

Effect of Chromosomal Instability on the Mutation-Selection Balance in Unicellular Populations

Eran Itan, Emmanuel Tannenbaum*

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

Abstract

This paper develops a mathematical model describing the evolutionary dynamics of a unicellular, asexually replicating population exhibiting chromosomal instability. Chromosomal instability is a form of genetic instability characterized by the gain or loss of entire chromosomes during cell division. We assume that the cellular genome is divided into several homologous groups of chromosomes, and that a single functional chromosome per homologous group is required for the cell to have the wild-type fitness. If the fitness is unaffected by the total number of chromosomes in the cell, our model is analytically solvable, and yields a mean fitness at mutation-selection balance that is identical to the mean fitness when there is no chromosomal instability. If this assumption is relaxed and the total number of chromosomes in the cell is not allowed to increase without bound, then chromosomal instability leads to a reduction in mean fitness. The results of this paper provide a useful baseline that can inform both future theoretical and experimental studies of chromosomal instability.

Citation: Itan E, Tannenbaum E (2012) Effect of Chromosomal Instability on the Mutation-Selection Balance in Unicellular Populations. PLoS ONE 7(5): e26513. doi:10.1371/journal.pone.0026513

Editor: Yamini Dalal, National Cancer Institute, United States of America

Received: July 1, 2011; **Accepted:** September 28, 2011; **Published:** May 23, 2012

Copyright: © 2012 Itan, 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: The authors have no funding or support to report.

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

* E-mail: emmanuel.tannenbaum@gmail.com

Introduction

Living systems have evolved a range of mechanisms to ensure the accurate transmission of their genetic information from one generation to the next [1]. When one or more of these mechanisms break down, the result is generally significantly increased mutation rates and variability in the distribution of genotypes in a population. Very often, the variability in the distribution of genotypes manifests itself through organizational changes to the genome itself, as opposed to point-mutations along the DNA gene sequences. When this happens, the genomes are said to exhibit *genetic instability* [1,2,3].

Genetic instability is a hallmark feature of cancer cells, and generally comes in one of two different forms: (1) Microsatellite INstability, or MIN, tumors, are characterized by elevated point-mutation rates and the accumulation of sequences of DNA in their chromosomes [1,2,3,4]. (2) Chromosomal INstability, or CIN, tumors, are characterized by a breakdown in chromosomal segregation mechanisms during cell division, leading to the gain or loss of entire chromosomes [1,2,3,4]. Such cells may no longer be diploid, with different homologous groups being characterized by different copy numbers, a phenomenon known as *aneuploidy* [1,2,3,4]. In addition to the gain or loss of entire chromosomes, CIN is also characterized by the fusion of parts of chromosomes to one another, leading to more complicated genome-wide rearrangements [1,2,3,4].

The CIN instability is by far more prevalent than the MIN instability [1,4]. Given the centrality of genetic instability to the progression of cancer, an understanding of the role of CIN on the evolutionary dynamics of a cellular population is important for developing accurate models of cancer progression, and for possibly assessing the efficacy of various treatment strategies.

Materials and Methods

To model the evolutionary dynamics associated with the CIN instability, we consider a unicellular population of asexually reproducing organisms. We also assume that the organismal genome is organized as follows: We let N denote the total number of homologous groups in the genome. Labelling the homologous groups $1, \dots, N$, we let n_i denote the number of chromosomes in homologous group i (by a homologous group, we mean a set of chromosomes that contain the same sequence of genes, though corresponding genes on different chromosomes within a group may differ from one another due to mutations).

We assume that each chromosome is a double-stranded DNA molecule, and that, when each chromosome replicates, the two strands unwind, and each strand serves as the template for the synthesis of the corresponding daughter strand (a process known as *semiconservative* replication).

During normal cell division, the two daughter chromosomes of a given parent chromosome segregate into separate daughter cells. However, with chromosomal instability, it is possible that two daughter chromosomes will segregate into the identical daughter cell. We let p_c denote the probability that the two daughter chromosomes of a given parent segregate into separate daughter cells, so that $1 - p_c$ is the probability that they co-segregate into the same daughter cell.

We assume that DNA replication is not error-free. Thus, we let p denote the probability that a given template strand of a given chromosome produces a daughter chromosome that is identical to the parent.

Following the single-fitness-peak approximation that is commonly used as a starting point in quasispecies and population genetics models, we assume that for each homologous group there

is a wild-type chromosome for which that chromosome is functional. Any mutation to the wild-type is assumed to render the chromosome non-functional.

If the chromosomes are taken to be sufficiently long, then we may make an approximation known as the *neglect of backmutations*, which states that a template strand that differs from the wild-type will produce a daughter that differs from the wild-type with probability 1. The basis for this assumption is that, for a sufficiently long chromosome, any new mutations will likely occur in a previously unmutated portion of the genome, so that the wild-type cannot be restored.

Finally, we assume that the first-order growth rate constant, or fitness, of a genome is determined by the number of functional chromosomes in each homologous group, as well as by the number of non-functional chromosomes in each homologous group. Thus, if n_i^+ denotes the number of functional chromosomes in homologous group i , and n_i^- denotes the number of non-functional chromosomes, then the fitness is given by $\kappa(\vec{n}^+; \vec{n}^-)$, where $\vec{n}^\pm \equiv (n_1^\pm, \dots, n_N^\pm)$.

When a given parent cell reproduces, it produces two daughters. Let us label these daughters as a “left” daughter cell and “right” daughter cell. Given a left daughter cell with genome characterized by (\vec{n}^+, \vec{n}^-) , we define parameters m_{ij}^\pm , where $i = 1, \dots, N$, $j = 1, \dots, 6$ for m_{ij}^+ and $j = 1, 2, 3$ for m_{ij}^- , as follows:

1. $m_{i,1}^+$ denotes the number of functional chromosomes in homologous group i of the parent cell that do not exhibit instability, and that send a functional daughter into the left daughter cell. The probability that a given functional chromosome produces daughters that segregate in this manner is $p_c p$.
2. $m_{i,2}^+$ denotes the number of functional chromosomes in homologous group i of the parent cell that do not exhibit instability, and that send a non-functional daughter into the left daughter cell. The probability that a given functional chromosome produces daughters that segregate in this manner is $p_c(1-p)$.
3. $m_{i,3}^+$ denotes the number of functional chromosomes in homologous group i of the parent cell that exhibit instability, and send two functional daughters into the left daughter cell. The probability that a given functional chromosome produces daughters that segregate in this manner is $(1-p_c)p^2/2$.
4. $m_{i,4}^+$ denotes the number of functional chromosomes in homologous group i of the parent cell that exhibit instability, and send one functional and one non-functional daughter into the left daughter cell. The probability that a given functional chromosome produces daughters that segregate in this manner is $(1-p_c)p(1-p)$.
5. $m_{i,5}^+$ denotes the number of functional chromosomes in homologous group i of the parent cell that exhibit instability, and send two non-functional daughters into the left daughter cell. The probability that a given functional chromosome produces daughters that segregate in this manner is $(1-p_c)(1-p)^2/2$.
6. $m_{i,6}^+$ denotes the remainder of the functional chromosomes in homologous group i of the parent cell. These chromosomes must exhibit instability and send the daughters into the right daughter cell. The probability that a given functional chromosome produces daughters that segregate in this manner is $(1-p_c)/2$.
7. $m_{i,1}^-$ denotes the number of non-functional chromosomes in homologous group i of the parent cell that do not exhibit instability. The probability that a given non-functional chromosome produces daughters that segregate in this manner is p_c , and

such a chromosome will produce one non-functional daughter that segregates into the left daughter cell, and one non-functional daughter that segregates into the right daughter cell.

8. $m_{i,2}^-$ denotes the number of non-functional chromosomes in homologous group i of the parent cell that exhibit instability, and send the daughters into the left daughter cell. The probability that a given non-functional chromosome produces daughters that segregate in this manner is $(1-p_c)/2$, and such a chromosome will produce two non-functional daughters that segregate into the left daughter cell.
9. $m_{i,3}^-$ denotes the number of non-functional chromosomes in homologous group i of the parent cell that exhibit instability, and send the daughters into the right daughter cell. The probability that a given non-functional chromosome produces daughters that segregate in this manner is $(1-p_c)/2$, and such a chromosome will produce two non-functional daughters that segregate into the right daughter cell.

Now, for $i = 1, \dots, N$, let $\vec{m}_j^+ = (m_{1,j}^+, \dots, m_{N,j}^+)$, where $j = 1, \dots, 6$, and $\vec{m}_j^- = (m_{1,j}^-, \dots, m_{N,j}^-)$, where $j = 1, 2, 3$. If \tilde{n}_i^+ , \tilde{n}_i^- denote the number of functional and non-functional chromosomes in homologous group i , respectively, then we have,

$$\begin{aligned} \tilde{n}_i^+ &= m_{i,1}^+ + m_{i,2}^+ + m_{i,3}^+ + m_{i,4}^+ + m_{i,5}^+ + m_{i,6}^+ \\ \tilde{n}_i^- &= m_{i,1}^- + m_{i,2}^- + m_{i,3}^- \end{aligned} \tag{1}$$

For the left daughter cell, we also have,

$$\begin{aligned} n_i^+ &= m_{i,1}^+ + 2m_{i,3}^+ + m_{i,4}^+ \\ n_i^- &= m_{i,2}^+ + m_{i,4}^+ + 2m_{i,5}^+ + m_{i,1}^- + 2m_{i,2}^- \end{aligned} \tag{2}$$

These two equations imply that $0 \leq m_{i,4}^+ \leq n_i^+$, and so we have that $m_{i,4}^+ = 0, \dots, \min\{n_i^+, n_i^-\}$. We then have that $m_{i,3}^+ = 0, \dots, \lfloor (n_i^+ - m_{i,4}^+)/2 \rfloor$, $m_{i,2}^- = 0, \dots, \lfloor (n_i^- - m_{i,4}^+)/2 \rfloor$, $m_{i,5}^+ = 0, \dots, \lfloor (n_i^- - m_{i,4}^+ - 2m_{i,2}^-)/2 \rfloor$, $m_{i,2}^+ = 0, \dots, n_i^- - m_{i,4}^+ - 2m_{i,2}^- - 2m_{i,5}^+$, $m_{i,6}^+ = 0, \dots, \infty$, $m_{i,3}^- = 0, \dots, \infty$.

Taking into account the transition probabilities listed above, as well as degeneracies in the ways we can choose the segregation patterns of the chromosomes consistent with given values of \vec{m}_j^\pm , we obtain,

$$\begin{aligned} \frac{dx(\vec{n}^+; \vec{n}^-)}{dt} &= -(\kappa(\vec{n}^+; \vec{n}^-) + \bar{\kappa}(t))x(\vec{n}^+; \vec{n}^-) + \\ &2 \sum_{m_{1,4}^+ = 0}^{\min\{n_1^+, n_1^-\}} \sum_{m_{1,3}^+ = 0}^{\lfloor \frac{n_1^+ - m_{1,4}^+}{2} \rfloor} \sum_{m_{1,2}^- = 0}^{\lfloor \frac{n_1^- - m_{1,4}^+}{2} \rfloor} \sum_{m_{1,5}^+ = 0}^{\lfloor \frac{n_1^- - m_{1,4}^+ - 2m_{1,2}^-}{2} \rfloor} \\ &\sum_{m_{1,2}^+ = 0}^{n_1^- - m_{1,4}^+ - 2m_{1,2}^- - 2m_{1,5}^+} \sum_{m_{1,6}^+ = 0}^{\infty} \sum_{m_{1,3}^- = 0}^{\infty} \dots \\ &\sum_{m_{N,4}^+ = 0}^{\min\{n_N^+, n_N^-\}} \sum_{m_{N,3}^+ = 0}^{\lfloor \frac{n_N^+ - m_{N,4}^+}{2} \rfloor} \sum_{m_{N,2}^- = 0}^{\lfloor \frac{n_N^- - m_{N,4}^+}{2} \rfloor} \sum_{m_{N,5}^+ = 0}^{\lfloor \frac{n_N^- - m_{N,4}^+ - 2m_{N,2}^-}{2} \rfloor} \end{aligned}$$

$$\frac{p_c^{-n_N^- - m_{N,4}^+ - 2m_{N,2}^- - 2m_{N,5}^+}}{\sum_{m_{N,2}^+ = 0}^{\infty} \sum_{m_{N,6}^+ = 0}^{\infty} \sum_{m_{1,3}^- = 0}^{\infty} \prod_{i=1}^N} \frac{(m_{i,1}^+ + m_{i,2}^+ + m_{i,3}^+ + m_{i,4}^+ + m_{i,5}^+ + m_{i,6}^+)! (m_{i,1}^- + m_{i,2}^- + m_{i,3}^-)!}{m_{i,1}^+! m_{i,2}^+! m_{i,3}^+! m_{i,4}^+! m_{i,5}^+! m_{i,6}^+!} \times$$

$$[p_c p]^{m_{i,1}^+} [p_c(1-p)]^{m_{i,2}^+} \left[\frac{(1-p_c)p^2}{2}\right]^{m_{i,3}^+} [(1-p_c)p(1-p)]^{m_{i,4}^+} \left[\frac{(1-p_c)(1-p)^2}{2}\right]^{m_{i,5}^+} \left[\frac{1-p_c}{2}\right]^{m_{i,6}^+} \times$$

$$p_c^{m_{i,1}^-} \left[\frac{1-p_c}{2}\right]^{m_{i,2}^-} \left[\frac{1-p_c}{2}\right]^{m_{i,3}^-} \kappa(\vec{m}_1^+ + \dots + \vec{m}_6^+; \vec{m}_1^- + \dots + \vec{m}_3^-) x(\vec{m}_1^+ + \dots + \vec{m}_6^+; \vec{m}_1^- + \dots + \vec{m}_3^-) \quad (3)$$

where $x(\vec{n}^+; \vec{n}^-)$ is the fraction of the population whose genome is characterized by \vec{n}^\pm , and $\bar{\kappa}(t)$ is the mean fitness of the population, or the average of all the first-order growth-rate constants weighted over the population distribution.

We assume that the fitness is either 1 or 0. The fitness is taken to be 1 if each homologous group contains at least one functional chromosome, and if the total number of chromosomes does not exceed some value N_{max} . Otherwise, the fitness is taken to be 0.

If $N_{max} = \infty$, then the fitness only depends on the number of functional chromosomes in each homologous group, so that we may write $\kappa(\vec{n}^+; \vec{n}^-) = \kappa(\vec{n}^+)$. If we define $x(\vec{n}^+)$ to be the total fraction of the population with n_i^+ functional chromosomes in homologous group i , where $i = 1, \dots, N$, then by summing over all possible values of \vec{n}^- , we obtain, after some manipulation,

$$\frac{dx(\vec{n}^+)}{dt} = -(\kappa(\vec{n}^+) + \bar{\kappa}(t))x(\vec{n}^+) + 2 \sum_{m_{1,2}^+ = 0}^{\infty} \sum_{m_{1,3}^+ = 0}^{\lfloor \frac{n_1^+}{2} \rfloor} n_1^+ - 2m_{1,3}^+ \sum_{m_{1,4}^+ = 0}^{\infty} \sum_{m_{1,5}^+ = 0}^{\infty} \sum_{m_{1,6}^+ = 0}^{\infty} \dots \sum_{m_{N,2}^+ = 0}^{\infty} \sum_{m_{N,3}^+ = 0}^{\lfloor \frac{n_N^+}{2} \rfloor} n_N^+ - 2m_{N,3}^+ \sum_{m_{N,4}^+ = 0}^{\infty} \sum_{m_{N,5}^+ = 0}^{\infty} \sum_{m_{N,6}^+ = 0}^{\infty} \prod_{i=1}^N \frac{(m_{i,1}^+ + m_{i,2}^+ + m_{i,3}^+ + m_{i,4}^+ + m_{i,5}^+ + m_{i,6}^+)!}{m_{i,1}^+! m_{i,2}^+! m_{i,3}^+! m_{i,4}^+! m_{i,5}^+! m_{i,6}^+!} \times$$

$$[p_c p]^{m_{i,1}^+} [p_c(1-p)]^{m_{i,2}^+} \left[\frac{(1-p_c)p^2}{2}\right]^{m_{i,3}^+} [(1-p_c)p(1-p)]^{m_{i,4}^+} \left[\frac{(1-p_c)(1-p)^2}{2}\right]^{m_{i,5}^+} \left[\frac{1-p_c}{2}\right]^{m_{i,6}^+} \times$$

$$\kappa(\vec{m}_1^+ + \dots + \vec{m}_6^+) x(\vec{m}_1^+ + \dots + \vec{m}_6^+) \quad (4)$$

Results and Discussion

We may analytically solve for the mutation-selection balance of Eq. (4). To do so, we define the quantity $\langle n_1^+ \dots n_N^+ \rangle$ via,

$$\langle n_1^+ \dots n_N^+ \rangle = \sum_{n_1^+ = 0}^{\infty} \dots \sum_{n_N^+ = 0}^{\infty} n_1^+ \dots n_N^+ x(n_1^+, \dots, n_N^+) \quad (5)$$

Using Eq. (4), we can show, after some manipulation, that,

$$\frac{d\langle n_1^+ \dots n_N^+ \rangle}{dt} = (2p^N - 1 - \bar{\kappa}(t)) \langle n_1^+ \dots n_N^+ \rangle \quad (6)$$

Assuming that the system converges to a steady-state, and assuming that $\langle n_1^+ \dots n_N^+ \rangle$ remains finite, we obtain that $\bar{\kappa} = 2p^N - 1$. Interestingly, this suggests that the mean fitness at steady-state does not depend on the value of p_c .

In Figure 1 we show a plot of the steady-state mean fitness, or $\bar{\kappa} \equiv \bar{\kappa}(t \rightarrow \infty)$, obtained from a stochastic simulation of reproducing organisms. Instead of plotting the mean fitness versus p , we plotted it versus $\mu \equiv N(1-p)$. Note the good agreement between our theory and the simulation results.

Although we do not show plots for smaller values of N_{max} , we do indeed find that, as N_{max} is decreased, the mean fitness drops, and may drop significantly below the $N_{max} \rightarrow \infty$ values.

Our result for the mean fitness hinges on the assumption that $\langle n_1^+ \dots n_N^+ \rangle$ converges to a finite value. However, for $p = 1$, if this is the case, then it follows that $\bar{\kappa} = 1$. This implies, however, that $\langle n_1^+ \dots n_N^+ \rangle \rightarrow \infty$ at mutation-selection balance for $p_c < 1$. To see this, note that since cells with some n_i^+ finite have a positive probability of producing a daughter without any chromosomes in homologous group i , it follows that if $x(\vec{n}^+) > 0$ with some n_i^+ finite, then the subsequent dynamics will lead to the production of genomes with zero fitness, resulting in $\bar{\kappa}(t) < 1$, contradicting the fact that the population is at steady-state.

The way this issue is resolved is if $\bar{\kappa} = 2p^N - 1$ for $p < 1$, so that, by continuity, $\bar{\kappa} = 1$ for $p = 1$. Also, $\langle n_1^+ \dots n_N^+ \rangle$ is finite for $p < 1$, but infinite for $p = 1$. As a result, the convergence of $\langle n_1^+ \dots n_N^+ \rangle$ as a function of N_{max} is not uniform, but rather must be increased as $p \rightarrow 1$. While we do not show plots of $\langle n_1^+ \dots n_N^+ \rangle$, our stochastic simulation results suggest that the analysis above is indeed correct.

We developed a simple mathematical model describing the evolutionary dynamics of a population exhibiting CIN. Our goal at this stage was not to model the dynamics of cancer progression, but rather to understand some basic features of the mutation-selection balance that is associated with CIN. We found that, under certain simplifying assumptions, that the mean fitness of the population did not depend on the extent of CIN. This is clearly an oversimplification, as we are neglecting additional features of CIN such as the fusion and deletion of parts of chromosomes, which can disrupt the normal sequence of genes and thereby affect the fitness of the population.

Another feature we are neglecting is the role of finite population size in our model, which can have a much stronger effect in the case of chromosomal instability. The reason for this is that, if a cell has a relatively large number of homologous pairs in the genome, and its initial ploidy level is low, then for a sufficiently high level of chromosomal instability it becomes likely that one daughter cell will not receive any chromosomes from some homologous group, while the other daughter cell will not receive any chromosomes from another homologous group. This will lead to a sharp initial decay in the population. If the population is sufficiently small, then

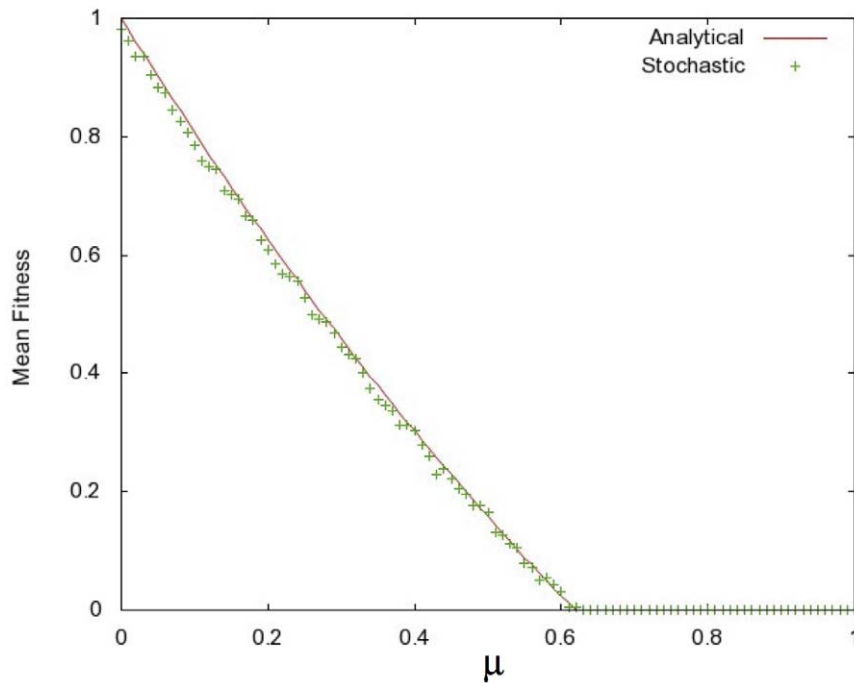


Figure 1. Comparison of the analytical expression for $\bar{\kappa}$ with the values obtained from stochastic simulations. We took $N=3$, $N_{max}=500$, and $p_c=0.5$. The stochastic simulation was run with a population size of 10,000, out to a time of 200 in time steps of 0.1, and with an initially clonal, wild-type population of haploids. doi:10.1371/journal.pone.0026513.g001

this will lead to the extinction of the population. With an infinite population, the population growth rate does initially drop sharply. However, a small fraction of the cells will accumulate chromosomes in every homologous group. As the number of chromosomes in the homologous groups increases, these cells become shielded from the effects of chromosomal instability, and the population recovers as the mean fitness settles into a steady-state value. Thus, for future research we will need to develop stochastic models that explicitly take into account the finite size of the population.

Finally, we should emphasize that we are not the first to develop evolutionary dynamics models that incorporate CIN [2,5,6,7]. Nevertheless, we believe that the model offered here provides additional perspective on the subject, and reveals new and

interesting properties of replicator dynamics as applied to unicellular populations.

Acknowledgments

This paper is dedicated to the memory of my friend, Marianna Tereschenko, who fought but succumbed to Ewing's Sarcoma (ESFT) in 2011.

Author Contributions

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

References

1. Attolini CSO, Michor F (2009) Evolutionary theory of cancer. *Ann New York Acad Sci* 1168: 23–51.
2. Brumer Y, Michor F, Shakhnovich EI (2006) Genetic instability and the quasispecies model. *J Theor Biol* 241: 216–222.
3. Nowak MA, Komarova NL, Sengupta A, Jallepalli PV, Shih le-M, et al. (2002) The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci USA* 99: 16226–16231.
4. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instability in human cancers. *Nature (London)* 396: 643–649.
5. Michor F, Iwasa Y, Nowak MA (2004) Dynamics of cancer progression. *Nature Reviews Cancer* 4: 197–205.
6. Michor F, Iwasa Y, Lengauer C, Nowak MA (2005) Dynamics of colorectal cancer. *Seminars in Cancer Biol* 15: 484–493.
7. Nowak MA, Michor F, Iwasa Y (2006) Genetic instability and clonal expansion. *J Theor Biol* 241: 26–32.