Effects of initial telomere length distribution on senescence onset and heterogeneity

Sarah Eugène[a,b], Thibault Bourgeron[c,d,*], Zhou Xu[e,*]

[a] *Sorbonne Universités, UPMC Université Pierre et Marie Curie, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005 Paris, France*
[b] *INRIA Paris, 2 rue Simone Iff, F-75012 Paris, France*
[c] *École Normale Supérieure de Lyon, UMR 5569, Unité de Mathématiques Pures et Appliquées, 69007 Lyon, France*
[d] *INRIA Numed, 46 allée d'Italie, 69007 Lyon, France*
[e] *Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 8226, Laboratoire de Biologie Moléculaire et Cellulaire des Eucaryotes, Institut de Biologie Physico-Chimique, 75005 Paris, France*

ABSTRACT

Replicative senescence, induced by telomere shortening, exhibits considerable asynchrony and heterogeneity, the origins of which remain unclear. Here, we formally study how telomere shortening mechanisms impact on senescence kinetics and define two regimes of senescence, depending on the initial telomere length variance. We provide analytical solutions to the model, highlighting a non-linear relationship between senescence onset and initial telomere length distribution. This study reveals the complexity of the collective behavior of telomeres as they shorten, leading to senescence heterogeneity.

1. Introduction

Telomeres, the ends of eukaryote chromosomes, are poised in a dynamic equilibrium controlled by two processes: limited telomere shortening at each cell division and elongation by telomerase, a dedicated holoenzyme able to generate *de novo* telomere sequence. When telomerase is not expressed, as in human somatic cells, or is experimentally mutated in model organisms such as *Saccharomyces cerevisiae* (Lundblad and Szostak, 1989), telomeres only shorten and after many divisions the cell enters replicative senescence, a permanent cell cycle arrest induced by short telomeres that elicit a DNA damage response. Replicative senescence is implicated in organismal ageing and is a potent barrier to cancer emergence, but its remarkable asynchrony and heterogeneity remain a challenge for investigating the exact relationship between initial telomere length distribution and senescence onset.

Telomere shortening is the unavoidable consequence of the end-replication problem (Olovnikov, 1973; Watson, 1972; Soudet et al., 2014). In most examined species, telomeres end with a 5′ to 3′ singled-stranded DNA overhang (Fig. 1) (Hemann and Greider, 1999; Henderson and Blackburn, 1989; Klobutcher et al., 1981; Makarov et al., 1997; McElligott and Wellinger, 1997; Raices et al., 2008; Riha et al., 2000; Wellinger et al., 1993). When the replication fork reaches the end of the chromosome, the processing of the last Okazaki fragment leaves a gap at the lagging strand, which recreates the single-stranded

overhang of the parental telomere (Fig. 1). On the leading strand, after replication, complex maturation steps involving resection and fill-in also regenerate the overhang structure (Larrivée et al., 2004; Faure et al., 2010; Chai et al., 2006; Wu et al., 2012; Soudet et al., 2014). Regardless of these maturation steps, the leading strand template for replication is shorter than the lagging strand one, thus generating after replication two new telomeres of different lengths, one unchanged compared to the parental telomere and the other shorter by exactly the length of the overhang, as illustrated in Fig. 1. Previous mathematical models of telomere shortening also based on the end-replication problem (Levy et al., 1992; Arino et al., 1995; Olofsson and Kimmel, 1999; Arkus, 2005) did not consider the maturation of the leading strand telomere that generates a 3′-end overhang identical to the one on the lagging strand. This maturation step is widely conserved throughout species with the notable exception of angiosperm plants that display a blunt end at the leading telomere (Riha et al., 2000). We also note that other mathematical models examined higher level structures such as t-loops (Griffith et al., 1999; Rodriguez-Brenes and Peskin, 2010), or additional telomere states or breaking mechanisms (Kowald, 1997; Rubelj and Vondracek, 1999; Proctor and Kirkwood, 2002, 2003), such as damage due to oxidative stress (von Zglinicki, 2002). In *S. cerevisiae*, however, oxidative stress does not significantly alter telomere length (Romano et al., 2013) and the end-replication problem is the main mechanism of telomere shortening.

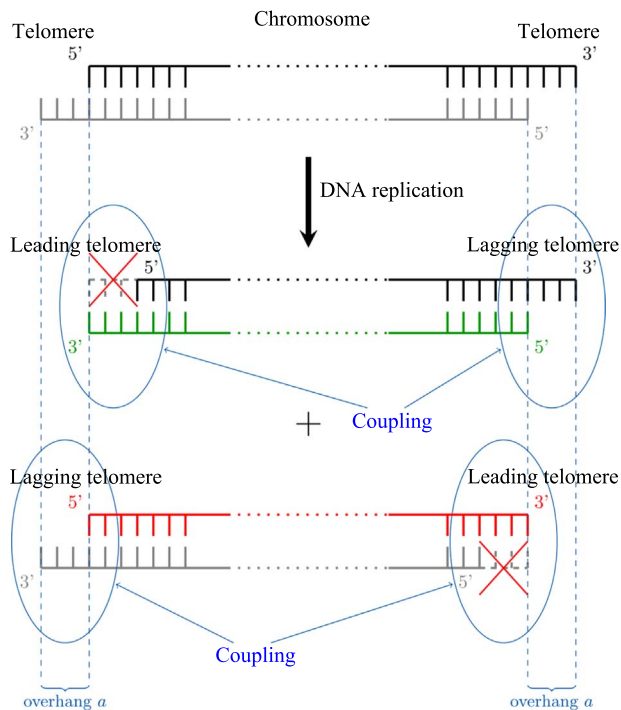Consistently, on average, telomeres shorten at a constant rate of

Fig. 1. Scheme of a chromosome bearing two telomeres and undergoing replication. Telomeres end with a 3′ overhang of length a (measured to be 5–10 nucleotides in yeast (Soudet et al., 2014), chosen here as a=1 or 7 for theoretical or numerical purposes, respectively). After DNA replication, each telomere generates, through either the leading or the lagging strand replication machineries, two new telomeres of different lengths. The coupling effect between the two ends of the same chromosome imposes that only one of the two is shortened while the other retains the parental length.

Fig. 2. Steady-state telomere length distribution in the presence of telomerase. (a) and (b) Probability of recruitment and action of telomerase, as modeled from Teixeira et al. (2004) with a length threshold L_s or simplified with a sharp switch occurring at i_s. (c) Simulation of the steady-state distribution of telomere length using either (a) (black) or (b) (grey) to describe telomerase recruitment. i_s was set so as to reach the same mean in the steady-state distribution (Appendix B).

exactly half of the overhang length per division. However, while studying the average is informative of the global regulation and homeostasis of telomere length, it misses important contributions of the asymmetry of telomere replication mechanism to the overall telomere length distribution and to the heterogeneity of the onset of senescence. Taking this asymmetry into account, the shortening of a telomere in a cell lineage, defined as a random succession of mitotically related cells (Xu et al., 2015), is probabilistic and follows a Bernoulli process. Additionally, if the two ends of a given chromosome are considered together, the 3′-end at one telomere belongs to the same DNA strand as the 5′-end on the other telomere of the same chromosome, implying that the asymmetry at one telomere is inverted compared to the other (Fig. 1). We define this relationship between the two ends of the same chromosome as a coupling mechanism, which adds another layer of constraint and will also be modeled here.

In this article, we study the consequences of the asymmetry and the coupling on the distribution and the dynamics of telomere length in two distinct phases: at steady state in the presence of telomerase and in a strictly shortening phase without telomerase. We show that the robustness of telomerase recruitment impacts on the variance of the steady-state distribution of telomere length. In turn, this variance defines different regimes of senescence. In a regime of low initial variance, senescence onset cannot be linearly inferred from the average telomere length or even the length of the shortest telomere and we provide an asymptotic expansion to account for this phenomenon. In contrast, a high variance implies a linear correlation between the initial shortest telomere and senescence onset. We provide analytical solutions to the different models we describe and suggest applications for the inference of the initial telomere length distribution from experimental measurements of senescence onset.
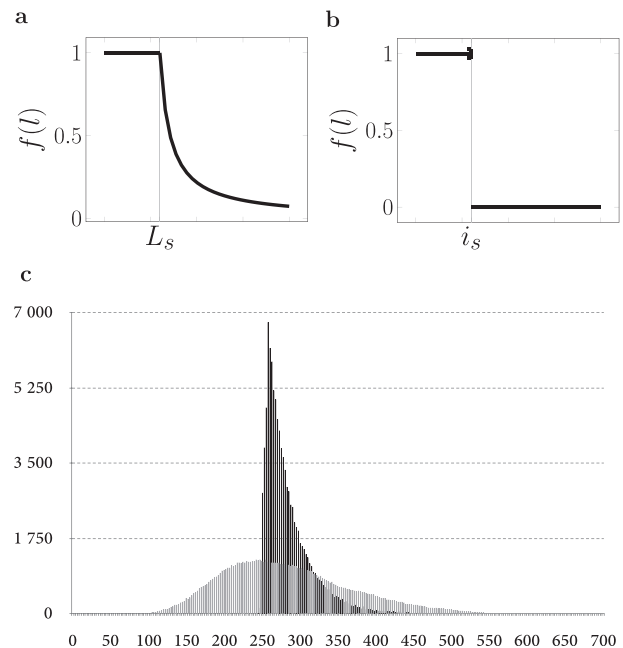
2. Telomeres evolving with telomerase

We first describe the most general model, corresponding to a lineage of haploid yeast cells dividing in the presence of active telomerase. The two telomeres of a given chromosome at generation n, called (L_n^1, L_n^2), are coupled as defined above and the 32 telomeres of the cell shorten according to a Bernoulli random variable B_n of parameter $1/2$: if $B_n = 1$, then L_n^1 is shortened by a nucleotides, whereas L_n^2 is preserved, and conversely if $B_n = 0$. Telomerase adds new telomere sequences preferentially to shorter telomeres (Teixeira et al., 2004; Britt-Compton et al., 2009), behavior that we capture by introducing C_n^i, $i \in \{1, 2\}$, Bernoulli random variables of parameter $f(L_n^i)$, according to Xu et al. (2013), where f has the shape shown in Fig. 2 (a) and L_n^i is the length of the telomere at the extremity i before replication. The shape of f is such that below a length threshold L_s, the Bernoulli random variable C_n^i equals 1, that is telomerase is always active. For a telomere longer than L_s, the probability of $\{C_n^i = 1\}$ decreases to zero, meaning that the longer the telomere, the less likely it is to be elongated by telomerase. Since the number of nucleotides added by telomerase is independent of the length of the telomere (Teixeira et al., 2004; Xu et al., 2013), we introduce $\mathcal{G}_n^1$ and $\mathcal{G}_n^2$ two independent geometric random variables of parameter p, independent of all the other quantities (including L_n^1, L_n^2), which correspond to the number of nucleotides added by telomerase. As a result, for any given chromosome the $\mathbb{N}^2$-valued process (L_n^1, L_n^2) follows:

$$\begin{pmatrix} L_{n+1}^1 \\ L_{n+1}^2 \end{pmatrix} = \begin{pmatrix} (L_n^1 - a \cdot B_n)^+ + C_n^1 \cdot \mathcal{G}_n^1 \\ (L_n^2 - a \cdot (1 - B_n))^+ + C_n^2 \cdot \mathcal{G}_n^2 \end{pmatrix}, \tag{1}$$

where $x^+ = \max(0, x)$ is x if $x > 0$ and 0 otherwise. Telomere length is nearly always positive because, in the presence of telomerase, a telomere shorter than $L_s > 0$ is elongated with probability 1, while in the absence of telomerase, senescence is triggered before the short

telomere reaches 0 (Bourgeron et al., 2015). Using the same Bernoulli random variable B_n for L_n^1 and L_n^2 mathematically defines the coupling between the two telomeres.

To characterize the steady state of telomere length distribution, we focus on one telomere—because as an approximation, telomeres of different chromosomes are assumed to be independent (Shampay and Blackburn, 1988)—, and consider the projection of the first coordinate of a chromosome in order to compute its equilibrium. We will analyze the coupling effect in more depth in the second regime without telomerase. Our model thus becomes:

$$L_{n+1} = (L_n - a \cdot B_n)^+ + C_n \cdot \mathcal{G}_n \tag{2}$$

where L_n is the length of a given telomere, L_{n+1} the length of one of the two daughter telomeres, $\mathcal{G}_n$ a geometric random variable of parameter $p \in (0, 1)$. An averaged version of this model has been studied in Xu et al. (2013); Dao Duc and Holcman (2013) and used in Bourgeron et al. (2015), where instead of being stochastic, telomere shortening was chosen to be deterministic with a constant value of $a/2$. To make our computations fully explicit without betraying the principles of the biological mechanism, instead of f, we consider a sharp threshold at a value i_s (Fig. 2b). Our model becomes:

$$L_{n+1} = (L_n - a \cdot B_n)^+ + \mathcal{G}_n \cdot \mathbb{1}_{\{L_n \leq i_s\}} \tag{3}$$

Independently of the value of i_s, the Markov chain $(L_n)_{n\geq0}$ defined by (3) has a unique equilibrium distribution L_∞ and the generating function of L_∞ is characterized by the equality:

$$\forall u \in \mathbb{C}, |u| \leq 1,$$
$$(u^a - 1)(1 - uq)\mathbb{E}(u^{L_\infty})$$
$$= -q(1 - u)(u^a + 1)S_{i_s}(u)$$
$$+ pS_{a-1}(1)u^a - pS_{a-1}(u), \tag{4}$$

where $\pi_j = \mathbb{P}(L_\infty = j)$, $S_k(u) = \mathbb{E}(u^{L_\infty}\mathbb{1}_{\{L_\infty \leq k\}}) = \sum_{j=0}^k \pi_j u^j$ and $q = 1 - p$.

Appendix A gives a proof of this result and explains how to get a fully explicit expression for the distribution L_∞ from Eq. (4). This calculation reveals how the parameters a and p affect the steady-state distribution.

The choice of the value of i_s according to biological experiments is explained in Appendix B. We find that the variance of the steady-state telomere length distribution obtained using the simplified model (3) is significantly smaller than the one with the complete model (2) (Fig. 2c in black and grey, respectively; 37 bp as compared to 101 bp), demonstrating that the residual recruitment of telomerase to rather long telomeres strongly contributes to the spread of the steady-state distribution of telomere length. In turn, the variance of this distribution is critical for determining the onset of senescence and its heterogeneity, as we show below. Thus, the mode of recruitment and activation of telomerase, dependent on the biochemical properties of the holoenzyme and on its interactions with telomeric proteins (Wellinger and Zakian, 2012), controls key features of senescence once telomerase is removed.

3. Telomeres evolving without telomerase

We then analyze the consequences of the steady-state distribution on the onset of senescence, meaning the number of generations undergone by a given cell lineage until it enters senescence. We simply call it time of senescence, denoted by T. One practical goal of this section is to derive the parameters of the initial distribution from the time of senescence, which is useful for experimentalists. In senescing cells, telomerase is inactive and when the shortest telomere reaches a threshold S, the cell enters replicative senescence and stops dividing (Abdallah et al., 2009; Lundblad and Szostak, 1989; Hemann et al., 2001; Zou et al., 2004; Armanios et al., 2009). A haploid yeast cell has 16 chromosomes and thus 32 telomeres. Mathematically, we consider the vector $(L_n^1, L_n^2, ..., L_n^{32})$ of these 32 telomere lengths at generation n.

Because each chromosome behaves independently (Shampay and Blackburn, 1988), we can start by studying one chromosome and the behavior of the 16 will easily follow. More precisely, the vector $(L_n^1, L_n^2, ..., L_n^{32})$ can be seen as a family $(X_n^i, Y_n^i)_{1\leq i\leq 16}$ of 16 independent identically distributed couples each representing the two telomeres of a chromosome, with $(X_0^i, Y_0^i) \overset{dist.}{\sim} \Pi$. The time of senescence is mathematically expressed as:

$$T = \inf\left\{n \geq 0, \min_{1\leq i\leq 16}[\min(X_n^i, Y_n^i)] < S\right\}.$$

Normalizations. As telomeres can only shorten, we consider the shortening length to be $a=1$. For numerical estimations of the time of senescence, we will divide our results by $a=7$ to obtain biologically relevant values. Moreover, we can choose $S=0$ by simply translating the initial state by S. These assumptions are made in all following calculations unless stated otherwise.

Distribution of the Time of Senescence. Under these normalization conditions, we find that the distribution of the random variable T is fully explicit:

$$\mathbb{P}(T > n) = \left[\sum_{k+l\geq n} \Pi(X_0 = k, Y_0 = l)2^{-n} \sum_{t=n-l}^k \binom{n}{t}\right]^{16}. \tag{5}$$

In particular, its expectation can be written as a function of π as follows:

$$\mathbb{E}(T) = \sum_{n=0}^\infty \left[\sum_{k+l\geq n} \Pi(X_0 = k, Y_0 = l)2^{-n} \sum_{t=n-l}^k \binom{n}{t}\right]^{16}. \tag{6}$$

See Appendix C for the proof. Because of the difficulty to invert this formula, we choose to study separately the influence of the mean and the variance of the initial state on the time of senescence. Thus, we first consider a deterministic and constant initial state $X_0 = x_0$:

$$k = 1, ..., 32, \quad L_0^k = \mathbb{E}(L_\infty) =: x_0$$

We define $T_{x_0}^1$ as the first time one of two coupled telomeres reaches zero both starting from x_0, and T_{x_0} as the time of senescence of the whole cell when the initial state is constant and equals x_0.

Almost surely, $x_0 \leq T_{x_0}^1 \leq 2x_0$, this implies that $\mathbb{P}(T_{x_0}^1 > n) = 0$ for $n \geq 2x_0$, and $\mathbb{P}(T_{x_0}^1 \geq n) = 1$ for $n < x_0$. For $x_0 \leq n \leq 2x_0 - 1$, the law of $T_{x_0}^1$ is given by:

$$\mathbb{P}(T_{x_0}^1 > n) = 2^{-n} \sum_{k=n-x_0}^{x_0} \binom{n}{k}.$$

The expected time of senescence is then:

$$\mathbb{E}(T_{x_0}) = x_0 + \sum_{n=x_0}^{2x_0-1} [\mathbb{P}(T_{x_0}^1 > n)]^{16} = x_0 + \sum_{n=x_0}^{2x_0-1} \left[2^{-n} \sum_{k=n-x_0}^{x_0} \binom{n}{k}\right]^{16}. \tag{7}$$

We then perform an asymptotic expansion of $\mathbb{E}(T_{x_0})$ for large values of x_0, which is numerically justified (Appendix B and Fig. 2c). At the first order the mean behavior prevails:

$$\mathbb{E}(T_{x_0}^1) \sim 2x_0. \tag{8}$$

Concerning the second order, we obtain the following convergence in distribution:

$$\frac{2x_0 - T_{x_0}^1}{\sqrt{x_0}} \xrightarrow[x_0 \to +\infty]{dist.} |N|, \quad \text{with } N \overset{dist.}{\sim} \mathcal{N}(0, 2). \tag{9}$$

See Appendix D for proofs of these results.

The asymptotic development of the time of senescence for one chromosome $T_{x_0}^1$ allows us to derive an approximation of the expected time of senescence (7) by replacing the law of $T_{x_0}^1$ by its asymptotic (9):
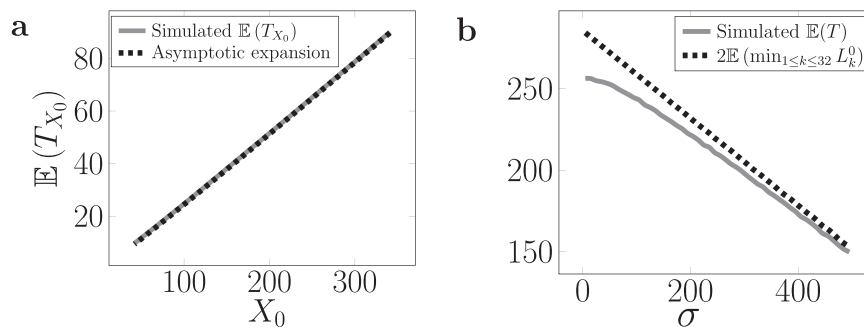
Fig. 3. Distinct effects of the mean and variance of the initial distribution on theoretical expressions and numerical simulations of the time of senescence. (a) Starting from a constant distribution $\mathbb{E}(L_\infty) := x_0$, the asymptotic expansion in Eq. (10) is computed and compared to numerical simulations (1000 independent simulations). (b) Starting from a uniform distribution of variance σ and mean $\mathbb{E}(L_\infty)$, the time of senescence is computed using Eq. (11), which takes only the mean behavior of the initial shortest telomere into account, and compared to numerical simulations (1000 independent simulations).

$$\mathbb{E}(T_{x_0}) \approx x_0 + \sum_{n=x_0}^{2x_0-1} [\mathbb{P}(2x_0 - \sqrt{x_0}|N| > n)]^{16}$$

$$= x_0 + \sum_{k=0}^{x_0-1} \left[\text{erf}\left(\frac{k}{2\sqrt{x_0}} \right) \right]^{16}, \tag{10}$$

where erf is the error function defined as:

$$\text{erf } x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

We find that the expansion (10) is hardly distinguishable from the theoretical process (6), as shown in simulations (compare grey and dashed black lines in Fig. 3 (a)) and can thus be directly used to estimate the mean of the initial state in experimental studies.

Influence of the initial variance on the time of senescence. Now, to study only the influence of the initial variance, we consider that each initial telomere is uniformly distributed in the interval $[\mathbb{E}(L_\infty) - \sigma, \mathbb{E}(L_\infty) + \sigma]$ and simulate the expected time of senescence as a function of σ (Fig. 3b). When σ has large values, there is a higher probability that the initial shortest telomere of $(L_0^1, \ldots, L_0^{32})$ is far from the mean $\mathbb{E}(L_\infty)$ and, thus, that it remains the shortest one until senescence. We therefore expect, for large enough value of σ (Fig. 3b), that the time of senescence is asymptotically equivalent to the time when the initial shortest telomere reaches zero. As, on average, the number of steps for a simple random walk starting from $M = \mathbb{E}[\min_{1 \leq k \leq 32} L_0^k]$ to reach zero is $M/(1/2) = 2 M$, we expect the following result:

$$\lim_{\sigma \to \mathbb{E}(L_\infty)} \mathbb{E}(T) = 2\mathbb{E}\left[\min_{1 \leq k \leq 32} L_0^k \right]. \tag{11}$$

We indeed find this asymptotic behavior by simulations (Fig. 3b), highlighting two regimes that depend on the initial variance. If σ has a small value, which is close to the deterministic initial state studied above (Eq. (10)), the time of senescence is much smaller than expected by just considering the shortest telomere because of the coupling effect. If σ has a large value, the time of senescence is mainly determined by the shortening of the initial shortest telomere.

We next ask which of these two regimes can be observed in simulated times of senescence from the telomere distributions described in Fig. 2, which are biologically more relevant than the previous distribution models. To do so, we simulate 1000 individual lineages of senescing cells and record their time of senescence, starting by randomly drawing their 32 telomeres from the biologically relevant distribution (large variance, in grey in Fig. 2c), from the simplified distribution (intermediate variance, in black in Fig. 2c) or from a constant distribution (no variance) (Fig. 4). We then compare these simulated times of senescence with those predicted either from the mean behavior of the shortest telomere (Eq. (11), dashed black lines in Fig. 4) or the asymptotic expansion on the mean of the initial distribution (Eq. (10), black lines in Fig. 4). The biologically relevant distribution gives simulated times of senescence that are fully predicted by computing the mean behavior of the average initial shortest telomere (Fig. 4a, compare grey and dashed black lines which are superimposed). In contrast, the constant distribution leads to a senescence onset dictated by the asymptotic expansion (Fig. 4c, compare grey and black lines), consistent with the results in Fig. 3a. The simplified distribution produces an intermediate result where the mean behavior of the shortest telomere and the asymptotic expansion lead to similar predictions, which lay close to the simulated times of senescence (Fig. 4b, compare grey, black and dashed black lines). These
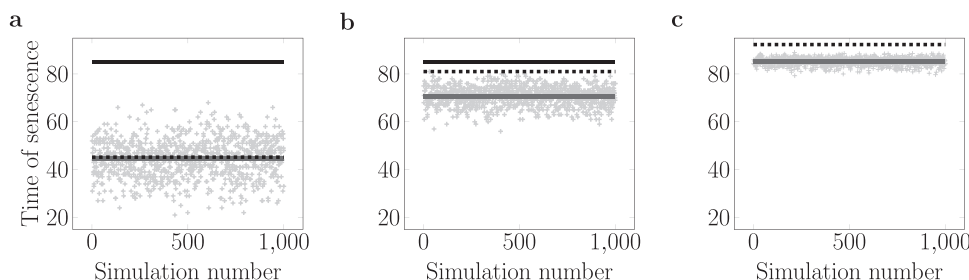

Fig. 4. Comparison between simulated times of senescence (grey dots) and predictions from Eqs. (11) (dashed black lines) and (10) (black lines). (a) 32 telomere lengths are randomly drawn from a biologically relevant distribution with a high variance (Fig. 2c, grey distribution) and the time of senescence is simulated to give one data point (grey dot). This process is repeated 1000 times and compared to the two predictions. The grey line represents the average simulated time of senescence. (b) and (c): as in (a), but starting with an intermediate level of variance for the initial telomere length distribution or no variance at all, respectively.

results show that defining two senescence regimes depending on the initial variance of telomere length distribution is critical for understanding the relevant dynamics of telomere shortening leading to senescence.

4. Conclusion

In summary, in this article, we isolated all the sources of fluctuations of the time of senescence that are dependent on telomere length. To do so, we modeled several molecular mechanisms that contribute at various levels to telomere length distribution and dynamics in *S. cerevisiae*, where they are the most exhaustively and quantitatively described. Among these mechanisms, we found that the asymmetry of telomere replication and the coupling between the two telomeres belonging to the same chromosome significantly contribute to senescence heterogeneity and we formally established their links. We also showed that the mode and robustness of telomerase recruitment control the variance of the steady-state telomere length distribution, which in turn defines two senescence regimes. With a low initial variance, the time of senescence is non-linearly related to the initial mean telomere length. In contrast, a high initial variance leads to a major role of the initial shortest telomere in controlling senescence. Because natural telomere length distributions can vary considerably, even within a species, we suggest that depending on the initial variance, the two regimes we describe may operate at the same time during

senescence. As the core mechanisms modeled here are conserved in most eukaryotes, we expect that our conclusions should also, in principle, apply to telomere-dependent senescence in human cells, although additional factors and mechanisms also contribute to senescence heterogeneity (Griffith et al., 1999; Rodriguez-Brenes and Peskin, 2010; Proctor and Kirkwood, 2002). This work uncovers a new layer of complexity in the relationship between senescence onset and telomere shortening explained by the asymmetry and coupling mechanisms, and proposes methods for assessing the time of senescence or conversely inferring parameters of the initial telomere length distribution.

Appendix

A1. Steady state of telomeres evolving with telomerase

First, we prove that both Markov chains defined by (2), (3) are ergodic and, second, we derive formula (4). The ergodicity is a direct consequence of Foster-Lyapunov criteria (Corollary 8.7 p. 214 in Robert (2003) or Proposition 1.3 in Hairer (2016), for instance). The chains are time-homogeneous and, for $L_0 > a$:

$$\mathbb{E}(L_1 - L_0) = \mathbb{E}(-a \cdot B_0 + C_0 \cdot \mathcal{G}_0) = -a/2 + f(L_0)\mathbb{E}(\mathcal{G}_0),$$

and this last quantity is negative for large enough values of L_0 because f tends to zero, either f is $f(l) = (1 + \beta(l - L_s))^{-1}$ or $f(l) = \mathbb{1}_{l \leq i_s}$. This proof of ergodicity works for any function f having a limit l at infinity which satisfies $l < \frac{a}{2\mathbb{E}(\mathcal{G}_0)}$.

Moreover, the Markov chains (L_n) defined by (2), (3) are irreducible and aperiodic. Therefore, in both cases, there exists a unique equilibrium distribution, denoted L_∞. For the model (3), it is characterized by the fact that L_∞ and $(L_\infty - a \cdot B_0)^+ + \mathcal{G}_0 \mathbb{1}_{\{L_\infty \leq i_s\}}$ have the same probability generating function.

To establish formula (4) we distinguish the three regimes: $L_\infty < a \cdot B_0$, $a \cdot B_0 \leq L_\infty \leq i_s$, $L_\infty > i_s$, and the cases $B_0 = 0$ or $B_0 = 1$. For u such that $|u| \leq 1$, we obtain:

$$
\begin{aligned}
2\mathbb{E}(u^{L_\infty}) &= 2\mathbb{E}(u^{(L_\infty - a \cdot B_0)^+ + \mathcal{G}_0 \mathbb{1}_{L_\infty \leq i_s}}) = 2\mathbb{E}(u^{\mathcal{G}_0}\mathbb{1}_{\{L_\infty < a \cdot B_0\}}) + 2\mathbb{E}(u^{L_\infty - a \cdot B_0 + \mathcal{G}_0}\mathbb{1}_{\{a \cdot B_0 \leq L_\infty \leq i_s\}}) + 2\mathbb{E}(u^{L_\infty - a \cdot B_0}\mathbb{1}_{\{L_\infty > i_s\}}) \\
&= \mathbb{E}(u^{\mathcal{G}_0})[\mathbb{P}(L_\infty < a) + \mathbb{P}(L_\infty < 0)] + \mathbb{E}(u^{\mathcal{G}_0})[\mathbb{E}(u^{L_\infty - a}\mathbb{1}_{\{a \leq L_\infty \leq i_s\}}) + \mathbb{E}(u^{L_\infty}\mathbb{1}_{\{0 \leq L_\infty \leq i_s\}})] + [\mathbb{E}(u^{L_\infty - a}\mathbb{1}_{\{L_\infty > i_s\}}) + \mathbb{E}(u^{L_\infty}\mathbb{1}_{\{L_\infty > i_s\}})] \\
&= \mathbb{E}(u^{\mathcal{G}_0})S_{a-1}(1) + \mathbb{E}(u^{\mathcal{G}_0})[u^{-a}(S_{i_s}(u) - S_{a-1}(u)) + S_{i_s}(u)] + [u^{-a}(\mathbb{E}(u^{L_\infty}) - S_{i_s}(u)) + (\mathbb{E}(u^{L_\infty}) - S_{i_s}(u))].
\end{aligned}
\tag{A.1}
$$

From (A.1), we get:

$$\mathbb{E}(u^{L_\infty})(1 - u^{-a}) = S_{i_s}(u)(\mathbb{E}(u^{\mathcal{G}_0}) - 1)(1 + u^{-a}) \qquad\qquad + \mathbb{E}(u^{\mathcal{G}_0})(S_{a-1}(1) - u^{-a}S_{a-1}(u)).$$

As the probability generating function of a geometric distribution is explicitly given by $\mathbb{E}(u^{\mathcal{G}_0}) = p(1 - uq)^{-1}$, with $q = 1 - p$, we obtain (4) after multiplication by $(1 - uq)u^a$.

In order to compute the π_k for all k, we identify the coefficients of the power series of each side of (4) distinguishing cases for the values of k. For simplicity we set $\pi_k = 0$ for $k < 0$. Using the identities: $(u^a - 1)(1 - uq) = -qu^{a+1} + u^a + qu - 1$, $(u - 1)(u^a + 1) = u^{a+1} - u^a + u - 1$, we get:

$$
\begin{aligned}
\sum_{k=0}^{\infty}(-q\pi_{k-a-1} + \pi_{k-a} + q\pi_{k-1} - \pi_k)u^k &= \sum_{k=0}^{\infty}\pi_k(-qu^{k+a+1} + u^{k+a} + qu^{k+1} - u^k) = (-qu^{a+1} + u^a + qu - 1)\mathbb{E}(u^{L_\infty}) \\
&= q(u^{a+1} - u^a + u - 1)S_{i_s}(u) + pS_{a-1}(1)u^a - pS_{a-1}(u) = q\left[\sum_{k=a+1}^{i_s+a+1}\pi_{k-a-1}u^k - \sum_{k=a}^{i_s+a}\pi_{k-a}u^k + \sum_{k=1}^{i_s+1}\pi_{k-1}u^k - \sum_{k=0}^{i_s}\pi_k u^k\right] \\
&\quad + pS_{a-1}(1)u^a - p\sum_{k=0}^{a-1}\pi_k u^k.
\end{aligned}
\tag{A.2}
$$

The following table gives the recurrence relations obtained after identification in the coefficients of (A.2).

k	relation
$[\![0, a-1]\!]$	ϕ
a	$\pi_a = 2\dfrac{1-p}{p}\pi_o - \sum \dfrac{a-1}{k=1}\pi_k$
$[\![a+1, i_s]\!]$	$\pi_k = -2\dfrac{1-p}{p}\pi_{k-a-1} + \dfrac{2-p}{p}\pi_{k-a}$
$i_s + 1$	$\pi_{i_s+1} = -2q\pi_{i_s-a} + (1+q)\pi_{i_s+1-a}$
$[\![i_s+2, i_s+a]\!]$	$\pi_k = -2q\pi_{k-a-1} + (1+q)\pi_{k-a} + q\pi_{k-1}$
$i_s + a + 1$	$\pi_{i_s+a+1} = -2q\pi_{i_s} + \pi_{i_s+1} + q\pi_{i_s+a}$
$> i_s + a + 1$	$\pi_k = -q\pi_{k-a-1} + \pi_{k-a} + q\pi_{k-1}$

Note that for k such that $0 \leq k \leq a-1$, the identification gives no information on the values of the π_k. These formulas show that all the π_k depend linearly on the a first states, π_k, $k = 0,\ldots,a-1$. Now, we indicate how to compute these a first values.

Hence, using (4), the generating function of L_∞ is only a function of the a first states. Dividing (4) by $(1-u)(1-uq)$, we can find $\psi: \mathbb{R}^a \times [0, 1] \to \mathbb{R}$ linear in the first a coordinates such that:

$$(1+\cdots+u^{a-1})\mathbb{E}(u^{L_\infty}) = \psi(\pi_0,\ldots,\pi_{a-1}, u). \tag{A.3}$$

The $a-1$ roots of $R(u) = 1+\cdots+u^{a-1}$ are the $u_k = e^{2i\pi k/a}$ for $1 \leq k \leq a-1$, which are in the unit disk. Therefore, the vector $(\pi_0,\ldots,\pi_{a-1})$ is solution of the system:

$$\psi(\pi_0,\ldots,\pi_{a-1}, u_k) = 0 \text{ for } 1 \leq k \leq a-1 \tag{A.4}$$

where π_0 is, as usual determined, by the normalization condition. If the $a \times a$ system (A.4) to which we added the normalization condition is invertible, then there exists a unique solution $(\pi_0,\ldots,\pi_{a-1})$. Having the vector $(\pi_0,\ldots,\pi_{a-1})$, the generating function $\mathbb{E}(L_\infty)$ follows from (A.3).

Finally, we want to explicitly determine the π_k for $k \geq a$. The previous table gives homogeneous linear recurrence relations with constant coefficients. For instance π_k, for k in $[\![a+1, i_s]\!]$, is a linear combination (independent of k) of the k-th powers of the roots of the (conjugate of the) characteristic polynomial $cX^{a+1} - (1+c)X^a + 1$, where $c = 2q/p \in (0, \infty)$. This polynomial is of degree $a+1$ and has 1 as a root. This last property is also true for the characteristic polynomials of the relations for the π_k for k in $[\![i_s+2, i_s+a]\!]$ or $> i_s + a + 1$. As the quartic equation has explicit solutions, the expressions of the π_k, $k \geq a$ are fully explicit if $a + 1 \leq 4 + 1$, that is $a \leq 4$. In particular for $a=1$, the $(\pi_k)_{k\in[\![1,i_s]\!]}$ and $(\pi_k)_{k>i_s}$ are two geometric progressions:

$$\forall\, k \in [\![1, i_s]\!],\ \pi_k = c^k\pi_0, \quad \forall\, k > i_s,\ \pi_k = p(1-p)^k\left(\frac{2}{p}\right)^{i_s+1}\pi_0. \tag{A.5}$$

For $a > 4$ the roots of these polynomials can be found numerically.

B1. Choice of i_s

To rigorously compare the variance of the simplified model (3) with the one of (2), we choose i_s so that the ceiling function of the mean of the equilibrium of (3) $\lceil \mathbb{E}(L_\infty)\rceil$ is the same as the one for (2), i.e. 342 bp in Xu et al. (2013). We take the biological parameters obtained in Teixeira et al. (2004); Soudet et al. (2014) and used in Xu et al. (2013):

$$a = 7, p = 0.026, L_s = 90, \beta = 0.045,$$

where β is a fitting parameter describing telomerase recruitment to telomeres (Xu et al., 2013) (Table E.1). In order to compute the corresponding mean of the equilibrium, for each choice of i_s we run 10^6 numerical simulations of (3). This mean is then computed and plotted as a function of i_s, cf. Fig. B.1. Finally, we chose the value of i_s that gives $\lceil \mathbb{E}(L_\infty)\rceil = 342$ bp. This procedure leads to i_s=308 bp.

C1. Expected time of senescence

The aim of this section is to establish formula (5). Before entering the details just note that taking $a=1$ is not strictly equivalent to make a change in the length unit as the lengths of the telomeres are integer-valued. But taking $a > 1$ only make the results more difficult to state without exhibiting a new behavior. For the sake of simplicity, we drop the superscript i in this section and start by studying a typical couple (X_n, Y_n). The shortening of these two telomeres can be mathematically translated into the following model:
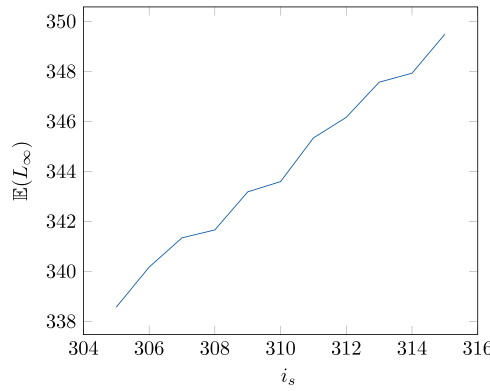
Fig. B.1. The expected length $\mathbb{E}(L_\infty)$ as function of i_s.

$$\begin{pmatrix} X_{n+1} \\ Y_{n+1} \end{pmatrix} = \begin{pmatrix} (X_n - a \cdot B_n)^+ \\ (Y_n - a \cdot (1 - B_n))^+ \end{pmatrix}, \tag{C.6}$$

where B_n is a Bernoulli random variable of parameter $1/2$, and $(X_0, Y_0) \overset{dist.}{\sim} \Pi$.

This process is an oriented simple random walk on $\mathbb{Z}^2$ until one of the coordinates reaches zero, and can be written explicitly:

$$X_n = X_0 - a \sum_{k=1}^{n} B_k = X_0 - aB(n, 1/2), \quad Y_n = Y_0 - a \sum_{k=1}^{n} (1 - B_k) = Y_0 - a(n - B(n, 1/2)), \tag{C.7}$$

where $B(n, 1/2)$ is a binomial distribution of parameters n and $1/2$. In this case, let us define the first time one of the coordinates reaches zero, T^1, as:

$$T^1 = \inf\{n \geq 0, \min(X_n, Y_n) < 0\}$$

Then, from (C.7) and for $a=1$, because (X_n) and (Y_n) are non-increasing, we get:

$$\mathbb{P}(T^1 > n) = \mathbb{P}(X_n \geq 0, Y_n \geq 0) = \mathbb{P}(n - Y_0 \leq B(n, 1/2) \leq X_0) = \sum_{\substack{k+l \geq n \\ k, l \geq 0}} \Pi(X_0 = k, Y_0 = l) 2^{-n} \sum_{t=n-l}^{k} \binom{n}{t}. \tag{C.8}$$

From here, we easily derive the distribution of the time of senescence by considering all 16 independent pairs of telomeres:

$$\mathbb{P}(T > n) = \mathbb{P}(\neg \{\text{senescence at the } n^{th} \text{ generation}\}) = \mathbb{P}\left(\min_{1 \leq k \leq 32} L_k^n \geq 0\right) = \mathbb{P}(\forall i \in [\![1, 16]\!], \min(X_n^i, Y_n^i) \geq 0) = \mathbb{P}(\min(X_n, Y_n) \geq 0)^{16} = \mathbb{P}(T^1 > n)^{16}. \tag{C.9}$$

Formulas (C.8) and (C.9) lead to (5), which gives the expected time of senescence (6) using $\mathbb{E}(T) = \sum_{n=0}^{\infty} \mathbb{P}(T > n)$.

D1. Asymptotics of the mean time of senescence

The aim of this section is to prove formulas (9) and (10). To do so we first note that, using a generalized version of the central limit theorem, for any fixed positive number t, and any sequence $u_n \sim tn$:

$$\frac{B\left(u_n, \frac{1}{2}\right) - \frac{1}{2}u_n}{\sqrt{n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \frac{1}{2}\mathcal{N}(0, t), \tag{D.10}$$

where $B(u_n, \frac{1}{2})$ is a binomial distribution with parameters u_n and $1/2$.

Let w be a real number. As in (C.8), we get, with $n(x_0) = \lfloor 2x_0 - w\sqrt{x_0} \rfloor \sim 2x_0$:

$$\mathbb{P}\left(\frac{2x_0 - T_{x_0}^1}{\sqrt{x_0}} < w\right) = \mathbb{P}(T_{x_0}^1 > 2x_0 - w\sqrt{x_0}) = \mathbb{P}(T_{x_0}^1 > n(x_0)) = \mathbb{P}\left(n(x_0) - x_0 \leq B\left(n(x_0), \frac{1}{2}\right) \leq x_0\right) = \mathbb{P}\left(\left|\frac{B\left(n(x_0), \frac{1}{2}\right) - \frac{n(x_0)}{2}}{\sqrt{x_0}}\right| \leq \frac{x_0 - \frac{n(x_0)}{2}}{\sqrt{x_0}}\right).$$

As x_0 tends to ∞, thanks to (D.10), we obtain:

$$\mathbb{P}\left(\frac{2x_0 - T_{x_0}^1}{\sqrt{x_0}} \leq w\right) \to \mathbb{P}\left(\frac{|N|}{2} \leq \frac{w}{2}\right),$$

where $N \sim \mathcal{N}(0, 2)$.

The speed of convergence is given by the Berry-Essen theorem in terms of the cumulative distribution function of $\frac{2x_0 - T_{x_0}^1}{\sqrt{x_0}}$, say F_{x_0}:

$$\forall x \in \mathbb{R}, \forall x_0 \in \mathbb{N} \qquad |F_{x_0}(x) - \text{erf } x| \leq C x_0^{-1/2},$$

Table E1
Parameters used in this study.

a	Length of the 3′-end overhang.
S	Threshold length of the shortest telomere inducing senescence.
L_s	Length threshold for the function f, below which telomerase is recruited to the telomere with probability 1.
i_s	Length threshold for the simplified model of telomerase recruitment.
p	Parameter of the geometric random variable $\mathcal{G}_n$.
β	Parameter of the function f, fitted on experimental data (Xu et al., 2013).

where C is an absolute constant. This inequality should provide a bound to the difference in Eq. (10):

$$\left| \mathbb{E}(T_{x_0}) - \left(x_0 + \sum_{k=0}^{x_0-1} \left[\mathrm{erf}\left(\frac{k}{2\sqrt{x_0}} \right) \right]^{16} \right) \right|.$$

E1. Parameters used in this study

See Table E1.

References

Abdallah, Pauline, Luciano, Pierre, Runge, Kurt W., Lisby, Michael, Géli, Vincent, Gilson, Eric, Teixeira, M. Teresa, 2009. A two-step model for senescence triggered by a single critically short telomere. Nat. Cell Biol. 11 (8), 988–993.

Arino, Ovide, Kimmel, Marek, Webb, Glenn F., 1995. Mathematical modeling of the loss of telomere sequences. J. Theor. Biol. 177 (1), 45–57.

Arkus, Natalie, 2005. A mathematical model of cellular apoptosis and senescence through the dynamics of telomere loss. J. Theor. Biol. 235 (1), 13–32.

Armanios, Mary, Alder, Jonathan K., Parry, Erin M., Karim, Baktiar, Strong, Margaret A., Greider, Carol W., 2009. Short telomeres are sufficient to cause the degenerative defects associated with aging. Am. J. Hum. Genet. 85 (6), 823–832.

Bourgeron, Thibault, Xu, Zhou, Doumic, Marie, Teixeira, M. Teresa, 2015. The asymmetry of telomere replication contributes to replicative senescence heterogeneity. Sci. Rep. 5.

Britt-Compton, Bethan, Capper, Rebecca, Rowson, Jan, Baird, Duncan M., 2009. Short telomeres are preferentially elongated by telomerase in human cells. FEBS Lett. 583 (18), 3076–3080.

Chai, Weihang, Sfeir, Agnel J., Hoshiyama, Hirotoshi, Shay, Jerry W., Wright, Woodring E., 2006. The involvement of the Mre11/Rad50/Nbs1 complex in the generation of G-overhangs at human telomeres. EMBO Rep. 7 (2), 225–230.

Dao Duc, Khanh, Holcman, David, 2013. Computing the Length of the Shortest Telomere in the Nucleus. Phys. Rev. Lett. 111 (22), 228104.

Faure, Virginie, Coulon, Stéphane, Hardy, Julien, Géli, Vincent, 2010. Cdc13 and telomerase bind through different mechanisms at the lagging-and leading-strand telomeres. Mol. Cell 38 (6), 842–852.

Griffith, Jack D., Comeau, Laurey, Rosenfield, Soraya, Stansel, Rachel M., Bianchi, Alessandro, Moss, Heidi, Lange, Titia De, 1999. Mammalian telomeres end in a large duplex loop. Cell 97 (4), 503–514.

Hairer, Martin, 2016. Convergence of Markov Processes. (accessed 19.06.16).

Hemann, Michael T., Greider, Carol W., 1999. G-strand overhangs on telomeres in telomerase-deficient mouse cells. Nucleic Acids Res. 27 (20), 3964–3969.

Hemann, Michael T., Strong, Margaret A., Hao, Ling-Yang, Greider, Carol W., 2001. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. Cell 107 (1), 67–77.

Henderson, E.R., Blackburn, E.H., 1989. An overhanging 3′ terminus is a conserved feature of telomeres. Mol. Cell. Biol. 9 (1), 345–348.

Klobutcher, Lawrence A., Swanton, Marshal T., Donini, Pierluigi, Prescott, David M., 1981. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3′ terminus. Proc. Natl. Acad. Sci. USA 78 (5), 3015–3019.

Kowald, A., 1997. Possible mechanisms for the regulation of telomere length. J. Mol. Biol. 273 (4), 814–825.

Larrivée, LeBel, Catherine, Wellinger, Raymund J., 2004. The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. Genes Dev. 18 (12), 1391–1396.

Levy, Michael Z., Allsopp, Richard C., Bruce Futcher, A., Greider, Carol W., Harley, Calvin B., 1992. Telomere end-replication problem and cell aging. J. Mol. Biol. 225 (4), 951–960.

Lundblad, Victoria, Szostak, Jack W., 1989. A mutant with a defect in telomere elongation leads to senescence in yeast. Cell 57 (4), 633–643.

Makarov, Vladimir L., Hirose, Yoko, Langmore, John P., 1997. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. Cell 88 (5), 657–666.

McElligott, Richard, Wellinger, Raymund J., 1997. The terminal DNA structure of mammalian chromosomes. EMBO J. 16 (12), 3705–3714.

Olofsson, Peter, Kimmel, Marek, 1999. Stochastic models of telomere shortening. Math. Biosci. 158 (1), 75–92.

Olovnikov, Alexeij M., 1973. A theory of marginotomy: the incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. J. Theor. Biol. 41 (1), 181–190.

Proctor, Carole J., Kirkwood, Thomas B.L., 2002. Modelling telomere shortening and the role of oxidative stress. Mech. Ageing Dev. 123 (4), 351–363.

Proctor, Carole J., Kirkwood, Thomas B.L., 2003. Modelling cellular senescence as a result of telomere state. Aging Cell 2 (3), 151–157.

Raices, Marcela, Verdun, Ramiro E., Compton, Sarah A., Haggblom, Candy I., Griffith, Jack D., Dillin, Andrew, Karlseder, Jan, 2008. C. elegans telomeres contain G-strand and C-strand overhangs that are bound by distinct proteins. Cell 132 (5), 745–757.

Riha, Karel, McKnight, Thomas D., Fajkus, Jiri, Vyskot, Boris, Shippen, Dorothy E., 2000. Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. Plant J. 23 (5), 633–641.

Robert, Philippe, 2003. Stochastic networks and queues. Stochastic Modelling and Applied Probability Series. Springer-Verlag, New York

Rodriguez-Brenes, Ignacio A., Peskin, Charles S., 2010. Quantitative theory of telomere length regulation and cellular senescence. Proc. Natl. Acad. Sci. USA 107 (12), 5387–5392.

Romano, Gal Hagit, Harari, Yaniv, Yehuda, Tal, Podhorzer, Ariel, Rubinstein, Linda, Shamir, Ron, Gottlieb, Assaf, Silberberg, Yael, Peer, Dana, Ruppin, Eytan, Sharan, Roded, Kupiec, Martin, 2013. Environmental stresses disrupt telomere length homeostasis. PLoS Genet. 9 (9), e1003721.

Rubelj, Ivica, Vondracek, Zoran, 1999. Stochastic mechanism of cellular aging; abrupt telomere shortening as a model for stochastic nature of cellular aging. J. Theor. Biol. 197 (4), 425–438.

Shampay, Janis, Blackburn, Elizabeth H., 1988. Generation of telomere-length heterogeneity in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 85 (2), 534–538.

Soudet, Julien, Jolivet, Pascale, Teixeira, M. Teresa, 2014. Elucidation of the DNA end-replication problem in Saccharomyces cerevisiae. Mol. Cell 53 (6), 954–964.

Teixeira, M. Teresa, Arneric, Milica, Sperisen, Peter, Lingner, Joachim, 2004. Telomere length homeostasis is achieved via a switch between telomerase -extendible and -nonextendible states. Cell 117 (3), 323–335.

von Zglinicki, Thomas, 2002. Oxidative stress shortens telomeres. Trends Biochem. Sci. 27 (7), 339–344.

Watson, 1972. Origin of concatemeric T7 DNA. Nat. New Biol. 239 (94), 197–201.

Wellinger, Raymund J., Wolf, Alexander J., Zakian, Virginia A., 1993. Saccharomyces telomeres acquire single-strand TG 1-3 tails late in S phase. Cell 72 (1), 51–60.

Wellinger, Raymund J., Zakian, Virginia A., 2012. Everything you ever wanted to know about Saccharomyces cerevisiae telomeres: beginning to end. Genetics 191 (4), 1073–1105.

Wu, Peng, Takai, Hiroyuki, de Lange, Titia, 2012. Telomeric 3′ overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. Cell 150 (1), 39–52.

Xu, Zhou, Dao Duc, Khanh, Holcman, David, Teixeira, M. Teresa, 2013. The Length of the Shortest Telomere as the Major Determinant of the Onset of Replicative Senescence. Genetics 194 (4), 847–857.

Xu, Zhou, Fallet, Emilie, Paoletti, Camille, Fehrmann, Steffen, Charvin, Gilles, Teixeira, M. Teresa, 2015. Two routes to senescence revealed by real-time analysis of telomerase-negative single lineages. Nat. Commun. 6, 7680.

Zou, Ying, Sfeir, Agnel, Gryaznov, Sergei M., Shay, Jerry W., Wright, Woodring E., 2004. Does a sentinel or a subset of short telomeres determine replicative senescence? Mol. Biol. Cell 15 (8), 3709–3718.