Solvable senescence model with positive mutations

J. B. Coe,[1] Y. Mao,[1,2] and M. E. Cates[3]
[1]*Cavendish Laboratory, Madingley Road, Cambridge CB3 OHE, United Kingdom*
[2]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*
[3]*Department of Physics and Astronomy, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom*
(Received 24 April 2004; published 17 August 2004)

We build upon our previous analytical results for the Penna model of senescence to include positive mutations. We investigate whether a small but nonzero positive mutation rate gives qualitatively different results to the traditional Penna model in which no positive mutations are considered. We find that the high-lifespan tail of the distribution is radically changed in structure, but that there is not much effect on the bulk of the population. The mortality plateau that we found previously for a stochastic generalization of the Penna model is stable to a small positive mutation rate.

DOI: 10.1103/PhysRevE.70.021907 PACS number(s): 87.23.−n, 87.10.+e

I. INTRODUCTION

The Penna model, introduced by Penna in 1995 [1], has been extensively studied through simulation [2,3]. It models senescence in an asexually reproducing population, with the lifetime of an individual encoded in a simple bit-string model of its genome. During reproduction, the bit-string is allowed to copy and mutate, so that the population evolves a distribution of genetic lifespans. The model is powerful enough to reproduce some aspects of observed real-life behavior, and adaptable enough to allow modifications aimed at greater realism [4,5].

One shortcoming of the standard Penna model is that a mutated genome is always worse than its parent [1]. While this is a reasonable assumption, given that the rate of beneficial mutations is small compared to that of harmful ones, very little work has been carried out on Penna models with positive mutations [6]. Here we develop an analytical solution to a Penna model with a small rate of positive mutation. Our analysis builds upon previous work where a solution to a generalized class of stochastic Penna models was presented [7]. All populations are considered in the thermodynamic limit of a large population where statistical fluctuations and effects of discretization can be ignored. The exact limiting behavior found below gives insight into any population large enough for these two effects to be unimportant.

In the standard (deterministic) Penna model, the bit-string determines the lifespan of an organism directly. One way (but not the only way [8]) to interpret this is to say that there is a set of hereditable diseases, which strike the organism at a set of fixed ages during its lifetime. Once it has developed a given number of these, the organism dies. Although the age of death is thus programmed from birth, an individual is in effect reading the bit-string sequentially through its life; it drops dead after encountering the specified number of defective bits. (We call these 1's; the rest are 0's.) So long as it lives, the organism produces offspring at a steady rate; these inherit the parental genome, with a small rate m of harmful mutation $(0\rightarrow 1)$ and, in this paper, an even smaller rate α for positive mutation, $1\rightarrow 0$.

The simplest of all Penna models is that in which an organism dies after developing a single disease. For clarity of argument, we confine our discussion to this version of the Penna model in Secs. II and III. However, the analysis of positive mutations that we present here can be adapted to any of the more sophisticated Penna model variants we have solved [9]. In Sec. IV, we address a stochastic variant showing a mortality plateau [7].

II. A SIMPLE PENNA MODEL

We consider first the single-disease simple Penna model, without positive mutation. The sites along the bit-string are numbered $x=0,1,2,\ldots$; an organism in whose bit-string the first 1 occurs on site l will live for exactly l time steps and thus has "genetic lifespan" l. Note that an organism with $l=0$ (i.e., a 1 at site $x=0$ in the string) will die instantly and never contributes to the population. Note also that x can be thought of as the age of an organism; so long as $x<l$, it remains alive.

The number of organisms with age x and genetic lifespan l at time step j is denoted by $n_j(x,l)$. Newborn organisms with genetic lifespan l can be produced either as copies of organisms with the same l, or as mutated copies of organisms of lifespan $l'>l$. In both cases, sites $0\cdots l-1$ must go unmutated. These dynamics give the following discrete evolution equation:

$$n_{j+1}(0,l) = be^{-\beta l}\sum_{x=0}^{\infty} n_j(x,l) + mbe^{-\beta l}\sum_{l'>l}^{\infty}\sum_{x=0}^{\infty} n_j(x,l'), \quad (1)$$

where $e^{-\beta}=1-m$, with m the (small) mutation rate and b the birth rate. The sum over ages of $n(x,l)$ is defined to be $n(l)$ and can be evaluated at steady state to give $n(l)=l\,n(0,l)$. At steady state, the size of any part of the population is time-independent. Manipulation of Eq. (1) then gives a recursion relation for the relative sizes of subpopulations at steady state,

$$\frac{n(l+1)}{n(l)} = \frac{l+1}{l}\frac{e^{\beta l}-bl}{e^{\beta(l+1)}-b(l+1)e^{-\beta}}. \quad (2)$$

For the population to remain finite, there must be a maximum sustainable genetic lifespan $l_{\max}$; for $l > l_{\max}$, $n(l)=0$ (see Sec. III B for further discussion). Any subpopulation with a smaller genetic lifespan is partly reliant on a flux, by mutation, from longer-lived subpopulations. These two conditions give restrictions on the choice of $l_{\max}$, b, and β, as explained, and confirmed by simulation, in [7],

$$l_{\max} < \frac{1}{1-e^{-\beta}}, \tag{3}$$

$$b = \frac{1}{l_{\max}} e^{\beta l_{\max}}. \tag{4}$$

III. SOLUTION WITH POSITIVE MUTATION

We now introduce a small positive mutation rate α into the simple one-disease Penna model just described. Just as a harmful mutation converts a 0 into a 1, a positive mutation rate α converts a 1 into a 0 with probability α. The rate α is taken to be sufficiently small that there is no chance of multiple positive mutations occurring on the same organism. Further, we assume that α is small compared to the rate β of harmful mutations. This ensures that after the first 1 in an organism's bit-string, the remaining bits are, to high accuracy, all 1's due to the accumulated effects of harmful mutations. In the absence of positive mutation this is clearly the case, as the accumulation of harmful mutations is irreversible and there is no evolutionary pressure for an organism to have healthy sites on its bit-string beyond the site $x=l$. Positive mutations allow this accumulation to be reversed, but so long as $\alpha \ll \beta$, there remains no evolutionary pressure on sites after the first 1; an organism's bit-string can thus be taken to consist of entirely 1's after the first 1. This is a very strong condition, and in steady state it allows (as in Sec. II) any bit-string to be characterized by a single number, the genetic lifespan l.

A. Dynamical equations

When considering mutations on an offspring, we impose that positive mutations take place first, then negative mutations. This is to further enforce the weak nature of positive mutations, preventing them from overriding harmful mutations in the same time step.

Introducing the positive mutation rate α then gives a slightly modified equation for $n_j(0,l)$ in place of Eq (1),

$$\begin{aligned} n_{j+1}(0,l) = {} & (1-\alpha) b e^{-\beta l} \sum_{x=0}^{\infty} n_j(x,l) + \alpha b e^{-\beta l} \sum_{x=0}^{\infty} n_j(x,l-1) \\ & + (1-\alpha) m b e^{-\beta l} \sum_{l'>l}^{\infty} \sum_{x=0}^{\infty} n_j(x,l') \\ & + \alpha m b e^{-\beta l} \sum_{l'' \geqslant l}^{\infty} \sum_{x=0}^{\infty} n_j(x,l''). \end{aligned} \tag{5}$$

The first term on the right corresponds to organisms with genetic lifespan l reproducing with no harmful mutations at sites $x<l$ and no positive mutation at $x=l$. (Note that there are l sites with $x<l$; the factor $e^{-\beta l}$ is the probability of no harmful mutation at any of these.) The second term corresponds to offspring from organisms with lifespan $l-1$ with one positive mutation (at site $l-1$) and no harmful mutations at sites $x<l$. The third term gives mutated offspring from longer-lived organisms (of lifespan $l'>l$) without positive mutation, without harmful mutation for sites $x<l$, but with harmful mutation at site $x=l$. The final term gives mutated offspring from longer-lived organisms of lifespan l'' where positive mutation occurs at site $x=l''$ but is negated by harmful mutation at l, with no harmful mutation for sites $x<l$.

Defining $n(l)=\Sigma_x^{\infty} n(x,l)$ as the total number of individuals of lifespan l, in the steady state we find

$$\begin{aligned} 0 = {} & \left((1-\alpha) b e^{-\beta l} + \alpha m b e^{-\beta l} - \frac{1}{l} \right) n(l) + \alpha b e^{-\beta l} n(l-1) \\ & + m b e^{-\beta l} \sum_{l'>l}^{\infty} n(l'). \end{aligned} \tag{6}$$

Writing a similar expression for $n(l+1)$ allows construction of a recursion relationship between $n(l)$, $n(l-1)$, and $n(l+1)$. [Note that $n(0)$ is known to be zero.] This can be written, for given β, b, and α, as

$$n(l+1) = \frac{l+1}{l} \times \frac{[e^{\beta l} - (1-\alpha-\alpha e^{-\beta}) b l] n(l) - \alpha b l n(l-1)}{e^{\beta(l+1)} - (1-\alpha) b (l+1) e^{-\beta}}. \tag{7}$$

Note that if α is set to zero, Eq. (2) is recovered.

B. Subtleties at large l

Deducing boundary conditions at large l, in the presence of positive muations, is nontrivial. The possibility that an infinitely long-lived organism can evolve from a population of shorter genetic lifespans means that no part of the population is uniquely self-sustaining. This is in contrast to the simple Penna model ($\alpha=0$) where the maximum sustainable genetic lifespan $l_{\max}$ remains finite in the thermodynamic limit of a large population; individuals of $l=l_{\max}$ reproduce unmutated offspring at a rate that precisely balances their own death rate. With positive mutation present, this cannot be true. Organisms of putative maximal lifespan $l_{\max}$ can produce positively mutated offspring with a longer lifespan, and in a large enough population there can be no $l_{\max}$.

Instead we insist that, for a steady-state distribution to be physical, the population must be finite (with the thermodynamic limit taken only at the end). The effects of this requirement can be seen by taking the limit of an arbitrarily small birth rate b. When taking this limit, we do not apply the steady-state conditions (3) and (4), which hold for the simple Penna model only. In the limit of a vanishing birth rate, both the simple Penna model and the Penna model with positive mutations return a recursion relation for the relative sizes of successive subpopulations as follows:

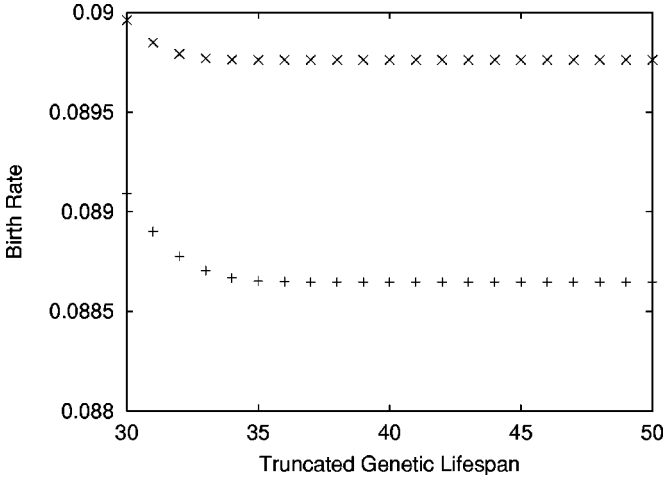

FIG. 1. Birth rate b is plotted against the trucated genetic lifespan l_c for $\alpha=0.001$ (×) and $\alpha=0.005$ (+). In both cases $\beta=\frac{1}{30}$. Genetic lifespan is measured in time steps.

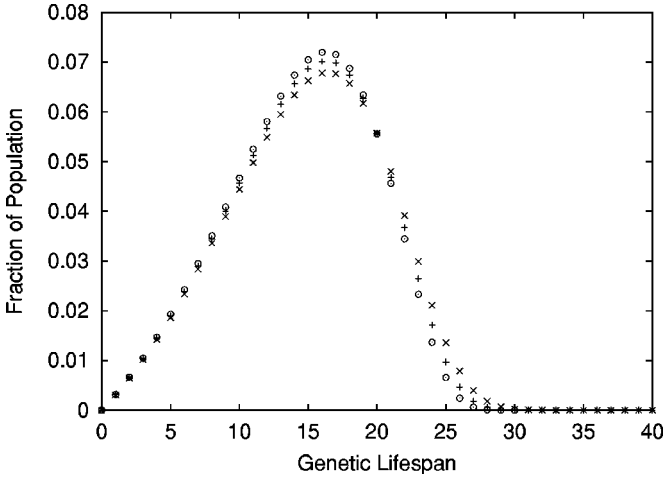

FIG. 2. Lifespan distributions for a Penna model with positive mutations in which $\alpha=0.005$ (×) and $\alpha=0.001$ (+) and without positive mutations (⊙). In each case $\beta=\frac{1}{30}$. Genetic lifespan is measured in time steps.

$$\frac{n(l+1)}{n(l)}=\frac{l+1}{l}e^{-\beta}. \tag{8}$$

This limiting expression is independent of the birth rate so that, as b tends to zero, the population can remain nonzero, with a distribution that is solely dependent on β. This nonintuitive result can be explained by the behavior of organisms with infinite l. An infinitely long-lived organism can reproduce during its lifetime despite an arbitrarily small birth rate b. Moreover, in the limit of $b\to 0$, the mutated offspring of this "super organism" make up the entire population. However, a population with even one super organism cannot be finite since, according to Eq. (8), $n(l\to\infty)/n(1)=0$. It is clear, therefore, that steady-state solutions of the recursion (7) involving the presence of a super organism are not physical, and should be discarded.

To summarize, in the simple Penna model of Sec. II the requirement of a finite population directly imposes an $l_{\max}$. When positive mutations are allowed there is no $l_{\max}$, but the population is still required to be finite. Thus the steady state must not contain super organisms if it is to represent a physical population, and the thermodynamic limit must be taken so as to exclude them.

C. Results

An acceptable steady state can be found by imposing an artificial maximum genetic lifespan l_c beyond which $n(l)$ is taken to be zero. With this artificial cutoff, for specified α and β, b can be found so as to satisfy the steady-state conditions. As l_c approaches infinity, the difference between this approximate steady state and the real steady state will vanish for a sufficiently large population. The convergence of b with increasing l_c is shown in Fig. 1. We require that at l_c the values of $n(l_c)$ and $n(l_c-1)$ predicted by the positive mutation recursion relation satisfy to high order

$$\frac{n(l_c)}{n(l_c-1)}=\frac{\alpha b l_c}{e^{\beta l_c}-\alpha m b l_c-(1-\alpha)b l_c}. \tag{9}$$

This ensures that contributions from longer-lived organisms vanish, preventing the existence of a super organism as l_c tends to infinity. This procedure resolves the paradoxes associated with the thermodynamic limit that were set out above, and allows us to find all the population properties from the recursion derived earlier.

Analytical results, found by this method, for a Penna model with a small but nonzero α are shown in Fig. 2. The population distributions with and without positive mutations are not dissimilar, but there are crucial differences. As discussed, with positive mutations allowed there is no maximum genetic lifespan. What is observed instead is a small but nonzero population beyond what would be $l_{\max}$ in the simple Penna model ($\alpha=0$). The remaining distribution closely resembles $\alpha=0$, with $l_{\max}$ taking its largest allowed value for the specified β.

Note that, depending on the size and history of the population, the maximum lifespan of a simple Penna population may be less than the largest permissible value of l_{max} [7], due to the effect of Muller's rachet [10]. In other words, if a fluctuation within a finite population causes the self-sustaining subpopulation of longest-lived individuals to die off, they can never return and the maximum lifespan is permanently reduced. With positive mutations this loss of fitness due to statistical fluctuations is reversible; positive mutations can act to restore the mean fitness of a population which has fallen below the maximum permitted by the simple Penna model.

IV. EFFECT ON MORTALITY PLATEAU

In [7] we demonstrate that a Penna model with a modified survival function can exhibit a mortality plateau at advanced ages [11,12]. The strict deterministic nature of the original Penna model is relaxed; organisms with genetic lifespan l do not necessarily die at age l, though this is their expected lifespan. The "step function" survival up to age l of the simple Penna model is softened to a Fermi-like function whose width parameter w controls the genetic indeterminacy in the age of death.

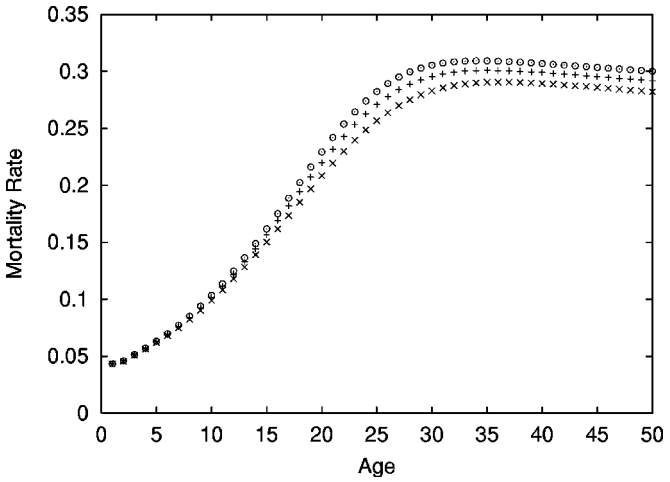

FIG. 3. Mortality rates for a Penna model with positive mutations and Fermi-like survivability function in which $w=0.12$, $\alpha=0.005$ ($\times$), and $\alpha=0.001$ (+) and without positive mutations ($\odot$). In each case $\beta=\frac{1}{30}$. Age is measured in time steps.

We observe that for a small positive mutation rate the mortality plateau is preserved (see Fig. 3). This plateau is the result of organisms living beyond l "on borrowed time" [7]. Even with $\alpha>0$, the majority of the organisms at advanced ages are rare survivors to ages far beyond l, and the mortality of these survivors is almost constant. The variation in mortality rates as α increases comes from the increased fraction of organisms with longer genetic lifespans (see Fig. 2). Since a positive mutation rate acts to increase the average lifespan of the population, a corresponding decrease in mortality is to be expected.

V. CONCLUSION

The nature of the positive mutations we have considered differs from that investigated by Oliveira *et al.* [6] in that it does not increase the mean fitness of the population over time. Our small positive mutation rate is a relatively weak effect, and while it is able to restore the mean fitness of a population if pushed away from steady state (offering limited protection from Muller's ratchet), it cannot improve the population's fitness in a sustained manner. Strong positive mutations, such as those considered by Oliveira *et al.*, are capable of sustained improvement in the fitness of a population but will take place over a far greater time scale than the weak positive mutations we have considered.

Our analysis is limited to a small positive mutation rate. If the positive mutation rate were to become comparable to the negative one [13,14] (it must remain smaller for there to be a possible steady state), then the assumption that there is no evolutionary pressure on bits after the terminal one would become invalid and our description of the Penna string would break down. However, it has long been accepted that positive mutation is extremely rare compared to harmful mutation; as such, this work addresses the relevant regime.

ACKNOWLEDGMENTS

The authors would like to thank D. Khemelnitski, M. Rutter, and J. Trail for useful advice and discussion.

[1] T. J. P. Penna, J. Stat. Phys. **78**, 1629 (1995).
[2] T. J. P. Penna and D. Stauffer, Int. J. Mod. Phys. A **6**, 233 (1995).
[3] S. Moss de Oliveira, Physica A **257**, 465 (1998).
[4] T. J. P. Penna, S. Moss de Oliveira, and D. Stauffer, Phys. Rev. E **52**, R3309 (1995).
[5] A. K. Altevolmer, Int. J. Mod. Phys. A **10**, 717 (1999).
[6] S. Moss de Oliveira, D. Stauffer, P. M. C. de Oliveira, and J. S. S. Martins, Physica A **332**, 380 (2004).
[7] J. B. Coe, Y. Mao, and M. E. Cates, Phys. Rev. Lett. **89**, 288103 (2002).
[8] J. B. Coe, Y. Mao, and M. E. Cates (unpublished).
[9] J. B. Coe and Y. Mao, Phys. Rev. E **67**, 061909 (2003).
[10] H. J. Muller, Mutat Res. **1**, 2 (1964).
[11] S. D. Pletcher and J. W. Curtsinger, Evolution (Lawrence, Kans.) **52**, 454 (1998), and citation therein.
[12] J. S. Weitz and H. B. Fraser, Proc. Natl. Acad. Sci. U.S.A. **98**, 15383 (2001), and citations therein.
[13] W. Hwang, P. L. Krapivsky, and S. Redner, Phys. Rev. Lett. **83**, 1251 (1999).
[14] W. Hwang, P. L. Krapivsky, and S. Redner, J. Math. Biol. **44**, 375 (2002).