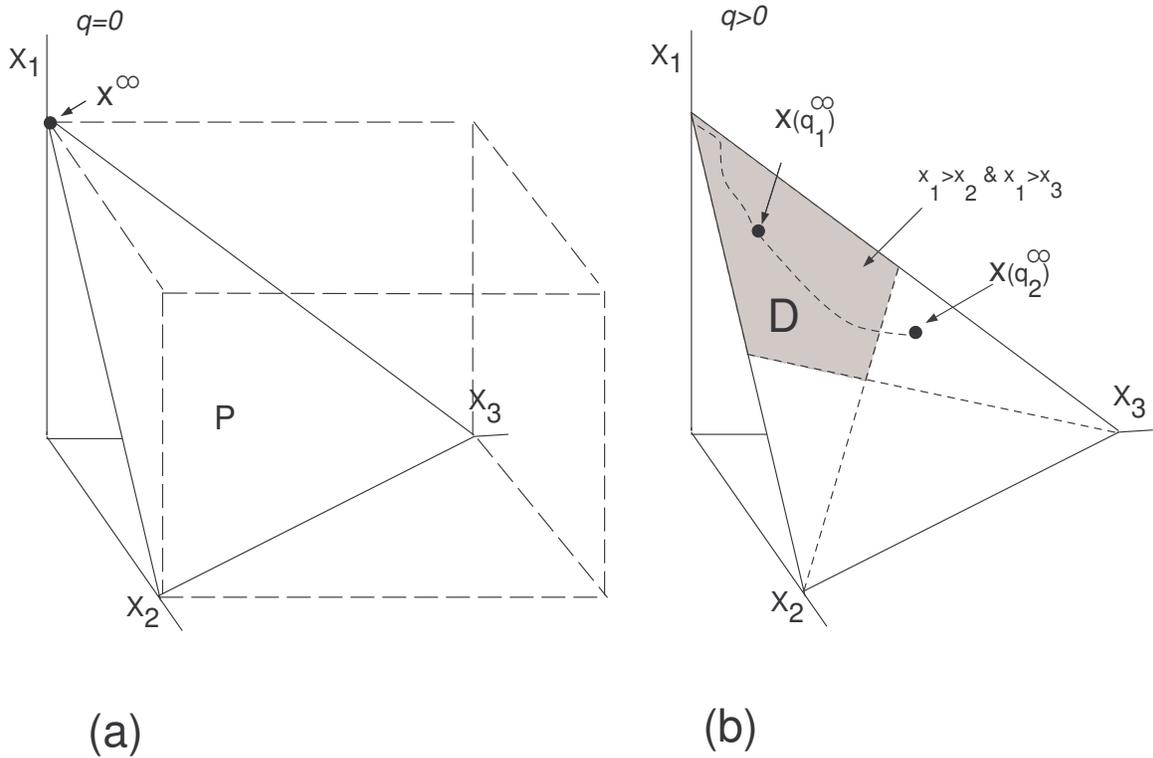


Figure 1: When q increases from 0, the unique positive equilibrium moves on a curve starting from the stable equilibrium $(1, 0, 0)$ for $q = 0$. See text.

(b) Case $q \approx 0$. Since we know that there is a unique positive equilibrium for each value of the mutation rate $q \in (0, 1)$, we now question what is the location of this equilibrium in the hypercube $\mathcal{C} = (0, 1)^N$. We show that when q increases gradually above zero, the positive equilibria $\bar{x}^\infty(q)$ lie on a curve starting at the globally stable equilibrium $(1, 0, \dots, 0)$. To prove this we first use the Implicit Function Theorem to show the existence of a curve of equilibria (not necessarily positive) and then prove that this curve lies in \mathcal{C} by verifying that $x_i^\infty(q) \in (0, 1)$ (Appendix D). We also show in appendix D that the frequency of the wild type gradually decreases while the frequencies of the other types increase when the mutation rate starts to increase from 0.

We illustrate this and further statements in Figure 1 where we have depicted a three dimensional sequence space. As $\sum_{i=1,2,3} x_i = 1$, the frequencies x_i are physically located on the part P of the plane defined by this equation and lying within the cube $(0, 1)^3$. When $q = 0$ the stable equilibrium is at $x_1 = 1, x_2 = 0, x_3 = 0$ (Figure 1(a)). When q starts increasing from 0, the family of equilibria $\bar{x}^\infty(q)$ lies on a curve starting at $(1, 0, 0)$, Figure 1(b). The curve is initially in the shaded part D of P which is characterized by the property $x_1 > x_2, x_3$, i.e. S_1 is dominant. When q grows further, there is the possibility that $\bar{x}^\infty(q)$ will move out of D and S_1 will lose dominance. There is also another possibility, that when q increases from 0 to 1, the curve of equilibria will never leave D and S_1 will remain dominant.

In short, in Appendix D we obtain the following result:

Proposition 3.1. *For $q \approx 0, q > 0$ the unique positive equilibrium $\bar{x}^\infty(q)$ of the system (3.2) lies on a curve starting, at $q = 0$, from the equilibrium $\bar{x}^\infty(0) = (1, 0, \dots, 0)^T$. For values of q in some small interval around $q = 0$, the wild type frequency x_1^∞ remains larger than the frequencies of the remaining types: $x_1^\infty > x_i^\infty, i = 2, \dots, N$ (i.e. S_1 is the dominant species for small mutation rates).*

Since the equilibrium frequency of the wild type x_1^∞ decreases when the mutation rate q increases from

0, two possibilities arise: either x_1^∞ becomes smaller than some other genotype frequencies for some value of q or it remains the highest frequency for all q .

Biological interpretation. For small values of the mutation rate, the wild type remains dominant, although its frequency in the population decreases when the mutation rate increases. The wild type may lose its dominance when the mutation rate increases further, or it can remain dominant independent of the mutation rate.

3.4 Analysis for high mutation rates

Further we investigate what happens with the curve of equilibria for large values of q .

3.4.1 A toy model with $N = 2$

We consider now an oversimplified "sequence" consisting of only one "nucleotide", which can assume only two values (0 or 1). The consideration of this example will be helpful in analyzing a more complex model with $N > 2$. In this case only two "genotypes" - $S_1 = \{0\}$ and $S_2 = \{1\}$ are possible. Alternatively, we may consider that S_1 and S_2 are two subpopulations of the quasispecies: one with higher nascent fitness than the other. Let their replication rates be A_1 and A_2 and their degradation rates be D_1 and D_2 . We assume again that S_1 has higher nascent fitness than S_2 :

$$A_1 - D_1 > A_2 - D_2. \quad (3.3)$$

The quasispecies model then takes the form:

$$\begin{aligned} \dot{x}_1 &= [A_1(1 - q) - D_1 - f(\vec{x})]x_1 + A_2qx_2 = \phi_1(\vec{x}, q) \\ \dot{x}_2 &= A_1qx_1 + [A_2(1 - q) - D_2 - f(\vec{x})]x_2 = \phi_2(\vec{x}, q) \end{aligned} \quad (3.4)$$

where $\vec{x} = (x_1, x_2)$ and $f(\vec{x}) = (A_1 - D_1)x_1 + (A_2 - D_2)x_2$.

According to the results in the previous section, when $q = 0$, the model has two equilibria: $(1, 0)$ and $(0, 1)$. From the latter originates the positive equilibrium $x_1^\infty(q), x_2^\infty(q)$ which is such that for small mutation rates q , $x_1^\infty(q) > x_2^\infty(q)$. The wild type (the one with the larger nascent fitness $A_1 - D_1$) is dominant for small mutation rates.

If x_1^∞ becomes equal to x_2^∞ at some threshold value $q = q_{min}$ and, $x_1^\infty < x_2^\infty$ when q grows further beyond the threshold, then S_1 loses dominance. We show in Appendix E that under certain conditions on the replication and degradation rates, the wild type never loses dominance.

Proposition 3.2. *If, in addition to (3.3), one of the following conditions is also satisfied:*

$$\begin{aligned} (a) & A_1 - A_2 \leq 0 \\ \text{or} & \\ (b) & A_1 - A_2 > 0 \text{ and } A_2 + D_2 \geq A_1 + D_1, \end{aligned} \quad (3.5)$$

the wild type remains dominant for all $q \in (0, 1)$; i.e. $x_1(q) > x_2(q)$ for all $q \in [0, 1]$.

Proof of this statement is reported in Appendix E.

Biological interpretation. Thus, for this toy case we proved that if the wild type (genotype S_1) has higher nascent fitness but lower replication rate than genotype S_2 , (i.e. (*) $A_2 - A_1 > 0$ and $A_1 - D_1 > A_2 - D_2$) it will remain dominant independently of the value of the mutation rate. This result points to the importance of considering differences in the degradation rates. In effect, if the degradation rates of S_1 and S_2 were equal, higher nascent fitness would imply higher replication rate: $A_1 - D_1 > A_2 - D_2$ if and only if $A_1 > A_2$ and the above pair of inequalities (*) would be invalid. Different rates of degradation can affect the rates of accumulation of genotypes as differences in the degradation rates can cause a species that replicates more slowly to have higher nascent fitness, as would be the case when the inequalities (*) hold together. If the more fit species has a lower replication rate, then it will produce mutants more slowly but will be degraded

less readily, whereas the mutant species will produce the wild type at a higher rate and be degraded faster at the same time.

We also showed that if the wild type has a higher replication rate than the mutant, it can still remain dominant for all values of the mutation rate, if $A_1 + D_1 \leq A_2 + D_2$. That is, if the wild type is degraded at a much lower rate than the mutant, so that the previous inequality holds, the wild type will remain dominant. However, this (i.e. condition (b)) is impossible if $D_1 = D_2$.

Thus, the cases for which we show that the wild type remains always dominant are impossible if the degradation rates of the two types are equal. Therefore if $D_1 = D_2$, the wild type will always lose dominance.

3.4.2 The dominant genotype for very large mutation rates ($q \approx 1$) and $N > 2$.

If the number N of genotypes N is larger than 2, it is no more possible to derive the conditions for the parameters of the model for which the wild type will be dominant independently of the value of the mutation rate $q \in [0, 1]$. However, we will derive conditions for the model parameters that will help us construct examples for which the wild type preserves its dominance as well as other examples in which we can predict the mutant types that emerge and remain dominant for sufficiently high mutation rates.

We consider binary sequences as in [38, 15, 44]. Thus if the sequence length is n , the number of possible sequences is $N = 2^n$. We assume that all genotypes have positive replication rates $A_i > 0$ and nonnegative degradation rates, so that $F_i = A_i - D_i > 0$. Finally, without loss of generality, we assume that type with index 1 is the wild type, i.e. $F_1 > F_i, i = 2, \dots, N$.

As we established in section 3.3, the wild type is dominant for sufficiently small values of the mutation rate q . We next investigate whether and under what conditions it would be possible for the wild type to be dominant, i.e. $x_1^\infty(q) > x_i^\infty(q), i \neq 1$ for very high mutation rate values, i.e. $q \approx 1$. For this purpose we shall show that for $q = 1$ there is a unique nonnegative asymptotically stable equilibrium $\vec{x}^\infty(1)$ and that for $q \approx 1$, the positive equilibrium \vec{x}^∞ branches out from $\vec{x}^\infty(1)$. Then, we will derive conditions that ensure that $x_1^\infty(1) > x_i^\infty(1), \forall i \neq 1$. We will finally construct, using these conditions, examples of quasispecies models that demonstrate that the wild type can remain dominant independently of the value of the mutation rate.

Case N=4 To ease understanding we demonstrate the approach in more detail for $N = 4$; however this approach works for $N = 2^m$. We consider a small quasispecies system with 4 species represented as $S_1 = (1, 1), S_2 = (1, 0), S_3 = (0, 1), S_4 = (0, 0)$ with replication and degradation rates A_i and D_i , respectively.

Since q is the probability of mutation of one nucleotide per one sequence replication, the mutation probability matrix Q is then

$$Q = \begin{pmatrix} (1-q)^2 & (1-q)q & (1-q)q & q^2 \\ (1-q)q & (1-q)^2 & q^2 & (1-q)q \\ (1-q)q & q^2 & (1-q)^2 & (1-q)q \\ q^2 & (1-q)q & (1-q)q & (1-q)^2 \end{pmatrix}. \quad (3.6)$$

The system for the equilibria is

$$(QA - D - f(\vec{x}^\infty)I_4)\vec{x}^\infty = 0, \quad (3.7)$$

where $A = \text{diag}(A_1, A_2, A_3, A_4), D = \text{diag}(D_1, D_2, D_3, D_4)$ and I_4 is the identity 4×4 matrix.

When $q = 1$, the matrix $QA - D - f(\vec{x}^\infty)I_4$ becomes an X - shaped matrix

$$\begin{pmatrix} -D_1 - f(\vec{x}^\infty) & 0 & 0 & A_4 \\ 0 & -D_2 - f(\vec{x}^\infty) & A_3 & 0 \\ 0 & A_2 & -D_3 - f(\vec{x}^\infty) & 0 \\ A_1 & 0 & 0 & -D_4 - f(\vec{x}^\infty) \end{pmatrix} \quad (3.8)$$

which is equivalent to the block - diagonal matrix

$$P = \begin{pmatrix} -D_1 - f(\vec{x}^\infty) & A_4 & 0 & 0 \\ A_1 & -D_4 - f(\vec{x}^\infty) & 0 & 0 \\ 0 & 0 & -D_2 - f(\vec{x}^\infty) & A_3 \\ 0 & 0 & A_2 & -D_3 - f(\vec{x}^\infty) \end{pmatrix}, \quad (3.9)$$

and is no longer irreducible. Thus, for $q = 1$, a positive equilibrium can exist only in the special case when $s(L_1) = s(L_2)$, where

$$L_1 = \begin{pmatrix} -D_1 & A_4 \\ A_1 & -D_4 \end{pmatrix}, L_2 = \begin{pmatrix} -D_2 & A_3 \\ A_2 & -D_3 \end{pmatrix}. \quad (3.10)$$

We set this special case aside and consider the case when

$$s(L_1) \neq s(L_2). \quad (3.11)$$

Let us assume that $s(L_1) > s(L_2)$.

Because of the block - diagonal structure of P , and because L_1, L_2 are quasipositive (Appendix A), no positive equilibrium solution exists, but there are exactly two nonnegative equilibrium solutions, which we denote as

$$\vec{y}_1^\infty = (\chi_1^\infty, \chi_4^\infty, 0, 0)^T \text{ and } \vec{y}_2^\infty = (0, 0, \chi_2^\infty, \chi_3^\infty)^T. \quad (3.12)$$

Here $(\chi_1^\infty, \chi_4^\infty)^T$ is the unique positive eigenvector of L_1 satisfying $f(\vec{y}_1^\infty) = s(L_1)$ and $(\chi_2^\infty, \chi_3^\infty)$ is the unique positive eigenvector of L_2 satisfying $f(\vec{y}_2^\infty) = s(L_2)$.

For this case we can prove the following

Proposition 3.3. *If $s(L_1) > s(L_2)$, \vec{y}_1^∞ is a (locally asymptotically) stable equilibrium and \vec{y}_2^∞ is unstable. Conversely, if $s(L_2) > s(L_1)$, \vec{y}_2^∞ is a locally asymptotically stable equilibrium and \vec{y}_1^∞ is unstable.*

Proof of this statement is reported in Appendix F.

Note. For the reader who is not acquainted with stability theory we would point out that an equilibrium is *locally asymptotically stable* if the solutions of the quasispecies model converge to it if the initial (at time 0) distribution is close to the equilibrium; it is *globally stable* if all solutions (independently of their initial conditions) converge to it.

Thus, so far we have shown that when $q = 1$, the equilibrium corresponding to the matrix (L_1 or L_2) with the larger spectral abscissa is the only nonnegative (locally asymptotically) stable equilibrium. Next we show that this equilibrium gives rise to the family of positive globally stable equilibria of the model when the mutation rate q varies with values close to 1.

Proposition 3.4. *Assume that $s(L_1) > s(L_2)$. When q changes, decreasing from 1, there exists a smooth curve of positive equilibria $\vec{x}^\infty(q)$ defined in an interval around $q = 1$ such that $\vec{x}^\infty(1) = \vec{y}_1^\infty$. Else, if $s(L_2) > s(L_1)$, there is a curve of positive equilibria, such that $\vec{x}^\infty(1) = \vec{y}_2^\infty$.*

It can be proved (proof not shown) that if q varies from 0 to 1, the positive globally stable equilibria of (2.2) form a curve with one end, for $q = 0$, at $(1, 0, \dots, 0)^T$, and the other, for $q = 1$, at the unique positive equilibrium (either \vec{y}_1^∞ , if $s(L_1) > s(L_2)$ or \vec{y}_2^∞ otherwise). To determine which is the dominant genotype for $q \approx 1$, we need to find which one of the coordinates $x_i^\infty(q)$ is the largest.

Let $q \approx 1$ be sufficiently close to 1. In view of the above analysis, if $s(L_1)$ is the largest spectral abscissa, then the equilibrium $\vec{x}^\infty(q) = (x_1^\infty(q), x_2^\infty(q), x_3^\infty(q), x_4^\infty(q))$ is such that $x_2^\infty(q) \approx 0$ and $x_3^\infty(q) \approx 0$ and either $x_1^\infty(q)$ (the frequency of the wild type) or $x_4^\infty(q)$ (the frequency of the wild type's "mirror" sequence) is the dominant species. Alternatively, if $s(L_2)$ is the largest spectral abscissa, then $x_1^\infty(q) \approx 0$ and $x_4^\infty(q) \approx 0$ and either $x_2^\infty(q)$ or $x_3^\infty(q)$ (the frequency of its "mirror" sequence) is the dominant species.

Thus, for $q \approx 1$, there is always a pair of dominating sequences which mirror each other while the remaining sequences are present at very low frequencies.

Additional conditions are needed to determine the dominant type for values of the mutation rate $q \approx 1$. These conditions are derived using the same analysis as in the case $N = 2$ above. Using proposition 3.2, we obtain the following result.

Proposition 3.5. *For the case $N = 4$, if the following conditions are fulfilled:*

- a) $A_i > 0, D_i \geq 0, i = 1, \dots, N$
- b) $F_1 = A_1 - D_1 > A_4 - D_4 = F_4 > 0,$
- c) $s(L_1) = -\frac{1}{2}(D_1 + D_4) + \frac{1}{2}\sqrt{(D_1 - D_4)^2 + 4A_1A_4} > s(L_2) = -\frac{1}{2}(D_2 + D_3) + \frac{1}{2}\sqrt{(D_2 - D_3)^2 + 4A_2A_3}$
- d) $A_1 - A_4 \leq 0$ or $A_1 - A_4 > 0$ and $A_1 + D_1 < A_4 + D_4,$

then the wild type is dominant for high mutation rate $q \approx 1$.

If a), b) and c) are true, but instead of d) the following is true: $A_1 - A_4 > 0$ and $A_1 + D_1 > A_4 + D_4$, the mirror sequence of the wild type is dominant for high mutation rates (sufficiently close to 1).

If the opposite inequality in c) is fulfilled, neither the wild type nor its mirror are dominant for high mutation rates ($q \approx 1$); the other two sequences dominate.

Note. If $D_i = D$ as assumed in [38, 44, 15], conditions a)-d) cannot hold together, as b) and d) will be contradictory.

3.5 General Case $N > 2$

Let us now consider the general case with the assumptions (3.1). Having the experience from the case $N = 4$, let us arrange the quasispecies sequences in the following way. The index of the wild type is 1 and the index of its mirror is 2. Next, take any sequence different from these two and give it an index 3 and give its mirror sequence index 4, and so on. Continuing in this way, we arrange all sequences.

It is then evident that the system defining an equilibrium \vec{x}^∞ for $q = 1$ can be brought in the block-diagonal form

$$L\vec{x}^\infty = \begin{pmatrix} L_1 - f(\vec{x}^\infty)I_2 & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & L_2 - f(\vec{x}^\infty)I_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & L_3 - f(\vec{x}^\infty)I_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \dots & L_{N/2} - f(\vec{x}^\infty)I_2 \end{pmatrix} \vec{x}^\infty = \vec{0}, \quad (3.13)$$

where

$$L_i = \begin{pmatrix} -D_{2i-1} & A_{2i} \\ A_{2i-1} & -D_{2i} \end{pmatrix}, \quad (3.14)$$

$\mathbf{0}$ is a 2-dimensional square zero matrix, I_2 is a 2-dimensional identity matrix and $\vec{0}$ is N-dimensional vector with zero entries.

Let k_0 be an index for which L_{k_0} has the largest spectral abscissa. We assume that

$$s(L_{k_0}) = \max_{i=1, \dots, N/2} s(L_i) > s(L_j), j \neq k_0. \quad (3.15)$$

Then (3.13) has a nonnegative equilibrium $\vec{\zeta}^\infty$ such that $\zeta_i^\infty = 0$, for $i \neq 2k_0 - 1$ or $2k_0$ and $\zeta_{2k_0-1}^\infty, \zeta_{2k_0}^\infty$ solve the system

$$L_{k_0} \begin{pmatrix} \zeta_{2k_0-1}^\infty \\ \zeta_{2k_0}^\infty \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (3.16)$$

Here $\begin{pmatrix} \zeta_{2k_0-1}^\infty \\ \zeta_{2k_0}^\infty \end{pmatrix}$ is a positive eigenvector, corresponding to $s(L_{k_0})$ and $f(\vec{\zeta}^\infty) = s(L_{k_0})$ (see Appendix A).

In addition, $\zeta_{2k_0-1}^\infty + \zeta_{2k_0}^\infty = 1$ and are both positive.

For mutation rates approaching 1, the dominant genotype is either the one indexed with k_0 or its mirror. The wild type will be dominant at high values of the mutation rate if $k_0 = 1$ and if $A_2 + D_2 > A_1 + D_1$. We state the following (proof in Appendix H).

Proposition 3.6. *Let the sequences comprising the quasispecies be indexed as described in the beginning of this section. If the following conditions are fulfilled*

- a) $A_i > 0, D_i > 0, i = 1, \dots, N$;
- b) $F_1 = A_1 - D_1 > A_2 - D_2 = F_2 > 0$;
- c) $s(L_1) = -\frac{1}{2}(D_1 + D_2) + \frac{1}{2}\sqrt{(D_1 - D_2)^2 + 4A_1A_2} >$
 $-\frac{1}{2}(D_{2i-1} + D_{2i}) + \frac{1}{2}\sqrt{(D_{2i-1} - D_{2i})^2 + 4A_{2i-1}A_{2i}} = s(L_i)$;
- d) $A_1 - A_2 \leq 0$ or $A_1 - A_2 > 0$ and $A_1 + D_1 < A_2 + D_2$,

the wild type will be dominant for high mutation rates.

If a), b) and c) are true, but instead of d) the following is true: $A_1 - A_2 > 0$ and $A_1 + D_1 > A_2 + D_2$, the mirror sequence of the wild type is dominant for high mutation rates.

If c) is not fulfilled, neither the wild type nor its mirror will be dominant for high mutation rates ($q \approx 1$). Then for the index $k_0 > 1$ satisfying (3.15), one of the sequences S_{2k_0-1} or S_{2k_0} will be dominant at high mutation rates.

Biological interpretation. We have found conditions that guarantee that in the quasispecies model any one of the genotype and mirror couples could dominate the "quasispecies" at very high mutation rates. In particular, we found conditions under which the wild type is dominant for both high and low mutation rates. These conditions are expressed in terms of relationships between the replication and degradation rates of the four participating genotypes.

If the degradation rates of all sequences are equal, these conditions cannot hold simultaneously, and the wild type would lose dominance. In addition, if the degradation rates of all sequences are equal and the wild type has the highest replication rate (single-peak fitness landscape) the mirror of the wild type is the dominant type at high mutation rates $q \approx 1$. These two arguments explain the observations in [38, 44, 15].

In the next section we create examples using the conditions of the above proposition to illustrate the theoretical results.

4 Examples

The quasispecies model with mutation probability matrix and parameter values given below was solved using Matlab's routine ode45 on a sufficiently long interval so that at the end of the interval the solution was in the vicinity of the equilibrium point and did not change up to the 16th digit. The value of the "per nucleotide" mutation rate q was varied from 0.01 to 0.99 and the equilibrium concentrations for each q were plotted.

Example 1. In this example, $n = 2, N = 2^2 = 4$ and we chose parameter values so that the conditions a) - d) of Proposition 3.5 are fulfilled. The parameter values are $D_1 = 0.1; D_2 = 1.3; D_3 = 1.5; D_4 = 1.3; A_1 = 2; A_2 = 1.5; A_3 = 1.8; A_4 = 1.4$; and $s(L_1) = 1.0776$, while $s(L_2) = 0.2462$, with L_1 and L_2 defined in (3.10).

Based on Proposition 3.5, it is expected that $x_1^\infty(q) > x_2^\infty(q)$ for $q \approx 1$. The mutation probability matrix has the form (3.6). As seen on Figure 2, the wild type remains dominant for all values of the mutation rate q . Also as expected, the mirror sequence of the wild type becomes second in dominance for sufficiently large q .

Example 2. In this example, $n = 3, N = 2^3 = 8$ and we chose parameter values so that the conditions a)-d) of proposition 3.6 are fulfilled. The parameters are

$D_1 = 0.1; D_2 = 1.4; D_3 = 1; D_4 = 0.8; D_5 = 0.5; D_6 = 1.3; D_7 = 1.1; D_8 = 0.9;$
 $A_1 = 2; A_2 = 1.5; A_3 = 1.2; A_4 = 1.3; A_5 = 0.9; A_6 = 1.5; A_7 = 1.2; A_8 = 1.8.$

Again, the dominance of the wild type remains valid for all values of the mutation rate and its mirror emerges as second to dominant for high mutation rates.

Example 3. In this example, $n = 4, N = 2^4 = 16$ and we chose parameter values so that the conditions a)-d) of proposition 3.6 are fulfilled. The parameters are:

$D_1 = 1; D_2 = 1.8; D_3 = 1.7; D_4 = 1.85; D_5 = 1.6; D_6 = 1.85; D_7 = 1.6; D_8 = 1.; D_9 = 1; D_{10} = 1.75; D_{11} = 1.85; D_{12} = 1.6; D_{13} = 1.6; D_{14} = 1.6; D_{15} = 1.75; D_{16} = 0.4;$
 $A_1 = 2; A_2 = 1.9; A_3 = 1.8; A_4 = 1.9; A_5 = 1.7; A_6 = 1.95; A_7 = 1.65; A_8 = 1.8; A_9 = 1.9; A_{10} = 1.8; A_{11} = 1.9; A_{12} = 1.65; A_{13} = 1.7; A_{14} = 1.7; A_{15} = 1.8; A_{16} = 0.5;$

As expected, the wild type is dominant for large values of q , seconded by its mirror. The wild type remains dominant for all values of $q \in (0, 1)$.

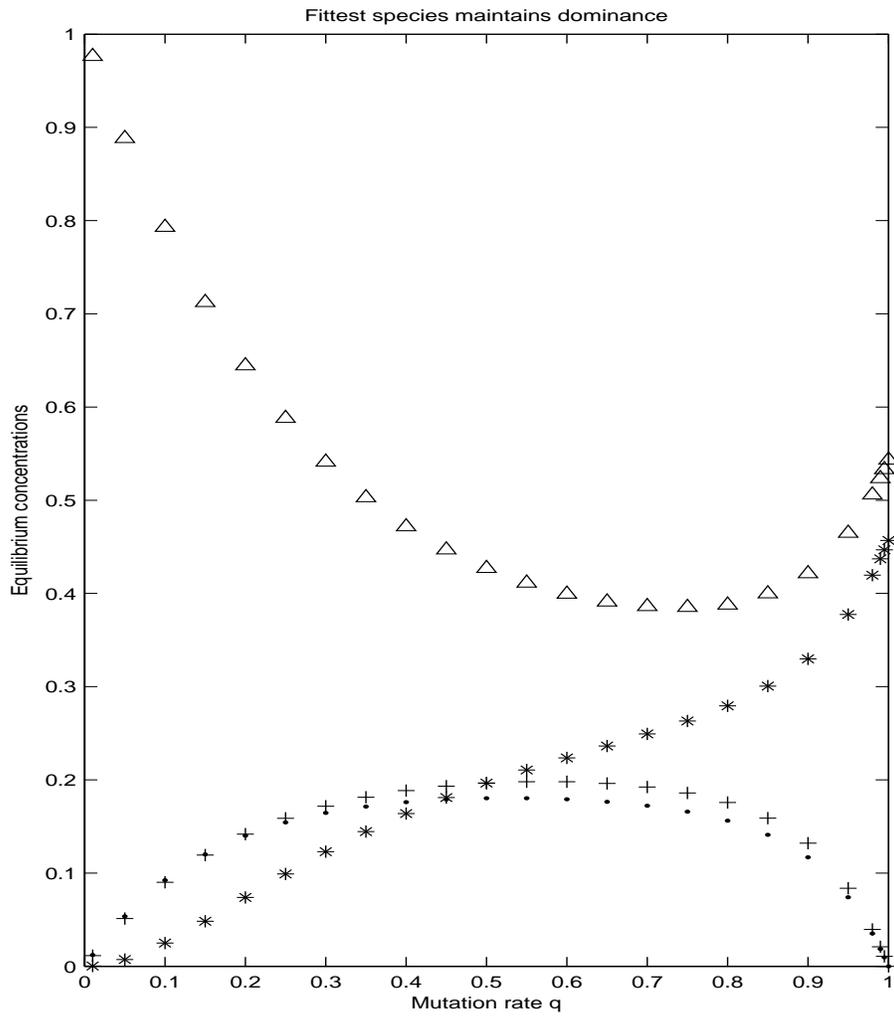


Figure 2: Equilibrium concentrations calculated for Example 1. " Δ " $\rightarrow x_1$, "*" $\rightarrow x_2$, "." $\rightarrow x_3$, "+" $\rightarrow x_4$.

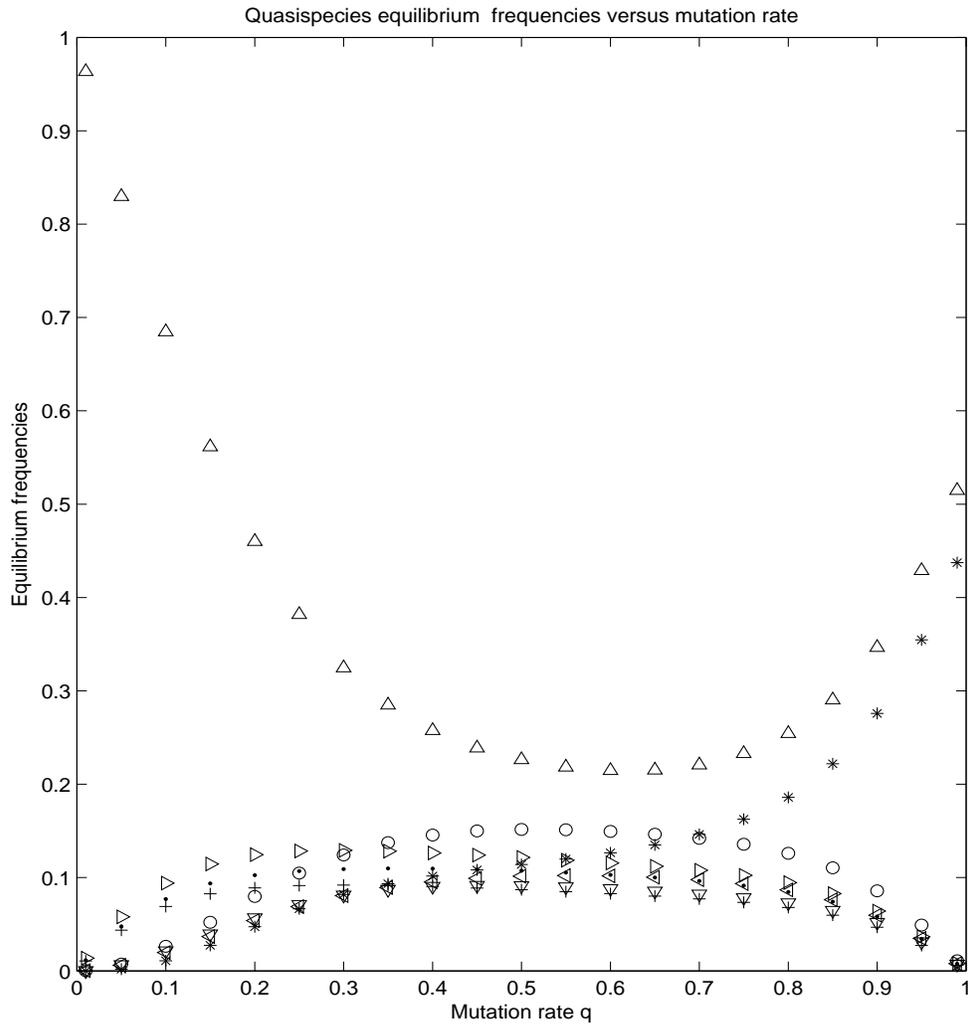


Figure 3: Equilibrium concentrations calculated for for Example 2. "△" → x_1 , "*" → x_8 , "+" → x_2 , "·" → x_3 , "◇" → x_4 , "○" → x_5 , "▽" → x_6 , "◁" → x_7

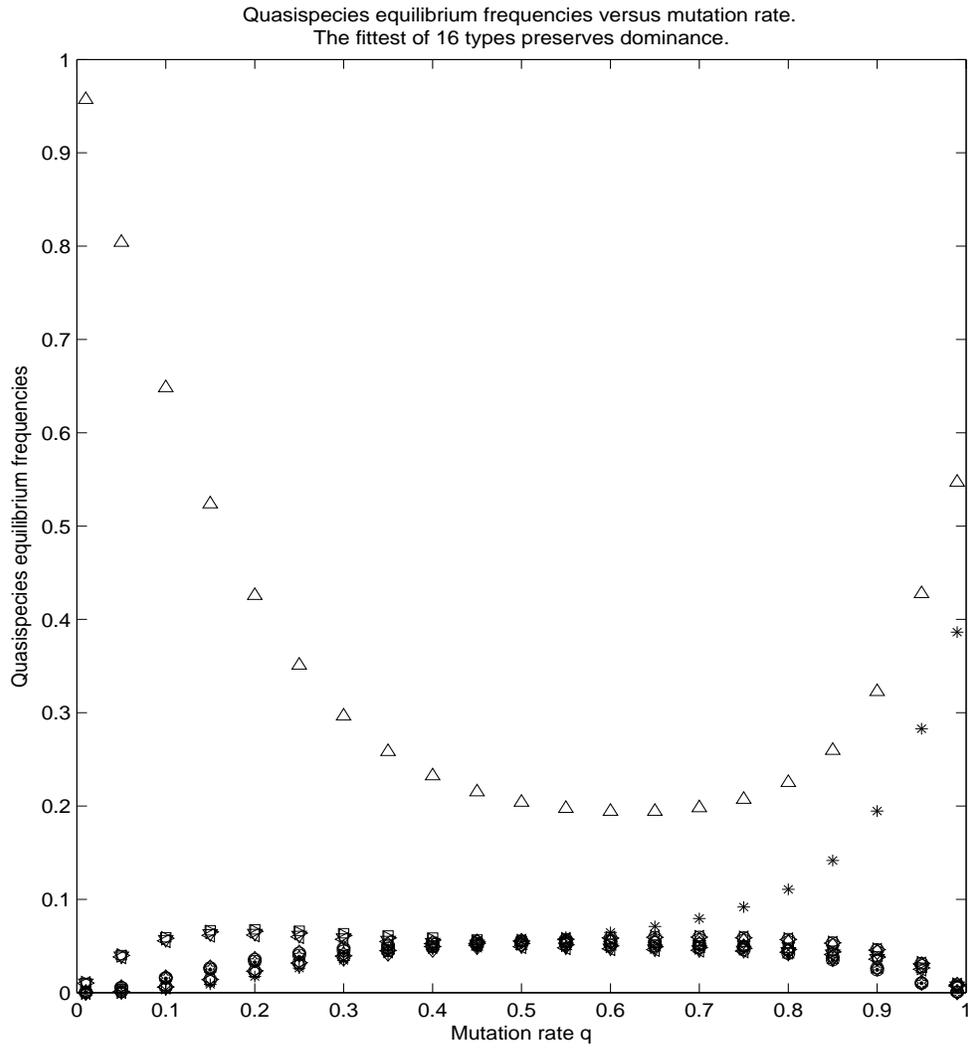


Figure 4: Equilibrium concentrations calculated for Example 3. " \triangle " $\rightarrow x_1$, " $*$ " $\rightarrow x_{16}$, the mirror of the wild type. The wild type remains dominant for all values of q . All other 14 frequencies are represented with various symbols that overlap on the plot.

Example 4. In this example, $n = 4$, $N = 2^4 = 16$ and we chose parameter values so that the condition c) of proposition 3.6 is not fulfilled. The parameters are:

$D_1 = 1; D_2 = 1.8; D_3 = 1.7; D_4 = 1.85; D_5 = 1.6; D_6 = 1.85; D_7 = 1.6; D_8 = 1.; D_9 = 1; D_{10} = 1.75; D_{11} = 1.85; D_{12} = 1.6; D_{13} = 1.6; D_{14} = 1.6; D_{15} = 1.75; D_{16} = 0.01;$
 $A_1 = 2; A_2 = 1.9; A_3 = 1.8; A_4 = 1.9; A_5 = 1.7; A_6 = 1.95; A_7 = 1.65; A_8 = 1.8; A_9 = 1.9; A_{10} = 1.8; A_{11} = 1.9; A_{12} = 1.65; A_{13} = 1.7; A_{14} = 1.7; A_{15} = 1.8; A_{16} = 0.05;$

In addition, $s(L_1) = 0.0824$ while $s(L_8)$ is the largest of all $s(L_i)$ and equal to 0.8493. As expected from Proposition 3.6, S_{14} is the dominant sequence for $q \approx 1$, seconded by its mirror sequence S_{15} (Figure 5). The wild type (depicted by triangle symbols) loses dominance for $q \approx 0.15$ and never regains it.

5 Discussion

As discussed in the Introduction, Eigen's model does not take into account the complex nature of natural viral replication which involves viral genome-encoded species-specific replication factors (such non-structural proteins and non-coding regions) that might enable preservation of the quasispecies and render impossible phenomena such as mutation-induced error catastrophe. However, even a simplified model, which does not capture the full complexity of a biological mechanism, may yet capture important salient features and, therefore, be worthy of further investigation.

In this paper we analyzed Eigen's quasispecies model, focusing on the effects of mutation rate on genotype dominance. Rigorous mathematical analysis is possible in the extremes at which the mutation rate is very small (near zero) or very large (close to one). At the first extreme, as expected, Eigen's model predicts that the fittest genotype (the wild type) - the one with the highest excess productivity - will dominate (with highest concentrations) in the quasispecies population. In the other extreme, we have proved that, depending on whether certain relations among the replication and degradation rates are fulfilled, any one of the genotypes could become the dominant type in the quasispecies. For values of the mutation rate between the extremes, two types of behavior are possible - either the wild type remains dominant or it loses dominance. Dominance can be exchanged among various genotypes for intermediate values of the mutation rate but eventually, one dominant genotype will emerge and it remain dominant as the mutation rate increases and approaches $q = 1$. We have derived formulae that enable exact prediction which genotype will emerge as dominant at high mutation rates. We then demonstrated on numerical examples the validity of the predictions. We used the developed theory to construct computational examples where the wild type remains dominant and where a new dominant type emerges.

Importantly, we have not restricted our analysis to the single-peaked fitness landscape, nor to the case, previously considered in simulations, with uniformly equal degradation rates. We considered and provided proofs for the general case with arbitrary values of the replication and degradation rates.

We note that the same genotype can have different fitnesses depending on certain conditions, e.g. for varying mutation rates. We introduced the concept of *nascent fitness*, i.e., the excess production rate of a replicating genotype when the replication occurs with perfect fidelity. Doing so, we emphasized that the fitness of a genotype is a relative concept, dependent not only on its replication and degradation rates but also on other factors, one of which, in this model, is the fidelity of replication. When the mutation rate increases, thus decreasing the fidelity of replication, the fitness of the genotype decreases.

From the mathematical perspective, we asked the question: given the equation for the equilibrium of a system, can we characterize the largest coordinate of the equilibrium? The main challenge here was to determine the stability of the possible equilibria in the extreme case when $q = 1$. This was accomplished by applying results from the theory of quasipositive matrices. We proved that in this case only one of the numerous equilibria is locally stable and that it has a continuation when q becomes less than 1, which is, therefore, the unique globally stable equilibrium of Eigen's model. This result allowed us to characterize the largest equilibrium coordinate when $q \approx 1$. For intermediate values of q this approach does not work, of course. This is why we contended to examine this case numerically and provide examples where the largest eigenvalue coordinate remains such for all values of the parameter q .

Further theoretical development in this area will require analysis of modifications of Eigen's model that relax the assumption of mass conservation and consider the case when non-viable genotypes are defined and have a zero replication rate. Another important question to consider is what artefacts may be introduced

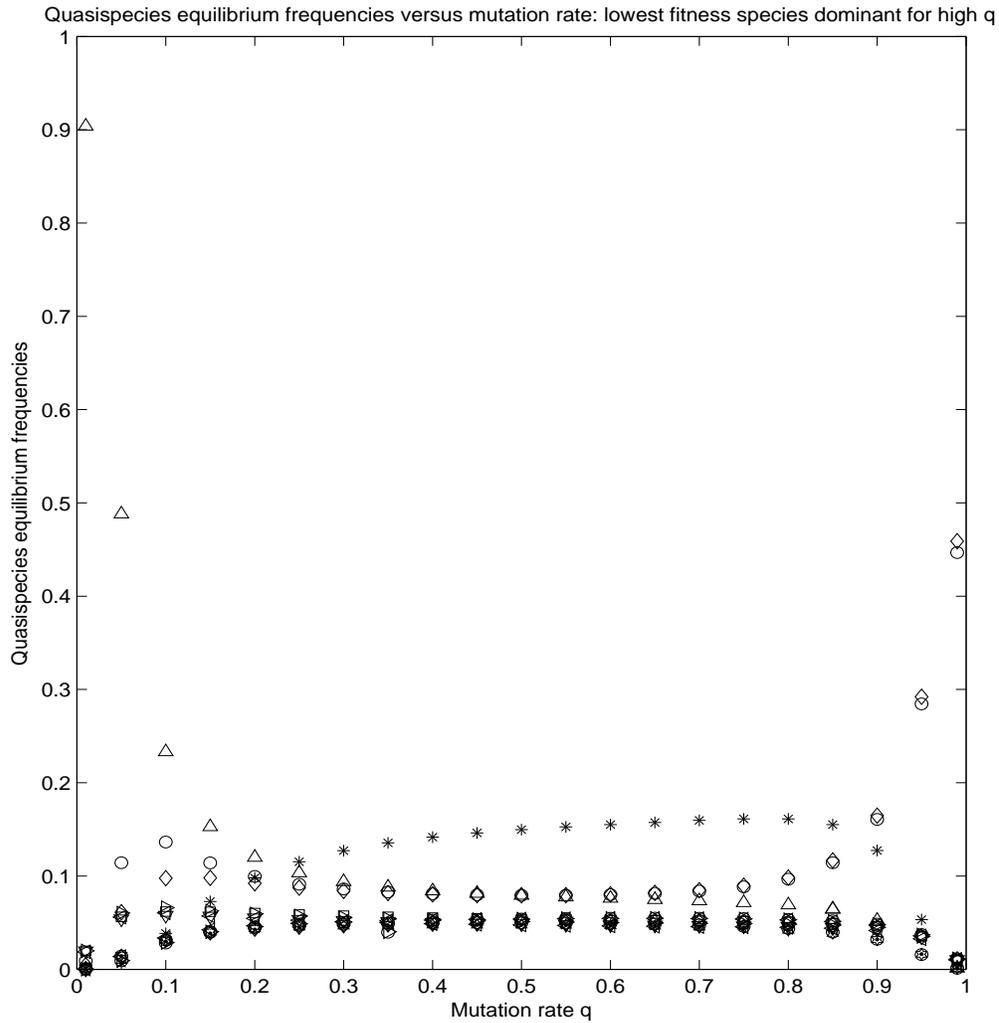


Figure 5: Equilibrium concentrations calculated for Example 4. "△" $\rightarrow x_1$, "*" $\rightarrow x_{16}$, the mirror of the wild type, "○" $\rightarrow x_3$, "◇" $\rightarrow x_{14}$. The wild type loses dominance for all values of q . All other 14 frequencies are represented with various symbols that overlap on the plot.

in constructing a model using four nucleotides, rather than the simplistic binary sequences used by all quasispecies model simulations to date. Including more realistic complexity in a quasispecies model would necessarily mean that the replication rates be functions of the concentrations, rather than constants. To give an example, the rate of viral RNA transcription depends on the availability of RNA dependent RNA polymerase, and the quantity of the latter is a function of available viral RNA templates. Thus, in the early stages of host cell infection, transcription rate would be slow and would accelerate proportionally to the growing number of available RNA templates. A viral replication rates that is calculated as a nonlinear function, possibly involving all viral sequences, for example of the form $A_i \sum_j \alpha_j x_j x_i$. Similar arguments can be put forward regarding the degradation rate as well. We plan to consider and analyse such modifications and complexity in further work.

6 Acknowledgments

This work was performed partially under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The work was supported by an LLNL-LLNS internally funded grant to T.K. and C.Z. through the Laboratory Directed Research and Development program. T.K. 's research was also funded by NSF. Any opinion, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

The authors express gratitude to Professor Raul Andino (UCSF) for many insightful discussions of the quasispecies phenomenon, which the helped in understanding the limitations of the quasispecies model.

Appendix A. Properties of quasipositive matrices.

(Spectral property) If A is an irreducible quasipositive matrix, it has a simple eigenvalue $s(A)$ such that $s(A) > \Re \lambda_i(A)$ for all other eigenvalues λ_i of A , and a positive eigenvector $\vec{w}(A)$ corresponding to $s(A)$ and any positive eigenvector of A is a multiple of \vec{w} . $s(A)$ is called the spectral abscissa of A .

(Majorization property) In addition, if A and B are quasipositive matrices and $A < B$, and $A + B$ is irreducible, then $s(A) < s(B)$.

These properties are easy corollaries of similar ones for nonnegative matrices derived in [3].

Appendix B. Existence of unique positive equilibrium of (2.2)

Any equilibrium of (2.2) satisfies the equation

$$[QA - D - f(\vec{x})I]\vec{x} = 0 \quad (6.1)$$

where $Q = (Q_{ij})$ is the mutation matrix, A is a diagonal matrix with elements A_i and D is a diagonal matrix with elements D_i . Then obviously, any equilibrium \vec{x} is an eigenvector of $QA - D$ and the corresponding eigenvalue is $f(\vec{x})$.

A substantial assumption is that the mutation matrix Q is irreducible and that $A_i > 0$. (For example, if Q is defined as in (2.3), and $0 < q < 1$, it is positive and therefore, irreducible.) Therefore, the matrix $QA - D$ is an irreducible quasipositive matrix. We now use the Spectral property of quasipositive matrices (Appendix A) to conclude that $QA - D$ has a unique positive eigenvector \vec{v} corresponding to the spectral radius $s(QA - D)$ and any other positive eigenvector is its multiple by a scalar. Because \vec{x}^∞ is positive by assumption, then $\vec{x}^\infty = \sigma \vec{v}$ and $f(\vec{x}^\infty) = s(QA - D)$. The scalar σ is obtained as follows: $f(\vec{x}^\infty) = \sigma \sum_i (A_i - D_i)v_i = s(QA - D)$, i.e. $\sigma = \frac{s(QA - D)}{\sum_i (A_i - D_i)v_i}$. Then, $\vec{x}^\infty = \frac{s(QA - D)}{\sum_i (A_i - D_i)v_i} \vec{v}$ is the unique positive solution of (6.1).

If Q is reducible, the quasispecies consists of at least two "subspecies", which cannot mutate into each other. In this case each subspecies has its own equilibrium and the dynamics of the system develops in two subspaces that do not intersect. We will not consider this case here.

Appendix C. The case $q = 0$.

The equilibria of (3.2) are eigenvectors of $F = \text{diag}(F_i)$ corresponding to eigenvalues $\lambda_i = F_i$. If all F_i are different, there are exactly N equilibria - the unit vectors $(1, 0, \dots, 0)$, $(0, 1, \dots, 0)$, \dots , $(0, 0, \dots, 1)$. If some of the nascent fitnesses are equal (for example $F_2 = F_3$), apart from them, the system also has other, infinitely many equilibria (in this case, all vectors satisfying $x_1 = 0, x_2 + x_3 = 1, x_i = 0, i = 4, \dots, N$). However, all equilibria, with the exception of $(1, 0, \dots, 0)$ are unstable because the Jacobian at each such equilibrium always has the eigenvalue $F_1 - \sum_{i=2}^N F_i x_i > 0$. We show next that the equilibrium $(1, 0, \dots, 0)$ is globally asymptotically stable.

System (3.2) can be solved to give $x_i(t) = x_i(0)e^{F_i t - \int_0^t f(\vec{x}(s)) ds}$. Using this equality, and the following one

$$\frac{d}{dt} e^{\int_0^t f ds} = \sum_i F_i x_i(0) e^{F_i t},$$

we solve for f :

$$f(\vec{x}(t)) = \frac{\sum F_i x_i(0) e^{F_i t}}{\sum F_i x_i(0) \int_0^t e^{F_i s} ds}$$

and take the limit at $t \rightarrow \infty$ to obtain

$$\lim_{t \rightarrow \infty} f(\vec{x}(t)) = F_1.$$

We see from (3.2) that $x_i(t) \rightarrow 0$ for all $i \neq 1$ and because of conservation of mass, it follows that $\lim_{t \rightarrow \infty} x_1 \rightarrow 1$.

Appendix D. For $q > 0$ and close to 0, $\vec{x}^\infty(q)$ is a continuation of $(1, 0, \dots, 0)$.

The existence of a continuation is easily established by the Implicit Function Theorem having in mind that $(1, 0, \dots, 0)$ is asymptotically stable and thus, the Jacobian is non-singular. However, it is important to verify that $x_i^\infty(q) \in (0, 1)$. This is easy to see by finding the derivatives of x_i^∞ at $q = 0$. Specifically, we find that

$$\frac{dx_1^\infty}{dq}(0) = \frac{1}{F_1} \left(-NA_1 - A_1 \sum_{j:d(1,j)=1} \frac{F_j}{F_1 - F_j} \right) < 0. \quad (6.2)$$

If $d(1, j) = 1$, i.e. the sequence with index j is only one mutation away from the wild type, we find that

$$\frac{dx_j^\infty}{dq}(0) = \frac{A_1}{F_1 - F_j} > 0 \quad (6.3)$$

Similarly, for the sequences that are 2 mutations away from the wild type, i.e. $d(1, j) = 2$, we calculate

$$\frac{d^2 x_j^\infty}{dq^2}(0) = \frac{2A_1}{F_1 - F_j} > 0 \quad (6.4)$$

while $\frac{dx_j^\infty}{dq}(0) = 0$, and in general, if $d(1, j) = k, k = 1, \dots, N$:

$$\frac{d^k x_j^\infty}{dq^k}(0) = k! \frac{A_1}{F_1 - F_j} > 0 \quad (6.5)$$

while all lower derivatives are equal to 0.

Using Taylor expansion, we see that for small values of q , x_1^∞ decreases when q increases ($\frac{dx_1}{dq}(0) < 0$) and for $j > 1$, $x_j^\infty(q)$ are increasing functions of q : $x_j^\infty \approx \frac{d^k x_j^\infty}{dq^k}(0) q^k$, where $k = d(1, j)$. Therefore, for small $q > 0$ the continuation $\vec{x}^\infty(q)$ of $\vec{x}^\infty(0)$ is positive and coincides with the unique positive equilibrium of (3.2).

Obviously, as x_j^∞ are continuous, it follows that in some interval $q \in (0, q_{min})$, $x_1 > x_i, i = 2, \dots, N$. Thus, we have proved Proposition 3.1.

Appendix E. Proof of Proposition 3.2.

Proof. Really, the equation for the equilibrium can be written as

$$(L - f(\vec{x}^\infty)I)\vec{x}^\infty = 0 \quad (6.6)$$

where

$$L = \begin{pmatrix} A_1(1-q) - D_1 & A_2q \\ A_1q & A_2(1-q) - D_2 \end{pmatrix}. \quad (6.7)$$

As $\vec{x}^\infty > 0$ and because L is quasipositive (see Appendix A) it follows that $\vec{x}^\infty > 0$ is the unique eigenvector corresponding to $s(L)$ and $s(L) = f(\vec{x}^\infty)$.

$$s(L) = \frac{1}{2}(S_1(q) + S_2(q)) + \sqrt{\left(\frac{S_1(q) - S_2(q)}{2}\right)^2 + A_1 A_2 q^2} \quad (6.8)$$

where $S_i(q) = A_i(1-q) - D_i$.

Therefore $f(\vec{x}^\infty) = S_1(0)x_1^\infty + S_2(0)x_2^\infty = s(L)$ and $x_1^\infty + x_2^\infty = 1$ from where

$$x_1^\infty = \frac{s(L) - S_2(0)}{S_1(0) - S_2(0)}. \quad (6.9)$$

Let us now find conditions (if any) for which $x_1^\infty > x_2^\infty$, i.e. $x_1 > \frac{1}{2}$, for all $q \in (0, 1)$. Using (6.9) and (6.8) after simple but tedious algebraic calculations we obtain

$$x_1 > \frac{1}{2} \iff [(A_1 - A_2) - (D_1 - D_2)][(A_1 - A_2) - (D_1 - D_2) - 2q(A_1 - A_2)] > 0. \quad (6.10)$$

Because of (3.3), the latter inequality is possible for all $q \in (0, 1)$ only if $(A_1 - A_2) - (D_1 - D_2) - 2q(A_1 - A_2) > 0$.

(a) If $A_1 - A_2 \leq 0$, the latter is always true.

(b) If $A_1 - A_2 > 0$, the inequality is possible only if for all $q \in (0, 1)$, $\frac{D_1 - D_2}{(A_1 - A_2)} < 1 - 2q$, which is valid if and only if $\frac{D_1 - D_2}{(A_1 - A_2)} < -1$ which is equivalent to $A_1 - A_2 > 0$ and $A_1 + D_1 < A_2 + D_2$.

Therefore, the wild type would remain dominant for all values of $q \in (0, 1)$ if and only if $A_1 - A_2 \leq 0$ or $A_1 - A_2 > 0$ and $A_1 + D_1 \leq A_2 + D_2$. \square

Appendix F. Proof of Proposition 3.3.

Proof. We first note that any equilibrium \vec{x} is an eigenvector of $QA - D$ corresponding to the eigenvalue $f(\vec{x})$ and, for $q \in (0, 1)$, since $QA - D$ is irreducible, the unique *positive* equilibrium \vec{x}^∞ is the eigenvector corresponding to $f(\vec{x}^\infty) = s(QA - D)$.

As L_1 and L_2 are irreducible and quasipositive, they have unique positive eigenvectors $\vec{v}_1 = (\chi_1^\infty, \chi_4^\infty)^T$, $\vec{v}_2 = (\chi_2^\infty, \chi_3^\infty)^T$ corresponding to $s(L_1)$ and $s(L_2)$. Because of the block - diagonal structure of P it has only two nonnegative eigenvectors (up to a scalar multiplication):

$$\vec{y}_1^\infty = (\chi_1^\infty, \chi_4^\infty, 0, 0)^T \text{ and } \vec{y}_2^\infty = (0, 0, \chi_2^\infty, \chi_3^\infty)^T. \quad (6.11)$$

Note that $\sum_i \chi_i^\infty = 1$. Therefore $\chi_i < 1$ which implies

$$\frac{df}{dx_i} \chi_i^\infty = F_i \chi_i^\infty < A_i. \quad (6.12)$$

The latter inequality is used in the proposition below.

Stability of y_1^∞ . The Jacobian at \vec{y}_1^∞ is equivalent to a block - diagonal matrix $J(\vec{y}_1^\infty) = \text{diag}(J_1, J_2)$ where

$$J_1 = \begin{pmatrix} -D_1 - \frac{\partial f}{\partial y_1} \chi_1^\infty - f(\vec{y}_1^\infty) & A_4 - \frac{df}{dx_4}(\vec{y}_1^\infty) \chi_4^\infty \\ A_1 - \frac{df}{dx_1}(\vec{y}_1^\infty) \chi_1^\infty & -D_4 - \frac{\partial f}{\partial x_4} \chi_4^\infty - f(\vec{y}_1^\infty) \end{pmatrix} < L_1 - f(\vec{y}_1^\infty) I_2. \quad (6.13)$$

and

$$J_2 = \begin{pmatrix} -D_2 - f(\vec{y}_1^\infty) & A_3 \\ A_2 & -D_3 - f(\vec{y}_1^\infty) \end{pmatrix}. \quad (6.14)$$

As J_1 and L_1 are both irreducible quasipositive matrices, using the Majorization property (Appendix A), we obtain:

$$s(J_1) < s(L_1 - f(\vec{y}_1^\infty) I_4) = 0. \quad (6.15)$$

Also, $s(J_2) = s(L_2) - f(\vec{y}_1^\infty) = s(L_2) - s(L_1) < 0$. Therefore, y_1^∞ is an asymptotically stable equilibrium.

Instability of y_2^∞ . Similarly, the Jacobian at \vec{y}_2 is equivalent to a block - diagonal matrix $J(\vec{y}_2) = \text{diag}(G_1, G_2)$ where

$$G_1 = \begin{pmatrix} -D_1 - f(\vec{y}_2^\infty) & A_4 \\ A_1 & -D_4 - f(\vec{y}_2^\infty) \end{pmatrix} \quad (6.16)$$

Obviously, $s(G_1) = s(L_1) - f(\vec{y}_2^\infty) = s(G_1) - s(G_2) > 0$. Therefore, \vec{y}_2^∞ is an unstable equilibrium.

Vice versa, if $s(L_2) > s(L_1)$, y_2^∞ is stable and y_1^∞ is unstable. \square

Appendix G. Proof of proposition 3.4.

Proof. The existence of $\vec{x}^\infty(q)$ is guaranteed by the Implicit Function Theorem because the eigenvalues of Jacobian at \vec{y}_1^∞ lie in the left half of the complex plane; thus $s(P) < 0$ and the Jacobian is nonsingular. We need only show that $x_i^\infty(q) \in (0, 1)$. Because $1 > \chi_1^\infty > 0, 1 > \chi_4^\infty > 0$, it follows that $x_1^\infty(q) \in (0, 1), x_1^\infty(q) \in (0, 1)$ for q in the vicinity of $q = 1$. Therefore, we need only show that $\frac{dx_i^\infty}{dq}(1) < 0, i = 2, 3$ (i.e., when q becomes smaller than 1, x_i^∞ would increase from 0 and be positive).

Differentiating the right hand side of (2.2) with respect to q , we obtain:

$$\sum_j A_j \frac{dQ_{ij}}{dq} x_j^\infty + \sum_j (A_j Q_{ij} - F_j) \frac{dx_j^\infty}{dq} - (D_i + f(\vec{x}^\infty)) \frac{dx_i^\infty}{dq} - \sum_j F_j \frac{dx_j^\infty}{dq}. \quad (6.17)$$

We next note that

$$\begin{aligned} Q_{ij} &= 0, \text{ if } d(i, j) \neq n; \\ Q_{ij} &= 1, \text{ if } d(i, j) = n, \end{aligned} \quad (6.18)$$

and

$$\begin{aligned} \frac{dQ_{ij}}{dq} &= n, \text{ if } d(i, j) = n \\ &- 1, \text{ if } d(i, j) = n - 1 \\ &0, \text{ otherwise,} \end{aligned} \quad (6.19)$$

and also

$$\begin{aligned} x_1^\infty, x_N^\infty &\neq 0 \\ x_j^\infty &= 0, \text{ otherwise.} \end{aligned} \quad (6.20)$$

Using (6.18,6.19,6.20) we get the following equations for $\frac{dx_i^\infty}{dq}(1)$:

$$(L_2 - f(\vec{y}_1^\infty)I_2) \left(\frac{dx_2^\infty}{dq}(1), \frac{dx_3^\infty}{dq}(1) \right)^T = (-A_1 \frac{dQ_{21}}{dq} \chi_1^\infty, -A_4 \frac{dQ_{24}}{dq} \chi_4^\infty)^T = (A_1 \chi_1^\infty, A_4 \chi_4^\infty)^T \gg 0. \quad (6.21)$$

We next note that the matrix $-(L_2 - f(\vec{y}_1^\infty)I_2)$ is an M-matrix and thus, its inverse is positive. It follows that

$$\left(\frac{dx_2^\infty}{dq}(1), \frac{dx_3^\infty}{dq}(1) \right)^T = -(L_2 - f(\vec{y}_1^\infty)I_2)^{-1} (A_1 \chi_1^\infty, A_4 \chi_4^\infty)^T < 0. \quad (6.22)$$

This concludes the proof of the proposition. \square

Appendix H. Proof of Proposition 3.6.

The Jacobian at $\vec{\zeta}$ is a block diagonal matrix $J(\vec{\zeta}) = \text{diag}(J_1, \dots, J_{N/2})$ where

$$J_{k_0} = \begin{pmatrix} -D_{2k_0-1} - F_{2k_0-1} \zeta_{2k_0-1}^\infty - f(\vec{\zeta}^\infty) & A_{2k_0} - F_{2k_0} \zeta_{2k_0}^\infty \\ A_{2k_0-1} - F_{2k_0-1} \zeta_{2k_0-1}^\infty & -D_{2k_0} - f(\vec{\zeta}^\infty) - F_{2k_0} \zeta_{2k_0}^\infty \end{pmatrix}. \quad (6.23)$$

and

$$J_i = \begin{pmatrix} -D_{2i-1} - f(\vec{\zeta}^\infty) & A_{2i} \\ A_{2i-1} & -D_{2i} - f(\vec{\zeta}^\infty) \end{pmatrix} \text{ for } i \neq k_0. \quad (6.24)$$

Using the same arguments as for the case $N = 4$ above, we find that ζ^∞ is asymptotically stable and has a continuation $\vec{x}^\infty(q)$ for $q < 1$. The proof that the continuation really satisfies the condition $x_i^\infty \in (0, 1), i = 1, \dots, N$ is more complicated than in the case $N = 4$. This is because if $K = d(k_0, j) > 1$, the first $K - 1$ derivatives of ζ_j^∞ with respect to q are zero at $q = 1$. In these cases we can take the K -th derivative at $q = 1$ and show that it is positive.

Thus, for high mutation rates, close to 1, the dominant genotype is either the one indexed with k_0 or its mirror. The wild type will be dominant at high values of the mutation rate if $k_0 = 1$ and if $A_2 + D_2 > A_1 + D_1$.

References

- [1] Alves, K., Canzian, M., Delwart, E. L. 2002. HIV type 1 envelope quasispecies in the thymus and lymph nodes of AIDS patients. *AIDS Res Hum Retroviruses* 18, 161-165.
- [2] Bascompte, J., and F. Rodriguez-Trelles, Eradication thresholds in epidemiology, conservation biology and genetics. *Journal of Theoretical Biology* 192, no. 4: 415-18
- [3] A. Berman, R.J. Plemmons, 1970. *Nonnegative matrices in the mathematical sciences*, Academic Press, New York, .
- [4] Bull, J.J., L.A. Meyers, M. Lachmann, 2005. Quasispecies made simple. *PLoS Comp Bio* 1: 450-460
- [5] Cases-Gonzalez C, Arribas M, Domingo E, Lazaro E., 2008. Beneficial effects of population bottlenecks in an RNA virus evolving at increased error rate. *J Mol Biol.*, Dec 31;384(5):1120-9. Epub 2008 Oct 14.
- [6] Crotty S., Cameron CE, Andino R, 2001. RNA virus error catastrophe: direct molecular test by using ribavirin, *PNAS* 98, 12, 6895-6900
- [7] Domingo E, Menendez-Arias L, Quinones-Mateu ME, Holguin A, Gutierrez-Rivas M, Martinez MA, Quer J, Novella IS, and Holland JJ, 1997. Viral quasispecies and the problem of vaccine-escape and drug-resistant mutants. *Prog Drug Res* 48:99 128.
- [8] Domingo, EC Biebricher, M. Eigen, and J. J. Holland, 2001. *Quasispecies and RNA virus evolution: principles and consequences*. Eureka.com/Landes Bioscience, Georgetown, Texas
- [9] Domingo E, Escarmis C, Lazaro E, Manrubia SC., 2005. Quasispecies dynamics and RNA virus extinction, *Virus Res.* 107(2):129-39.
- [10] Domingo E, 2007. *Virus Evolution*, In: *Fields Virology, Fifth Edition*, Edited by DM Knipe, PM Howley, DE Griffin, et al., Philadelphia, Lippincott Williams and Wilkins
- [11] Drake JW, 1993. Rates of spontaneous mutation among RNA viruses, *PNAS* 90:4171-4175
- [12] Drake JW, Holland JJ, 1999 Mutation rates among RNA viruses, *PNAS* , 96, 24, 13910-13913
- [13] Eigen M, 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465-523
- [14] Eigen M., Schuster P, 1977. A principle of natural self-organization, *Die Naturwissenschaften*, 64, 541-565
- [15] Eigen, M., J. McCaskill, and P. Schuster, 1989. The molecular quasispecies, *Adv. Chem. Phys.* 75 149-263
- [16] Eigen M, 2000. Natural selection: a phase transition?, *Biophysical chemistry*, 85 101-123
- [17] Fahrig, L. Effect of habitat fragmentation on the extinction threshold: a synthesis. *Ecological Applications* 12:346-353, 2002
- [18] Farci P, Shimoda A, Coiana A et al., 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies, *Science* 288, 339 - 344.
- [19] Ghosh Z., Bibekanand M., Jayprokas C., 2009. Cellular versus viral microRNAs in host-virus interaction, *Nucleic Acids Research*, 37, 4 1035-1048
- [20] Holmes EC, 2009. *The evolution and emergence of RNA viruses*, Oxford Series in Ecology and Evolution, Oxford Press, New York
- [21] Iranzo J, Manrubia SC, 2009. Stochastic extinction of viral infectivity through the action of defectors, *Europhysics Letters* 85, 1, 18001 (p1-p6)

- [22] Jerzak G, Bernard KA, Kramer LD, Ebel GD, 2005. Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection *Journal of General Virology*, 86, 2175-2183
- [23] Jonsson CB, Miligan BC, Arterburn JB, 2005. Potential importance of error catastrophe to the development of antiviral strategies for hantaviruses, *Virus Research* 107, 195-205
- [24] Jones, L. R., Zandomeni, R., Weber, EL, 2002. Quasispecies in the 59 untranslated genomic region of bovine viral diarrhoea virus from a single individual. *J Gen Virol* 83, 2161-2168.
- [25] Kostova, T., Carlsen T., 2005. The effect of small-size habitat disturbances on population density and time to extinction of the prairie vole, *Nonlinear Analysis: Real World Applications*, v. 6, no. 4, 731-746
- [26] Lee CW, Jackwood MW, 2000. Evidence of genetic diversity generated by recombination among avian coronavirus IBV, *Arch Virol.* 145: 2135-2148
- [27] Mas A, Ulloa E, Bruguera M, Furcic I, Garriga D, Fabregas S, Andreu D, Saiz JC, Diez J., 2004. Hepatitis C virus population analysis of a single-source nosocomial outbreak reveals an inverse correlation between viral load and quasispecies complexity, *J Gen Virol.* 85(12):3619-26.
- [28] Ovaskainen O., Hanski I., 2003. Extinction threshold in metapopulation models, *Annales zoologici Fennici*, v. 40, no. 2, p. 81-97
- [29] Pult I, Abbott N, Zhang Y-Y, Summers J, 2001. Frequency of spontaneous mutations in an avian hepadnavirus infection, *J. Virol.*, 75, 20, 9623-9632
- [30] Peery T, Mathews MB, 2007. Viral conquest of the host cell, In: *Fields Virology, Fifth Edition*, Edited by DM Knipe, PM Howley, DE Griffin, et al., Philadelphia, Lippincott Williams and Wilkins
- [31] Pfeiffer J, Kirkegaard K, 2003. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity, *PNAS*, 100, 12, 72897294
- [32] Plyusnin, A., Cheng, Y., Lehtvaslaiho, H., Vaheri, A., 1996. Quasispecies in wild-type Tula hantavirus populations. *J Virol* 70, 9060-9063.
- [33] Schikora BM, Shih LM, Hietala SK, 2003. Genetic diversity of avian infectious bronchitis virus California variants isolated between 1988 and 2001, based on the S1 subunit of the spike glycoprotein, *Archives of Virology*, 148, 115-136
- [34] Schneider WL, Roosnick MJ, 2001. Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions, *J. Virol.*, 75, 14, 6566-6571
- [35] Sierra M, Airaksinen A, Gonzalez-Lopez C, Agudo R, Arias, Domingo E, 2006. Foot-and-mouth disease virus mutant with decreased sensitivity to ribavirin: implications for error catastrophe, *J.Virol.* 81, 4, 2012-2024
- [36] Sole RV, Sardanyes J, Diez J, Mas A, 2006. Information catastrophe in RNA viruses through replication thresholds, *Journal of Theoretical Biology* 240: 353-359
- [37] Summers J, Litwin S, Examining the theory of error catastrophe, 2006. *J. Virol*, 80, 1: 20-26
- [38] Swetina J. Schuster P, 1982. Self-replication with errors – a model for polynucleotide replication , *Biophys. Chem.* 16, 329-345.
- [39] Umbach JL, Cullen BR, 2009. The role of RNAi and microRNAs in animal virus replication and antiviral immunity, *Genes & Development* 23:1151-1164
- [40] Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R., 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population, *Nature*, 439,19, 344-348
- [41] Vignuzzi M., Wendt E., Andino R., 2008. Engineering attenuated virus vaccines by controlling replication fidelity, *Nature Medicine*, v. 14, 2: 154-161

- [42] Vignuzzi M, Stone JK, Andino A., 2005. Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications, *Virus Research* 107, 173-181
- [43] Wagner GP, Krall P, 1993. What is the difference between models of error threshold and Muller's ratchet? *J. Math. Biol.* 32:33-44
- [44] Wilke, CO, Ronnewinkel C. and Martinetz T., 2000. Dynamic fitness landscapes in molecular evolution, [arXiv:physics/9912012v2](https://arxiv.org/abs/physics/9912012v2)
- [45] Williams J., Young H., Austin P. , 1975. Complementation of human adenovirus type 5ts mutants by human adenovirus type 12, *J. Virology*, 15, 675-678