4-7-2010

Modeling Growth and Telomere Dynamics in Saccharomyces Cerevisiae

Peter Olofsson
Trinity University, polofsso@trinity.edu

Alison A. Bertuch

Modeling growth and telomere dynamics in *Saccharomyces cerevisiae*

Peter Olofsson* and Alison A. Bertuch†

September 15, 2009

Abstract

A general branching process is proposed to model a population of cells of the yeast *Saccharomyces cerevisiae* following loss of telomerase. Previously published experimental data indicate that a population of telomerase-deficient cells regain exponential growth after a period of slowing due to critical telomere shortening. The explanation for this phenomenon is that some cells engage telomerase-independent pathways to maintain telomeres that allow them to become "survivors." Our model takes into account random variation in individual cell cycle times, telomere length, finite lifespan of mother cells, and survivorship. We identify and estimate crucial parameters such as the probability of an individual cell becoming a survivor, and compare our model predictions to experimental data.

Keywords: Telomere, branching process, yeast, Fibonacci sequence
AMS 2000 Mathematics Subject Classification Codes: 60G99, 60K99, 62P10, 92D25

*Corresponding author. Trinity University, Mathematics Department, One Trinity Place, San Antonio, TX 78212. Email:polofsso@trinity.edu

†Baylor College of Medicine, Department of Pediatrics, One Baylor Plaza, Houston, TX 77030

1 Introduction

Telomeres are specialized structures found at the ends of linear chromosomes, which consist of protein bound, tandem repeats of short DNA sequences called telomeric repeats (Denchi, 2009). Telomeres function to prevent the factors that detect and repair chromosome breaks from acting upon natural chromosome ends. When telomere function is lost, either through the loss of telomeric repeats or certain proteins that associate with them, telomeres become substrates of the DNA double strand break repair machinery, which results in telomere-telomere fusions and genomic instability.

During semi-conservative DNA replication of linear chromosomes, there is loss of terminal telomeric DNA due to the so-called end replication problem (Baird, 2008). To offset this loss, certain cell types, such as germ cells, express telomerase, which catalyzes the addition of telomeric repeats onto chromosome termini. Most human somatic cells, however, express insufficient or undetectable levels of telomerase and as a consequence, telomeres shorten in these cells with each cell cycle. Eventually, one or more telomeres reach a critical length that triggers an irreversible halt in cellular proliferation known as cellular senescence. Rarely, cells bypass the induction of senescence, but go on to die due to the eventual loss of telomeric function and secondary genomic instability. A rare population of cells, however, may bypass this crisis by reactivating telomerase or engaging a recombination-based mechanism of telomere maintenance, allowing for continued proliferation. These processes of telomere dysfunction followed by telomere stabilization are thought to represent important steps in the development of cancer, and thus are of substantial interest.

The budding yeast, *Saccharomyces cerevisae*, is a genetically tractable model organism for the study of telomere biology. Telomerase-deficient strains can be readily generated and telomere loss rates of 3–5 base pairs have been determined in the absence of telomerase (Lundblad and Szostak, 1989; Singer and Gottschling, 1994). As in higher eukaryotes, telomerase deficiency in yeast results in eventual cellular senescence, and rare cells that evade this fate can be isolated (Lundblad and Blackburn, 1993). Two major classes of these telomerase-independent/post-senescence "survivors" have been characterized, which can be distinguished in part by their telomere structures and growth properties. Chromosome ends in type I survivors have extensively amplified subtelomeric Y' elements and terminate with short tracts of telomeric repeats, whereas the chromosome ends in type II survivors exhibit

a much lesser degree of Y' amplification and terminate with heterogeneous and long telomeric repeat tracts (Lundblad and Blackburn, 1993; Teng and Zakian, 1999). Type II telomeres are akin to those that are observed in cancer cells that don't activate telomerase, but rather use an alternative lengthening of telomeres (ALT) pathway for telomere maintenance (Bryan *et al.*, 1997). Although type I survivors arise more frequently than type II survivors, they exhibit severely impaired growth (Teng and Zakian, 1999). Type II survivors, in contrast, grow comparable to wild type cells, and consequently predominate with serial propagation of telomerase-deficient strains in liquid culture. Although many of the proteins that are required for or influence one or both of the pathways have been identified, the molecular events that drive the telomere recombination choice and the rate they occur are poorly defined.

Previous mathematical models of the process of telomere loss include Levy *et al.*, 1992; Arino *et al.*, 1995 and 1997; Olofsson and Kimmel, 1999 and 2005; Rubelj *et al.*, 1999; Tan, 1999; Olofsson, 2000; Op Den Buijs *et al.*, 2004; Dyson *et al.*, 2007; and Portugal *et al.*, 2008. In the present paper we develop a model of cellular proliferation in response to telomere dynamics in yeast that takes into account the facts that mothers and daughters are distinguishable individuals, that the proliferative potential of a cell is limited by factors other than telomere length, and, in particular, that population growth can be restored after a period of slowing due to the bypass mechanisms mentioned above.

2 The branching process model

We propose a general branching process (Haccou *et al.*, 2005; Jagers and Nerman, 1984) where individuals reproduce by budding. The times between consecutive budding events (i.e., the cell cycle times) are assumed to be independent random variables with the same distribution. This is where the general branching process comes into play: a mother cell produces daughter cells at different times during her life which is relevant to budding yeast and different from, for example, the binary fission of bacterial reproduction that can be modeled by simpler branching processes. As we are also keeping track of loss of telomeric DNA, we are in fact using a multitype process (Jagers, 1992) where type accounts for telomere length, however, specific results for multitype processes will not be needed. An individual cell contains 16 chromosomes so it has 32 telomeres. In cells that express telomerase, the

number of telomeric repeats present varies to a certain extent from end to end resulting in a distribution of telomere lengths around a strain-specific point (Shampay and Blackburn, 1988). In the absence of telomerase, the rate of loss of telomeric DNA follows a probability distribution over a range from 3–5 base pairs per end per cell division (Lundblad and Szostak, 1989; Singer and Gottschling, 1994). We therefore define a "telomere unit" to be 4 base pairs and, for the time being, assume that one telomere unit is lost per division. We let the type of a cell be the number of remaining telomere units.

For increased clarity, we will first describe a simplified model and successively extend it to incorporate all the features we need, leading up to the description in Proposition 2.2. The population starts from a single telomerase-deficient cell of type n and upon completion of the cell cycle, this cell has produced one daughter cell. As all telomeres have shortened by one unit in the preceding round of DNA replication and are randomly allocated to mother and daughter cell, it is reasonable to assume that both mother and daughter, after division, have type $n-1$. For simplicity, we refer to the two cells as the first generation. When a cell reaches type 0 (the critical telomere length mentioned in Section 1), it stops dividing and becomes senescent, unless one of the alternative mechanisms for telomere maintenance described in Section 1 is established. We will address this situation later. Senescent cells remain in the population but do not further reproduce. As the ultimate fate of a senescent cell is likely to be death, we could let each senescent cell remain in the population for a time following some probability distribution. However, as the mean in this distribution is likely far to exceed the duration of the experiment, senescent cells can for practical purposes be considered simply nondividing. Again, it is not difficult to adjust the model but any possible gain would hardly outweigh the increased complexity of the model.

We assume that cell cycle times are independent random variables with the common cumulative distribution function (cdf) F. The main quantity of interest is the number of cells in the population at time t, which we denote by Z_t. The population starts from a single cell of type n at time $t = 0$. In this basic model, the expected value of Z_t is

$$E[Z_t] = 1 - F(t) + \sum_{k=1}^{n-1} 2^k \left(F^{*k}(t) - F^{*(k+1)}(t) \right) + 2^n F^{*n}(t)$$

where F^{*k} denotes k-fold convolution of F, that is, the cdf of the sum of k

cell cycle times. The factor 2^k is the expected number of cells in the kth generation (recall our use of "generation" mentioned above). Now note that each cell in the kth generation is present in the population if the sum of k cell cycle times is less than t but the sum of $k+1$ cell cycle times is greater than t. As the probability of this event is $F^{*k}(t) - F^{*(k+1)}(t)$, the expected number of cells from the kth generation that are present at t equals

$$2^k \left(F^{*k}(t) - F^{*(k+1)}(t) \right)$$

and, noting that cells of type 0 do not further reproduce, summing over k gives the expression for $E[Z_t]$. Note that $E[Z_t] \to 2^n$ as $t \to \infty$ so 2^n is the final number of cells.

Additional factors influence the replicative capacity of budding yeast cells, such that a given cell will undergo a finite number of cell divisions even when telomerase is expressed and telomere length is maintained (D'Mello and Jazwinski, 1991). We denote this maximum number of divisions by n_0 and assume for now that it is constant. Thus, if a cell has telomere length k, it will produce $\min(k, n_0)$ more daughter cells, which means that we must break up the expression for $E[Z_t]$ above in a term for $k \leq n_0$ and one for $k > n_0$. The general expression is

$$E[Z_t] = 1 - F(t) + \sum_{k=1}^{n-1} m(k) \left(F^{*k}(t) - F^{*(k+1)}(t) \right) + m(n) F^{*n}(t)$$

where $m(k)$ replaces 2^k as the expected number of cells in the kth generation.

As long as $k \leq n_0$, we still have $m(k) = 2^k$ and for $k > n_0$ we describe a recursive scheme that enables us to compute $m(k)$. To that end, in any given generation, let k_j be the number of cells that are able to reproduce j times for $j = 0, 1, ..., n_0$. Each cell with $j > 1$ produces a daughter cell that is able to reproduce n_0 times and is then itself able to reproduce another $j-1$ times. Cells with $j = 0$ remain unchanged. For $n_0 \leq k \leq n$, the transition from generation $k-1$ to generation k is therefore as follows:

$$\begin{cases} \text{Generation } k-1: & (k_0, k_1, ..., k_{n_0-1}, k_{n_0}) \\ \text{Generation } k: & (k_0 + k_1, k_2, ..., k_{n_0}, \sum_{i=1}^{n_0} k_i) \end{cases}$$

and after relabeling in generation k to $(k_0, k_1, ..., k_{n_0})$ again, we have

$$m(k) = \sum_{j=0}^{n_0} k_j$$

The initial configuration in generation 0 is $(0, 0, ..., 0, 1)$ since there is one cell that is able to divide n_0 times. For $k > n$, all cells are senescent so $m(k)$ stays constant and we have $m(k) = m(n)$ for $k \geq n$. It is interesting to note that the sequence $\{k_1, ..., k_{n_0}\}$ (excluding k_0) constitutes an "n_0-nacci" sequence, that is, a generalized Fibonacci sequence where each new number is obtained by adding the previous n_0 numbers, starting from $(0, 0, ..., 0, 1)$. Hence, it is possible to obtain explicit expressions for the k_j (Flores, 1967). We state this observation as a proposition that is interesting in its own right; for our computational purposes, however, the recursive formula above is sufficient.

Proposition 2.1 *Consider the vector $(k_0, k_1, ..., k_{n_0})$ in the kth generation of the branching process above for $k \leq n$. Let F_j denote the jth n_0-nacci number and set $F_0 = F_1 = ... = F_{n_0-1} = 0$, $F_{n_0} = 1$. Then $(k_0, k_1, ..., k_{n_0})$ equals*

$$\left(\sum_{j=0}^{k} F_j, F_{k+1}, ..., F_{k+n_0} \right)$$

To account for the observed phenomenon of restored exponential growth, we assume that cells that have reached type 0 have the possibility to turn into "survivors" or become senescent. The expression for $m(k)$ remains the same for $k \leq n$, but for $k > n$ it changes since cells of type 0 may now escape senescence and keep reproducing. We assume that a cell of type 0 becomes a survivor with probability p and that the survivor status is inherited by all of its daughter cells. Thus, each survivor starts a population of survivors where telomere length is generally maintained and we assume the only limiting factor is the lifespan n_0. The nonsurvivors turn senescent.

To arrive at an expression for the expected number of cells in generation k for $k > n$, we first need to figure out how many cells in generation n that are eligible to become survivors, and for each of those, how many more times it is able to reproduce. There is a total of $m(n)$ cells, but many of those cells have ceased reproducing after reaching the threshold imposed by the maximum number of offspring n_0. For now, we assume that the only cells

in generation n that are eligible to become survivors are those that have reached type 0 and are still able to reproduce at least once, an assumption that will later be relaxed. To account for those cells, we use the recursive scheme above and in generation n arrive at the vector $(k_0, k_1, ..., k_{n_0})$. We then have $k_1 + ... + k_{n_0}$ cells that are eligible to become survivors. The cells that do not become survivors become senescent, joining the k_0 cells that are no longer able to reproduce. Thus, in generation n there is an expected number of $k_0 + (1-p)(k_1 + ... + k_{n_0})$ senescent cells and an expected number of $p(k_1 + ... + k_{n_0})$ survivors. Since the survivors differ in the number of possible daughter cells, they do not contribute equally to the population and we need to deal with each k_j separately.

Let us start with the k_1 cells that may reproduce once more. Each such cell produces one daughter cell that starts a population now capable of maintaining telomere length, limited only by the reproductive limit n_0. The mother cell turns senescent. Thus, in generation $n + j$, the expected size of the population stemming from the original survivor is $1 + m(j-1)$. The k_2 cells that may reproduce twice more contribute an expected number of $1 + m(0) = 2$ cells in generation $n + 1$ and an expected number of $1 + m(1) + m(0)$ cells in generation $n + 2$. After that, the mother turns senescent and, in generation $n + j$, the expected size stemming from an initial survivor of this kind is $1 + m(j-1) + m(j-2)$. Generally, a survivor that can reproduce i more times contributes an expected number of $m_{i,j}$ cells to generation $n + j$ where

$$m_{i,j} = 1 + \sum_{l=1}^{\min(i,j)} m(j-l)$$

Further, let $s(n)$ be the number of cells in generation n that are eligible to become survivors ($s(n) = k_1 + ... + k_{n_0}$ in the notation above) and let $d(n)$ be the number of cells in generation n that are already senescent from before ($d(n) = k_0$ in the notation above). We are now ready for the final expression.

Proposition 2.2 *Let $M(k)$ denote the expected number of cells in generation k in a branching process with survivors as described above. Then*

$$M(k) = \begin{cases} 2^k & \text{for } k \leq n_0 \\ m(k) & \text{for } n_0 < k \leq n \\ d(n) + (1-p)s(n) + p\sum_{i=1}^{n_0} m_{i,k-n} & \text{for } k > n \end{cases}$$

and the expected number of cells in the population at time t is

$$E[Z_t] = 1 - F(t) + \sum_{k=1}^{\infty} M(k)\left(F^{*k}(t) - F^{*(k+1)}(t)\right)$$

Note that all senescent cells are included in the expression for $M(k)$ even though they have ceased to reproduce. However, the expression for $E[Z_t]$ is still valid by the telescoping properties of the terms in the sum; once a cell is senescent, it stays in the population forever which is mathematically equivalent to saying that it "replaces itself" at time intervals with cdf F.

3 Data, estimation, and model fitting

To fit and calibrate our model, we use data from Bertuch and Lundblad (2004). In this study, telomerase-deficient haploid cells (containing one set (1n) of chromosomes) were generated from telomerase-proficient diploid (2n) cells, in which one of the two copies of the gene that encodes the catalytic subunit of telomerase (*EST*2) was replaced by a selectable marker (*est*2Δ). Upon meiosis of a parental diploid cell, four daughters would be formed - two that would inherit the wild-type *EST*2 copy and two that would inherit the deleted *est*2Δ copy, which would be telomerase-proficient and telomerase-deficient, respectively. The parental diploid strain also contained one wild-type and one deleted copy of a gene that encodes an enzyme that degrades dysfunctional telomeres, *EXO1*. These similarly segregated 2:2 to the offspring haploids. The individual cells were separated from one another on solid media and allowed to grow into colonies for 4 days. After four days, the individual colonies were resuspended in their entirety in 5 milliliters (ml) media and allowed to grow for 1 day. Cells counts were performed and the cultures subsequently diluted into fresh media at a concentration of 10^5 cells/ml. The cultures were grown for 22 hours after which time cell counts

were obtained. The dilution, growth, and cell counting protocol was repeated every 22 hours for a total of 9 days. The genotypes of the cultures (*EST*2 vs. *est*2Δ and *EXO1* vs. *exo1*Δ) were ultimately determined.

We convert the growth rate data to logarithmic population size data. For example, if the initial count by day 1 is 10^8 cells per ml, the logarithmic population size is $\log(5 \cdot 10^8) \approx 20.0$. Next, suppose the count by day 2 is again 10^8. This means that the 10^5 cells (per ml) that were harvested at the end of day 1 grew to 10^8 cells (per ml) by day 2, that is, a 1000-fold increase. We then estimate the logarithmic population size to be $20.0 + \log(1000) \approx 26.9$, and so on and so forth.

First, we estimate cell cycle parameters. For this purpose, we use the data for the wild-type, telomere-proficient strain. Since the data indicate that this population is in steady exponential growth, we use asymptotic results for exponentially growing branching processes to estimate the mean and variance of the population (Haccou *et al.*, 2005; Jagers and Nerman, 1984). For this purpose we need to describe the reproduction process $\xi(t)$ which gives the number of daughter cells a mother cell has up to age t. Let the times between consecutive births (i.e., cell cycle times) be independent random variables with the common cdf F. Denote the time of the kth birth by T_k and assume there is a total of n_0 daughter cells. The reproduction process becomes

$$\xi(t) = \sum_{k=1}^{n_0} I\{T_k \leq t\} \tag{3.1}$$

where I denotes indicator function. The mean of $\xi(t)$ is

$$\mu(t) = \sum_{k=1}^{n_0} P(T_k \leq t) = \sum_{k=1}^{n_0} F^{*k}(t)$$

where F^{*k} denotes the k-fold convolution of F with itself, that is, the cdf of T_k. For simplicity, we assume that n_0 is constant, however, a random n_0, call it N_0, is easily dealt with. The reproduction process remains

$$\xi(t) = \sum_{k=1}^{N_0} I\{T_k \leq t\}$$

and its mean is

$$\mu(t) = \sum_{n_0=1}^{\infty} \sum_{k=1}^{n_0} F^{*k}(t) P(N_0 = n_0)$$

We henceforth assume that n_0 is constant (see the discussion after (3.5) for some justification of this assumption). Note that we can write

$$\mu(t) = \int_0^t \mu(du)$$

where

$$\mu(du) = \sum_{k=1}^{n_0} f^{*k}(u)du$$

if F is a continuous distribution with probability density function (pdf) f. The Laplace transform of ξ is defined as

$$\widehat{\xi}(s) = \int_0^\infty e^{-st}\xi(dt)$$

which has expected value

$$\widehat{\mu}(s) = E[\widehat{\xi}(s)] = \int_0^\infty e^{-st}\mu(dt)$$

and the asymptotic growth rate of the process is given by the Malthusian parameter α which is defined through the relation $\widehat{\mu}(\alpha) = 1$. Finally, let

$$\beta = \int_0^\infty te^{-\alpha t}\mu(dt)$$

The asymptotic mean and variance of Z_t are then given by

$$E[Z_t] \sim e^{\alpha t}\frac{1}{\alpha\beta} \tag{3.2}$$

and

$$\mathrm{Var}[Z_t] \sim e^{2\alpha t}\frac{\mathrm{Var}[\widehat{\xi}(\alpha)]}{\alpha^2\beta^2(1-\widehat{\mu}(2\alpha))} \tag{3.3}$$

For technical details and conditions, see Jagers and Nerman, 1984. In order to estimate cell cycle parameters, we make the assumption that cell cycle times follow a gamma distribution. This distribution is a flexible two-parameter family that is commonly used to model lifetimes (Oprea and Kepler, 2001; Larsson *et al.*, 2008). Specifically, if the parameters are a and b, the pdf is

$$f(t) = e^{-bt} b^a \frac{t^{a-1}}{\Gamma(a)}$$

where $\Gamma(a)$ is the gamma function. The Laplace transform of the $\Gamma(a, b)$ distribution is

$$\widehat{f}(\alpha) = \int_0^\infty e^{-\alpha t} f(t) dt = \left(\frac{b}{\alpha + b}\right)^a$$

As T_k is the sum of k independent $\Gamma(a, b)$ variables, we get $T_k \sim \Gamma(ka, b)$ and as the Laplace transform of the sum of independent random variables equals the product of the individual Laplace transforms, (3.1) gives

$$E[\widehat{\xi}(\alpha)] = \sum_{k=1}^{n_0} \left(\frac{b}{\alpha + b}\right)^{ka} = \frac{1 - \left(\frac{b}{\alpha+b}\right)^{a(n_0+1)}}{1 - \left(\frac{b}{\alpha+b}\right)^a} - 1 \tag{3.4}$$

Our sample contains 18 observed values of Z_t where $t = 22$ hours, the sample mean is $\bar{Z} = 2136$, and sample variance is $s_Z^2 = 261952$. Using (3.2) and (3.3) we can solve the equations

$$E[Z_{22}] = 2136 \quad \text{and} \quad \text{Var}[Z_{22}] = 261952$$

together with $E[\widehat{\xi}(\alpha)] = 1$, to get estimates of α, a, and b. Note that there is also a fourth parameter n_0 so there is no unique solution to the 3 equations. However, we know that n_0 is on average about 25 (Sinclair *et al.*, 1997 and 1998). We also know that the mean cell cycle time (which is a/b in the gamma distribution) is on the order of 90–120 minutes (depending on strain background, growth conditions, etc) and with these restrictions we can get approximate estimates for α, a, and b. Our goal is not to get as precise estimates as possible, rather, we wish to examine whether our model can qualitatively reproduce experimental results. As the expressions above are quite complicated, we suggest the following simplification.

The lifespan n_0 of a cell is 25 and at the beginning of each 22-hour time period, cells are distributed according to Proposition 2.1. By our day 0, there have been about 30 population doublings and by day 8, there have been about 100 doublings; hence, we have k in the range 30–100 (n can be considered infinite as telomere length is maintained in the wild-type population). With n_0 and k in these ranges, Proposition 2.1 shows that the fraction of cells that are not able to keep producing daughter cells for the full 22 hours is very

small. For example, more than 99% of cells are able to produce at least 18 daughter cells which with overwhelming probability takes far more than 22 hours. In fact, almost 50% of cells are newborn, another 25% are second generation, and so on. In conclusion, almost all cells are able to produce daughter cells for the full 22-hour period so we can view the population as a binary splitting process, also known as a Bellman–Harris process, where lifetimes are $\Gamma(a, b)$ and each cell produces 2 daughter cells (one of which is actually the mother cell).

In such a Bellman–Harris process, the mean reproduction process simplifies to

$$\mu(t) = 2F(t)$$

which has Laplace transform

$$\widehat{\mu}(\alpha) = \frac{2b^a}{(\alpha + b)^a}$$

and setting this expression equal to 1 gives an explicit solution for the Malthusian parameter as

$$\alpha = b(2^{1/a} - 1)$$

Note that this relation implies that

$$\left(\frac{b}{\alpha + b}\right)^a = \frac{1}{2}$$

which, by (3.4) gives

$$E[\widehat{\xi}(\alpha)] = 1 - \left(\frac{1}{2}\right)^{n_0} \tag{3.5}$$

As we have $E[\widehat{\xi}(\alpha)] < 1$ for any n_0, the Bellman–Harris approximation in fact overestimates α, which makes intuitive sense since it disregards the restrictions imposed by the finite lifespan of cells. However, for realistic values of n_0, we have $E[\widehat{\xi}(\alpha)] \approx 1$, confirming that our approximation is reasonable. Equation (3.5) also provides some justification for assuming a constant n_0; over the range of values that have been experimentally determined (Sinclair *et al.*, 1997 and 1998) the effect of varying n_0 is negligible.

Another source of overestimation of α is the fact that the cells in our data set of daily counts are not newborn, but have lived for a while at the

time of harvesting. It is possible to use stable population theory (Jagers and Nerman, 1984) to establish that the age of a harvested cell has pdf given by

$$f(t) = ce^{-\alpha t}(1 - F(t))$$

where F is the cdf of the cell cycle time and c is a normalizing constant. However, such detailed modeling is not likely to yield much improvement considering all the other sources of error and uncertainly in the data sets. It is enough for our purposes to get a rough estimate of α, knowing that it is likely to be a slight overestimate which also has bearing on the estimates of a and b.

In the Bellman–Harris approximation we also get the explicit expression

$$\beta = 2\int_0^\infty te^{-\alpha t}f(t)dt = \frac{2ab^a}{(\alpha + b)^{a+1}}$$

and equations (3.2) and (3.3) become

$$E[Z_t] \sim e^{\alpha t}\frac{1}{2\alpha\beta}$$

$$\mathrm{Var}[Z_t] \sim e^{2\alpha t}\frac{2\widehat{\mu}(2\alpha) - 1}{4\alpha^2\beta^2(1 - \widehat{\mu}(2\alpha))}$$

for details see Haccou *et al.*, 2005 or the classic Harris, 1963. We have thus eliminated n_0 and have been able to express α as a function of a and b, hence, the equations $E[Z_{22}] = 2669$ and $\mathrm{Var}[Z_{22}] = 409300$ can be solved to yield the moment estimates $a = 17$ and $b = 9$, rounded to nearest integers. The mean cell cycle time is estimated to $a/b = 1.9$ hours with an approximate 95% confidence interval 1.9 ± 0.2 and the estimate of the Malthusian parameter is $\alpha = 0.37$.

As a side note, let us point out that the Bellman–Harris approximation is bad for small values of n_0. For example, if $n_0 = 2$, properties of the regular Fibonacci sequence shows that the distribution in Proposition 2.1 quickly converges to the proportions $(1, \phi, 1)$ where $\phi \approx 0.62$ is the golden section. Thus, in that case, as many as 38% of cells have reached the end of their lifespan and are no longer able to produce daughter cells.

We are now ready to compare our model to data. For that purpose, we use the data for the *est*2Δ *EXO1* strain where estimated cell counts exist for 9 consecutive days (22-hour periods). The first such count is after 5 days (see the description of the experiment above) so we have data for days 5 to

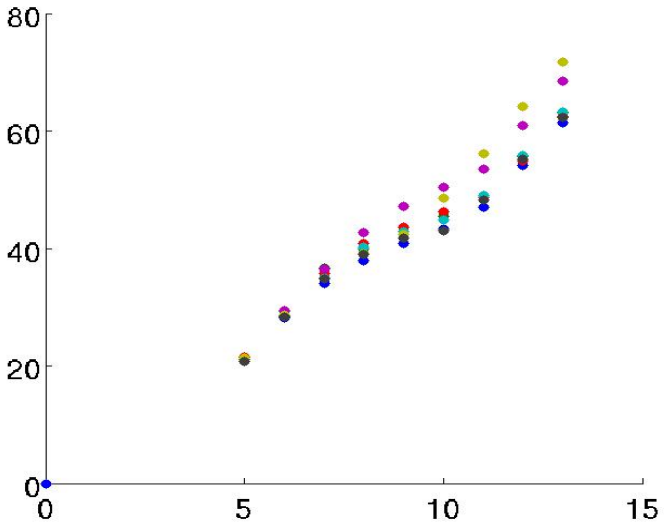

Figure 1: Cell counts of 7 yeast cultures for 13 days (logarithmic scale).

13, and as each population starts from one single cell, we also have a data point for day 0. Figure 1 displays the 7 data sets on a logarithmic scale. It is noticeable how the growth is slower in the beginning. Between days 0 and 5 there have been about 30 population doublings ($21/\log 2 \approx 30$), whereas it only takes a little over 3 days for an additional 30 doublings. Possible reasons for this phenomenon are a lag in the time from when a single cell is plated onto rich media and resumes mitotic cell growth, and differences in growth on solid (the first 4 days) as compared to liquid media (the subsequent days). All of these factors are unknown so rather than trying to model the initial 5 days, we relabel day 5 as day 0 and rescale the data so that each data set has one cell on day 0 (by simply dividing each data set by the cell count on its day 0). These relabeled and rescaled data are displayed in Figure 2 together with the mean (solid line). Note that the line is the logarithm of the average population counts which is why it is not "in the middle" of the dots (the logarithm of the average is always greater than the average of the logarithms, by Jensen's inequality). Hence, we want our model to give a result such that $\log E[Z_t]$ resembles the solid line in Figure 2.

Other than the already estimated n_0, a, and b, our model also contains the

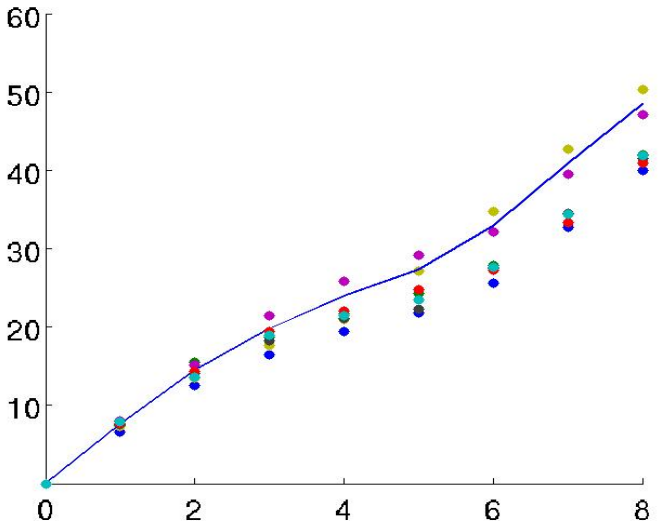

Figure 2: Data from Figure 1 rescaled, with mean.

parameters n and p. As for n, the initial telomere length at day 0, note that the population by then has passed through about 30 population doublings and 10^5 cells are harvested to start the population which means that rather than one value of n, we have a distribution over many different values. There is also random variation due to the fact that telomere loss might not be constant, and due to the asynchronous cell division. Figure 2 indicates that telomere shortening starts significantly influencing population growth around days 3–4, when there has been about 30–35 population doublings so we let the average number of telomere units on day 0 be in that range.

Another potential source of variation comes from the way in which cells turn into survivors. It is currently unknown how survivorship is related to telomere length, but it is certainly possible that a cell may achieve survivor status not only when it is about to become senescent, but also while it still proliferating. Thus, in our model we can let cells become survivors not only when their type is 0, but also at type $1, 2, 3$, etc. Rather than incorporating this directly into the model, we can view a population where cells may become survivors at type j as a population starting with type $n - j$, and average in some suitable way over such populations. This is possible as long as we deal

with the expected value $E[Z_t]$; if we were also to compute the variance of Z_t, we lose linearity and with it, many of he suggested simplifications.

Thus, all of the sources of variation mentioned above can be incorporated by letting n have a probability distribution over some suitable range of values and compute the expected value $E[Z_t]$ when n is chosen accordingly. There is much uncertainty and many unknown factors; we shall assume that n has a probability distribution on some range of integers where the mean is in the range $30 - 35$.

The final unknown parameter is p, the probability that a cell becomes a survivor, for which we try different values to examine the fit to data. Figure 3 displays the 7 data sets (dots), the mean (dashed red line), and 3 lines computed from our model where n has a uniform distribution with mean 35. The solid line, which presents the best fit, has $p = 2 \cdot 10^{-5}$ whereas the 2 dashed lines have $p = 10^{-3}$ (upper) and $p = 2 \cdot 10^{-6}$ (lower), respectively. In Figure 4, n instead has a binomial distribution with mean 35.

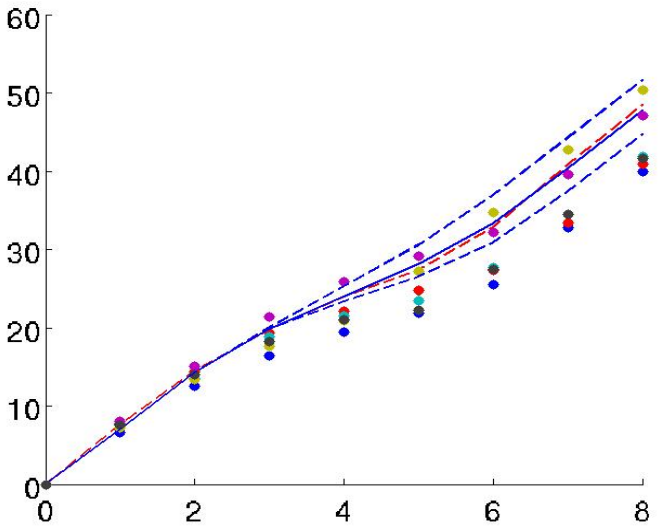

Figure 3: Cell count data sets with mean (dashed red line) and model predictions for different values of p (blue lines) with uniform n.

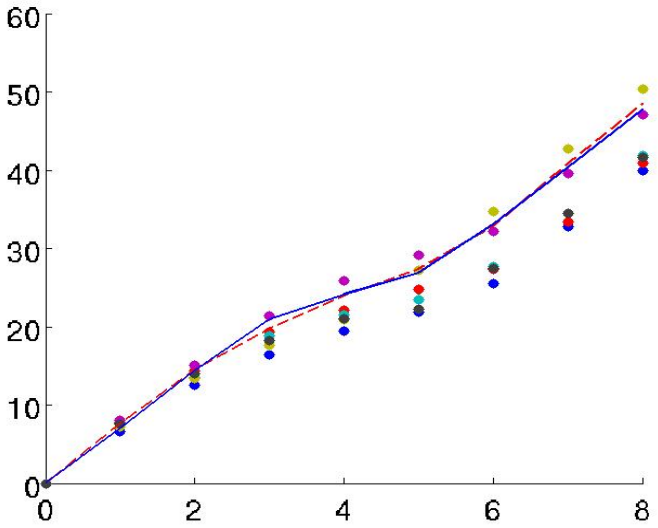

Figure 4: Data, mean, and model prediction with binomial n.

4 Discussion

We used a general branching process to model growth and telomere dynamics in the yeast *S. cerevisiae*. The model accounts for random variability in cell cycle times and shortening of telomeres by "telomere units" of 4 bp. Cells also have the possibility to turn into survivors and we assumed that this happens independently with probability p in cells that are approaching a critical telomere length. The model also takes into account the known fact that cells undergo a finite number of cell divisions even when telomerase is expressed and telomere length is maintained. The general branching process easily deals with this problem since it allows a mother cell to repeatedly give birth to daughter cells until the maximum number is reached.

Cell cycle times were assumed to follow a gamma distribution whose parameters were estimated from data on telomerase-proficient yeast by applying asymptotic branching process results. We then used these estimates together with values of initial cell lifespan and telomere length, taken from the literature, to predict the development of mean population size over time. Different values of the remaining parameter p were tried and the best fit to data occurred for p on the order of $2 \cdot 10^{-5}$. It should be emphasized that our goal

was not to get as precise estimates as possible; rather, we aimed at building a stochastic model that incorporates the necessary parameters to describe cellular proliferation in response to telomere attrition and maintenance. Our model agrees with data very well, both qualitatively and quantitatively, but in order to perform a more sophisticated analysis, a new experimental design needs to be implemented to eliminate some of the uncertainties present in the current data.

We have made several simplifications in the model. Here, we considered the average replicative life span of pre- and post-senescent survivors to be equivalent, however, recent work on survivors indicates that this limit on replicative life span is reduced in type II survivors (Chen *et al.*, 2009). We also disregarded the possibility that cell cycle times slow down progressively during senescence progression or replicative aging (Sinclair *et al.*, 1998). In addition, we viewed population counts as actual counts rather than estimates based on sequential sampling. A more thorough analysis could take into account sampling errors when the 10^5 cells are harvested from much larger populations to initiate another day of growth.

The approach taken here allows for future evaluation of the contribution of specific factors on growth dynamics and survivor formation in response to telomerase-deficiency. For example, the effects of an *exo1*Δ deletion, which has been shown to slow senescence progression and favor the formation of type I survivors (Bertuch and Lundblad, 2004; Maringele and Lydall, 2004) or the effects of a *rad50*Δ mutation, which eliminates the formation of type II survivors (Teng *et al.*, 2000), can be mathematically modeled.

5 References

Arino, O., Kimmel, M. and Webb, G.F., 1995. Mathematical modeling of the loss of telomere sequences. *J. Theor. Biol.* **177**, 45–57

Arino, O., Sanchez, M. and Webb, G.F., 1997. Polynomial growth dynamics of telomere loss in a heterogeneous cell population. *Dynam. Contin. Discrete Impuls. Systems* **3**, 262–282

Baird, D.M., 2008. Telomere dynamics in human cells. *Biochimie* **90**, 116–121

BERTUCH, A.A. and LUNDBLAD, V., 2004. *EXO1* contributes to telomere maintenance in both telomerase-proficient and telomerase-deficient *Saccharomyces cerevisiae*. *Genetics* **166**, 1651–1659

BRYAN, T.M., DALLA-POZZA, L., DUNHAM, M.A. and REDDEL, R.R., 1997. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med* **3**, 1271–1274

CHEN, X.F., MENG, F.L. and ZHOU, J.Q., 2009. Telomere recombination accelerates cellular aging in *Saccharomyces cerevisiae*. *PLoS Genet* **5**, e1000535

D'MELLO, N.P. and JAZWINSKI, S.M., 1991. Telomere length constancy during aging of *Saccharomyces cerevisiae*. *J Bacteriol* **173**, 6709–6713

DENCHI, E.L., 2009. Give me a break: How telomeres suppress the DNA damage response. *DNA Repair* (Amst) **8**, 1118–1126

DYSON, J., VILLELLA-BRESSAN, R., and WEBB, G.F., 2007. Stabilization of telomeres in nonlinear models of proliferating cell lines. *J. Theoret. Biol.* **244(3)**, 400–408

FLORES, I., 1967. Direct calculation of k-generalized Fibonacci numbers. *The Fibonacci Quarterly* **5**, 259–266

GILLESPIE, C.S., PROCTOR, C.J., BOYS, R.J., SHANLEY, D.P., WILKINSON, D.J. and KIRKWOOD, T.B.L., 2004. A mathematical model of ageing in yeast. *J. Theor. Biol* **229(2)**, 189–196

HACCOU, P., JAGERS, P. and VATUTIN, V., 2005. Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge university press.

HARRIS, T.E., 1963. The Theory of Branching Processes. Springer, Berlin.

HEMANN, M.T., STRONG, M.A., HAO, L.Y. and GREIDER, C.W., 2001. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107(1)**, 67–77

JAGERS, P. and NERMAN, O., 1984. The Growth and Composition of Branching Populations. *Adv. Appl. Prob.* **16**, 221-259

JAGERS, P., 1992. Stabilities and instabilities in population dynamics. *J. Appl. Prob.* **29**, 770–780

LARSSON, S., RYDEN, T., HOLST, U., OREDSSON, S. and JOHANSSON, M., 2008. Estimating the total rate of DNA replication using branching processes. *Bulletin of Mathematical Biology* **70(8)**, 2177–2194

LEVY, M.Z., ALLSOP, R.C., FUTCHER, A.B., GREIDER, C.W. and HARLEY, C.B., 1992. Telomere end-replication problem and cell aging. *J. Molec. Biol.* **225**, 951–960

LUNDBLAD, V. and SZOSTAK, J.W., 1989. A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633–643

MARINGELE, L. and LYDALL, D., 2004. EXO1 plays a role in generating type I and type II survivors in budding yeast. *Genetics* **166**, 1641–1649

OLOFSSON, P. and KIMMEL, M., 1999. Stochastic models of telomere shortening. *Math. Biosci.* **1**, 75-92

OLOFSSON, P., 2000. A Branching Process Model of Telomere Shortening. *Communications in Statistics-Stochastic Models* **16(1)**, 167-177

OLOFSSON, P. and KIMMEL, M., 2005. Mathematical modeling of telomere shortening. In: P. Haccou, P. Jagers and V. Vatutin *Branching Processes: Variation, Growth, and Extinction of Populations.* Cambridge university press, 225–231

OPREA, M. and KEPLER, T.B., 2001. Improved inference of mutation rates: II. Generalization of the Luria-Delbruck distribution for realistic cell-cycle time distributions. *Theoretical Population Biology* **59(1)**, 49–59

OP DEN BUIJS, J., VAN DEN BOSCH, P.P.J., MUSTERS, M.W.J.M. and VAN RIEL, N.A.W., 2004. Mathematical modeling confirms the length-dependency of telomere shortening. *Mechanisms of Ageing and Development* **125(6)**,

437–444

PORTUGAL, R.D., LAND, M.G.P. and SVAITER, B.F., 2008. A computational model for telomere-dependent cell-replicative aging. *Biosystems* **91**, 262–267

RUBELJ, I. and VONDRACEK, Z., 1999. Stochastic Mechanism of Cellular Aging-Abrupt Telomere Shortening as a Model for Stochastic Nature of Cellular Aging. *J. Theor Biol.* **197**, 425–438

SHAMPAY, J. and BLACKBURN, E.H., 1988. Generation of telomere-length heterogeneity in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **85**, 534–538

SHAMPAY, J., SZOSTAK, J.W. and BLACKBURN, E.H., 1984. DNA sequences of telomeres maintained in yeast. *Nature* **310**, 154–157

SINCLAIR, D.A. and GUARENTE, L., 1997. Extrachromosomal rDNA circles - A cause of aging in yeast. *Cell* **91(7)**, 1033–1042

SINCLAIR, D.A., MILLS, K. and GUARENTE, L., 1998. Aging in Saccharomyces cerevisiae. *Annu. Rev. Microbiol.* **52**, 533–560

SINGER, M.S. and GOTTSCHLING, D.E., 1994. TLC1: template RNA component of Saccharomyces cerevisiae telomerase. *Science* **266**, 404–409

TAN, Z., 1999. Intramitotic and Intraclonal Variation in Proliferative Potential of Human Diploid Cells: Explained by Telomere Shortening. *J. Theor Biol.* **198**, 259–268

TENG, S.C., CHANG, J., MCCOWAN and AAKIAN, V.A., 2000. Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. *Mol Cell* **6**, 947–952