8-2018

Quantifying the effect of dietary restriction on lifespans in Saccharomyces cerevisiae

Kathryn Rouse

University of Tennessee at Chattanooga, nmb846@mocs.utc.edu

Follow this and additional works at: https://scholar.utc.edu/honors-theses

Part of the Computer Sciences Commons

Recommended Citation

This Theses is brought to you for free and open access by the Student Research, Creative Works, and Publications at UTC Scholar. It has been accepted for inclusion in Honors Theses by an authorized administrator of UTC Scholar. For more information, please contact scholar@utc.edu.
Quantifying the Effect of Dietary Restriction on Lifespans in *Saccharomyces cerevisiae*

Kathryn P. Rouse

Departmental Honors Thesis

The University of Tennessee at Chattanooga

Computer Science Department

Examination Date: July 5, 2018

Dr. Hong Qin

Associate Professor of Computer Science and Engineering

Thesis Director

Dr. Mina Sartipi

UC Foundation Professor or of Computer Science and Engineering

Department Examiner
Abstract

Dietary restriction (DR) increases lifespan in many organisms. The effect of DR is best studied in the budding yeast *Saccharomyces cerevisiae*; however, previous studies of the DR effect on yeast lifespan tend to be qualitative. Here, we quantitatively study the effect of DR on lifespan for a selection of gene deletions in a 168-strain set of *Saccharomyces cerevisiae*. Lifespans of these yeast mutants were measured in YPD with 2% glucose and 0.05% glucose conditions. To quantify the aging process and its influence by DR, we fit yeast lifespan data with a three-parameter network aging model that is informative of the gene network changes during aging. We estimated three parameters – initial mortality rate, initial virtual age, and average gene interactions per essential gene from yeast lifespans. We compared these parameters for mutants that can shorten or extend lifespan under DR conditions. Data-mining approaches using R-based programming software were performed to find statistically significant associations between the network aging parameters and genotype-dependent responses to DR. In both cases of shortening and extending lifespan, the genes which had a major effect upon removal resulted in an increase in average gene interactions under DR conditions. It was also observed that for a majority of genes causing noteworthy changes in lifespan, there existed a directly proportional relationship between shortening or extending lifespan and how that gene positively or negatively influences initial mortality rate as well as initial virtual age. A few exceptional genes to these trends were identified, discussed, and noted as potential subjects of future study.
Table of Contents

Abstract .............................................................................................................................. 1
List of Figures ..................................................................................................................... 3
Introduction ....................................................................................................................... 4
Research Background ....................................................................................................... 5
  Dietary Restriction and Yeast.......................................................................................... 5
  Cellular Aging .................................................................................................................. 6
  Goals and Aims ............................................................................................................... 9
Methodology .................................................................................................................... 10
  Analysis of Replicative Lifespan .................................................................................. 12
Results and Discussion ................................................................................................... 14
  Appropriateness of the Binomial Aging Model .............................................................. 17
  Comparison of Binomial Parameters .............................................................................. 21
  Contouring the Effect of Binomial Parameters ................................................................ 30
Conclusion ....................................................................................................................... 35
  Recommendations for Further Work ............................................................................. 37
References ....................................................................................................................... 38
Appendix .......................................................................................................................... 40
  Appendix A: Dataset from Collaborators ...................................................................... 41
  Appendix B: Source Code for Binomial Fitting ............................................................... 45
  Appendix C: Source Code for Bar Graphs ...................................................................... 48
  Appendix D: Source Code for Fitting Probability Density Curves ................................... 54
  Appendix E: Probability Density Profiles ..................................................................... 59
  Appendix F: Source code for Contour Plots .................................................................. 65
List of Figures

Figure 1: Bar graph of Genes with Largest RLS change ...........................................14
Figure 2: Bar graph of Genes with Known Effect on RLS ......................................16
Figure 3: Probability Density Profile of Gene: vma2 ..............................................19
Figure 4: Probability Density Profile of Gene: fob1 ..............................................20
Figure 5: Bar graph of change in Mortality Rate ....................................................22
Figure 6: Bar graph of change in Initial Virtual Age ..............................................23
Figure 7: Bar graph of change in no. of Non-essential Genes ...............................23
Figure 8: Bar graph of extended genotypes change in gene interactions ..........24
Figure 9: Bar graph of extended genotypes change in Mortality Rate ...............25
Figure 10: Bar graph of extended genotypes change in Initial Virtual Age ....25
Table 1: Table of parameter changes for extending genes .................................26
Figure 11: Bar graph of shortening genotypes change in gene interactions ......27
Figure 12: Bar graph of shortening genotypes change in Mortality Rate ..........27
Figure 13: Bar graph of shortening genotypes change in Initial Virtual Age ....28
Table 2: Table of parameter changes for shortening genes ..............................28
Figure 14: Contour plot for extending genes .......................................................30
Figure 15: Contour plot for shortening genes ......................................................32
Introduction

The ageing process is a fundamental subject in biology. Over the last few decades, the concept of cellular aging as it relates to organism lifespan has prompted major discussion and debate. Despite continued research efforts to further understanding on the subject, the mechanisms behind aging at molecular levels remain unclear [1].

Recent exploration has determined that cellular aging is strongly influenced by the complex interactions of coordinated gene networks. Determining the role of aging as an emergent property of gene networks requires a series of unique statistical analysis and predictive modeling to further the understanding and identification of the mechanisms behind lifespan extension [2].

One of the most widely studied environmental manipulations used to analyze lifespan extension is the condition of Dietary Restriction (DR). It is largely known that DR increases the lifespan in yeast cells, worms, fruit flies, mice, humans, and various other organisms; however, the detailed genetic mechanisms for this response are still unknown [3].
Research Background

Dietary Restriction and Yeast

Dietary Restriction is the reduction of specific or total nutrient availability without resulting in a state of malnutrition [4]. In order to investigate the underlying means by which DR slows aging at the molecular level, numerous research studies have been conducted on model organisms, with one of the most popular research subjects being *Saccharomyces cerevisiae*, more commonly known as budding yeast [5, 6].

*S. cerevisiae* are unicellular and eukaryotic, meaning that their internal cell compartmentalization is structured similarly to that of mammals [3]. Additionally, *S. cerevisiae* yeast cells are used as a popular model for aging because they allow for the separate study of both mitotic aging and post-mitotic aging. In other words, budding yeast cells demonstrate aging in cells that are still performing mitosis, actively dividing or replicating, or cells that are no longer capable of undergoing mitosis [7].

The cellular aging of yeast is referred to as mitotic, or replicating aging, and is commonly measured in replicative lifespan (RLS), defined as the quantity of daughter cells created by parent cells before division ceases [3]. This can be found by assessing and quantifying the manual dissection of yeast daughter cells as they are taken away from the mother cells [8].

Post-mitotic aging is referred to as chronological aging where it is measured in chronological lifespan (CLS) defined by the length of time that a cell can survive in a non-dividing state [3, 8]. CLS is often evaluated by the quantification of colony-forming units rather than a set of parent and daughter cells [3].

While both RLS and CLS are viable options for measuring aging, we will be using RLS as our unit of measurement for cellular aging as this study focuses on the
genetic influences of aging in individual strains. This is more appropriately assessed by RLS because the daughter cells expectedly share the same genome as the mother cells which will aid in the determination of emergent properties, those found at the system level rather than at the component level [2]. Furthermore, the most common method of DR for yeast cells is reducing the culture medium's initial glucose concentration, which, appropriately, is how RLS is most often assessed in a majority of published studies on the subject [9, 10].

The methodology for studying RLS under the condition of dietary restriction is based on maintaining the yeast cells in a nutrient-rich, growth-supporting medium. This culture contains yeast extract, peptone, and glucose (also known as dextrose), which is referred to as YPD. It is the glucose component of the control YPD culture medium that is manipulated for performing dietary restriction [3, 11].

**Cellular Aging**

There exist numerous, hypothetical models of cellular aging that attempt to more accurately define the aging process with identifiable association factors. The use of aging models can help determine biologically significant associations between dietary restriction and the RLS of yeast [11].

This study utilizes the definition of aging that relates the age of the organism with the increasing probability of death at any point and deteriorating cellular functions [12]. This increasing probability can be designated by the organism’s mortality rate, $m(t)$,
which is the decreasing rate of biological viability, $S(t)$, over time $t$. This relationship is represented by the following equation:

$$Mortality\ rate: \ m(t) = -\frac{1}{S(t)} \frac{dS(t)}{dt}$$

Aging will occur when the mortality rate is both a positive and increasing function as it will indicate the growing chance of dying as an organism ages [9]. The biological viability, $S(t)$, is also known as the survival function and can be derived from the mortality rate if $m(t)$ is already known [9]. The actual death of a cell occurs when a gene that is considered to be essential to cellular function loses all of its interactions with other genes as if the essential gene was deleted from the genome [2].

In order to create a general theory of species aging and longevity, considerable investigation has been done towards developing a fundamental quantitative theory of aging. The leading theory for biological aging is based on the Gompertz law, which states that mortality rates increases exponentially with age. Because the Gompertz law is observed in so many diverse biological species, an explanation for the phenomenon is highly desired [9, 13].

From the Gompertz law derives the two-parameter Gompertz model. It defines the mortality rate $m(t)$, at time $t$, as an exponential function where $R_0$ is the initial mortality at $t = 0$, and $G$ is the Gompertz parameter, or the rate of aging in unites of 1/ time. The Gompertz coefficient describes the acceleration of $m(t)$ over time thus making it a measure for the rate of aging[2]. The Gompertz model is as follows:

$$mortality\ rate: \ m(t) = R_0 e^{Gt}$$

The initial mortality rate, $R$, describes an organism or cell's inherent proneness to dying without considering external factors. The Gompertz parameter is a coefficient that
determines the accelerating rate of the mortality rate over a period of time making it an appropriate representation for the rate of aging [14].

The basis for this biological aging model comes from Gavrilov & Gavrilova 2001 Journal of Theoretical Biology [9]. It states that for biological aging, the Gompertz model is generally adopted and hypothesizes the above equations related to mortality rate and viability of the Gompertz model for a data set. It is based on the findings that this exponential increase of mortality rate is a ubiquitous feature of biological aging found in organisms as simple as yeast and fruit flies to more complex mice and humans [3, 9].

The model offers insight into the mortality, R, and Gompertz, parameters, which are utilized for deriving parameter distributions and influences to aging. The following are some of the more important equations derived from the Gompertz model:

The viability of the Gompertz Model at time $t$ is represented as $s(t)$:

$$
\text{network viability: } s(t) = e^{\left(\frac{R_0}{G} (1-e^{Gt})\right)}
$$

The median lifespan can be defined as $s(t_{\text{median}}) = 0.5$ or 50% where:

$$
t_{\text{median}} = \frac{\ln(1 + \ln(2)) G}{G}
$$

and where viability is estimated with the following equation, normalized from 0 to 1, and a derivative equation for mortality R at that time:

$$
R_i = \frac{G_i \cdot \ln(2)}{e^{G_i t_{\text{median}}}} - 1
$$

The central idea of this study is that if cellular aging emerges as a property of gene networks and if DR can improve the reliability of the gene interactions, then our
hypothesized network model of aging can provide a hypothesis for how DR actually extends lifespan [2]. The gene deletion strands are non-essential genes, in that these genes are not considered critical for the survival of the cell. The influence of their absence will allow for the mapping of the relationship between genes as a network.

In regards to the binomial model for gene networks concerning mortality rate, our network model can be summarized by a three-parameter binomial function:

\[
\text{Network Viability: } s(t) = \exp \left( - \int_{t=0}^{t} u_{\text{net}}(t) \right) = \exp \left( \frac{Rt_0(1-t/t_0)^n}{n} \right)
\]

\[
\text{Network Mortality Rate: } u_{\text{net}}(t) = R \left( 1 + \frac{t}{t_0} \right)^{n-1}
\]

where \( t_0 = \frac{1-p}{p\lambda} \)

In the binomial model, \( n \) represents the average number of gene interactions per essential gene, \( t_0 \) represents the initial virtual age, \( p \) represents the probability that a gene interaction is active, and \( \lambda \) represents the decay rate of interaction efficacy. Both \( \lambda \) and \( t_0 \) relate to the reliability of gene interactions [1, 14]. The main difference between the binomial model and the Gompertz model is that the binomial model is able to further partition the Gompertz coefficient \( G \) into \( n \) and \( t_0 \), thus allowing for the interpretation of the number of gene interactions per essential gene. The binomial model will be the mathematical focus for the ensuing investigation into the influence of DR on the RLS of yeast.

\textit{Goals and Aims}
In this study, we investigate the interconnection between Dietary Restriction and cellular aging in the budding yeast *Saccharomyces cerevisiae*. Our aim is to investigate whether the lifespan extension effect of DR is associated with its effect on gene network changes as reflected by the three parameters of binomial aging model.

**Methodology**
The research that will be the basis for this study comes from the Kaeberlein dataset from the Schleit 2013 Aging Cell paper [4]. Primarily, we are utilizing the data from this study where the lifespans of yeast cells in a controlled environment and a restricted caloric environment were recorded. Each dataset contains S. cerevisiae cells where a different non-essential gene, one that is not critical to cell survival, has been removed. The purpose of this removal is to understand the effect that each individual, non-essential gene has on aging and its response to dietary restriction [15].

Here, we are determining the effect of DR on lifespan for 168 gene deletion strains in Saccharomyces cerevisiae with sample sizes greater than 30 cells. These genetic backgrounds have been used to describe association factor in yeast cells between 2% glucose () and 0.05% glucose (a form of DR) conditions. The data worked with is from a parent study that measures the RLS of yeast under DR conditions. This data concerns yeast cells with different gene deletion strands, meaning that each yeast cell’s response to DR and the control is identified by its lack of one specific gene. The absence of single gene allows us to analyze that gene’s influence on yeast aging and genotype-dependent responses to DR.

The data was analyzed using biology based, statistical models related to the binomial aging model that are able to determine biologically significant associations between dietary restriction and the RLS of yeast. The graphical modeling, correlation testing, and statistical significance testing were done through probability density profiles, viability plotting, and statistical modeling software utilizing different R graphing and statistics libraries.
Each analysis, numerical calculation, and diagram was performed using the open source programming software: RStudio (version 1.1.442), which utilized the R programming language and console (version 3.4.4). The following libraries were downloaded for free from their online sources and loaded into RStudio for use in code: ‘ggplot2’, ‘flexsurv,’ ‘survival,’ ‘stats,’ and ‘sm.’ The function for calculating binomial aging was provided by Dr. Qin’s “lifespan.r” source code.

**Analysis of Replicative Lifespan**

The replicative lifespan of yeast cells with each gene deletion strand was recorded under two conditions, one with a normal growth medium and 2% glucose concentration, the YPD condition, and the other in a dietary restricted medium of 0.05% glucose concentration, labeled as DR or 05D. The number of samples for each gene deletion strand varied, so in order to ensure there was a large enough sample size for each gene, only those with more than 20 samples were considered during analysis.

To prepare the dataset for analysis, the data was separated by measurements done under YPD and DR conditions. Within each category were the 168 gene strands that each had their own .csv file containing the recorded replicative lifespans for the number of samples of yeast. The dataset was now converted into two folders, YPD and DR, that each contained 168 .csv files named after the specific gene deletion strand.

The next step was to take the two, sorted datasets and explore the RLS fitting outcomes of the Gompertz and binomial models. The sample size, average RLS and standard deviation of the RLS of each gene strain were calculated and then used for the binomial and Gompertz fitting.
The RLS data sets were fit by strains to the Gompertz model with the use of the R library ‘flexsurv’ which has a flexsurvreg() function that takes in the recorded RLS samples for a single strain and outputs parameters of the Gompertz Aging model. These include the initial mortality R-value, Gompertz parameter G, and the Logistic Likelihood, LogLikely, and AIC (Akaike information criterion). Both the LogLikely and the AIC provide an estimate for the relative quality of statistical models for a given dataset and means for selecting the most likely fitting models [16].

Each strain was then fit to the proposed binomial aging model by selecting 6 as a constant for the average number of gene interactions per essential gene, using the initial mortality R from the Gompertz fit, and then calculating the initial virtual age, t0, using n and the G value from the Gompertz fit. These three values were the input for the binomial fitting based on Dr. Qin’s function for binomial aging, ‘llh.binomialMortality.single.run,’ which output its own values for R, t0, and n. All of these calculated values were then placed into two data frame from the YPD and DR values and are listed in Appendix A.

In order to determine which genes are being influenced by the dietary restricted environment, the average lifespan of the yeast under DR and YPD conditions needs to be compared for each individual gene deletion. To determine which genes are having the largest change, the difference in average RLS of each gene was divided by RLS under YPD. This way, the data is normalized to find the genes with the largest percent change when placed under the DR conditions. The source code and some sample output for this portion of the investigation is provided in Appendix B and C.
Results and Discussion

The 10 genes with the largest percent change were selected for evaluation meaning their absence would have had the largest effect on the replicative lifespan. The genes with the largest percent changes due to their absence were genotypes: nhx1, pdb1, phb1, phb2, por1, rad6, sod2, vma13, vma2, and vma5. A comparison of the average RLS under DR and YPD conditions for each of these genes, which can now be referred to as the “top ten” genes, is shown in Figure 1 below:

Figure 1: A bar graph of average RLS’s of 10 genotypes with largest percent change

While all of these genes demonstrate a clear change in RLS under different conditions, the direction change can be differentiated into two categories: one of RLS extension and one of RLS shortening. In other words, the absence of a particular gene, for example phb1, causes a lifespan extension under DR conditions. On the other hand, the
absence of the sod2 gene causes a substantial shortening in lifespan under DR conditions. The genes that have the “extending” effect are nhx1, pdb1, phb1, phb2, and por1 while the genes that have the “shortening” effect are rad6, sod2, vma13, vma2, and vma5.

Extensive work has already been done to identify the genes in yeast that are responsible for aging and has led to investigations into determining the molecular mechanisms underlying genotype-specific responses to dietary restriction. The result of looking into characterizing the mechanisms of lifespan extension by DR is the identification of several single gene deletions that have a significant contribution to life extension.

For example, previous studies on how aging occurs at the molecular level have consistently focused on the ‘sod1’, ‘sod2’, and ‘pda1’ genes for how they relate to the mechanisms responsible for life-span extension, and in particular how they function under periods of dietary restriction [1, 17, 18]. Highlighted in the Schleit 2013 Aging Cell paper, the base study for this investigation, also indicate the ‘phb1’ and ‘phb2’ genes as showing the most positive response to dietary restriction [1]. Additionally, a report from the 2015 Cell Metabolism Journal contains one of the most comprehensive yeast data sets of aging, containing 4698 deletions tested strains. The study indicates that two of the most widely studied and well-known genes to have shown direct effects on replicative lifespan include the ‘fob1’ and ‘sch9’ genes [8], in addition to the previously listed genes.

Given the extensive literature on the role of particular genes in cellular aging, the following seven genes have been included in the comparison along with the ten genes with the largest percent change in average RLS. These genes include: fob1, pda1, phb1
and phb2 (which also have two of the largest percent changes as shown above), sch9, sod1 and sod2. A comparison of the average RLS under DR and YPD conditions for these seven genes signified as playing a meaningful role in the influence of dietary restriction is shown in the bar graph of Figure 2 below and will be referred to as the “studied” genes:

**Figure 2: Average RLS's of genotypes with known lifespan influencing properties**

Out of the “studied” genes identified in previous literature, the genes that have the “extending” effect are fob1, pda1, phb1, phb2 while the genes that have the “shortening” effect are sch9, sod1, and sod2.
From what is known concerning the general trend in dietary restriction, the expected result is for lifespan to increase. The extending effect on a genotype would indicate that that gene’s absence allows for the dietary restrictive effects to take place, while the shortening effect on a genotype indicates that the gene’s absence hinders the effect of dietary restriction on the lifespan of the cell. Therefore, it can be assumed that the genes that create the shortening and extension effects play important roles in the effectiveness of dietary restriction. Determining the relationship between the genes and dietary restriction is done by breaking down the parameters responsible for aging, such as the initial mortality rate, number of interactions per essential gene, and initial virtual age and comparing how these parameters changed under DR and YPD conditions.

*Appropriateness of the Binomial Aging Model*

Before determining the effect of individual parameters, the appropriateness of the binomial model for the data set needs to be statistically verified. This is to not only ensure that the data has sample sets that are large enough to be statistically viable, but to also see that the binomial model is an appropriate model for biological aging and deterring the effect dietary restriction has on the aging process. To do this, the recorded RLS values for each gene were plotted as probability density points and fit to the Gompertz and binomial models. The Gompertz model is included as a comparison for the binomial model since it is already known that the Gompertz model is an appropriate model for biological aging and will be used to reference if the binomial model is also appropriate.

The formula for the Gompertz probability density function was derived from the Gompertz equations for network viability and mortality rate. Upon accepting the R-value,
G value, and frequency of the RLS values for a particular gene, represented by the variable X, it results in an appropriate probability density. The derived equation for this function is as follows:

\[ \text{Gompertz Probability Density} = R \cdot e^{Gx - \frac{(R)}{2}} \cdot (e^{Gx} - 1) \]

The formula for the binomial probability density function was derived from the definition of the three-parameter binomial network model itself where the probability density function for lifespan is the product of the network mortality rate and the network viability [10]. The result is as follows:

\[ \text{Network viability} S_m(t) \ast \text{Network Mortality Rate: } m_{net}(t) = \]

\[ \text{Binomial probability density} = \exp \left( R \frac{t_0 (1-t/t_0)^n}{n} \right) \ast R \left( 1 + \frac{t}{t_0} \right)^{n-1} \]

To visually and computationally represent the application of these functions, the replicative lifespan of each gene was plotted against its probability density with the data points being fitted to the Gompertz and binomial models respectively. The regression error was calculated using the cor.test() function of the R ‘stats’ library which tests for the association and correlation between paired samples. This error is represented by a p-value between 0 and 1 where a small p-value, less than 0.05 indicates strong evidence for the fit of the model. In the calculations, values were rounded to the nearest thousandth, so for where it indicates a p-value = 0 to, it might not necessarily be 0 in value but it is smaller than 0.05 and therefore supports the fit of the model.

A probability density profile was created for each gene where the genes are compared under YPD and DR conditions and further compared based on their fits to the Gompertz and Binomial models. The source code for the creation of these plots is located
in Appendix D. For the purpose of creating a general impression as to the appropriateness of the models, several of these probability density profiles are shown in Figures 3 and 4, while the probability density profiles of the remaining 14 total evaluated genes are provided Appendix E.

**Probability Density Profile of Gene: vma2**

*Figure 3: These are Probability Density Graphs for fitting the appropriateness of the Gompertz model (left) and binomial model (right) for modeling the aging process of the vma2 gene. Each data point represents the RLS of a sample of vma2 plotted against the frequency of that RLS value among all the samples for the set of vma2, in other words, the probability density.*

The probability density profile of the vma2 gene represents the fit of the binomial model on a lifespan-shortening gene. The binomial function visually fits the data points,
with a better fit under DR, as the average RLS of the data and the calculated average RLS based on the binomial model are relatively close with a values of 2.3 and 2.7, respectively, compared to 5.6 and 2.8 for the YPD fit. These values are comparable to those of the Gompertz fit on the left of each graph. Additionally, the p-values are all less than 0.05 supporting the binomial model as a statistically viable option for predicting biological aging.

**Probability Density Profile of Gene: fob**

*Figure 4: These are Probability Density Graphs for fitting the appropriateness of the Gompertz model (left) and binomial model (right) for modeling the aging process of the fob1 gene. Each data point represents the RLS of a sample of fob1 plotted against the frequency of that RLS value among all the samples for the set of fob1, in other words, the probability density.*
The probability density profile of the fob1 gene represents the fit of the binomial model on a lifespan-extending gene. The binomial function visually fits the data points, and compares closely to each respective Gompertz fit, again with a closer fit for the DR data. Also, the p-values are all less than 0.05 supporting the binomial model as a statistically viable option for predicting biological aging.

An important component to distinguish between Figure 3 and 4 is the shape of each probability density function. Figure 3 is representative of the lifespan-shortening effect and has a decreasing curve where the probability of a longer lifespan decreases as the RLS increases while Figure 4, representative of the lifespan-extending effect, has a much more normally distributed curve with the highest probability being towards the median of the RLS values. This is noteworthy because it once again distinguishes the difference between genotypes in having a shortening and extending effect, this time on a probability based level. Another importance of the probability density function curves is to also support the confidence in the calculated parameters and values that will be presented and the focus of further discussion.

**Comparison of Estimated Parameters**

The probability density fittings demonstrate that the three-parameter binomial function is appropriate for modeling biological aging for the set of 168 gene-deletion strands under dietary restricted conditions. In order to examine the influence of the three distinct binomial parameters on cellular aging and the gene’s absence, each parameter was compared under DR and YPD conditions for the “top ten” largest changes in RLS
genes and the “studied” gene deletions. The process for these comparisons is similar to that done for the average RLS, where the genes are plotted in DR and YPD conditions but the average RLS is replaced with each binomial parameter: the initial mortality rate $R$, the initial virtual age $t0$, and the average number of gene interactions per essential gene $n$.

Before being sorted by lifespan extension and shortening, the parameters as they are influenced by normal YPD conditions and DR 0.5% glucose conditions were calculated for the “top ten” genes so the overall changes for each parameter can be effectively visualized. The comparison of $R$, $t0$, and $n$ are displayed below in Figures 5, 6, and 7 respectively:

*Figure 5: A bar graph of the mortality rate $R$ of 10 genotypes with largest RLS change.*

*Note that the genotypes that resulted in lifespan extension have visibly smaller values in $R$, such that putting them on the same scale as the shortening genotypes makes it difficult to see the changes being made. This is another motive for separating the two groups.*
Examination of Figures 5, 6, and 7 indicates that there are some differences between the extending and shortening groups for the R, t0 and n values. Regarding the mortality rate R and the initial virtual age t0, the genes that are included in the lifespan...
extension group (nhx1, pdb1, phb1, phb2, and por1) have consistently smaller values than
the genes included in the lifespan shortening group (rad6, sod2, vma13, vma2, and
vma5). The one exception to this trend is with rad6, which is a part of the shortening
group but has R and t0 results similar to those of the lifespan extension group. The
number of interactions between essential genes did not seem to be influenced by if the
genes were responsible for lifespan extension or reduction as the values remained
relatively similar across both extending and shortening groups. To further investigate the
roles of t0, R, and n between YPD and DR conditions, two groups were made that
combined the “top ten” greatest changes in RLS genes with the most “studied” genes in
order to distinguish the activity of the parameters in relation to lifespan extension and
shortening. The lifespan extension group consists of 7 the genes: nhx1, pdb1, phb1, phb2,
por1, fob1, and pda1. The lifespan shortening group consists of the 7 genes: rad6, sod2,
vma13, vma2, vma5, sch9, and sod1. The comparison of n, R, and t0, for the Lifespan
Extending Genotypes are displayed below in Figures 8, 9, and 10 respectively:

![Bar graph showing n of RLS Extended Genotypes](image)

**Figure 8:** A bar graph of n no. of gene interactions per essential gene for RLS extending
genotype
Figure 9: A bar graph of change in initial mortality rate $R$ for RLS extending genotype

Figure 10: A bar graph of change in initial virtual age $t_0$ for RLS extending genotypes

The three previous figures of the RLS extending genotypes demonstrate the changes among the parameters from YPD to DR conditions. Figure 8 shows that every lifespan extending gene also had an extension of $n$ when in DR conditions. Likewise,
almost all of the genes displayed a general extension of $R$ and $t0$ from YPD to DR with the exception of small shortening in $R$ of phb1, small shortening $t0$ of por2, and a major shortening of $t0$ for pda1.

To compare these increases and decreases, the changes were recorded with each extension or shortening gene denoted by percent change of a each parameter under DR conditions. The following table, Table 1, not only displays these observations, but and also indicates the general trend of the group where the change in $R$ is positive (extending), $t0$ is positive (extending), and $n$ is positive (extending). It highlights the gene pda1 because it has the largest deviation from this trend with $t0$ largely negative (shortening) instead of extending as the majority of the rest of the genes

<table>
<thead>
<tr>
<th>Extension Change</th>
<th>$R$</th>
<th>$t0$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>nhx1</td>
<td>+ 32.96%</td>
<td>+ 219.9%</td>
<td>+ 94.0%</td>
</tr>
<tr>
<td>pdb1</td>
<td>+ 452.2%</td>
<td>+ 141.8%</td>
<td>+ 110.9%</td>
</tr>
<tr>
<td>phb1</td>
<td>- 3.9%</td>
<td>+ 676.1%</td>
<td>+ 97.8%</td>
</tr>
<tr>
<td>phb2</td>
<td>+ 69.4%</td>
<td>+ 428.5%</td>
<td>+ 102.3%</td>
</tr>
<tr>
<td>por1</td>
<td>+ 293.5%</td>
<td>- 2.9%</td>
<td>+ 13.8%</td>
</tr>
<tr>
<td>fob1</td>
<td>+ 42.3%</td>
<td>+ 5.8%</td>
<td>+ 1.1%</td>
</tr>
<tr>
<td>pda1</td>
<td>+ 125.1%</td>
<td>- 21.3%</td>
<td>+ 4.9%</td>
</tr>
</tbody>
</table>

Table 1: Comparison of the binomial parameters amongst lifespan extending genes. The percentages were calculated by: subtracting the values for each parameter under YPD conditions by the values under DR conditions, dividing this result by the values in YPD conditions, and multiplying by one hundred. This way, the percent change is in the direction from the control conditions to the DR conditions.
The comparison of $n$, $R$, and $t_0$ for the Lifespan Shortening Genotypes are displayed below in Figures 11, 13, and 13 respectively:

**Figure 11:** A bar graph of $n$ no. of gene interactions per essential gene for RLS shortening genotypes

**Figure 12:** A bar graph of change in mortality rate $R$ for RLS shortening genotype. There was a change in $R$ for rad6 and sch9 but they were very small and are included in the summary Table 2 on the following page.
The three previous figures of the RLS shortening genotypes demonstrate the changes among the parameters from YPD to DR conditions, and their results are summarized in Table 2 below with the percent changes of each parameter.

<table>
<thead>
<tr>
<th>Shortening Change</th>
<th>R</th>
<th>t0</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>rad6</td>
<td>-3.1%</td>
<td>-60.1%</td>
<td>+29.5%</td>
</tr>
<tr>
<td>sod1</td>
<td>-24.5%</td>
<td>-25.9%</td>
<td>+3.7%</td>
</tr>
<tr>
<td>vma13</td>
<td>-70.1%</td>
<td>0.0%</td>
<td>+22.4%</td>
</tr>
<tr>
<td>vma2</td>
<td>-61.3%</td>
<td>-20.2%</td>
<td>+4.8%</td>
</tr>
<tr>
<td>vma5</td>
<td>-36.2%</td>
<td>-98.2%</td>
<td>-28.2%</td>
</tr>
<tr>
<td>sch9</td>
<td>-0.1%</td>
<td>-23.1%</td>
<td>+2.7%</td>
</tr>
<tr>
<td>sod2</td>
<td>-95.7%</td>
<td>+169.5%</td>
<td>-46.9%</td>
</tr>
</tbody>
</table>

*Table 2: Comparison of binomial parameters amongst lifespan shortening genes. The percentages were calculated the same way as Table 1 so that the percent change is in the direction from the control conditions to the DR conditions.*
Table 2 shows that, in general, the lifespan shortening genes had a negative change (shortening) of R, a shortening in t0, and a positive change (extension) of n when in DR conditions. This is with the exception of the noteworthy increase in t0 for the sod2 gene and the decreases in n for the sod2 and vma5 genes. Because of the deviation from the trend for the rest of the shortening genes, sod2 has been highlighted.

Because the number of essential gene interactions, n, was almost always increasing, while staying within a small margin of about 4 to 8 interactions, for both lifespan extending and shortening genes, looking into the relationship between R and t0 will likely yield more insight as to which of these two parameters of the binomial equation play a larger role in changing lifespan under dietary restriction.
Contouring the Effect of Binomial Parameters

To determine what is causing the shortening or extending effect on replicative lifespan under dietary restriction, a contour map can be used to detail the influence of the mortality, R, and the initial virtual age \( t_0 \) on individual strains.

A contour map is useful for determining the parameter with the greater effect because it is able to plot the contours for the same median lifespan aims to plot contours for the same median lifespan (e.g., \( t_{\text{median}} = 5, 15, 20, 40, 60, ... \)). This was done by first, defining a set of \( t_0 \) values (x-axis), then the R (in log scale) values that serve as the y-axis are calculated via given \( t_0 \) input and the contour values (median lifespan \( t_{\text{median}} \)) via the following formula, derived from inputting the median lifespan as resulting in an R of 0.5 or 50%.

\[
R = \frac{n \cdot \ln(2)}{t_{\text{median}} \cdot ((1 + \frac{t_{\text{median}}}{t_0})^{n-1} - 1)}
\]

Using this derived function, a series of points \((t_0, \log R)\) will be calculated, and the contour will be based on these constants. The source code for the creation of these plots is located in Appendix F. Determining the direction that each genotype travels towards when crossing contour lines shows which factor, \( R \) or \( t_0 \), has a stronger influence on the changes behind each genotype. For example, if a gene is almost a vertical line and crosses three contour lines, there has been great change but with only a small amount due to the changes in \( t_0 \) and a much larger amount due to changes in R. Similarly, a nearly horizontal line would demonstrate change being the product of \( t_0 \) and less a result of \( R \).

The contour map shown in *Figure 14* demonstrates the effect of the initial virtual age \( t_0 \)
and mortality rate, R, have on individual strains that show evident extension in RLS under DR.

Figure 14: Contour Map of Binomial Parameters on individual strains that show sizeable extension in RLS under DR. The starting point of each arrow represents the t0 plotted by the natural log of R for each gene under YPD conditions, while the arrowhead represents these values under DR conditions. The contour lines plot contours for the same median lifespans where \( t_{\text{median}} = 5, 10, 20, 30, 40, 50, \) and 70, while the direction of the arrow represents the extension of RLS based on the increasing change across the \( t_{\text{median}} \) contour lines.
This figure can be interpreted by separating the extending genotypes into three categories: where the mortality rate $R$ has the greatest influence, where the initial virtual age $t0$ has the greatest influence, or where both $R$ and $t0$ have a relatively equal effect for their roles in extending RLS. Based on those three groups, the pdb1, hx1, and fob1 genes can be grouped together as having relatively equally strong $R$ and $t0$ parameters.

The phb2 and phb1 genes do not move greatly across the $y$-axis, but they move across several contour lines in the positive $x$-direction, which would suggest that the initial virtual age $t0$ has a major role in the action of these genes. Inversely, the por1 gene has little change across the $x$-axis, in fact it is even in the negative direction, and great change across the positive $y$-axis, suggesting that the mortality rate $R$ has a major role in how the por1 influences lifespan.

Lastly, the pdal gene is actually pointed in almost the opposite direction of the other genes, with a large decrease in initial virtual age with a similar change in mortality rate to that of por1. This could be the subject of future investigation, with the question of why is this gene in the life-extending category the only one to exhibit a substantial decrease in initial virtual age while still increasing in mortality rate.

The contour map shown in Figure 15 demonstrates the effect of the initial virtual age $t0$ and mortality rate, $R$, have on individual strains that show major reduction in replicative lifespan under DR.
Figure 15: Contour Map of Binomial Parameters on individual strains that show sizeable reduction in RLS under DR. The starting point of each arrow represents the t0 plotted by the natural log of R for each gene under YPD conditions, while the arrowhead represents these values under DR conditions. The contour lines plot contours for the same median lifespans where $t_{median} = 2.5, 5, 10, 20, 40, 75,$ and $150$, while the direction of the arrow represents the reduction of RLS based on the decreasing change across the $t_{median}$ contour lines.

Figure 15 can be interpreted by separating the shortening genotypes into three categories similar to that of the extending genotypes. Unlike the extending genotypes that
had greater variety, all but two of the shortening genotypes show noteworthy change in the negative direction of the initial virtual age, with little change, and negative if any, in the mortality rate. These gene vma13 had sizeable movement in a negative mortality rate and almost no change in \( t0 \).

Finally, the sod2 gene was the only gene to have shown a change in \( t0 \) that was not only large but also positive in addition to having a large change in mortality rate. Based on this contour map in addition to the observations made in Figure 13, it is likely that the fact that the sod2 gene was one of the only genes out of all the shortening and extending genes that, in its absence, had a decrease in non-essential gene interactions, \( n \), from normal to dietary restricted conditions. This prompts future investigation into the role of the binomial parameters and their relationship with the sod2 gene. It also supports the findings from other research studies that have focused exclusively on superoxide dismutase or SOD [19]. These studies have identified the idea of oxygen free radicals and hydro peroxides as being casual factors in cellular aging based on the role of sod1 and sod2 in producing related proteins. Looking more into these factors could prompt an investigation into how they relate to the binomial parameters and being modeled.
Conclusion

The findings in this study attempt to fill in some of the voids for the mechanisms responsible for cellular aging as they relate to specific gene deletion strands and affect on aging dietary restriction conditions. In addition to the data set from the Aging Cell 2013 paper, the set of tested gene deletion strands is constantly expanding, covering the single-gene deletion of nearly all non-essential genes [1, 8].

To build off of another popular model of representing biological aging, the Gompertz model, this study approached biological aging with a new model, the three-parameter binomial model. The main advantage of this model over the Gompertz is that the Gompertz model is comprised of into two parameters, the initial mortality rate $R$ and the Gompertz parameter $G$ which indicates the rate of aging, while the three-parameter binomial model takes this a set further and breaks the Gompertz parameter $G$ into the initial virtual age $t_0$ and average number of gene interactions per essential gene $n$. The added parameters allow for a more in depth look into what is happening to the aging process when dietary restriction is at play. After identifying genes that when removed result in major lifespan extension or reduction, the effect of each binomial parameter was examined.

Regarding the parameter, $n$, for the average number of gene interactions per essential gene, the most striking finding regards a development in both extending and shortening cases where it was observed that almost all genes increased in gene interactions $n$ when under dietary restriction. These positive values ranged from a slight change of 1.1% to at most a 110.9% increase. The two observed genes that resulted in
shortening RLS and had a negative change in \( n \) from YPD to DR conditions were vma5 and sod2, with vma5 at a -28.2% decrease and sod2 at a -46.9%.

Additionally, while for some genes the initial mortality rate \( R \) did influence the strength of the gene’s effect on aging, in a vast majority, particularly the genes that resulted in RLS shortening upon deletion, the initial virtual age \( t_0 \) played a larger role. This was determined from plotting the parameters on a contour map to see which parameter had the stronger influence.

In the extending genes, all of the genes exhibited a large positive increase in \( R \), with the exception of phb1 which had a -3.9% change, while in the shortening genes, all of the genes exhibited a negative change in \( R \). This could suggest the conclusion that the way a gene influences the initial mortality rate, \( R \), could be directly proportional to that genes influence on aging under dietary restricted conditions. This same conclusion can also be suggested by the \( t_0 \) where in almost all cases of lifespan extending genes, the \( t_0 \) increased, and for almost all the lifespan shortening genes the \( t_0 \) was decreasing.

The exceptions to these trends were both regarding \( t_0 \) where the sod2 gene that results in RLS shortening had a large positive percent change in \( t_0 \) with a 169.5% increase, and where the pda1 gene resulting in RLS extension had a negative percent change of -21.3%. These exceptions to the general trends of their DR influenced counterparts could provide the basis for future investigation regarding what role these genes have in gene networks.

Researching how DR extends lifespan concurrently with investigating increasing reliability of gene interactions provides a basis for extensive network modeling and potential relationships derived from these models.
Recommendations for Further Work

Regardless of how dietary restriction influences the lifespan, the gene interactions almost always increased, a property, which could prompt future investigation into the influence that dietary restriction has on gene interactions and gene networks [20]. One particular gene to look into is the case of the sod2 gene, which actually noticeably decreased in gene interactions and actually resulted in an increase in initial virtual age. The pda1 gene was also noteworthy in its deviation from genes that created similar extending effects, in that it was the only one that actually had a decrease in initial virtual age and an increase in mortality rate. This information on both the pda1 and sod2 genes could provide insight as to their role and influence in gene networks [21].

Also, the parameter that had the most notable influence on shortening the lifespan in the absence of a gene was the initial virtual age \( t_0 \). This is a parameter that is not found directly from the Gompertz model, and demonstrates the benefit of the three-parameter binomial model, and the potential and need for further investigation into the role that initial virtual age has on aging and how it is influenced by particular genes.
References

[1] Qin, H., Lu M., "Natural variation in replicative and chronological life spans of *Saccharomyces cerevisiae*


Appendix

Appendix A: Dataset from Collaborators ................................................................. 40

Table A1: Binomial and Gompertz Fitting of YPD Data ......................... 40

Table A2: Binomial and Gompertz Fitting of DR Data ...................... 42

Appendix B: Source Code for Binomial Fitting ................................................ 44

Appendix C: Source Code for Bar Graphs ................................................. 47

Appendix D: Source Code for Fitting Probability Density Curves .......... 53

Appendix E: Probability Density Profiles ................................................. 58

Figure E1: Probability Density Curves of nhx1 gene ......................... 58

Figure E2: Probability Density Curves of pdb1 gene ......................... 58

Figure E3: Probability Density Curves of phb1 gene ......................... 59

Figure E4: Probability Density Curves of phb2 gene ......................... 59

Figure E5: Probability Density Curves of por1 gene ......................... 60

Figure E6: Probability Density Curves of pda1 gene ......................... 60

Figure E7: Probability Density Curves of rad6 gene ......................... 61

Figure E8: Probability Density Curves of sod2 gene ......................... 61

Figure E9: Probability Density Curves of vma13 gene ...................... 62

Figure E10: Probability Density Curves of vma5 gene ....................... 62

Figure E11: Probability Density Curves of sod1 gene ....................... 63

Figure E12: Probability Density Curves of sch9 gene ....................... 63

Appendix F: Source code for Contour Plots ........................................... 64
## Appendix A: Dataset from Collaborators

### Table A1: Binomial and Gompertz Fitting of YPD Data

<table>
<thead>
<tr>
<th>Well</th>
<th>samp.site</th>
<th>R</th>
<th>n</th>
<th>genotype</th>
<th>avgS</th>
<th>std.err</th>
<th>CV</th>
<th>GenompitFits</th>
<th>GenompitFits</th>
<th>GenompitFits</th>
<th>GenompitFits</th>
<th>GenompitFits</th>
<th>GenompitFits</th>
<th>GenompitFits</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>400.53353</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>27.088588</td>
<td>0.034003</td>
<td>0.076380</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>410.00000</td>
<td>13.82901</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>420.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>430.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>440.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>450.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>460.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>470.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>480.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>490.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>500.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A12</td>
<td>510.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A13</td>
<td>520.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A14</td>
<td>530.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
<tr>
<td>A15</td>
<td>540.00000</td>
<td>131.83089</td>
<td>0.929544</td>
<td>0.957810</td>
<td>24.072727</td>
<td>0.008268</td>
<td>0.034003</td>
<td>4.925245</td>
<td>0.072834</td>
<td>0.083620</td>
<td>0.010000</td>
<td>0.010000</td>
<td>0.010000</td>
<td></td>
</tr>
</tbody>
</table>

*Note: The table continues with similar entries.*

---

**Appendix B:**

[Further content related to the dataset provided.]
Table A2: Binomial and Gompertz Fitting of DR Data

<table>
<thead>
<tr>
<th>binomial</th>
<th>gompertz</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter</td>
<td>parameter</td>
</tr>
<tr>
<td>estimate</td>
<td>estimate</td>
</tr>
<tr>
<td>standard error</td>
<td>standard error</td>
</tr>
<tr>
<td>0.00242939</td>
<td>0.83674251</td>
</tr>
<tr>
<td>0.51897351</td>
<td>0.74477617</td>
</tr>
<tr>
<td>0.11474033</td>
<td>0.00149213</td>
</tr>
<tr>
<td>0.162</td>
<td>1.15434454</td>
</tr>
<tr>
<td>0.00390781</td>
<td>0.00283707</td>
</tr>
<tr>
<td>0.08757725</td>
<td>0.00304281</td>
</tr>
<tr>
<td>0.00243793</td>
<td>0.00276327</td>
</tr>
<tr>
<td>0.15164403</td>
<td>0.05282732</td>
</tr>
<tr>
<td>0.00224379</td>
<td>0.00234741</td>
</tr>
<tr>
<td>0.00927918</td>
<td>0.00468786</td>
</tr>
<tr>
<td>0.00276327</td>
<td>0.00304281</td>
</tr>
<tr>
<td>0.00234741</td>
<td>0.00254372</td>
</tr>
<tr>
<td>0.00217883</td>
<td>0.00224379</td>
</tr>
<tr>
<td>0.00208431</td>
<td>0.00204704</td>
</tr>
<tr>
<td>0.00204704</td>
<td>0.00208431</td>
</tr>
<tr>
<td>0.00198215</td>
<td>0.00195222</td>
</tr>
<tr>
<td>0.00195222</td>
<td>0.00198215</td>
</tr>
<tr>
<td>0.00189215</td>
<td>0.00186521</td>
</tr>
<tr>
<td>0.00186521</td>
<td>0.00189215</td>
</tr>
<tr>
<td>0.00181631</td>
<td>0.00179413</td>
</tr>
<tr>
<td>0.00179413</td>
<td>0.00181631</td>
</tr>
<tr>
<td>0.00174437</td>
<td>0.00172303</td>
</tr>
<tr>
<td>0.00172303</td>
<td>0.00174437</td>
</tr>
<tr>
<td>0.00169345</td>
<td>0.00167223</td>
</tr>
<tr>
<td>0.00167223</td>
<td>0.00169345</td>
</tr>
<tr>
<td>0.00166207</td>
<td>0.00164076</td>
</tr>
<tr>
<td>0.00164076</td>
<td>0.00166207</td>
</tr>
<tr>
<td>0.00164003</td>
<td>0.00161871</td>
</tr>
<tr>
<td>0.00161871</td>
<td>0.00164003</td>
</tr>
<tr>
<td>0.00162857</td>
<td>0.00160725</td>
</tr>
<tr>
<td>0.00160725</td>
<td>0.00162857</td>
</tr>
<tr>
<td>0.00161871</td>
<td>0.00159739</td>
</tr>
<tr>
<td>0.00159739</td>
<td>0.00161871</td>
</tr>
<tr>
<td>0.00163421</td>
<td>0.00161289</td>
</tr>
<tr>
<td>0.00161289</td>
<td>0.00163421</td>
</tr>
<tr>
<td>0.00165043</td>
<td>0.0016291</td>
</tr>
<tr>
<td>0.0016291</td>
<td>0.00165043</td>
</tr>
<tr>
<td>0.00166692</td>
<td>0.00164556</td>
</tr>
<tr>
<td>0.00164556</td>
<td>0.00166692</td>
</tr>
<tr>
<td>0.00169345</td>
<td>0.00167223</td>
</tr>
<tr>
<td>0.00167223</td>
<td>0.00169345</td>
</tr>
</tbody>
</table>
Appendix B: Source Code for Binomial Fitting

# Katie Rouse - Thesis code
# June 04, 10:30am

# Binomial fitting for RLS

# clear environment
rm(list = ls())

llh.binomialMortality.single.run <- function(RtOn, lifespan) {
  I = RtOn[1];
  t0 = RtOn[2];
  n=RtOn[3];
  my.data = lifespan[lis.na(lifespan)];
  log_e = (I * t0 /n )*(1 - (1 + my.data/t0)^n);
  log_m = log(I) + (n-1) * log(1 + my.data/t0 );
  my.lh = sum(log_e) + sum(log_m);
  print(RtOn); # trace the convergence
  ret = -my.lh # because opsin seems to maximize
}

# install.packages("flexsurv")
# install.packages("survival")
# install.packages("gplots")
library(flexsurv)

## Loading required package: survival

library(gplots)

##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##   lowess

setwd("-/github/rouse_thesis/rouse_code/")

## Parse strains from files
files = list.files(path = "05d/", pattern = "csv")

tmp1 = gsub("\d{6}.", "", files)
strains_noCSV = gsub(".csv", "", tmp1)

1
genotype = files

## create a dataframe to hold the calculated contents that the size of the genotypes
df = c()
for (i in 1:168)
  df = c(df, files[grep(genotype[i], genotype)])

df <- sort(unique(df))

report = data.frame(chbind(files))
report$SampleSize = NA
report$R = NA
report$U = NA
report$SN = NA
report$NO = NA
report$Genotype = NA
report$SampleSize = NA
report$AV = NA
report$SEM = NA
report$CV = NA

report$GmpLogLikFlex = NA
report$GmpSFL = NA
report$GmpFLEX = NA
report$Gmp4ICFlex = NA

report$WestLogLikFlex = NA
report$WestRateFlex = NA
report$WestShapeFlex = NA
report$West4ICFlex = NA

setwd("~/github/rousse_thesis/rousse_code/")
#setwd("~/github/rousse_thesis/rousse_code/05d")

for (i in 1:168)
{
  my.files = files[grep(strains[i] \( ~ \) files)]
  #reportingfilenames[i] = paste(my.files, collapse = "::");
  tb = read.table(paste("~/github/rousse_thesis/rousse_code/05d/",my.files[i], sep=""), sep="\t", head=FALSE)

  ## DO AGAIN WITH VPD
  # tb = read.table(paste("~/github/rousse_thesis/rousse_code/ypd/",my.files[i], sep=""), sep="\t", Head=FALSE)
  if(length(my.files) > 1)
    for (fi in 2:length(my.files)) {
      tmp.tb = read.table(paste("~/github/rousse_thesis/rousse_code/05d/",my.files[fi], sep=""), sep="\t")
      tmp.tb = read.table(paste("~/github/rousse_thesis/rousse_code/ypd/",my.files[fi], sep=""), sep="\t")
      tb = rbind(tb, tmp.tb)
    }
}
```r
# my.data = read.csv('genotype[.i], header=TRUE)
# rls <- my.data$r1s
# GenoCov <- genotype[,j]
# report$genotype[i] = strains_noCSV[i]
# report$sampleSize[i] = length(rls)

maximum <- max(rls)
minimum <- min(rls)
report$avgLS[i] <- mean(rls)
se <- median(rls)
report$stdLS[i] <- sd(rls)

CompFit <- flexsurvreg(formula=Surv(rls) ~ 1, dist = "gompertz")
report$CompLogLidFlex[i] <- round(CompFit$loglik, 1)
report$CompRFlex[i] <- CompFit$res[,1]
report$CompDFlex[i] <- CompFit$res[,1]
report$CompIIMFlex[i] <- round(CompFit$DIC)

WebFit <- flexsurvreg(formula=Surv(rls) ~ 1, dist = "weibull")
report$WebLogLidFlex[i] <- round(WebFit$loglik, 1)
report$WebRShapeFlex[i] <- WebFit$res[,1]
report$WebIShapeFlex[i] <- WebFit$res[,1]
report$WebIIMFlex[i] <- round(WebFit$DIC)

nhat = report$CompRFlex[i];
Ghat = report$CompDFlex[i];
hat = 6;
t0 = (nhat-1)/Ghat;
fitBinom = optim(c(nhat, 0, nhat), llh.binomialMortality.single.run,
  list(b=tb[,i],
  method = 'LM-BG2-B',
  method.d = 'GANN',
  control = list(fnscale = -1),
  lower = c(1E-10, 1, 4), upper = c(1, 200,20));
report[i, c("n", "z0", "m")] = fitBinom$par[1:3]
report$[i] = (report$z[i] -1) / report$t0[i]

# fit[i,j] <- rbinding(c(Genocov, PosSise, maximum, minimum, average, se, standarddev, CompLog, CompRate, G)
```

```
## [1] 0.09280715 6.6005307391 6.000000000
## [1] 0.00280715 6.6005307391 6.000000000
## [1] 0.00180715 6.6005307391 6.000000000
## [1] 0.00280715 6.6005307391 6.000000000
## [1] 0.00280715 6.6005307391 6.000000000
## [1] 0.002280715 6.6005307391 6.001000000

3
```
Appendix C: Source Code for Bar Graphs

DRvsYPD.R

Katie Rouse
Fri Jun 22 13:00:00 2018

## Katie Rouse - Thesis code
## June 06, 12:30am

# Compare ALS of DR and YPD
# Compare the R, t0, and n

# clear environment
rm(list = ls())

library(flexsurv)

## Loading required package: survival

library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:stats':
##
## loess

library(ggplot2)

setwd("~/github/rouse_thesis/rouse_code/")
reportDR <- read.csv("CSV Files/05d_binomFit_June05.csv", header = TRUE, stringsAsFactors = FALSE)
reportYPD <- read.csv("CSV Files/ypd_binomFit_June05.csv", header = TRUE, stringsAsFactors = FALSE)
tmp1 = gsub("(\d(6.").", "", reportDR$files)
strainNames = gsub(".csv", ",", tmp1)
strainNames

## [1] "act1" "odb3" "atf3" "ahp1" "sim21" "sim29" "sim44"
## [8] "ain3" "alg12" "alt1" "atg1" "atg31" "atg33" "atg32"
## [15] "atg8" "atp1" "atp11" "atp5" "ben4" "bre5" "BY4741"
## [22] "BY4742" "cph20" "ctpl" "ctq2" "cgo8" "cgo4" "cttl1"
## [29] "cup2" "cyt2" "dbp3" "dil1" "dot5" "erg6" "ext3"
## [36] "exr1" "fob1" "fze1" "gcd1" "gcd2" "gcd20" "gcd3"
## [43] "gcn4" "gef1" "gpx1" "gpx2" "gtt2" "hac1" "hasp4"
## [50] "hca1" "hna1" "hmx1" "hop1" "kor7" "hasp104" "hst4"
## [57] "hxk2" "hyr1" "idh1" "ifs1" "1nh1" "1nh2" "ins61"
## [64] "insp53" "isz1" "isz2" "let1" "mcr1" "md1" "mef1"
## [71] "mgm1" "mrt1" "mrt22" "mrp11" "mrp34" "mth1" "mcr1"
## [78] "ndel" "ndi1" "nml1" "nha1" "nhx1" "nup170" "nup84"
```r
# [85] "pda1"  "pdb1"  "pdx1"  "pex13"  "pex19"  "phb1"  "phb2"
# [92] "phk2"  "pmc1"  "pmr1"  "pmt1"  "pnc1"  "por1"  "pos5"
# [99] "ptc6"  "put3"  "rad23"  "rad50"  "rad51"  "rad52"  "rad55"
# [106] "rad57"  "rad6"  "rgd1"  "rip1"  "rit1"  "rip19a"
# [113] "rlp120b"  "rlp121b"  "rlp131a"  "rlp16b"  "rlp19a"  "rlp4"  "rlp17a"
# [120] "rps26b"  "rtg2"  "sac6"  "sam1"  "sch9"  "sgf73"  "sgs1"
# [127] "sin4"  "sir2"  "sir3"  "sis2"  "sod1"  "sod2"  "spf1"
# [134] "spt4"  "stf1"  "stf2"  "stv1"  "stf1"  "stf2"  "tjf4631"
# [141] "tsi11"  "tom6"  "tom70"  "tom71"  "tor1"  "trx1"  "tsa1"
# [148] "tsm2"  "uwp4"  "ugd1"  "ung1"  "vsc1"  "vma15"  "vma2"
# [155] "vma21"  "vma5"  "vma6"  "vph1"  "vps3"  "vps4"  "vtc4"
# [162] "whi5"  "yap1"  "ybr238c"  "ygr130c"  "ylb47"  "yme1"  "ypt6"

compare = data.frame(cbind(strainNames))
genotype <- strainNames
compare$avgRLS.DR <- reportDR$avgLS
compare$avgRLSYPD <- reportYPD$avgLS
compar$R.YPD <- reportD$R
compar$R.DR <- reportD$RD
compar$0.YPD <- reportYPD$0
compar$0.DR <- reportD$0
compar$YPD <- reportYPD$Y
compar$DR <- reportD$DR
compar$changeRLS <- abs(compare$avgRLS.DR - compare$avgRLSYPD)
compar$percentChange <- (compare$changeRLS / compare$avgRLSYPD) * 100
compar$num.DR <- reportDR$samplesize
compar$num.YPD <- reportYPD$samplesize
setwd("~/github/rouse_thesis/rouse_code/figures")
write.csv(compare,"DRvsYPD.csv")
ordered <- compare[order(-compare$percentChange),]

#my.data <- ordered[1:10,]

## REPEAT FOR SIGNIFICANT GENES ##
significant = c("fob1", "sch9", "pda1", "sod1", "sod2", "phb1", "phb2")
my.data <- compare[(compare$strainNames %in% significant),]

### TOP 10 RLS ###
rls <- c(my.data$avgRLSYPD, my.data$avgRLS.DR)
geno <- rep(my.data$strainNames,2)
bars <- data.frame(geno, rls)
#condition <- c(rep("YPD",10),rep("DR",10))
condition <- c(rep("YPD",7),rep("DR",7))
```
```r
# pdf("./Fig1.pdf", width=12, height=8, paper='special')
pdf("./RLS_SigGenes.pdf", width=12, height=8, paper='special')

ggplot(bars, aes(geno, rls)) +
  geom_bar(stat="identity", aes(fill=condition), position="dodge") +
  xlab("Selected Genotype") + ylab("Average RLS") + theme_bw() +
  theme(text=element_text(size=22, face="italic"),
        axis.title=element_text(size=25, face="bold"))
dev.off()

## pdf
## 2

### SHOW DIFF IN R VALUE ####

r.value <- c(my.data$R.YPD, my.data$R.DR)
geno <- rep(my.data$strainNames, 2)
bars <- data.frame(geno, r.value)

condition <- c(rep("YPD", 10), rep("DR", 10))

pdf("./CompareR_1.pdf", width=12, height=8, paper='special')
pdf("./CompareR_SigGenes.pdf", width=12, height=8, paper='special')

ggplot(bars, aes(geno, r.value)) +
  geom_bar(stat="identity", aes(fill=condition), position="dodge") +
  xlab("Selected Genotype") + ylab("R Value") + theme_bw() +
  theme(text=element_text(size=22, face="italic"),
        axis.title=element_text(size=25, face="bold"))
dev.off()

## pdf
## 2

### SHOW DIFF IN t0 Value ####

t0.value <- c(my.data$t0.YPD, my.data$t0.DR)
geno <- rep(my.data$strainNames, 2)
bars <- data.frame(geno, t0.value)

condition <- c(rep("YPD", 10), rep("DR", 10))

pdf("./Compare.t0_1.pdf", width=12, height=8, paper='special')
pdf("./Compare.t0_SigGenes.pdf", width=12, height=8, paper='special')

ggplot(bars, aes(geno, t0.value)) +
  geom_bar(stat="identity", aes(fill=condition), position="dodge") +
  xlab("Selected Genotype") + ylab("Initial virtual age t0") + theme_bw() +
  theme(text=element_text(size=22, face="italic"),
        axis.title=element_text(size=25, face="bold"))
dev.off()

## pdf
## 2
```

3
### Compare n value ###

n.value <- c(my.data$n.YPD, my.data$n.DR)
geno <- rep(my.data$strainNames,2)
bars <- data.frame(geno, n.value)
#condition <- c(rep("YPD",10),rep("DR",10))
condition <- c(rep("YPD",7),rep("DR",7))

#pdf("./Compare_n_1.pdf", width=12, height=8, paper='special')
pdf("./Compare_n_SigGenes.pdf", width=12, height=8, paper='special')
ggplot(bars,aes(geno,n.value))*
  geom_bar(stat="identity",aes(fill=condition), position="dodge") +
  xlab("Selected Genotype") + ylab("No. of interactions/ essential gene") + theme_bw() +
  theme(text=element_text(size=22,face="italic"),
        axis.title=element_text(size=25,face="bold"))
dev.off()

##
## 2

### Break genes into RLS extension and shortening under DR conditions ###

extend = c("nhx1", "pdb1", "phb2", "por1", "fob1", "pda1")
shorten = c("rad6", "sod2", "sod1", "vma13", "vma2", "vma5", "sch9", "sod1")

df.extend <- compare[(compare$strainNames %in% extend),]
df.shornten <- compare[(compare$strainNames %in% shorten),]

rls.extend <- c(df.extend$avgRLS.YPD, df.extend$avgRLS.DR)
rls.shornten <- c(df.shornten$avgRLS.YPD, df.shornten$avgRLS.DR)

t0.extend <- c(df.extend$t0.YPD, df.extend$t0.DR)
t0.shornten <- c(df.shornten$t0.YPD, df.shornten$t0.DR)

n.extend <- c(df.extend$n.YPD, df.extend$n.DR)
n.shornten <- c(df.shornten$n.YPD, df.shornten$n.DR)

R.extend <- c(df.extend$R.YPD, df.extend$R.DR)
R.shornten <- c(df.shornten$R.YPD, df.shornten$R.DR)

ngen <- rep(df.extend$strainNames,2)
genon <- rep(df.shornten$strainNames,2)

bars <- data.frame(geno, rls.extend)
bars.s <- data.frame(geno.s, rls.shornten)

bars.t0.e <- data.frame(geno, t0.extend)
bars.t0.s <- data.frame(geno.s, t0.shornten)

bars.n.e <- data.frame(geno, n.extend)
bars.n.s <- data.frame(geno.s, n.shornten)
bars.R.e <- data.frame(geno, R.extend)
bars.R.s <- data.frame(geno, R.shorten)

condition <- c(rep("YPD",7),rep("DB",7))

pdf("./extend_rls.pdf",height=8,width=12,paper='special')
ggplot(bars,aes(geno,rls.extend)) +
  geom_bar(stat="identity",aes(fill=condition),position="dodge") +
  xlab("Genotype") + ylab("Average RLS") + theme_bw() + ggsave("OR Extension of RLS Genotypes") +
  theme(text=element_text(size=22,face="italic"),
        axis.title=element_text(size=25,face="bold"))
dev.off()

## pdf
## 2

pdf("./shorten_rls.pdf",height=8,width=12,paper='special')
ggplot(bars,aes(geno,rls.extend)) +
  geom_bar(stat="identity",aes(fill=condition),position="dodge") +
  xlab("Genotype") + ylab("Average RLS") + theme_bw() + ggsave("OR Extension of RLS Genotypes") +
  theme(text=element_text(size=22,face="italic"),
        axis.title=element_text(size=25,face="bold"))
dev.off()

## pdf
## 2

pdf("./extend_t0.pdf",height=8,width=12,paper='special')
ggplot(bars.t0.e,aes(geno,t0.extend)) +
  geom_bar(stat="identity",aes(fill=condition),position="dodge") +
  xlab("Genotype") + ylab("t0") + theme_bw() + ggsave("t0 of RLS Extended Genotypes") +
  theme(text=element_text(size=22,face="italic"),
        axis.title=element_text(size=25,face="bold"))
dev.off()

## pdf
## 2

pdf("./shorten_t0.pdf",height=8,width=12,paper='special')
ggplot(bars.t0.s,aes(geno.t0.shorten)) +
  geom_bar(stat="identity",aes(fill=condition),position="dodge") +
  xlab("Genotype") + ylab("t0") + theme_bw() + ggsave("t0 of RLS Shortened Genotypes") +
  theme(text=element_text(size=22,face="italic"),
        axis.title=element_text(size=25,face="bold"))
dev.off()

## pdf
## 2
Appendix D: Source Code for Fitting Probability Density Curves

density.R
katierouse
Fri Jun 22 12:47:38 2018

```r
## Katie Rouse – Thesis code
## June 07, 12:30am

##
library(ggplot2)
library(flexsurv)

## Loading required package: survival

library(survival)
## source("./lifespan.r")
## setwd("~/github/rouse_thesis/rouse_code/

### Function for probability density
### for the probability density profiles y_i = \rho(x_i), where x is RLS
### the average RLS is ARLS = \sum(y_i \times x_i) / \sum(y_i)
### the fitted RLS using pd.plot gives f_i = \hat{\rho}(x_i)
### and the fitted ARLS = \sum(f_i \times x_i) / \sum(f_i)
### which sum(y_i) or sum(f_i) should be 1 over (0,inf)

### Binomial Mortality
lh.binomialMortality.single.run <- function(R0n, lifespan, debug = 0)
{
  I = R0n[1];
t0 = R0n[2];
N = R0n[3];
my.data = lifespan[!is.na(lifespan)];
log_s = (I + t0 / n) * (1 - (1 + my.data/t0)^-1);
log_m = log(I) + (n-1) * log(1 + my.data/t0);
my.lh = sum(log_s) + sum(log_m);
if(debug) { print (my.lh ); } # trace the convergence
  ret = -my.lh # because optim seems to maximize
}

### Binomial probability density
pd.bin = function(R, t0, n, rls)
{
  # visibility $net

  I = R;
lifespan = rls
my.data = lifespan[!is.na(lifespan)];
log_s = (I + t0 / n) * (1 - (1 + my.data/t0)^-1);
log_m = log(I) + (n-1) * log(1 + my.data/t0);
```

1
s = exp(log_x)
m = exp(log_x)
ret = s*m
#print(ret)

### Gompertz

gomp = function(R, C, x)
{  
  # R is the rate and C is the Shape  
  return(R*exp(G*x-(R/C)*exp((G*x)-1)))
}

### multilplot###
multilplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL)
{
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  if (is.null(layout)) {
    # Make the plot
    # ncols: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols = ceiling(numPlots/cols)),
                      ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$rov,
                       layout.pos.col = matchidx$col))
    }
  }
}
### Issue

```r
my.dataDF <- read.csv("CSV Files/05d_binomFit_June06.csv", header = TRUE, stringsAsFactor = FALSE)
my.dataPD <- read.csv("CSV Files/05d_binomFit_June06.csv", header = TRUE, stringsAsFactor = FALSE)

clean <- my.dataDF$genotype
list <- my.dataDF$sfiles

###nov the loops
###
for (i in 1:length(list)) {
  ###reading the rls file from DR
  title <- paste("-\github\rouse_thesis\rouse_code\Figures\GOMPvBISOM/", names[i], ".fitting", ".pdf"
  dr.rls.file <- paste("05d/", list[i], sep = "")
  dr.rls.data <- read.csv(dr.rls.file, header = TRUE, stringsAsFactor = FALSE)
  dr.rls <- dr.rls.data$srls
  dr.freq <- count(dr.rls)
  dr.x <- dr.freq%>%
  dr.y <- dr.freq%>%length(dr.rls)
  dr.gomp.fit <- flexsurvreg(formula = Surv(dr.rls) ~ 1, dist = "gompertz")
  dr.gomp.B <- dr.gomp.fit%>%res[2,1]
  dr.gomp.G <- dr.gomp.fit%>%res[1,1]

  cr.g <- pd.gomp(dr.gomp.B, dr.gomp.G, dr.x)
  dr.ave.raw <- round(sum(dr.x*dr.y)/sum(dr.y), digits=1)
  dr.ave.g <- round(sum(dr.x*dr.g)/sum(dr.g), digits=1)
  dr.ave <- paste(dr.ave.raw,"/\",dr.ave.g, sep="") # *ave for Gomperts fittings

### parameters R, G and RLS age are needed ###

Rhat = dr.gomp.R;
Ghat = dr.gomp.G;
\( t_0 \) = (nhat-1)/Ghat;

## Using the binomial equation to find R, t0, and n
fitBinom = optim(c(Rhat, t0, nhat), lln.binomialMortality.single.run, 
  lifespan = dr.rls, 
  method = "L-BFGS-B", 
  method = "SANN", 
  lower = list(f=0.1, \( \alpha = 0.1 \)), upper = c(1, 200, 200))

dr.b.R <- fitBinom$par[1]

### Run Binning


```
dr.ave.b <- round(sum(dr.x*dr.b)/sum(dr.b), digits = 1)
dr.aveb <- paste(dr.ave.raw, "/", dr.ave.b, sep="")

dr.g.err <- round(dr.ave.g-dr.ave.raw, digits=3)
corr <- cor.test(dr.g,dr.y)
p.val <- round(corr$p.value,digits=3)
dr.ave <- paste(dr.ave.raw,"/",dr.ave.g,"",p="",p.val,sep="") # .ave for Gompertz fittings

dr.b.err <- round(dr.ave.b-dr.ave.raw, digits=3)
corr.b <- cor.test(dr.b,dr.y)
p.val.b <- round(corr.b$p.value,digits=3)
dr.aveb <- paste(dr.ave.raw,"/",dr.ave.b,"",p="",p.val.b,sep="") # .ave for binomial fittings

### reading rls from YPD ###
ypd.rls.data <- read.csv(ypd.rls.file, header = TRUE, stringsAsFactor = FALSE)
ypd.rls <- ypd.rls.data$rls
ypd.freq <- count(ypd.rls)
yp.x <- ypd.freq$x
yp.y <- ypd.freq$freq/length(ypd.rls)
ypd.gomp.fit <- flexsurvreg(formula=Surv(ypd.rls) ~ 1, dist = "gompertz")
ypd.gomp.R <- ypd.gomp.fit@res[2,1]
ypd.gomp.G <- ypd.gomp.fit@res[1,1]

Rhat = ypd.gomp.R;
Ghat = ypd.gomp.G;

zhat = 6;
t0 = (nhat-1)/Ghat;
fitBinom = optim(c(Rhat, t0, zhat), llh.binomialMortality.single.run, 
  lihespan = dr.rls, 
  method = "L-BFGS-B", 
  #method = "SANN", 
  #control=list(fnscale=-1), 
  lower=c(1E-10, 1, 4), upper = c(1, 200,20))

ypd.b.R <- fitBinom@par[1]
ypd.b.t0 <- fitBinom@par[2]
ypd.b.n <- fitBinom@par[3]
ypd.b.G <- (ypd.b.n -1) / (ypd.b.t0)

ypd.g <- pdc.gomp(ypd.gomp.R, ypd.gomp.G, ypd.x)
ypd.b <- pdc.bin(ypd.b.R, ypd.b.t0, ypd.b.n, ypd.x)
ypd.ave.raw <- round(sum(ypd.x*ypd.y)/sum(ypd.y), digits=1)
ypd.ave.g <- round(sum(ypd.x*ypd.g)/sum(ypd.g), digits=1)
ypd.ave.b <- round(sum(ypd.x*ypd.b)/ sum(ypd.b), digits =1)
ypd.ave <- paste(ypd.ave.raw,"/",ypd.ave.g, sep="")
ypd.aveb <- paste(ypd.ave.raw,"/",ypd.ave.b, sep="")

ypd.g.err <- round(ypd.ave.g-ypd.ave.raw, digits=3)
corr.gypd <- cor.test(ypd.g,ypd.y)
p.val.gypd <- round(corr,gypd$p.value,digits=3)
ypd.ave <- paste(ypd.ave.raw,"/",ypd.ave.g"," p="",p.val.gypd,sep="") # *ave for Gompertz fittings

ypd.b.err <- round(dr.ave-b.dr.ave.raw,digits=3)
corr.bYPD <- cor.test(ypd,b.pval)
p.val.bYPD <- round(corr.bYPD$p.value,digits=3)
ypd.aveb <- paste(ypd.ave.raw,"/",ypd.ave.b," p="",p.val.bYPD)

### set figure file
dr.all <- data.frame(dr.x,dr.y,dr.g)
ypd.all <- data.frame(ypd.x,ypd.y,ypd.g)

dr.title.g <- paste(List[1]," DR (Gompertz)",sep="")
dr.title.b <- paste(List[1]," DR (Binomial)",sep="")
ypd.title.g <- paste(List[1]," YPD (Gompertz)",sep="")
ypd.title.b <- paste(List[1]," YPD (Binomial)",sep="")

pg.dr <- ggplot(dr.all,aes(dr.x)) + theme_bw() + xlab("Gompertz") +
  geom_point(aes(y=dr.y)) + geom_line(aes(y=dr.g), linetype="dashed") + ggtitle(dr.title.g) +
  annotate("text", x=(max(dr.x)*0.8),y=(max(dr.y)*0.8),label=dr.ave,fontface="italic",color="red")

pg.ypd <- ggplot(ypd.all,aes(ypd.x)) + theme_bw() + xlab("Gompertz") +
  geom_point(aes(y=ypd.y)) + geom_line(aes(y=ypd.g), linetype="dashed") + ggtitle(ypd.title.g) +
  annotate("text", x=(max(ypd.x)*0.8),y=(max(ypd.y)*0.8),label=ypd.ave,fontface="italic",color="red")

# pd.drbinom <- ggplot(dr.all,aes(log(dr.a))) + theme_bw() + xlab("Binomial") +
#  geom_point(aes(y=dr.y)) + geom_line(aes(y=dr.binom), linetype="dashed") + ggtitle(dr.title.b) +
#  annotate("text", x=(max(dr.x)*0.8),y=(max(dr.y)*0.8),label=dr.binom,fontface="italic",color="red")

pdf(title="plot",width=8,height=6,paper="special")
multiplot(pg.dr,pg.ypd,pd.dr,pd.ypd,col=2)
dev.off()

## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
Appendix E: Probability Density Profiles

Figure E1: Probability Density Curves for the binomial and Gompertz models of nhx1

Figure E2: Probability Density Curves for the binomial and Gompertz models of pdb1
Figure E3: Probability Density Curves for the binomial and Gompertz models of phb1

Figure E4: Probability Density Curves for the binomial and Gompertz models of phb2
Figure E5: Probability Density Curves for the binomial and Gompertz models of por1

Figure E6: Probability Density Curves for the binomial and Gompertz models of pda1
Figure E7: Probability Density Curves for the binomial and Gompertz models of rad6

Figure E8: Probability Density Curves for the binomial and Gompertz models of sod2
Figure E9: Probability Density Curves for the binomial and Gompertz models of vma13

Figure E10: Probability Density Curves for the binomial and Gompertz models of vma5
Figure E11: Probability Density Curves for the binomial and Gompertz models of sod1

Figure E12: Probability Density Curves for the binomial and Gompertz models of sch9
Appendix F: Source code for Contour Plots

```r
### Katie Rouse ###
## Thesis - contour plots to see what is more significant. R or t0
## June 16th, 2018 011:00am

### Shortening contour plot

```text
library(flexsurv)

# Loading required package: survival
library(ggplot2)
library(plyr)
setwd("~/github/rouse_thesis/rouse_code/CSV Files/")

###
my.data <- read.csv("shorten.csv", header=TRUE, stringsAsFactor=FALSE)
Genotype <- my.data$strainNames

### DR fitting
DR.ARLS <- my.data$avgRLS.DR
DR.H <- my.data$DR

###
my.data$DR.G
DR.t0 <- my.data$t0.DR
DR.lnR <- log(DR.H)

### YPD fitting
YPD.ARLS <- my.data$avgRLS.YPD
YPD.R <- my.data$RP.YPD
YPD.lnR <- log(YPD.R)
YPD.t0 <- my.data$t0.YPD

### median lifespan equation
con.test = function(t0, n, t)
{
  #log = ((log((t1/t0))^(n-1))*log(2)
  top = n*log(2)
  bottom = t-(((1-t)/t0)^(n-1))
  return(top/bottom)
}
```
s <- c(1:length(DR.R))
#0. cont <- (s-1)*(200/(length(DR.R)-1))+1
0. cont <- c(10,30,50,75,100,150,200)
#n. cont <- c((s-1)*(4/(length(DR.R)))+3.8)
n. cont <- c(rep(6,7))
t. cont <- c((s-1)*(35/(length(DR.R)-1))+.1)

### Gompertz, G for contour
gg2 <- log(com.test(t. cont, n. cont[1], t. cont[1]))
0g5 <- log(com.test(t. cont, n. cont[2], t. cont[2]))
gg10 <- log(com.test(t. cont, n. cont[3], t. cont[3]))
gg15 <- log(com.test(t. cont, n. cont[4], t. cont[4]))
gg20 <- log(com.test(t. cont, n. cont[5], t. cont[5]))
gg25 <- log(com.test(t. cont, n. cont[6], t. cont[6]))
gg36 <- log(com.test(t. cont, n. cont[7], t. cont[7]))

sc <- t. cont
xsc <- c(208)
ysc <- c(gg2[7],gg5[7],gg10[7],gg15[7],gg20[7],gg25[7],gg36[7])

data.frame(DR.t0,DR.lnR,YPD.t0,YPD.lnR,gg2,gg5,gg10,gg15,gg20,gg25,gg36)

xcor.DR <- DR.t0-2
xcor.DR[6] = xcor.DR[6]-7
ycor.DR <- DR.lnR - 0.2

### Groups based on Int0 and R values####

pdf("./shorten_contour.pdf", width=6, height=6, paper='special')
ggplot(all.con, aes(DR.t0)) + ggtitle(label = "Shortening Effect of R and t0 on RLS") +
  theme_bw() + xlab("t0") + ylab("lnR") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_line(colour="black"),
        axis.title=element_text(size=14,face="bold"),
        axis.text=element_text(size=12)) +
  geom_line(aes(x=t0, y=gg2), col="grey", linetype=2) +
  geom_line(aes(x=t0, y=gg5), col="grey", linetype=2) +
  geom_line(aes(x=t0, y=gg10), col="grey", linetype=2) +
  geom_line(aes(x=t0, y=gg15), col="grey", linetype=2) +
  geom_line(aes(x=t0, y=gg20), col="grey", linetype=2) +
  geom_line(aes(x=t0, y=gg25), col="grey", linetype=2) +
  geom_line(aes(x=t0, y=gg36), col="grey", linetype=2) +
  #geom_line(aes(x=7, y=gg50), col="grey", linetype=2) +
  geom_point(aes(x=DR.t0, y=DR.lnR), size=1, color="black", shape=1) +
  geom_point(aes(x=YPD.t0, y=YPD.lnR), size=1, color="black") +
  xlim(c(0,210))+ylim(c(-10,2))+
  annotate("text",x=xcor.DR, y=ycor.DR, label=Genotype, fontface="italic", size=3)+
  annotate("text",x=xsc, y=ysc, label=ac, col="black", size=3)+
  geom_segment(aes(x=YPD.t0, y=YPD.lnR, xend=DR.t0, yend=DR.lnR),
               linetype=1, arrow=arrow(length=unit(0.15,"cm")), size=0.4, col="brown")
### Katie Rouse ###

#### Thesis - contour plots to see what is more significant, R or t0
#### June 15th, 2018 01:11am

```r
library(flexsurv)

## Loading required package: survival

library(ggplot2)
library(plyr)

setwd("~/github/rousse_thesis/rousse_code/CSV Files/")

###

my.data <- read.csv("extend.csv", header=TRUE, stringsAsFactors=FALSE)
Genotype <- my.data$strainFazes

#DR fitting
DR.ALS <- my.data$avgRLS.DR
DR.R <- my.data$R.DR
#DR.C <- my.data$DR.G
DR.t0 <- my.data$t0.DR
DR.IlnA <- log(DR.R)

#YPD fitting
YPD.ALS <- my.data$avgRLS.YPD
YPD.R <- my.data$R.YPD
YPD.lnR <- log(YPD.R)
YPD.G <- my.data$YPD.G
YPD.t0 <- my.data$t0.YPD

### median lifespan equation

cen.test <- function(t0,n,t)
{
  #top = ((log(1+(t/t0)))^(n-1))#log(2)
top = n*log(2)
  bottom = t-((1-(t/t0))^(n-1))
return(top/bottom)
}

###

u <- c(1:length(DR.R))
```
#t0.cont <- (a-1)*(100/(length(DR.R)-1))+5
t0.cont <- c(5,15,20,40,60,80,100)
n.cont <- c((a-1)*(6/(length(DR.R)-1))+4.2)
t.cont <- c((a-1)*(38/(length(DR.R)-1))+8)

### Computa, C for contour

gg2 <- log(com.test(t0.cont, n.cont[1], t.cont[1]))
gg5 <- log(com.test(t0.cont, n.cont[2], t.cont[2]))
gg10 <- log(com.test(t0.cont, n.cont[3], t.cont[3]))
gg15 <- log(com.test(t0.cont, n.cont[4], t.cont[4]))
gg20 <- log(com.test(t0.cont, n.cont[5], t.cont[5]))
gg27 <- log(com.test(t0.cont, n.cont[6], t.cont[6]))
gg36 <- log(com.test(t0.cont, n.cont[7], t.cont[7]))

sc <- t0.cont
xsc <- c(17,43,rep(103,3))
ysc <- c(-2.1,-2.3,gg10[7],gg15[7],gg20[7],gg27[7],gg36[7])

all.con <- data.frame(DR.t0,DR.lnR,YPD.t0,YPD.lnR,gg2,gg5,gg10,gg15,gg20,gg27,gg36)

xcor.DR <- DR.t0 + 3.5
ycor.DR <- DR.lnR - 0.08

#### groups based on lnR and R values####

pdf("./extend_contour.png", width=6, height=6, paper='special')
ggplot(all.con, aes(DR.t0))+
ggtitle(label = "Extending Effect of R and t0 on HLS") +
theme_bw() +
  xlab("t0") +
  ylab("lnR") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_line(colour="black"),
        axis.title = element_text(size=14,facess="bold"),
        axis.text = element_text(size=12)) +
geom_line(aes(x=t0.cont, y=gg2),col="grey",linetype=2) +
geom_line(aes(x=t0.cont, y=gg5),col="grey",linetype=2) +
geom_line(aes(x=t0.cont, y=gg10),col="grey",linetype=2) +
geom_line(aes(x=t0.cont, y=gg15),col="grey",linetype=2) +
geom_line(aes(x=t0.cont, y=gg20),col="grey",linetype=2) +
geom_line(aes(x=t0.cont, y=gg27),col="grey",linetype=2) +
geom_line(aes(x=t0.cont, y=gg36),col="grey",linetype=2) +
# geom_line(aes(x=G,y=gg50),col="grey",linetype=2) +
geom_point(aes(x=DR.t0,y=DR.lnR),size=1,color="black",shape=1) +
geom_point(aes(x=YPD.t0,y=YPD.lnR),size=1,color="black") +
  xlim(c(0,100))+
  ylim(c(-7,-2)) +
  annotate("text",x=xcor.DR,y=ycor.DR,label=Genotype,fontface="italic",size=3) +
  annotate("text",x=xsc,y=ysc,label=ss,col="black",size=3) +
  geom_segment(aes(x=YPD.t0,y=YPD.lnR),xend=DR.t0, yend=DR.lnR, linetype=1, arrow=arrow(length=unit(0.15,"cm")),size=0.4,col="brown")

## Warning: Removed 5 rows containing missing values (geom path).