1. **Previous and Preliminary results**

**4.1 Quantifying the aging process based on the Gompertz model**

We quantified yeast survival curves using the two-parameter Gompertz aging model:

|  |  |
| --- | --- |
|  | (Eq 1) |
|  | (Eq 2) |

where, *m* is mortality rate; *s* is the survival fraction of a population; *t* is time. Here, mortality rate is defined as the normalized declining rate of *s*. The initial mortality rate, *I*, describes the innate susceptibility to dying. The Gompertz coefficient, *G*, determines acceleration rate of the mortality rate over time and is viewed as the rate of aging. Examples of Gompertz survival curves can be seen in Fig 2.

Given a Gompertz model, the average life span is defined as:

|  |  |
| --- | --- |
| . | (Eq 3) |

This equation shows that for a given mean life span, there are many pairs of *I* and *G* values, as shown by the contour lines in Figure 3.

For a given set of measured life spans of N cells (), the likelihood function based on the Gompertz model is:

.

Hence, the log-likelihood is:

.

Maximization of the above log-likelihood function will yield the maximal likelihood estimations of the Gompertz parameters for a data set of life span. We have implemented the maximization procedure in the R-statistics language and environment. This likelihood approach enables us to design nested models to statistically compare *I* and *G* values of different strains (see section 4.3).

The Gompertz model can be extended by the introduction of a morality component that is independent of time, represented by a constant as an extra term inserted to Equation 1. This extension will then use three parameters and is often called Gompertz-Makeham model [[1](#_ENREF_1), [2](#_ENREF_2)]. Environmental factors, such as nutrients, radiations, and toxic influences, are known to influence the parameters in both the Gompertz model [[3](#_ENREF_3)] and the Gompertz-Makeham model [[4](#_ENREF_4)]. Other alternative models include the Weibull model based on a power function [[2](#_ENREF_2)] and the logistical model [[5](#_ENREF_5)].

|  |
| --- |
| C:\Users\hqin\Documents\manuscripts_n_proposals\RUI-network-aging\ExpGenrontology\figures and tables in work\gompertz-space.072805.jpg  ***I***  ***G***  **Fig 3**. Replicative aging described in the Gompertz parameter space. The ARLS values are represented by the dashed contour lines with the values labeled at the ends. Solid black circles are natural isolates. White circles are laboratory strains. |

Departure from the Gompertz model is often observed at late age when the acceleration of mortality rate slows down [[5](#_ENREF_5), [6](#_ENREF_6)]. In other words, a small percentage of long-lived individuals are often important to determine whether the Gompertz model or other aging models is the best-fit model. As a result, large sample sizes are often required to examine the differences of various aging models. The Gompertz model is often indistinguishable from other models when sample sizes are less than 100 [[5](#_ENREF_5), [6](#_ENREF_6)]. The sample sizes of most yeast RLS assays range from 30 to 60. Therefore, the Gompertz model is a reasonable choice to study the yeast replicative aging.

**4.2 Quantifying the aging processes in a collection of yeast strains**

We have systematically investigated the life span in a collection of yeast lab strains and wild isolates (Table 1) [[7](#_ENREF_7)]. We observed a considerable amount of RLS variation among the 14 natural isolates. The average RLS (ARLS) values of the isolates follow a normal distribution. The mean ARLS for the 14 natural isolates is 31.7 cell divisions, and the standard deviation is 5.8 cell divisions. RM11 has the largest ARLS value at 44.5 cell divisions, and SGU57 has the smallest value at 23.6 cell divisions. The two extreme values differ by 1.9-fold.

The pattern of natural RLS variation described by the Gompertz model is presented by a scatter plot of *I* ~ *G* (Fig. 3). The *G* values, which are related to the rate of aging, follow a normal distribution. The mean and standard deviation of the *G* value distribution are 0.12 and 0.025, respectively. The *I* values, the initial mortality rates, follow a lognormal distribution. The mean and standard deviation of the natural-logarithm transformations of *I* are -6.3 and 0.72, respectively. Using multiple regression analysis, negative correlations are observed between ARLS ~ *I* , ARLS ~ G, and *I* ~ G, with partial correlation coefficients at -0.90 (p-value = 2.7 x 10-5), -0.86 (p-value = 1.8 x 10-4), and -0.87 (p-value = 1.1 x 10-3), respectively. Intuitively, large *I* and *G* values ought to shorten the average life span. The negative partial correlation between *I* (or ln *I*) and *G* is often interpreted as tradeoff from the evolutionary perspective. This correlation can be further verified by the straightforward regression between ln *I* and G. Linear regression shows that correlation between ln *I* ~ *G* has a p-value of 0.065 and an R2 value of 0.26. This negative correlation is often called Strehler-Mildvan correlation and is a characteristic of aging in higher organisms [[8](#_ENREF_8)]. This observation demonstrates that yeast replicative aging shares similar characteristics of aging in higher organisms at the population level.

We used linear regression models to evaluate genotypic influence on RLS. In these linear models, we investigated whether the RLS values of individual cells could be predicted by variables corresponding to their genotypes, experimental groups, and plates. We found that genotypic variation is a significant factor on RLS and contributes to ~ 22% of the total RLS variation observed at the cellular level. This genotypic influence on life span is comparable to the estimation made in *Drosophila melanogaster* and in human population based on identical twins [[9-11](#_ENREF_9)].

**Table 1**. Replicative and chronological life spans in yeast strains

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strain | ARLS | *I* | *G* | CLS |
| 101S | 31.3±0.8 | 0.0012±0.0007 | 0.14±0.02 | 3.4±0.2 |
| M1-2 | 27.9±1.1 | 0.0024±0.0011 | 0.13±0.02 | 10.4±3.2 |
| M13 | 26.7±1.2 | 0.0030±0.0011 | 0.13±0.01 | 16.3±3.7 |
| M14 | 36.6±1.5 | 0.0019±0.0007 | 0.09±0.01 | 7.2 ±0.2 |
| M22 | 31.8±1.3 | 0.0024±0.0009 | 0.11±0.01 | 5.2±2.1 |
| M2-8 | 24.8±0.8 | 0.0042±0.0010 | 0.12±0.01 | 4.1±0.7 |
| M32 | 28.1±0.8 | 0.0018±0.0005 | 0.14±0.01 | 6.4±0.8 |
| M34 | 26.8±1.0 | 0.0013±0.0007 | 0.16±0.02 | 5.2±0.4 |
| M5 | 36.7±1.0 | 0.0040±0.0008 | 0.07±0.01 | 4.9±0.5 |
| M8 | 35.2±0.9 | 0.0004±0.0002 | 0.16±0.02 | 10.5±0.1 |
| SGU57 | 23.6±1.5 | 0.0080±0.0022 | 0.09±0.01 | 9.3±0.7 |
| RM11 | 44.6±1.5 | 0.0010±0.0004 | 0.09±0.01 | 12.7±2.9 |
| YPS128 | 35.0±1.2 | 0.0011±0.0005 | 0.12±0.01 | 3.3 ±0.1 |
| YPS163 | 34.4±0.8 | 0.0008±0.0003 | 0.14±0.01 | 4.2±1.1 |
| BY4743 | 33.2±0.9 | 0.0040±0.0013 | 0.08±0.01 | 9.7±1.7 |
| SK1 | 22.0±1.3 | 0.011 ±0.003 | 0.08±0.01 | 5.0±0.9 |
| W303 | 18.7±0.6 | 0.0034±0.0011 | 0.20±0.02 | 17.2±3.9 |

ARLS (Average replicative life span), CLS (Chronological life span). Standard deviations are estimated by bootstrapping. The numbers of bootstraps equal the sample sizes. These strains are a unique and important resource of this project.

**4.3 Diploid cells tend to age faster than haploid cells based on likelihood ratio tests of nested models.**

A particular challenge in studying variation in life span is its stochasticity, which demands rigorous quantitative modeling and analysis. Our likelihood approach enables us to statistically test *I* and *G* values of different yeast strains, which is often sidestepped by other investigators in yeast aging. We developed likelihood ratio tests based on nested models, implemented in the R-language and environment.

Diploid cells are much larger in size than haploid cells, indicating that there more protein molecules (network components) in gene/protein networks in diploid cells. Hence, networks are more robust in diploid cells than those in haploid ones. Our hypothesis will then predict that diploid cells should age fasters than haploid cells. We found that haploid wildtype strains (BY4741 and 4742) share the same rate of aging (Gompertz coefficient) (Table 2A). We then found that diploid wildtype strain (BY4743) has significantly higher rate of aging than do the two haploid wildtype strains (Table 2B). Therefore, these findings are in agreement with our hypothesis (see Fig 2). Interestingly, the diploid strain lives longer despite higher rate of aging, because of its lower initial mortality rate. This kind of tradeoff is known as the Strehler-Mlidvan correlation [[8](#_ENREF_8)].

|  |  |  |
| --- | --- | --- |
| **Table 2A**  Rates of aging are the same in haploid BY4741 and BY4742 based on nest model tests. | | |
| Models | lnL |  |
| H0 *I*BY4741= *I*BY4742, *G*BY4741= *G*BY4742 | -587.3 |  |
| H1i *I*BY4741≠ *I*BY4742, *G*BY4741= *G*BY4742 | -579.4 | Best parsimonious model  p-value=7.2x10-5(chi-square, df=1)  *G*BY4741= *G*BY4742= 0.0653 |
| H1g *I*BY4741= *I*BY4742, *G*BY4741≠*G*BY4742 | -581.4 |  |
| H2 *I*BY4741≠ *I*BY4742, *G*BY4741≠*G*BY4742 | -579.4 |  |

|  |  |  |
| --- | --- | --- |
| **Table 2B**  Aging rate is faster in diploid BY4743 than in haploid strains based on nest model tests. | | |
| Models | lnL |  |
| H0 *I*BY4741 = *I*BY4742 = *I*BY4743, *G*BY4741= *G*BY4742= *G*BY4743 | -952.7 |  |
| H1g2 *I*BY4741 = *I*BY4742 = *I*BY4743, *G*BY4741= *G*BY4742≠*G*BY4743 | -946.1 |  |
| H1i2 *I*BY4741 = *I*BY4742 ≠*I*BY4743, *G*BY4741= *G*BY4742 = *G*BY4743 | -939.4 |  |
| H2ig2 *I*BY4741 = *I*BY4742 ≠ *I*BY4743, *G*BY4741= *G*BY4742≠*G*BY4743 | -937.2 |  |
| H3i3ig2 *I*BY4741≠ *I*BY4742≠ *I*BY4743, *G*BY4741= *G*BY4742≠*G*BY4743 | -929.3 | Best parsimonious model  p-value=3.6x10-10 (chi-square, df=3)  *G*BY4741= *G*BY4742 = 0.0653  *G*BY4743 = 0.0786 |
| H6 *I*BY4741≠*I*BY4742≠*I*BY4743, *G*BY4741≠*G*BY4742≠*G*BY4743 | -929.3 |  |