

Effect of the SOS Response on the Mean Fitness of Unicellular Populations: A Quasispecies Approach

Amit Kama, Emmanuel Tannenbaum*

Department of Chemistry, Ben-Gurion University of the Negev, Be'er-Sheva, Israel

Abstract

The goal of this paper is to develop a mathematical model that analyzes the selective advantage of the SOS response in unicellular organisms. To this end, this paper develops a quasispecies model that incorporates the SOS response. We consider a unicellular, asexually replicating population of organisms, whose genomes consist of a single, double-stranded DNA molecule, i.e. one chromosome. We assume that repair of post-replication mismatched base-pairs occurs with probability λ , and that the SOS response is triggered when the total number of mismatched base-pairs is at least l_S . We further assume that the per-mismatch SOS elimination rate is characterized by a first-order rate constant κ_{SOS} . For a single fitness peak landscape where the master genome can sustain up to l mismatches and remain viable, this model is analytically solvable in the limit of infinite sequence length. The results, which are confirmed by stochastic simulations, indicate that the SOS response does indeed confer a fitness advantage to a population, provided that it is only activated when DNA damage is so extensive that a cell will die if it does not attempt to repair its DNA.

Citation: Kama A, Tannenbaum E (2010) Effect of the SOS Response on the Mean Fitness of Unicellular Populations: A Quasispecies Approach. PLoS ONE 5(11): e14113. doi:10.1371/journal.pone.0014113

Editor: Wayne M. Getz, University of California, United States of America

Received: April 17, 2008; **Accepted:** October 26, 2010; **Published:** November 30, 2010

Copyright: © 2010 Kama, Tannenbaum. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by an Alon Fellowship, Israel Science Foundation (<http://www.isf.org.il>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: emanuelt@bgu.ac.il

Introduction

Genetic repair is an essential component of cellular genomes. Without mechanisms for repairing damaged and mutated DNA, genomes could not achieve sufficient information content to code for the variety and complexity of modern terrestrial life [1].

Genetic repair mechanisms fall into two main categories: Those that correct base mis-pairings during the replication cycle of a cell, and those that repair mutated and damaged DNA during the growth (G) phase of the cellular life cycle [1].

Two important examples of the first class of repair mechanisms are DNA proofreading and mismatch repair (MMR). DNA proofreading is a repair mechanism that is built into the DNA replicases themselves. During daughter strand synthesis, an erroneously matched base is excised, and a second attempt at a base pairing is made [1]. Mismatch repair also removes erroneous bases from the daughter strand, but does this shortly after daughter strand synthesis [1].

Two important examples of the second class of repair mechanisms are Nucleotide Excision Repair (NER) and the SOS response [1]. NER protects a cell from damage due to radiation, chemical mutagens, and metabolic free radicals by removing damaged portions of the DNA strand and using the other, presumably undamaged strand as a template for re-synthesis of the excised region [1].

The SOS response is a genomic repair mechanism that only activates when there is extensive damage to the cellular genome. When DNA damage is sufficiently extensive, the cell stops growing, and the SOS repair pathways attempt to restore complementarity to the genome [1]. The SOS response only takes effect when DNA damage is so extensive that it may be impossible to use undamaged template strands to correctly re-synthesize damaged portions of the

genome. Thus, although this means that the SOS repair mechanism is highly error prone, it is evolutionary advantageous for the cell to repair the genome and risk fixing deleterious mutations, than it is to leave the damaged genome unrepaired [1–4].

In recent work with quasispecies models of evolutionary dynamics, quasispecies models [5–7] considering the first class of repair mechanisms have been studied [8–11]. In addition, semiconservative replication, including semiconservative replication with imperfect lesion repair (i.e. not all base-pair mismatches are eliminated), has been considered [12–15]. Additional effects, such as multiply-gened genomes, as well as multiply chromosomed genomes, have been considered as well [16,17].

This paper continues the theme of incorporating various details characteristic of cellular genomes by developing a quasispecies model that takes into consideration the SOS repair mechanism. The model is highly simplified, and therefore only a first step in developing proper evolutionary dynamics equations with SOS repair. Nevertheless, because our model is analytically tractable, we believe it is a useful and important initial approach to mathematically modeling the evolutionary aspects of the SOS repair pathway. A proper modeling of the SOS response is an essential component of developing a quantitative theory of mutation-propagation in cellular organisms, which is important for understanding phenomena such as the emergence of antibiotic drug resistance in bacteria, and cancer in multicellular organisms [2–4].

Materials and Methods

Definitions and Model Set-Up

We consider a unicellular population of asexually replicating organisms, whose genomes consist of a single DNA molecule, i.e.

one chromosome. The genome may then be denoted by $\{\sigma, \sigma'\}$, where σ, σ' denote the two strands of the DNA molecule. If the genome is of length L , then we may write $\sigma = b_1 \dots b_L$, $\sigma' = b'_1 \dots b'_L$ where each base b_i, b'_i is chosen from an alphabet of size S (usually $=4$). If \bar{b}_i denotes the base complementary to b_i (for the standard Watson-Crick bases, the pairings are *Adenine(A)–Thymine(T)*, *Guanine(G)–Cytosine(C)*), and $\bar{\sigma}$ denotes the strand complementary to σ , then $\bar{\sigma} = \bar{b}_L \dots \bar{b}_1$. This follows from the antiparallel nature of double-stranded DNA [1].

We let $n_{\{\sigma, \sigma'\}}$ denote the number of organisms with genome $\{\sigma, \sigma'\}$, and we assume that replication occurs with a genome-dependent, first-order rate constant, denoted $\kappa_{\{\sigma, \sigma'\}}$. The set of all $\kappa_{\{\sigma, \sigma'\}}$ defines the *fitness landscape*. It should be emphasized that fitness in this model only refers to replication rate. In particular, this model does not consider cell death.

The semiconservative replication of the DNA genomes happens in three stages:

1. Strand separation, whereby each strand of the chromosome separates to act as a template for daughter strand synthesis.
2. Daughter strand synthesis. We assume a genome and base-independent mismatch probability ε . This error probability ε includes all error correction mechanisms, such as proofreading and mismatch repair, that are active during the replication phase of the cell.
3. Lesion repair, where any post-replication mismatches are removed. Here, there is no longer the parent-daughter strand discrimination that was available during daughter strand synthesis, so in contrast to DNA proofreading and mismatch repair, lesion repair has a 50% chance of removing the mutation, and a 50% chance of communicating it to the parent strand and fixing the mutation in the genome (the lesion repair can occur via either Base or Nucleotide Excision Repair) [1]. We also do not assume that lesion repair is perfectly efficient, so that we consider a genome and base-independent probability λ of removing a mismatch. We call λ the lesion repair efficiency.

In our simplified model, the SOS response is triggered if a given genome has at least l_S mismatches. The replication rate of all cells undergoing SOS repair is zero. We assume that removal of mismatches is catalyzed by an enzyme that binds to a mismatch and then eliminates the mismatch at a rate characterized by a first-order rate constant κ_{SOS} . Therefore, the probability that a given mismatch is eliminated over an infinitesimal time interval dt is given by $\kappa_{SOS} dt$ (see Figure 1).

In this paper, we will consider the behavior of the model in the limit of infinite sequence length. If $\mu \equiv \varepsilon L$ is held constant as $L \rightarrow \infty$, then the probability of an error-free daughter strand synthesis is given by $(1 - \varepsilon)^L \rightarrow e^{-\mu}$. Therefore, fixing μ in the infinite sequence length limit is equivalent to fixing the per-genome replication fidelity. It should be noted that μ is the average number of mismatches produced per DNA strand per replication cycle.

The assumption of infinite sequence length is a common assumption in quasispecies theory, because it is the mathematical formalization of the long genome-length regime that makes the neglect of backmutations exact. While finite genome length effects need to be considered in dynamic fitness landscapes, where adaptation to specific genomes is important [18], for static landscapes (like the one being considered in this paper), good agreement with the infinite sequence length results may be obtained with genomes as short as ten bases.

Finally, we assume that the fitness landscape is defined by a master genome $\{\sigma_0, \bar{\sigma}_0\}$. Specifically, we define a genome $\{\sigma, \sigma'\}$

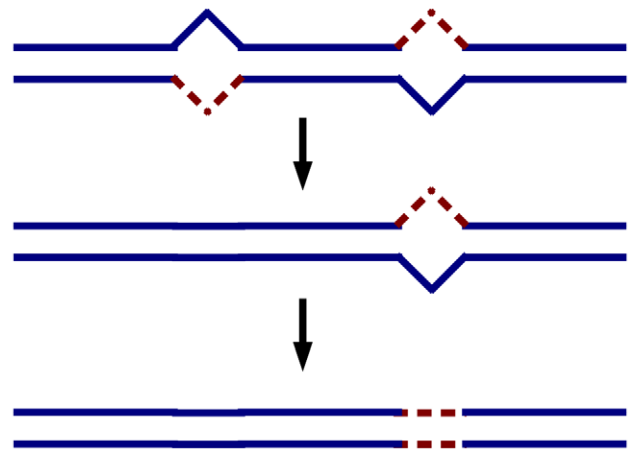


Figure 1. (Color online) Illustration of the SOS repair mechanism being considered in this paper. A DNA genome with two base-pair mismatches is restored to a fully complementary genome in two repair steps, where during each step a single mismatch (i.e. lesion) is eliminated. The first lesion is repaired correctly, so that the original base-pair of the master genome strands (solid blue lines) is restored, while the second lesion is repaired incorrectly, so that a mutation (dotted red lines) becomes fixed in the genome.

doi:10.1371/journal.pone.0014113.g001

to be viable, with a first-order growth rate constant $k > 1$, if it has fewer than l mismatches, and if it does not differ from $\{\sigma_0, \bar{\sigma}_0\}$ by any fixed mutations. Otherwise, the genome is unviable, with a first-order growth rate constant of 1. We recognize that this terminology is somewhat inappropriate, since a genome with a first-order growth rate constant of 1 may still replicate. However, this is standard terminology from quasispecies theory, where “viable” and “unviable” are taken to be synonymous with “higher fitness” and “lower fitness” respectively.

The justification for this choice of fitness landscape is as follows: If a genome has a fixed mutation, then neither DNA strand corresponds to either of the master strands $\sigma_0, \bar{\sigma}_0$. As a result, the genome does not contain all of the information corresponding to a viable organism, hence the organism is unviable. While this assumption is clearly extreme and oversimplified, it is the analogue of the single-fitness-peak landscape for single-stranded genomes.

However, in the case of a mismatch where one of the bases is the same as the corresponding base in one of the master strands, the information contained in the master genome is still preserved in one of the strands, so that the organism is assumed to remain viable if the total number of such mismatches does not exceed some cutoff value l .

For convenience, Table 1 summarizes the main parameters of the model.

Symmetrized Population Distribution

We can develop the infinite sequence length equations for our model, assuming an initially prepared clonal population consisting entirely of the wild-type (mutation-free) genome $\{\sigma_0, \bar{\sigma}_0\}$, i.e. a population consisting entirely of the fastest replicating genotype. Because, during replication, only a finite number of mutations are possible, at any time the population will consist of a distribution of genomes $\{\sigma, \sigma'\}$ where σ, σ' differ from either σ_0 and $\bar{\sigma}_0$ in at most a finite number of spots. Thus, given two gene sequences σ_1, σ_2 , if we let $D_H(\sigma_1, \sigma_2)$ denote the Hamming distance [19] between σ_1 and σ_2 (i.e. the number of sites where σ_1 and σ_2 differ), then either $D_H(\sigma, \sigma_0)$ and $D_H(\sigma', \bar{\sigma}_0)$ are finite, or $D_H(\sigma, \bar{\sigma}_0)$ and $D_H(\sigma', \sigma_0)$ are finite.

Table 1. The various parameters and their definitions in our model.

Parameter	Definition
$\{\sigma, \sigma'\}$	General notation for a genome
$\{\sigma_0, \bar{\sigma}_0\}$	The master genome
S	Alphabet size
L	Genome length
ϵ	Per-base mismatch probability during daughter strand synthesis
μ	$L\epsilon$
k	Fitness of the master genome
λ	Lesion repair probability
l	Maximum number of mismatches a genome can tolerate and still remain viable
l_S	The minimum number of mismatches required to trigger the SOS response
κ_{SOS}	First-order rate constant characterizing the rate of SOS repair

doi:10.1371/journal.pone.0014113.t001

The Hamming distance between two sequences σ_1 and σ_2 is simply equal to the number of positions by which they differ. It may be readily shown that the Hamming distance is a metric over the space of sequences [19], so that, in particular, the Hamming distance satisfies the Cauchy-Schwartz Inequality: Given three sequences σ_1 , σ_2 , and σ_3 , we have $D_H(\sigma_1, \sigma_2) + D_H(\sigma_2, \sigma_3) \geq D_H(\sigma_1, \sigma_3)$.

Now, in the limit of infinite sequence length, it may be shown that, with probability 1, that the Hamming distance between σ_0 and its complement $\bar{\sigma}_0$ is infinite [12]. Therefore, if $D_H(\sigma, \sigma_0)$ and $D_H(\sigma, \bar{\sigma}_0)$ were both finite, we would obtain $\infty = D_H(\sigma_0, \bar{\sigma}_0) \leq D_H(\sigma_0, \sigma) + D_H(\sigma, \bar{\sigma}_0) < \infty \Rightarrow \Leftarrow$, and so σ cannot simultaneously be of finite Hamming distance to σ_0 and $\bar{\sigma}_0$. Similarly, σ' cannot simultaneously be of finite Hamming distance to σ_0 and $\bar{\sigma}_0$.

As a result, we can define a strand ordering (σ, σ') for a genome $\{\sigma, \sigma'\}$, where it is understood that σ is a finite Hamming distance from σ_0 and σ' is a finite Hamming distance from $\bar{\sigma}_0$.

A given genome (σ, σ') may then be characterized by four parameters l_C, l_L, l_R , and l_B . We let l_C denote the number of sites where σ and σ' are both complementary, yet differ from the corresponding bases in σ_0 and $\bar{\sigma}_0$. We let l_L denote the number of sites where σ differs from σ_0 , but σ' is identical to $\bar{\sigma}_0$. We let l_R denote the number of sites where σ is identical to σ_0 , but σ' differs from $\bar{\sigma}_0$. Finally, we let l_B denote the number of sites where σ and σ' differ from σ_0 and $\bar{\sigma}_0$, but are not complementary (for an illustration of these parameters, see [7,13]).

Note that the fitness landscape depends only on l_C, l_L, l_R , and l_B , and hence the fitness of a given organism may be denoted by $\kappa_{(l_C, l_L, l_R, l_B)}$, where for our single-fitness-peak landscape we have $\kappa_{(l_C, l_L, l_R, l_B)} = k$ if $l_C = 0$ and $l_L + l_R + l_B \leq l$, and 1 otherwise. The condition $l_C = 0$ means that there are no mutations fixed in the genome, while the condition $l_L + l_R + l_B \leq l$ means that there are fewer than l lesions.

By the symmetry of the fitness landscape, and by the symmetry of the initial population distribution, we can group all genomes of identical l_C, l_L, l_R , and l_B , and derive the dynamical equations of the symmetrized population distribution. We therefore let $n_{(l_C, l_L, l_R, l_B)}$ denote the total number of organisms in the population whose genomes are characterized by the parameters l_C, l_L, l_R , and l_B , and we let $n_{(l_C, l_L, l_R, l_B)}^{(SOS)}$ denote the total number of organisms in

the population undergoing the SOS response, whose genomes are similarly characterized by the parameters l_C, l_L, l_R , and l_B . The corresponding population fractions are denoted $z_{(l_C, l_L, l_R, l_B)}$ and $z_{(l_C, l_L, l_R, l_B)}^{(SOS)}$, respectively.

Dynamical Equations

To develop the dynamical equations for both the $z_{(l_C, l_L, l_R, l_B)}$ and the $z_{(l_C, l_L, l_R, l_B)}^{(SOS)}$ quantities, we begin by considering a genome (σ, σ') , characterized by the parameters l_C, l_L, l_R , and l_B .

We first consider the case where this genome is not undergoing the SOS response. Then, due to the semiconservative nature of DNA replication, this genome is being destroyed at a rate given by $-\kappa_{(l_C, l_L, l_R, l_B)} n_{(l_C, l_L, l_R, l_B)}$. This genome, however, is produced by other genomes in the population, as a result of replication. So, consider some other genome (σ'', σ''') which produces (σ, σ') upon replication. This can either occur via the σ'' template strand, the σ''' template strand, or both.

If the (σ'', σ''') genome is characterized by the parameters l'_C, l'_L, l'_R , and l'_B , then σ'' differs from σ_0 in $l'_C + l'_L + l'_B$ bases. Because sequence lengths are infinite, the probability of a mismatch in one of these bases during daughter strand synthesis is 0. In the remaining sites, let l'_1 denote the number of mismatches that are not corrected, and l'_2 denote the number of mismatches that are repaired, but fixed as a mutation in the genome. Then the resulting genome (σ, σ') is characterized by:

1. $l_C = l'_C + l'_L + l'_B + l'_2$
2. $l_L = 0$
3. $l_R = l'_1$
4. $l_B = 0$

The probability of a given set of mutations corresponding to l'_1, l'_2 , is $e^{l'_1 + l'_2} (1 - \lambda)^{l'_1} (\lambda/2)^{l'_2} (1 - \epsilon + \epsilon\lambda/2)^{L - l'_C - l'_L - l'_B - l'_1 - l'_2}$. The term $(1 - \epsilon + \epsilon\lambda/2)^{L - l'_C - l'_L - l'_B - l'_1 - l'_2}$ arises as a probability that the remaining $L - l'_C - l'_L - l'_B - l'_1 - l'_2$ sites on σ'' remain identical to σ_0 , and the corresponding daughter strand sites are identical to $\bar{\sigma}_0$. The per-site probability of this is the probability of error-free daughter strand synthesis, $1 - \epsilon$, plus the probability of a mismatch, times λ , the probability that complementarity is restored, times $1/2$, the probability that complementarity is restored correctly. It is assumed that complementarity is restored by various DNA repair mechanisms, such as Nucleotide Excision Repair (NER) and Base Excision Repair (BER) [1]. However, because NER and BER do not distinguish between parent and daughter strands, the probability of correctly removing a mismatch via these mechanisms is $1/2$.

The degeneracy is given by $(L - l'_C - l'_L - l'_B)! / (l'_1! l'_2! (L - l'_C - l'_L - l'_B - l'_1 - l'_2)!)$, so in the limit of infinite sequence length the total probability becomes,

$$\begin{aligned} & \frac{(L - l'_C - l'_L - l'_B)!}{l'_1! l'_2! (L - l'_C - l'_L - l'_B - l'_1 - l'_2)!} \\ & \times e^{l'_1 + l'_2} (1 - \lambda)^{l'_1} \left(\frac{\lambda}{2}\right)^{l'_2} (1 - (1 - \lambda/2)\epsilon)^{L - l'_C - l'_L - l'_B - l'_1 - l'_2} \\ & \rightarrow \frac{1}{l'_1! l'_2!} [\mu(1 - \lambda)]^{l'_1} \left(\frac{\mu\lambda}{2}\right)^{l'_2} e^{-(1 - \lambda/2)\mu} \end{aligned} \tag{1}$$

If (σ, σ') is generated by σ''' , then we have,

1. $l_C = l'_C + l'_R + l'_B + l'_2$

2. $l_L = l'_1$
3. $l_R = 0$
4. $l_B = 0$

We also obtain an overall transition probability of $1/(l'_1!l'_2!)[\mu(1-\lambda)]^{l'_1}(\mu\lambda/2)^{l'_2}e^{-(1-\lambda/2)\mu}$.

It is important to note from the σ'' and σ''' results that genomes with $l_B > 0$ cannot be generated during replication. Since SOS repair eliminates mismatches, it follows that a population where l_B is initially 0 for all genomes will always have a population where $l_B = 0$. Therefore, we may assume in subsequent derivations that l_B, l'_B are 0.

Furthermore, note that strands σ'' that are a finite Hamming distance away from σ_0 can only generate daughter genomes where $l_L = 0$, while strands σ''' that are a finite Hamming distance away from $\bar{\sigma}_0$ can only generate daughter genomes where $l_R = 0$.

Then for the genomes (σ, σ') generated by σ'' , we have $l_C = l'_C + l'_L + l'_2$, and $l_R = l'_1$. Therefore, the restriction on (σ'', σ''') is that $0 \leq l'_2 \leq l_C$, $0 \leq l'_L \leq l_C - l'_2$, and $l'_C = l_C - l'_L - l'_2$. Note that there is no restriction on l'_R .

Then for the population number $n_{(l_C, 0, l_R, 0)}$, we have a contribution from the σ'' strands of

$$\begin{aligned} & \frac{1}{l_R!} [\mu(1-\lambda)]^{l_R} e^{-\mu(1-\lambda/2)} \\ & \times \sum_{l'_2=0}^{l_C} \frac{1}{l'_2!} \left(\frac{\mu\lambda}{2}\right)^{l'_2} \\ & \times \sum_{l'_L=0}^{l_C-l'_2} \sum_{l'_R=0}^{\infty} \kappa_{(l_C-l'_L-l'_2, l'_L, l'_R, 0)} n_{(l_C-l'_L-l'_2, l'_L, l'_R, 0)} \end{aligned} \quad (2)$$

A similar expression is obtained for the population number $n_{(l_C, l_L, 0, 0)}$, except l_R is replaced with l_L , and the roles of l'_L and l'_R are exchanged.

It should also be noted that, by the symmetry of the fitness landscape, we have that $n_{(l_C, l_L, l_R, l_B)} = n_{(l_C, l_R, l_L, l_B)}$. Another way to note this is that, for a given genome (σ, σ') , if we change the ordering of the strands so that the first strand is of finite Hamming distance to σ_0 , and the second strand is of finite Hamming distance to $\bar{\sigma}_0$, then the genome $\{\sigma, \sigma'\}$ must be represented as (σ', σ) , and is characterized by the parameters l_C, l_R, l_L , and l_B . If $\bar{n}_{(l_C, l_L, l_R, l_B)}$ denotes the number of genomes characterized by l_C, l_L, l_R , and l_B , with respect to the $(\bar{\sigma}_0, \sigma_0)$ strand ordering, then since there is a one-to-one correspondence between genomes (σ, σ') with parameters l_C, l_L, l_R, l_B with respect to the first ordering, and genomes (σ, σ') with parameters l_C, l_R, l_L, l_B with respect to the second ordering, it follows that $\bar{n}_{(l_C, l_L, l_R, l_B)} = n_{(l_C, l_R, l_L, l_B)}$. However, since the fitness landscape is invariant under strand ordering, we have $n_{(l_C, l_L, l_R, l_B)} = \bar{n}_{(l_C, l_L, l_R, l_B)}$, so that $n_{(l_C, l_L, l_R, l_B)} = n_{(l_C, l_R, l_L, l_B)}$.

Taking into consideration the contribution to $n_{(l_C, 0, 0, 0)}$, we may put everything together and obtain, after changing variables from population numbers to population fractions, the differential equations governing the time evolution of the various population fractions. These equations are,

$$\begin{aligned} \frac{dz_{(l_C, 0, 0, 0)}}{dt} = & -(\kappa_{(l_C, 0, 0, 0)} + \bar{\kappa}(t))z_{(l_C, 0, 0, 0)} \\ & + \kappa_{SOS} z_{(l_C, 0, 1, 0)}^{(SOS)} + (1 - \delta_{l_C 0})z_{(l_C-1, 0, 1, 0)}^{(SOS)} \end{aligned}$$

$$+ 2e^{-\mu(1-\lambda/2)} \sum_{l_1, C=0}^{l_C} \sum_{l_1=0}^{l_C-l_1, C} \sum_{l_2=0}^{\infty} \frac{1}{l_1, C!} \left(\frac{\mu\lambda}{2}\right)^{l_1, C}$$

$$\kappa_{(l_C-l_1, C-l_1, l_1, l_2, 0)} z_{(l_C-l_1, C-l_1, l_1, l_2, 0)}$$

$$\frac{dz_{(l_C, 0, l_R > 0, 0)}}{dt} = -(\kappa_{(l_C, 0, l_R, 0)} + \bar{\kappa}(t))z_{(l_C, 0, l_R, 0)}$$

$$+ \frac{1}{l_R!} [\mu(1-\lambda)]^{l_R} e^{-\mu(1-\lambda/2)}$$

$$\sum_{l_1, C=0}^{l_C} \sum_{l_1=0}^{l_C-l_1, C} \sum_{l_2=0}^{\infty} \frac{1}{l_1, C!} \left(\frac{\mu\lambda}{2}\right)^{l_1, C} \kappa_{(l_C-l_1, C-l_1, l_1, l_2, 0)} z_{(l_C-l_1, C-l_1, l_1, l_2, 0)}$$

for $l_R = 1, \dots, l_S - 1$

$$\begin{aligned} \frac{dz_{(l_C, 0, l_R > 0, 0)}^{(SOS)}}{dt} = & \kappa_{SOS} \left[\frac{l_R+1}{2} z_{(l_C, 0, l_R+1, 0)}^{(SOS)} + (1 - \delta_{l_C 0})z_{(l_C-1, 0, l_R+1, 0)}^{(SOS)} - l_R z_{(l_C, 0, l_R, 0)}^{(SOS)} \right] \\ & - \bar{\kappa}(t) z_{(l_C, 0, l_R, 0)}^{(SOS)} \end{aligned}$$

for $l_R = 1, \dots, l_S - 1$

$$\begin{aligned} \frac{dz_{(l_C, 0, l_R > 0, 0)}^{(SOS)}}{dt} = & \kappa_{SOS} \left[\frac{l_R+1}{2} z_{(l_C, 0, l_R+1, 0)}^{(SOS)} + (1 - \delta_{l_C 0})z_{(l_C-1, 0, l_R+1, 0)}^{(SOS)} - l_R z_{(l_C, 0, l_R, 0)}^{(SOS)} \right] \\ & - \bar{\kappa}(t) z_{(l_C, 0, l_R, 0)}^{(SOS)} \end{aligned}$$

$$+ \frac{1}{l_R!} [\mu(1-\lambda)]^{l_R} e^{-\mu(1-\lambda/2)}$$

$$\sum_{l_1, C=0}^{l_C} \sum_{l_1=0}^{l_C-l_1, C} \sum_{l_2=0}^{\infty} \frac{1}{l_1, C!} \left(\frac{\mu\lambda}{2}\right)^{l_1, C} \kappa_{(l_C-l_1, C-l_1, l_1, l_2, 0)} z_{(l_C-l_1, C-l_1, l_1, l_2, 0)}$$

for $l_R \geq l_S$ (3)

where $\bar{\kappa}(t) \equiv \sum_{l_C=0}^{\infty} \sum_{l_L=0}^{\infty} \sum_{l_R=0}^{\infty} \kappa_{(l_C, l_L, l_R, 0)} z_{(l_C, l_L, l_R, 0)}$
 $= \sum_{l_C=0}^{\infty} \left(\kappa_{(l_C, 0, 0, 0)} z_{(l_C, 0, 0, 0)} + 2 \sum_{l=1}^{\infty} \kappa_{(l_C, 0, l, 0)} z_{(l_C, 0, l, 0)} \right)$ is the mean fitness of the population. It should also be noted that δ_{ij} is the Kronecker delta function, so that $\delta_{ij} = 1$ if $i=j$, and 0 otherwise.

Note that we do not write down the dynamical equations for $z_{(l_C, l, 0, 0)}$ or $z_{(l_C, l, 0, 0)}^{(SOS)}$, since they are redundant.

The factor of 1/2 appearing in the SOS terms arises from the fact that when a mismatch is removed, it either corrects the daughter strand synthesis error, or it fixes the mismatch as a mutation in the genome. In the former case, the value of l_C remains unchanged, while in the latter case it is incremented by 1.

It should be noted that this factor is missing in the contribution to $z_{(l_C,0,0,0)}$ from SOS repair. The reason for this is that this contribution comes from $z_{(l_C,0,1,0)}^{(SOS)}$, $z_{(l_C,1,0,0)}^{(SOS)}$, $z_{(l_C-1,0,1,0)}^{(SOS)}$, and $z_{(l_C-1,1,0,0)}^{(SOS)}$. However, because $z_{(l_C,0,1,0)}^{(SOS)} = z_{(l_C,1,0,0)}^{(SOS)}$, and $z_{(l_C-1,0,1,0)}^{(SOS)} = z_{(l_C-1,1,0,0)}^{(SOS)}$, we may combine identical terms and eliminate the factor of 1/2.

The factor of $l+1$ and l in front of the κ_{SOS} rate constant arises from the fact that the fraction of genomes whose SOS enzymes are bound to a mismatch is proportional to the total number of mismatches, hence the resulting SOS rate constant is proportional to the total number of mismatches.

Results and Discussion

Steady-State Behavior

Definitions and basic equations. To obtain the steady-state behavior of our model, we begin by introducing some definitions that will allow us to simplify the calculations.

1. $z_1 = z_{(0,0,0,0)}$.
2. $z_2 = \sum_{l'=1}^l z_{(0,0,l',0)}$.
3. $z_3 = \sum_{l'=l+1}^{l_S-1} z_{(0,0,l',0)}$.
4. $z_4 = \sum_{l_C=1}^\infty z_{(l_C,0,0,0)}$.
5. $z_5 = \sum_{l_C=1}^\infty \sum_{l'=1}^l z_{(l_C,0,l',0)}$.
6. $z_6 = \sum_{l_C=1}^\infty \sum_{l'=l+1}^{l_S-1} z_{(l_C,0,l',0)}$.
7. $z_{0l}^{(SOS)} = z_{(0,0,l,0)}^{(SOS)}$.
8. $z_{1l}^{(SOS)} = \sum_{l_C=0}^\infty z_{(l_C,0,l,0)}^{(SOS)}$.
9. $z_0^{(SOS)} = \sum_{l=1}^\infty z_{0l}^{(SOS)}$.
10. $z^{(SOS)} = \sum_{l=1}^\infty z_{1l}^{(SOS)}$.

where we set $l=l_S-1$ whenever l was previously defined as $\geq l_S$. The differential equations for $z_1, z_2, z_3, z_4, z_5,$ and z_6 are readily derived. From the equations,

$$\sum_{l_2=0}^\infty \kappa_{(0,0,l_2,0)} z_{(0,0,l_2,0)} = kz_1 + kz_2 + z_3 \tag{4}$$

and

$$\begin{aligned} & \sum_{l_C=0}^\infty \sum_{l_1,C=0}^{l_C} \sum_{l_1=0}^{l_C-l_1,C} \sum_{l_2=0}^\infty \frac{1}{l_1,C!} \left(\frac{\mu\lambda}{2}\right)^{l_1,C} \\ & \times \kappa_{(l_C-l_1,C-l_1,l_1,l_2,0)} z_{(l_C-l_1,C-l_1,l_1,l_2,0)} \\ & = e^{\mu\lambda/2} [kz_1 + 2kz_2 + 2z_3 + z_4 + 2z_5 + 2z_6] \end{aligned} \tag{5}$$

we obtain,

$$\begin{aligned} \frac{dz_1}{dt} &= -(k + \bar{\kappa}(t))z_1 + 2e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \\ &+ \kappa_{SOS} z_{01}^{(SOS)} \\ \frac{dz_2}{dt} &= -(k + \bar{\kappa}(t))z_2 \\ &+ (f_l(\mu,\lambda) - 1)e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \\ \frac{dz_3}{dt} &= -(1 + \bar{\kappa}(t))z_3 \\ &+ (f_{l_S-1}(\mu,\lambda) - f_l(\mu,\lambda))e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \\ \frac{dz_4}{dt} &= -(1 + \bar{\kappa}(t))z_4 \\ &+ 2e^{-\mu(1-\lambda/2)} \left[e^{\mu\lambda/2} (kz_1 + 2kz_2 + 2z_3 + z_4 + 2z_5 + 2z_6) \right. \\ &\left. - (kz_1 + kz_2 + z_3) \right] + \kappa_{SOS} \left[2z_{11}^{(SOS)} - z_{01}^{(SOS)} \right] \\ \frac{dz_5}{dt} &= -(1 + \bar{\kappa}(t))z_5 \\ &+ (f_l(\mu,\lambda) - 1)e^{-\mu(1-\lambda/2)} \times \\ &\left[e^{\mu\lambda/2} (kz_1 + 2kz_2 + 2z_3 + z_4 + 2z_5 + 2z_6) \right. \\ &\left. - (kz_1 + kz_2 + z_3) \right] \\ \frac{dz_6}{dt} &= -(1 + \bar{\kappa}(t))z_6 \\ &+ (f_{l_S-1}(\mu,\lambda) - f_l(\mu,\lambda))e^{-\mu(1-\lambda/2)} \times \\ &\left[e^{\mu\lambda/2} (kz_1 + 2kz_2 + 2z_3 + z_4 + 2z_5 + 2z_6) \right. \\ &\left. - (kz_1 + kz_2 + z_3) \right] \end{aligned} \tag{6}$$

We also have,

$$\begin{aligned} \frac{dz_{0l}^{(SOS)}}{dt} &= \kappa_{SOS} \frac{l+1}{2} z_{0l+1}^{(SOS)} - (l\kappa_{SOS} + \bar{\kappa}(t))z_{0l}^{(SOS)} \\ &\text{for } l = 1, \dots, l_S - 1 \\ \frac{dz_{1l}^{(SOS)}}{dt} &= \kappa_{SOS} \frac{l+1}{2} z_{0l+1}^{(SOS)} - (l\kappa_{SOS} + \bar{\kappa}(t))z_{1l}^{(SOS)} \\ &+ \frac{1}{l!} [\mu(1-\lambda)]^l e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \\ &\text{for } l \geq l_S \\ \frac{dz_{1l}^{(SOS)}}{dt} &= \kappa_{SOS} (l+1) z_{1l+1}^{(SOS)} - (l\kappa_{SOS} + \bar{\kappa}(t))z_{1l}^{(SOS)} \\ &\text{for } l = 1, \dots, l_S - 1 \\ \frac{dz_{1l}^{(SOS)}}{dt} &= \kappa_{SOS} (l+1) z_{1l+1}^{(SOS)} - (l\kappa_{SOS} + \bar{\kappa}(t))z_{1l}^{(SOS)} \\ &+ \frac{1}{l!} [\mu(1-\lambda)]^l e^{-\mu(1-\lambda)} [kz_1 + 2kz_2 + 2z_3 + z_4 + 2z_5 + 2z_6] \\ &\text{for } l \geq l_S \end{aligned} \tag{7}$$

We can add these equations to obtain,

$$\begin{aligned} \frac{dz^{(SOS)}}{dt} = & -\kappa_{SOS}z_{11}^{(SOS)} - \bar{\kappa}(t)z^{(SOS)} \\ & + \left(1 - e^{-\mu(1-\lambda)}f_{l_S-1}(\mu, \lambda)\right) \times \\ & [kz_1 + 2kz_2 + 2z_3 + z_4 + 2z_5 + 2z_6] \end{aligned} \tag{8}$$

For the purposes of computing the mean fitness at steady-state, we can simplify the system of equations somewhat by defining $\bar{z}_4 = z_4 + 2z_5 + 2z_6$. We obtain,

$$\begin{aligned} \frac{d\bar{z}_4}{dt} = & -(1 + \bar{\kappa}(t))\bar{z}_4 + 2e^{-\mu(1-\lambda/2)}f_{l_S-1}(\mu, \lambda) \times \\ & \left[e^{\mu\lambda/2}(kz_1 + 2kz_2 + 2z_3 + \bar{z}_4) - (kz_1 + kz_2 + z_3)\right] \\ & + \kappa_{SOS} \left[2z_{11}^{(SOS)} - z_{01}^{(SOS)}\right] \end{aligned} \tag{9}$$

For consistency of notation, in what follows we shall simply let z_4 denote \bar{z}_4 .

Determining the population fractions $z_{01}^{(SOS)}$, $z_{11}^{(SOS)}$, and $z^{(SOS)}$. To obtain the steady-state behavior of this system of equations, we begin by first solving for the steady-state of the population undergoing SOS repair.

For $l=0, \dots, l_S-1$ we have at steady-state that,

$$z_{0l+1}^{(SOS)} = \frac{2}{l+1} \left(l + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}\right) z_{0l}^{(SOS)} \tag{10}$$

which gives,

$$z_{0l_S}^{(SOS)} = \frac{2^{l_S-1}}{l_S!} \left[\prod_{l=1}^{l_S-1} \left(l + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}\right)\right] z_{01}^{(SOS)} \tag{11}$$

For $l \geq l_S$, we have,

$$\begin{aligned} z_{0l+1}^{(SOS)} = & \frac{2}{l+1} \left(l + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}\right) z_{0l}^{(SOS)} \\ & - \frac{2}{\kappa_{SOS}(l+1)!} [\mu(1-\lambda)]^l \times \\ & e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \end{aligned} \tag{12}$$

This expression has the form of the recursion relation, $x_{n+1} = a_n x_n - b_n$. Using mathematical induction, it is possible to prove that $x_n = a_{n-1} \times \dots \times a_0 x_0 - a_{n-1} \times \dots \times a_1 b_0 - a_{n-1} \times \dots \times a_2 b_1 - \dots - a_{n-1} b_{n-2} - b_{n-1}$. Therefore,

$$\begin{aligned} z_{0l}^{(SOS)} = & \frac{2^{l-1}}{l!} \sum_{l'=1}^{l-1} \left(l' + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}\right) \\ & \times \left[z_{01}^{(SOS)} - \frac{2}{\kappa_{SOS}} e^{-\mu(1-\lambda/2)} (kz_1 + kz_2 + z_3) \right. \\ & \left. \times \prod_{l'=1}^{l_S} \frac{\mu(1-\lambda)}{2 \left(l' + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}\right)} \times \sum_{k=0}^{l_S-1} \prod_{l'=1}^k \frac{\mu(1-\lambda)}{2 \left(l_S + l' + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}\right)} \right] \end{aligned} \tag{13}$$

where we define $\prod_{i=1}^0 a_i = 1$.

If we define $g_l(\mu, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS}) = \prod_{l'=1}^l \frac{\mu(1-\lambda)}{l' + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}} \times \sum_{k=0}^{\infty} \prod_{l'=1}^k \frac{\mu(1-\lambda)}{l+l' + \frac{\bar{\kappa}(t=\infty)}{\kappa_{SOS}}}$, then imposing the requirement that $\lim_{l \rightarrow \infty} z_{0l}^{(SOS)} = 0$ gives, at steady-state, that,

$$\kappa_{SOS} z_{01}^{(SOS)} = 2e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \times g_{l_S}(\mu/2, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS}) \tag{14}$$

Using a similar argument, we obtain,

$$\kappa_{SOS} z_{11}^{(SOS)} = e^{-\mu(1-\lambda)} [kz_1 + 2kz_2 + 2z_3 + z_4] \times g_{l_S}(\mu, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS}) \tag{15}$$

For the steady-state value of $z^{(SOS)}$, we have, using the identity $\bar{\kappa}(t) = kz_1 + 2kz_2 + 2z_3 + z_4$,

$$z^{(SOS)} = 1 - e^{-\mu(1-\lambda)} \times \left(f_{l_S-1}(\mu, \lambda) + g_{l_S}(\mu, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS})\right) \tag{16}$$

Computing the steady-state mean fitness $\bar{\kappa}(t=\infty)$. Plugging our expressions for $\kappa_{SOS} z_{01}^{(SOS)}$ and $\kappa_{SOS} z_{11}^{(SOS)}$ into the steady-state population fractions equations, we obtain,

$$\begin{aligned} 0 = & -(k + \bar{\kappa}(t=\infty))z_1 \\ & + 2e^{-\mu(1-\lambda/2)} \left(1 + g_{l_S}\left(\frac{\mu}{2}, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS}\right)\right) \times [kz_1 + kz_2 + z_3] \\ 0 = & -(k + \bar{\kappa}(t=\infty))z_2 + (f_l(\mu, \lambda) - 1)e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \\ 0 = & -(1 + \bar{\kappa}(t=\infty))z_3 \\ & + (f_{l_S-1}(\mu, \lambda) - f_l(\mu, \lambda))e^{-\mu(1-\lambda/2)} [kz_1 + kz_2 + z_3] \\ 0 = & -(1 + \bar{\kappa}(t=\infty))z_4 \\ & + 2e^{-\mu(1-\lambda)} (f_{l_S-1}(\mu, \lambda) + g_{l_S}(\mu, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS})) \\ & \times [kz_1 + 2kz_2 + 2z_3 + z_4] \\ & - 2e^{-\mu(1-\lambda/2)} (f_{l_S-1}(\mu, \lambda) + g_{l_S}\left(\frac{\mu}{2}, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS}\right)) \\ & \times [kz_1 + kz_2 + z_3] \end{aligned} \tag{17}$$

From these equations we may derive the equality,

$$\begin{aligned} k(z_1 + z_2) + z_3 = & [k(z_1 + z_2) + z_3] e^{-\mu(1-\lambda/2)} \\ & \times \left[k \frac{1 + 2g_{l_S}(\mu/2, \lambda; \bar{\kappa}(t=\infty), \kappa_{SOS}) + f_l(\mu, \lambda)}{k + \bar{\kappa}(t=\infty)} + \frac{f_{l_S-1}(\mu, \lambda) - f_l(\mu, \lambda)}{1 + \bar{\kappa}(t=\infty)} \right] \end{aligned} \tag{18}$$

Below the error catastrophe, when z_1, z_2, z_3 are not all 0, we may cancel $k(z_1 + z_2) + z_3$ from both sides of the equation and re-arrange to obtain,

$$\bar{\kappa}(t = \infty)^2 - A(\mu, \lambda; \bar{\kappa}, \kappa_{SOS})\bar{\kappa}(t = \infty) - B(\mu, \lambda; \bar{\kappa}, \kappa_{SOS}) = 0 \quad (19)$$

where,

$$\begin{aligned} & A(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) \\ = & k \left[e^{-\mu(1-\frac{\lambda}{2})} (1 + f_i(\mu, \lambda)) + 2g_{I_S} \left(\frac{\mu}{2}, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS} \right) - 1 \right] \\ & + e^{-\mu(1-\frac{\lambda}{2})} (f_{I_S-1}(\mu, \lambda) - f_i(\mu, \lambda)) - 1 \quad (20) \\ & B(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) \\ = & k \left[e^{-\mu(1-\frac{\lambda}{2})} (1 + f_{I_S-1}(\mu, \lambda)) + 2g_{I_S} \left(\frac{\mu}{2}, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS} \right) - 1 \right] \end{aligned}$$

Beyond the error catastrophe, the mutation rate is sufficiently high that the selective advantage for remaining localized about the $l_C = 0$ genomes disappears, so that $z_1, z_2,$ and z_3 drop to 0. The relevant steady-state equation is then,

$$\begin{aligned} 0 = & -(1 + \bar{\kappa}(t = \infty))z_4 + 2e^{-\mu(1-\lambda)} \\ & \times \left(f_{I_S-1}(\mu, \lambda) + g_{I_S}(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) \right) z_4 \quad (21) \end{aligned}$$

which may be solved for $\bar{\kappa}(t = \infty)$ to give,

$$\bar{\kappa}(t = \infty) = 2e^{-\mu(1-\lambda)} \times \left[f_{I_S-1}(\mu, \lambda) + g_{I_S}(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) \right] - 1 \quad (22)$$

The error catastrophe occurs at the mutation rate for which the two expressions for the mean equilibrium fitness become equal. As with previous quasispecies models, the error catastrophe here also corresponds to a localization to delocalization transition over sequence space [5–7].

Limiting Cases. We now proceed to consider the behavior of the steady-state mean fitness for a number of limiting cases, in order to better understand our model. We consider the following cases: (1) $\lambda = 1$, corresponding to perfect lesion repair, so that there are non-complementary genomes in the population. (2) $I_S = \infty$, corresponding to the case where no genome ever undergoes the SOS response. (3) $\kappa_{SOS} \rightarrow \infty$, corresponding to the case where SOS repair happens rapidly, so that there is a negligible fitness penalty associated with undergoing the SOS response.

Case 1: $\lambda = 1$ When $\lambda = 1$, we get for $I_S > 0$ that $g_{I_S}(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = 0$, and that $f_{I_S-1}(\mu, \lambda) = 1$. Therefore, above the error catastrophe, we obtain $\bar{\kappa}(t = \infty) = 1$. Below the error catastrophe, we have $A(\mu, 1; \bar{\kappa}, \kappa_{SOS}) = k(2e^{-\mu/2} - 1) - 1$, $B(\mu, 1; \bar{\kappa}(t = \infty), \kappa_{SOS}) = k(2e^{-\mu/2} - 1)$, giving $\bar{\kappa}(t = \infty) = k(2e^{-\mu/2} - 1)$. These results are in agreement with the solution of the semiconservative quasispecies equations with perfect lesion repair [12].

Case 2: $I_S = \infty$ When $I_S = \infty$, then $g_{I_S}(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = 0$. Below the error catastrophe, we have $A(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = k[e^{-\mu(1-\lambda/2)}(1 + f_i(\mu, \lambda)) - 1] - f_i(\mu, \lambda)e^{-\mu(1-\lambda/2)} + e^{-\mu\lambda/2} - 1$, and $B(\mu, \lambda; \bar{\kappa}, \kappa_{SOS}) = k(e^{-\mu(1-\lambda/2)} + e^{-\mu\lambda/2} - 1)$. Above the error catastrophe, we have $\bar{\kappa}(t = \infty) = 1$. Both results are in agreement with the semiconservative quasispecies equations with arbitrary lesion repair efficiency [13].

Case 3: $\kappa_{SOS} \rightarrow \infty$ When $\kappa_{SOS} \rightarrow \infty$, then $g_{I_S}(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = e^{\mu(1-\lambda)} - f_{I_S-1}(\mu, \lambda)$. Above the error catastrophe, we get that $\bar{\kappa}(t = \infty) = 1$. Below the error catastrophe, we obtain that, $A(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = k[e^{-\mu(1-\lambda/2)}(1 + f_i(\mu, \lambda)) + 2e^{\mu(1-\lambda)/2} - 2f_{I_S-1}(\mu/2, \lambda)] - 1$ and $B(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = k[e^{-\mu(1-\lambda/2)}(1 + f_{I_S-1}(\mu, \lambda)) + 2e^{\mu(1-\lambda)/2} - 2f_{I_S-1}(\mu/2, \lambda)] - 1$.

Taking $I_S = 1$ for $\kappa_{SOS} \rightarrow \infty$ gives $A(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = k[2e^{-\mu/2} - 1] - 1$, and $B(\mu, \lambda; \bar{\kappa}(t = \infty), \kappa_{SOS}) = k[2e^{-\mu/2} - 1]$, so that $\bar{\kappa}(t = \infty) = k[2e^{-\mu/2} - 1]$ below the error catastrophe. This result is identical with the semiconservative quasispecies equations with perfect lesion repair, which makes sense, since here we assume that any lesion is eliminated instantaneously [13].

Optimal cutoff. If we assume that $k \gg 1$, and $\kappa_{SOS} \rightarrow \infty$, then it is possible to find the value of I_S which maximizes the steady-state mean fitness $\bar{\kappa}(t = \infty)$. To do this, we define a normalized mean fitness ϕ to be equal to $\bar{\kappa}(t = \infty)/k$, and if we divide Eq. (19) by k^2 , we obtain that ϕ is the solution to,

$$\phi^2 - \alpha(\mu, \lambda; \phi, \kappa_{SOS})\phi - \frac{1}{k}\beta(\mu, \lambda; \phi, \kappa_{SOS}) = 0 \quad (23)$$

where, $\alpha(\mu, \lambda; \phi, \kappa_{SOS}) = e^{-\mu(1-\lambda/2)}[1 + f_i(\mu, \lambda) + 2e^{\mu(1-\lambda)/2} - 2f_{I_S-1}(\mu/2, \lambda)] - 1 + \frac{1}{k}[e^{-\mu(1-\lambda/2)}(f_{I_S-1}(\mu, \lambda) - f_i(\mu, \lambda)) - 1]$, and $\beta(\mu, \lambda; \phi, \kappa_{SOS}) = e^{-\mu(1-\lambda/2)}[1 + f_{I_S-1}(\mu, \lambda) + 2e^{\mu(1-\lambda)/2} - 2f_{I_S-1}(\mu/2, \lambda)] - 1$.

Therefore, for large k we obtain that $\phi \rightarrow \lim_{k \rightarrow \infty} \alpha(\mu, \lambda; \phi, \kappa_{SOS})$, which gives,

$$\phi = e^{-\mu(1-\lambda/2)} + 2e^{-\mu/2} - 1 + e^{-\mu(1-\lambda/2)}(f_i(\mu, \lambda) - 2f_{I_S-1}(\mu/2, \lambda)) \quad (24)$$

so that maximizing ϕ is equivalent to maximizing $f_i(\mu, \lambda) - 2f_{I_S-1}(\mu/2, \lambda)$.

Now, because l must be re-set to $I_S - 1$ whenever we take $I_S \leq l$, we can only vary I_S independently of l whenever $I_S > l$. In this regime, the expression $f_i(\mu, \lambda) - 2f_{I_S-1}(\mu/2, \lambda)$ is maximized whenever $I_S = l + 1$.

In the regime where $I_S \leq l$, l is re-set to $I_S - 1$, and so,

$$\begin{aligned} & f_i(\mu, \lambda) - 2f_{I_S-1}(\mu/2, \lambda) \\ = & f_{I_S-1}(\mu, \lambda) - 2f_{I_S-1}(\mu/2, \lambda) \quad (25) \\ = & -1 + \mu(1-\lambda) \times \sum_{k=1}^{I_S-2} \frac{[\mu(1-\lambda)]^k}{(k+1)!} \left(1 - \frac{1}{2^k} \right) \end{aligned}$$

and so this expression is equal to -1 for $I_S = 1, 2$, and then increases with successive values of I_S .

Now, because l is re-set to $I_S - 1$ for $I_S \leq l$, it follows that we take $l = I_S - 1$ for $I_S \leq l + 1$. For $l = 0$, we then obtain that ϕ is maximized over $I_S \leq l + 1$ for $I_S = 1$, while when $l = 1$, we obtain that ϕ is maximized over $I_S \leq l + 1$ for $I_S = 1, 2$. For $l \geq 2$, we obtain that ϕ is maximized over $I_S \leq l + 1$ for $I_S = l + 1$.

Therefore, in any case, we can maximize ϕ over $I_S \leq l + 1$ by taking $I_S = l + 1$. Since we can maximize ϕ over $I_S \geq l + 1$ by setting $I_S = l + 1$, it follows that ϕ is maximized when $I_S = l + 1$.

We reach the conclusion that, when the fitness penalty for having a non-viable genome is sufficiently great, the SOS response will confer a

maximum selective advantage if it is activated when and only when the genome has sustained sufficient genetic damage so that it will be unviable without SOS repair. However, it should be emphasized that this only holds when μ is not near the error catastrophe, so that ϕ is sufficiently larger than 1 for large k that the above analysis holds.

Stochastic Simulations

We developed stochastic simulations of a unicellular population capable of undergoing the SOS response, in order to numerically test the analytical predictions of our model. We consider a constant population of genomes that is cycled over every time step. During each cycle, every genome is allowed to replicate with a probability $\kappa_{\{\sigma,\sigma'\}}\Delta t$, where $\kappa_{\{\sigma,\sigma'\}}$ is the first-order growth rate constant of genome $\{\sigma,\sigma'\}$, and Δt is the length of the time step. We take Δt to be sufficiently small so that the probability of a given genome replicating more than once during a cycle is negligible.

We assume that the population initially consists of a clonal population of wild-type (mutation-free) genomes. The fitness of a given genome $\{\sigma,\sigma'\}$ is determined by assigning l_C, l_L, l_R, l_B parameters to the ordered-pairs (σ,σ') , (σ',σ) with respect to the ordered-pair (σ_0,σ_0) . For each set of l_C, l_L, l_R , and l_B parameters, a fitness is assigned based on the fitness landscape defined previously. The fitness of the genome is then taken to be the larger of the two calculated fitnesses. In the limit of infinite sequence length, this prescription for calculating fitnesses becomes identical to the method used in the analytical solution of our model.

If a genome replicates during a cycle, then it is removed from the population, and the two daughters are added to the population of genomes. To maintain a constant population size, another, randomly chosen genome is removed from the population as well. Because this approach is simply the stochastic implementation of the quasispecies dynamics of the system, it converges to the infinite population, continuous time result as the population size gets larger and the time steps get smaller.

If a daughter genome is produced that has at least l_S lesions, then it enters the SOS response, and is assigned a replication probability

of 0. A genome that has initiated the SOS response continues to undergo SOS repair until all lesions have been removed, and a complementary genome has been restored. During every time step, a genome that is undergoing the SOS response has its lesions scanned, and each lesion is repaired with probability $\kappa_{SOS}\Delta t$. In addition to being chosen small enough so that the probability of a given genome replicating more than once during a cycle is negligible, we also choose Δt to be sufficiently small so that the probability that a given genome undergoing the SOS response has more than one lesion repaired during a cycle is also negligible.

The stochastic simulation is allowed to run for a sufficient number of time steps so that the mean fitness of the population does not change significantly, at which point the system is assumed to be at steady-state.

Figures 2 and 3 show plots comparing the mean fitness obtained from the analytical solution to the mean fitness obtained from the stochastic simulations. As can be seen from the figures, the agreement between the analytical solution and the stochastic simulation is excellent.

Conclusions and Future Research

This paper developed a quasispecies approach for describing the evolutionary dynamics of a unicellular population that incorporated a simplified model of the SOS response. The model was a generalization of the single-fitness-peak landscape that is often used in quasispecies theory to study various problems in evolutionary dynamics. The model was shown to be analytically solvable, and it was found that the solution led to a maximal selective advantage to the SOS response in a manner that is broadly consistent with the behavior of actual organisms. Specifically, we showed that the SOS response should only be activated in a cell with a sufficiently damaged genome that it will be unviable if the SOS response is not activated. In such a situation, the cell has “nothing to lose,” meaning that it is better to attempt to repair the genome and risk introducing deleterious mutations, than it is to leave a highly damaged genome alone.

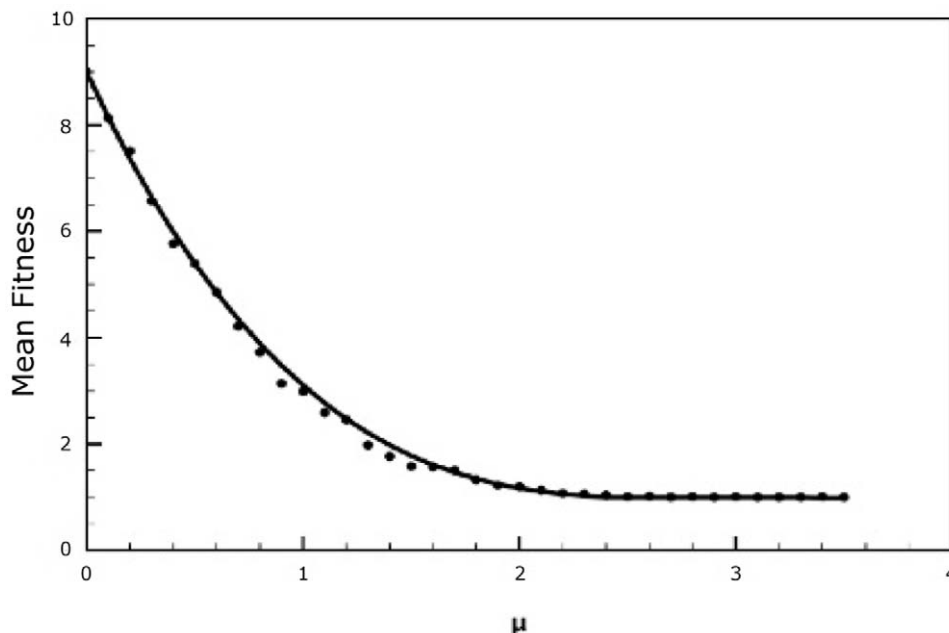


Figure 2. Comparison of the mean fitnesses obtained from both stochastic simulations (dots) and the analytical solution (solid line) of our model. Parameter values are $k=9$, $l=4$, $l_S=5$, $\lambda=0.08$, $\kappa_{SOS}=100$, $L=100$. The population size was set at 1,000. doi:10.1371/journal.pone.0014113.g002

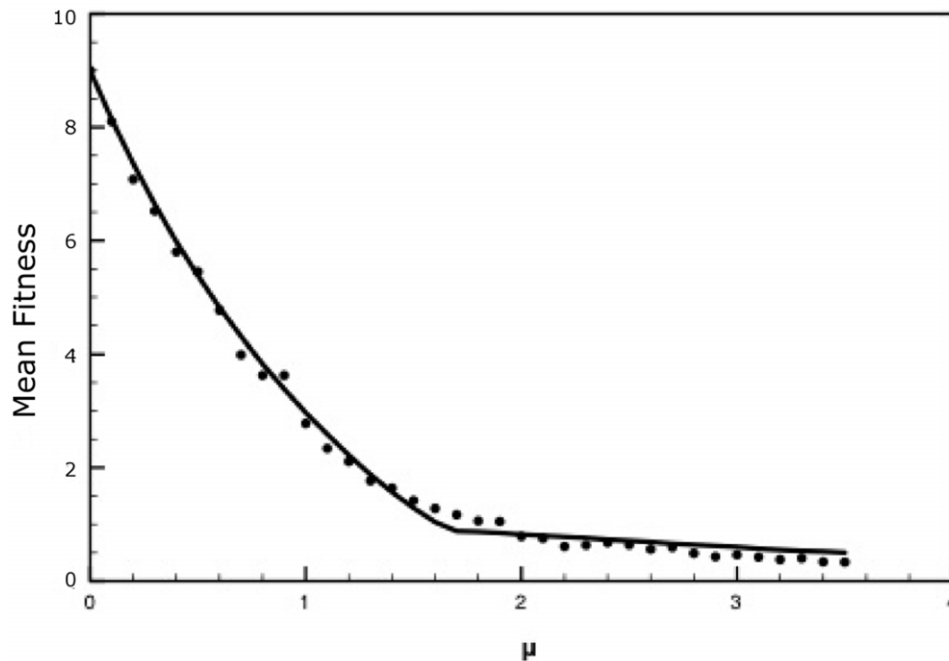


Figure 3. Comparison of the mean fitnesses obtained from both stochastic simulations (dots) and the analytical solution (solid line) of our model. Parameter values are $k=9$, $l=4$, $l_S=5$, $\lambda=0.08$, $\kappa_{SOS}=10$, $L=100$. The population size was set at 1,000. doi:10.1371/journal.pone.0014113.g003

For future research, it will be important to consider more realistic models that will allow for quantitative models that can be used in collaboration with experiment. Because the SOS response is a genetic repair pathway that works in conjunction with other cellular repair pathways, a proper understanding of the SOS response is important for developing a coherent theory of mutation-propagation that will be useful for understanding the

emergence of antibiotic drug resistance in bacteria, and cancer in multicellular organisms [2–4].

Author Contributions

Conceived and designed the experiments: EDT. Performed the experiments: EDT. Analyzed the data: AK EDT. Contributed reagents/materials/analysis tools: AK EDT. Wrote the paper: AK EDT.

References

- Voet D, Voet JG (2004) Biochemistry, 3rd ed. Hoboken, NJ: John Wiley and Sons, Inc.
- Beaber JW, Hochhut B, Waldor MK (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature (London)* 427: 72–74.
- Miller C, Thomsen LE, Gaggero C, Mosseri R, Ingmer H, et al. (2004) SOS response induction by β -lactams and bacterial defense against antibiotic lethality. *Science* 305: 1629–1631.
- McKenzie GJ, Harris RS, Lee PL, Rosenberg SM (2000) The SOS response regulates adaptive mutation. *Proc. Natl. Acad. Sci. USA* 97: 6646–6651.
- Bull JJ, Meyers LA, Lachmann M (2005) Quasispecies made simple. *PLoS Computational Biology* 1: e61.
- Wilke CO (2005) Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology* 5: 44.
- Tannenbaum E, Shakhnovich EI (2005) Semiconservative replication, genetic repair, and many-gene genomes: Extending the quasispecies paradigm to living systems. *Physics of Life Reviews* 2: 290–317.
- Tannenbaum E, Deeds EJ, Shakhnovich EI (2003) Equilibrium distribution of mutators in the single-fitness-peak model. *Physical Review Letters* 91: 138105.
- Tannenbaum E, Shakhnovich EI (2004) Error and repair catastrophes: A two-dimensional phase diagram in the quasispecies model. *Physical Review E* 69: 011902.
- Sasaki A, Nowak MA (2003) Mutation Landscapes. *The Journal of Theoretical Biology* 224: 241–247.
- Kessler DA, Levine H (1998) Mutator dynamics on a smooth evolutionary landscape. *Physical Review Letters* 80: 2012–2015.
- Tannenbaum E, Deeds EJ, Shakhnovich EI (2004) Semiconservative replication in the quasispecies model. *Physical Review E* 69: 061916.
- Tannenbaum E, Shakhnovich EI (2004) Imperfect DNA lesion repair in the semiconservative quasispecies model: Derivation of the Hamming class equations and solution of the single-fitness-peak landscape. *Physical Review E* 70: 061915.
- Brumer Y, Shakhnovich EI (2004) Host-parasite co-evolution and optimal mutation rates for semiconservative quasispecies. *Physical Review E* 69: 061909.
- Brumer Y, Shakhnovich EI (2004) Importance of DNA repair in tumor suppression. *Physical Review E* 70: 061912.
- Tannenbaum E, Shakhnovich EI (2004) Solution of the quasispecies model for an arbitrary gene network. *Physical Review E* 70: 021903.
- Tannenbaum E, Sherley JL, Shakhnovich EI (2006) Semiconservative quasispecies equations for polysomic genomes: The haploid case. *The Journal of Theoretical Biology* 241: 791–805.
- Nilsson M, Snoad N (2000) Error thresholds for quasispecies on dynamic fitness landscapes. *Physical Review Letters* 84: 191–194.
- Ash RB (1990) Information Theory. Dover Publications.