2012

Mathematical Modeling and Disease Related Applications: A New Method of Estimating Bacterial Mutation Rates, Dynamics of Killer Yeast in a Chemostat, and Other Problems

Nicholas Fitzgerald McClure
The University of Montana

Mathematical Modeling and Disease Related Applications: A New Method of Estimating Bacterial Mutation Rates, Dynamics of Killer Yeast in a Chemostat, and Other Problems.

By

Nicholas McClure

B.A., St. John's University/College of St. Benedict, Collegeville, MN, 2005
M.A., The University of Montana, Missoula, MT, 2007

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctorate of Philosophy
in Mathematics

The University of Montana
Missoula, MT

May 2012

Approved by:

Sandy Ross, Associate Dean of The Graduate School
Graduate School

Dr. Leonid Kalachev, Chair
Mathematical Sciences

Dr. Frank Rosenzweig
Division of Biological Sciences

Dr. John Bardsley
Mathematical Sciences

Dr. Emily Stone
Mathematical Sciences

Dr. Scott Miller
Division of Biological Sciences

Contents

Abstract

Among the main mechanisms of evolution we find adaptation, genetic drift, gene flow, mutation, natural selection and speciation. The following thesis explores bacterial mutation rate, mutants, mutators, and how competition acts as a driving force of adaptation.

In chapter 1, the system used in reference for mutation is the development of *Pseudomonas aeruginosa* antibacterial resistance in the Cystic Fibrosis patients. Current methods for determining mutation rate are quite involved mathematically or complicated experimentally. A new method for estimating mutation rate is discussed and compared with two other widely used methods.

In chapter 2, a microarray data set of different strains are compared. We find that the data set has strains that have sections of deleted genes. The two categories are analyzed across two different tests and genetic deletions are compared. The strains are also known to be mutators or non-mutators and the data will be used for strain classification.

In chapter 3, a competition model of killer and non-killer yeast is used to predict the outcome of competition in a continuous culture under nutrient limitation (e.g. a chemostat). This research explores possible detrimental effect on fitness when yeast harbor the killer virus under chemostat conditions and when a non-killer yeast strain can out compete a killer-strain of yeast.

Introduction

Mutation is a process that is the ultimate source of genetic diversity. It is the original source of genetic variation [41]. Mutation rate is defined as the probability of a mutation to be resistant to a measurable challenge occurring in one generation, whereas mutation frequency is just the proportion of mutated cells in a population at any specified time point. Mutation is the altering of genetic structure by insertions, deletions, amplifications, inversions, translocations and individual base pair changes. Since these changes in a genome may have widely different effects, this research confines itself to mutation rate in respect to a reproductive challenge, such as antibiotics. This restriction determines possible experiments for measuring the mutant frequency in a population. Mutations are assumed to be random along the genome [10]. Since most mutations are not beneficial, higher mutation rates result in a higher chance of harmful mutations, which leads to an increased likelihood of cell death. Lower mutation rates result in less genetic variation. Less genetic variation may lead to a decreased chance of a population surviving when faced with a new external challenge [14].

An example of an environment where bacterial populations develop a subpopulation with a high mutation rate are in the lungs of Cystic Fibrosis patients [28]. Cystic Fibrosis is a recessive genetic disease that results in a defective CFTR gene. The CFTR gene is responsible for choloride ion transport across cell membranes [29]. Loss of this function can result in thicker and stickier mucus and chronic lung infections, among other symptoms. Antibiotics are only a treatment and not a cure for the lung infections as bacteria accumulate resistance over time due to mutations. One of the main bacterial strains responsible for developing antibiotic resistance, among others, is *Pseudomonas aeruginosa*. In the

Cystic Fibrosis lung, this opportunistic pathogen develops a subpopulation of high mutation rate. Determining the mutation rate of *Pseudomonas aeruginosa* samples from Cystic Fibrosis patients is very important for identifying the severity of the infection.

Currently measuring mutation frequency in a bacterial population at a given time is one of the few ways of determining mutation rate. There are two main methods to measure mutation frequency that determine the mutation rate. They are fluctuation analyses [24] and mutant accumulation in a chemostat [21]. These methods have certain advantages and disadvantages. The method that is proposed and analyzed in Chapter 1, in this thesis, aims to combine the advantages of both methods and have fewer disadvantages.

To determine if there are any genes responsible for the trait of mutator or non-mutator, we compare aCGH microarray data between the two categories. Different methods of classification are explored based on a microarray test of 48 different strains in Chapter 2.

Some protocols call for multiple microarray tests to be performed for accuracy. It may be more efficient to be able to predict the presence and absence of genes from one microarray test from values obtained for another microarray test. To analyze this, nine strains were compared on two different microarray tests to determine how different microarray tests are correlated.

In Chapter 3, we outline a situation related to modeling competing yeast populations, where a susceptible population can outcompete an infectious population. The specific system that is described in this part of the thesis involves the killer-virus in yeast and the uninfected yeast interactions. The killer virus infects yeast and gives it the ability to produce a toxin which kills uninfected yeast cells. It is natural to suspect that the upkeep of a virus and the production of toxin detracts from the resources available for growth. Cell density in a chemostat is determined by its growth rate constant and the dilution rate constant of the system. If the dilution rate term is larger than the growth rate term, then the population is “washed out” from the chemostat. The possibility of setting a dilution rate high enough to wash out the killer yeast, but not the susceptible (non-killer) yeast, is explored using a mathematical model, and

a region in a parameter space where “wash out” can happen is determined from the data.

Chapter 1

Mutants and Mutation Rate Estimation

Modeling Mutation: Approaches to Estimating Mutation Rate Constants

1.1 Background

Mutation is a process that generates genetic diversity. Measuring mutation rate is important because mutation dictates how fast the genetic code of an organism changes. Mutation, in conjunction with other processes, is responsible for evolution [27, 9], speciation [37], disease [26, 6], and population survival among many others. In order to understand these processes, we must be able to measure mutation rate accurately. It is also important to understand the diversity of mutation rate within a population. Mutation rate variation in a single population may play an important role in chronic infectious diseases, such as Cystic Fibrosis.

Cystic Fibrosis is a genetic disease caused by a defect in the CFTR gene, which leads to the mucoidal output of the respiratory tract to become thicker than normal. Mucus in the lung provides protection against infection. Without such a barrier, Cystic Fibrosis patients are more susceptible to bacterial

infection by such opportunistic pathogens as *Pseudomonas aeruginosa*. In chronic lung infections, bacterial populations are more likely to develop antibacterial resistance due to the presence of mutators in the population. Mutators are cells with mutation rates much higher than expected. Experimentally, they will be more likely to develop antibacterial resistance. There may be a measurably larger amount of mutators in a Cystic Fibrosis lung infection than a non- Cystic Fibrosis infection. This is thought to happen due to the changing environment of antibiotics applied, which may select for a faster mutating population. With the changing antibiotics used, antibiotic resistant mutations come under positive selection. Due to this, mutators (cells with a high mutation rate) are thought to "hitch hike" along with generating beneficial mutations. In the following models, we will only consider a population with a fixed mutation rate constant, resulting in a mutated population from a non-mutated population. Since, in the presence of unlimited resources, bacterial populations will grow exponentially, in order to work with smaller numbers, we analyze the ratio of mutated cells to the total population or non-mutated cells to the total population instead of actual population numbers. These ratios provide insight to the model parameters which are important in determining the mutation rate constant.

1.2 Discrete Model

It is easy to think about bacteria replicating discretely: starting with one bacteria, there will be two bacteria after one generation, and four after two generations and so on. The problem with this approach is that normal sized bacterial colonies consist of millions or billions bacteria, so it is hard to envision all of them replicating synchronously. Asynchronous replication of small numbers of bacteria are handled using different statistical approaches [34]. These methods will be compared in later sections, but it is worth mentioning that populations of *Pseudomonas aeruginosa* sampled in the Cystic Fibrosis lung do not come from small or synchronous populations.

For our simple and introductory (illustrative) model we will consider only two populations, the non-mutated and mutated cell populations. In this chapter we are interested in determining the mutation

rate constant of a certain bacterial population by measuring the non-mutated and mutated populations. To formulate a discrete model, we must assume that each replication of a cell produces one more cell at each time step. The simplest model for the non-mutated and mutated populations may be formulated as follows:

$$\begin{aligned} n_{t+1} &= (2-\mu)n_t - d_n n_t, \\ N_{t+1} &= \mu_n n_t + 2N_t - d_N N_t. \end{aligned} \tag{1.1}$$

Here, n and N stand for the population of the non-mutated and mutated bacteria, respectively. Also, μ_n stands for the mutation rate constant, and d_n and d_N are the respective death rate constants. The discrete model assumes generation by generation doubling (synchronously). It also assumes the same growth rate and yield for the mutated and non-mutated bacteria. However, the mutated bacteria most likely have a different growth rate and yield. These assumptions (synchronicity, etc.) may be inaccurate.

We can rewrite system (1.1) to get:

$$\begin{aligned} n_{t+1} &= an_t, \\ N_{t+1} &= bn_t + gN_t. \end{aligned} \tag{1.2}$$

where $a = (2-\mu) - d_n$, $b = \mu$, and $g = 2 - d_N$. We can solve for n_t explicitly and substitute the solution back into the equation for N_{t+1}:

$$\begin{aligned} n_t &= C_1 a^t, \\ N_t &= bC_1 a^t + gN_t. \end{aligned} \tag{1.3}$$

Here C_1 is defined by initial conditions for n_t. We look for a solution of (1.3) in the form

$$N_t = C_2 g^t + A a^t. \tag{1.4}$$

Here C_2 is defined by initial conditions for N_t, and A is a yet unknown constant. The term Aa^t is the particular solution of the second, non-homogeneous, equation in (1.3). To find A, we need to substitute Aa^t into system (1.3). We obtain:

$$Aa = bC_1 + gA,$$

and thus,

$$A = \frac{bC_1}{a - g}. \tag{1.5}$$

Finally, substituting expression (1.5) for A into (1.4), we obtain for the mutated populations:

$$N_t = C_2 g^t + \frac{bC_1}{a - g} a^t \tag{1.6}$$

We will look at the ratio of the mutated to the whole population and the non-mutated to the whole population. We write

$$\frac{N_t}{n_t + N_t} = \frac{C_2 g^t + (bC_1)/(a - g)a^t}{C_2 g^t + (bC_1)/(a - g)a^t + C_1 a^t}, \tag{1.7}$$

or

$$\frac{N_t}{n_t + N_t} = \frac{(C_2 g^t)/(C_1)a^t + b/(a-g)}{(C_2 g^t)/(C_1)a^t + b/(a-g) + 1}. \tag{1.8}$$

We assume that under conditions without antibiotics, non-mutated cells are more viable compared to mutated cells [19], so we must have $a > g$. If $a > g$, then the g^t/a^t terms go to zero as time tends to infinity in Eq. (1.8), and the fraction approaches a fixed ratio (see Figure (1.1)),

$$\lim_{t \longrightarrow \infty} \frac{N_t}{n_t + N_t}\bigg|_{a>g} = \frac{b}{b + a - g} = \frac{\mu}{d_N - d_n}. \tag{1.9}$$

Similarly, we obtain the ratio $n_t/(n_t + N_t)$ by subtracting the Eq. (1.9) from 1.

$$\lim_{t \longrightarrow \infty} \frac{n_t}{n_t + N_t}\bigg|_{a>g} = \frac{a-g}{b + a - g} = \frac{\mu - d_n + d_N}{d_N - d_n}. \tag{1.10}$$

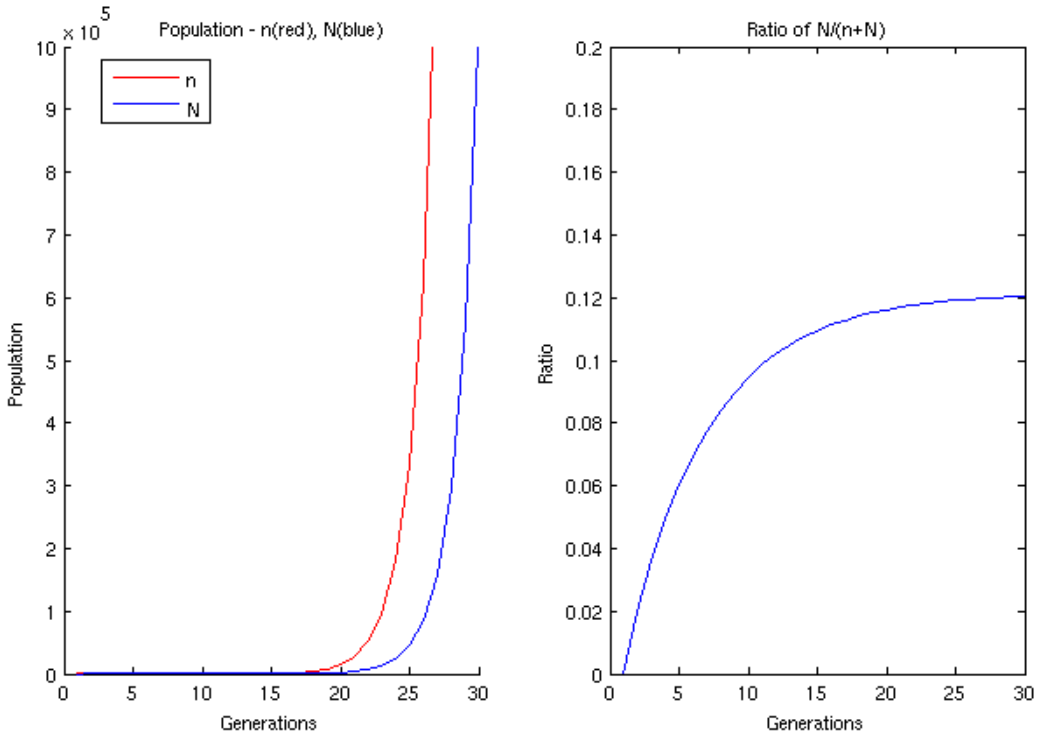

Figure 1.1: The discrete model: actual population values (left) and a plot of the ratio of $N/(n+N)$ (right). The parameter values were taken as follows: $\mu = 0.01$, $d_n = 0.1$, and $d_N = 0.4$.

During an infection, a patient may be treated with antibiotics to help reduce the bacterial population. This model predicts a ratio of mutated to the whole population that is independent of initial conditions. Because of this, if the antibacterial treatment is stopped, the population will return to the same ratio as before as time increases (see Figure (1.2)).

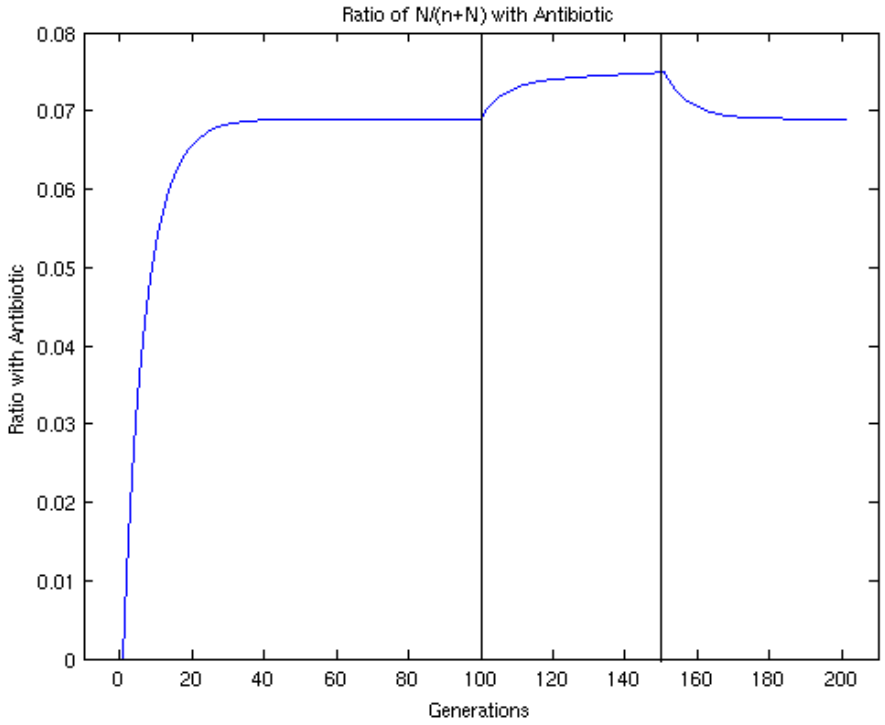

Figure 1.2: The discrete model, with an antibiotic treatment applied during generations 100-150. The parameter values were, $\mu = 0.01$, $d_n = 0.1$, and $d_N = 0.4$. During antibiotic treatment (100 generations to 150 generations) we change the death rate constants of the non-mutated by the effect, $c_{dn} = -0.01$, and the mutated death rate constant by a positive effect, $c_{dN} = +0.01$ on d_N.

During antibiotics, we assume that the death rate of the non-mutated population is increased and the death rate of the mutated population is decreased by a certain amount (c_{dn} and c_{dN}, respectively) (see legend of Figure 1.2). Because of this assumption, the ratio of mutated in the population does not go to one, but it does increase.

This approach naturally leads to question the importance of synchronicity and similar growth rates in our assumptions. We now move on to a continuous model, which allows one to more easily incorporate different growth rates and naturally describes bacterial population growth for asynchronous division.

1.3 Continuous Model

The continuous model is formulated under the assumption that reproduction is asynchronous. We do not make the assumption that growth rate and death rate constants are the same. The model has the form:

$$\begin{aligned} \frac{dn}{dt} &= k_n n - \mu n - d_n n, \\ \frac{dN}{dt} &= k_N N + \mu n - d_N N. \end{aligned} \tag{1.11}$$

Here, n and N stand for the populations of the non-mutated and mutated bacteria, respectively. Also, μ_n stands for the mutation rate constant and d_n and d_N are the respective death rate constants. In the continuous model k_n and k_N are the respective growth rate constants.

We introduce the notations $a = k_n - \mu - d_n$, $b = \mu k_n$ and $g = k_N - d_N$. Then, the system (1.11) can be written as

$$\begin{aligned} \frac{dn}{dt} &= an, \\ \frac{dN}{dt} &= bn + gN. \end{aligned} \tag{1.12}$$

Now we solve explicitly for $n(t)$ and $N(t)$.

$$\begin{aligned} n &= C_1 e^{at}, \\ N &= C_2 e^{gt} + \frac{bC_1}{a-g} e^{at}. \end{aligned} \tag{1.13}$$

Here C_1 is dependent on the initial population of n, and C_2 depends on the initial population of N.

We are interested in the ratio of $N/(n+N)$ or $n/(n+N)$, the ratio of mutated to the whole bacterial

population or non-mutated to the whole bacterial population. We can write:

$$\frac{N}{n+N} = \frac{C_2 e^{gt} + (bC_1)/(a-g)e^{at}}{C_2 e^{gt} + (bC_1)/(a-g)e^{at} + C_1 e^{at}}.$$

This simplifies to

$$\frac{N}{n+N} = \frac{(C_2/C_1)e^{(g-a)t} + b/(a-g)}{1 + (C_2/C_1)e^{(g-a)t} + b/(a-g)}. \tag{1.14}$$

Under normal conditions, we can assume that the growth rate of the mutated population is lower than the non-mutated [19], which means that $a > g$. When this condition holds, the exponential terms in Eq. (1.14) tend to zero as time goes to infinity, and we are left with a fixed ratio (see Figure (1.3)).

$$\lim_{t\longrightarrow\infty} \frac{N}{n+N}\bigg|_{a>g} = \frac{b}{b+a-g} = \frac{\mu}{k_n - k_N + d_N - d_n} \tag{1.15}$$

Similarly, for the ratio $n/(n+N)$, we get,

$$\lim_{t\longrightarrow\infty} \frac{n}{n+N}\bigg|_{a>g} = \frac{a-g}{b+a-g} = \frac{k_n - k_N - \mu + d_N - d_n}{k_n - k_N + d_N - d_n}. \tag{1.16}$$

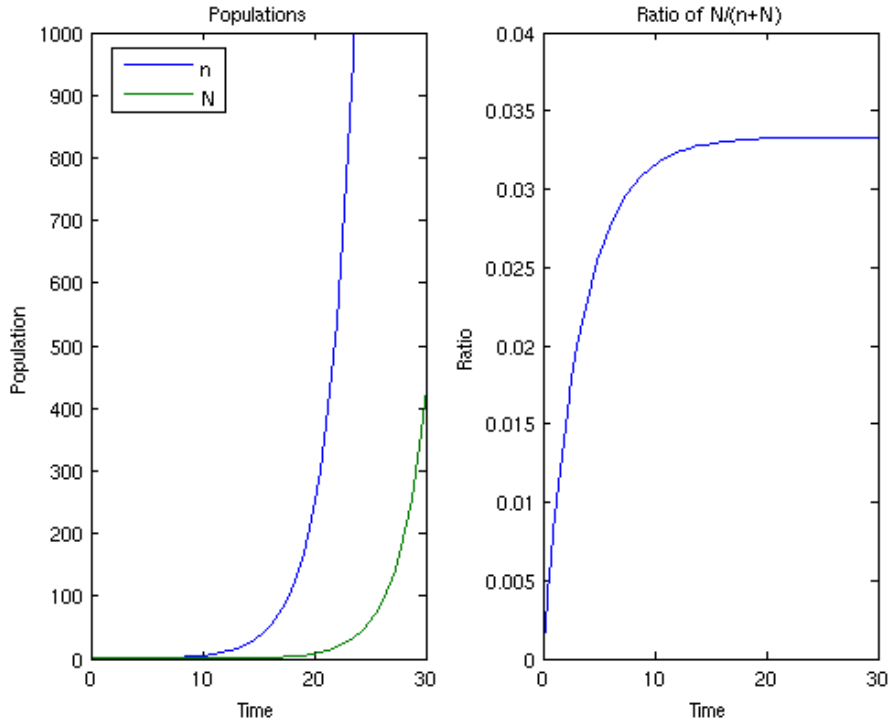

Figure 1.3: The continuous model. Actual population values (left) and a plot of the ratio of $N/(n+N)$ (right). The parameter values were chosen as follows: $\mu = 0.01$, $d_n = 0.1$, $d_N = 0.4$, and $k_n = k_N = 0.5$.

We now consider the case when $a = g$. The system in (1.12) becomes

$$\begin{aligned} \frac{dn}{dt} &= an, \\ \frac{dN}{dt} &= bn + aN. \end{aligned} \tag{1.17}$$

The explicit solution to (1.17) is

$$\begin{aligned} n(t) &= n_0 \mathrm{e}^{at}, \\ N(t) &= e^{at}\left(n_0 bt + N_0\right). \end{aligned} \tag{1.18}$$

The ratio, $N(t)/(N(t) + n(t))$ then becomes

$$\lim_{t\longrightarrow\infty} \frac{n}{n+N}\bigg|_{a=g} = \frac{N_0 + n_0 bt}{N_0 + n_0 bt + n_0} = 1.$$

This ratio approaches one because although they grow at the same rate, there is always continual production of mutated cells.

During an infection a patient may be treated with antibiotics to help reduce the bacterial population in the lungs. Here we define the effect of antibiotics as follows: they increase the death rate of the non-mutated population and decrease the death rate of the mutated population by a fixed amount. The factors of corresponding increase and decrease are described using c_{dn} and c_{dN}, respectively. This assumption means that the antibiotic does not kill off all the non-mutated instantly. It is made to assure that the non-mutated population does not go completely extinct, because there is always non-mutated cells (sensitive) in the lung. Alternatively, if we choose to kill all of the sensitive population, the model just becomes a growth model for mutated cells. However, since this model predicts a ratio of mutated to non-mutated that is independent of initial conditions, if the antibacterial treatment is stopped, the population will return to the same ratio as before (see Figure 1.3). The ratio from Eq. (1.15) under antibiotic administration is changed due to the change in the death rates.

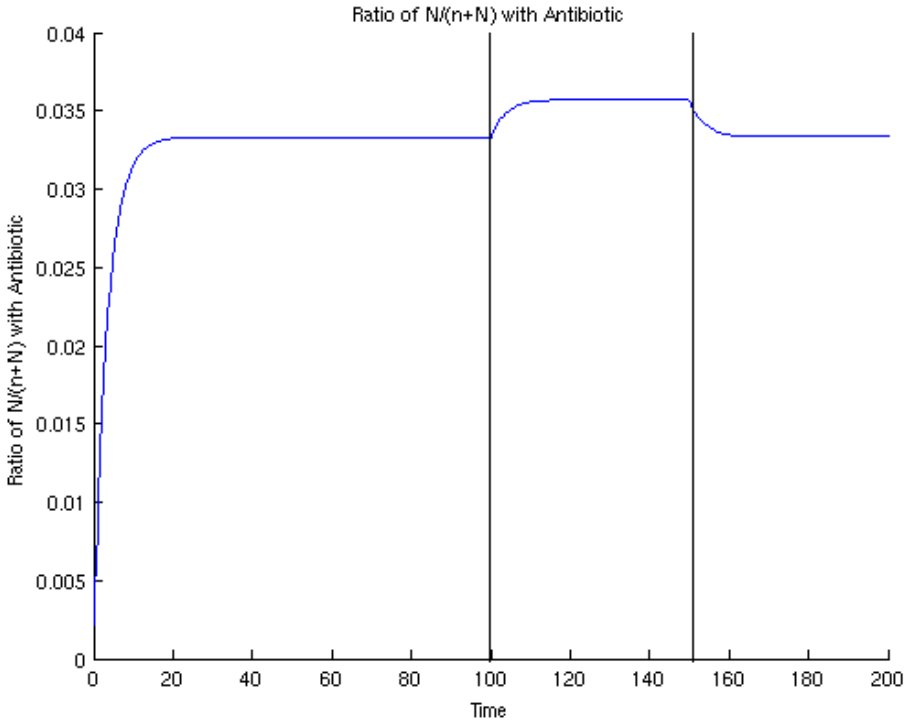

Figure 1.4: The continuous model with an antibiotic treatment applied during the times 100-150. The parameter values were chosen to be $\mu = 0.01$, $d_n = 0.1$, $d_N = 0.4$, and $k_n = k_N = 0.5$ During antibiotic treatment (100 generations to 150 generations) we change the death rate constants of the non-mutated by the effect, $c_{dn} = -0.01$, and the mutated death rate constant by a positive effect, $c_{dN} = +0.01$ on d_N.

The derivation of Equations (1.15) and (1.16) relied on the assumption that $k_N < k_n$. If this is not the case, then we re-derive under the new assumption that $k_N < k_n$ does not hold. We now assume that $a < g$ and proceed to look at the ratio $n/(n + N)$.

$$\frac{n}{n+N} = \frac{C_1 e^{at}}{C_1 e^{at} + (bC_1)/(a-g)e^{at} + C_2 e^{gt}}.$$

This simplifies to

$$\frac{n}{n+N} = \frac{C_1 e^{(a-g)t}}{C_1 e^{(a-g)t} + (bC_1)/(a-g)e^{(a-g)t} + C_2}. \tag{1.19}$$

For $a < g$ we write:

$$\lim_{t \longrightarrow \infty} \frac{n}{n+N}\bigg|_{a<g} = 0. \tag{1.20}$$

And similarly,

$$\lim_{t \longrightarrow \infty} \frac{N}{n+N}\bigg|_{a>g} = 1. \tag{1.21}$$

These ratios do not allow the estimation of the mutation rate μ. To estimate μ, we must fit the model to dynamic data using non-linear fitting procedure applied to the system (1.13).

1.4 Model with Competition

The previous model assumes that there are unlimited resources available to both populations, we assumed this to see the expected ratios in exponential phase. Exponential growth is not accurate over all time since the bacterial populations are limited because resources are not unlimited in the environment. Specifically, the case of sharing a limited resource through competition will be considered in the model that includes competition for the resource (space, food, etc...). We must also make the assumption of the Law of Mass Action, as there are interacting population densities. The model system has the form:

$$\begin{aligned} \frac{dn}{dt} &= k_n n\left(1 - \frac{pn + qN}{M}\right) - \mu n - d_n n, \\ \frac{dN}{dt} &= k_N N\left(1 - \frac{pn + qN}{M}\right) + \mu n - d_N N. \end{aligned} \tag{1.22}$$

Here, n and N are the populations of the non-mutated and mutated bacteria, respectively, μ_n stands for the mutation rate constant, d_n and d_N are the respective death rate constants, and k_n and k_N are the respective growth rate constants. M/p is the maximum population of n in the absence of N, and M/q is the maximum population of N in the absence of n. Since each population may use the resource differently, p and q represent the resource coefficient constant for the respective populations.

Let us find the steady states by setting the derivatives in the system (1.22) equal to zero.

$$\begin{aligned} 0 &= k_n n\left(1 - \frac{pn + qN}{M}\right) - \mu n - d_n n, \\ 0 &= k_N N\left(1 - \frac{pn + qN}{M}\right) + \mu n - d_N N. \end{aligned} \tag{1.23}$$

We have the trivial steady state where

$$(\bar{n}_1, \bar{N}_1) = (0, 0).$$

If we assume that $\bar{N} \neq 0$ and $\bar{n} = 0$, we arrive at the following equation:

$$k_N \bar{N}\left(1 - \frac{q\bar{N}}{M}\right) - d_N \bar{N} = 0.$$

Solving this for $\bar{N}$, we arrive at

$$(\bar{n}_2, \bar{N}_2) = \left(0, \frac{M}{qk_N(k_N - d_N)}\right).$$

If we assume that $\bar{N} \neq 0$ and $\bar{n} \neq 0$, we have:

$$\begin{aligned} 0 &= k_n\left(1 - \frac{pn + qN}{M}\right) - \mu - d_n, \\ 0 &= k_N\left(1 - \frac{pn + qN}{M}\right) + \mu - d_N. \end{aligned} \tag{1.24}$$

Rearranging the first equation of (1.24),

$$\left(1 - \frac{p\bar{n} + q\bar{N}}{M}\right) = \frac{\mu + d_n}{k_n}, \tag{1.25}$$

Plugging the first equation of (1.25) into the equation for mutated populations in system (1.23), we arrive at

$$0 = \frac{k_N}{k_n}(\mu + d_n)\bar{N} + \mu\bar{n} - d_N\bar{N}.$$

Thus,

$$\bar{N}[k_n d_N - k_N(\mu + d_n)] = (\mu k_n)\bar{n}.$$

Solving the above for $\bar{N}$,

$$\bar{N} = \frac{\mu k_n}{(k_n d_N - k_N(\mu + d_n))}\bar{n}. \tag{1.26}$$

Now we solve for $\bar{n}$, using the same procedure as above for the non-mutated population in system (1.23),

$$Mk_n - k_n(p\bar{n} + q\bar{N}) = M(\mu + d_n).$$

Plugging in (1.26), we get:

$$Mk_n - k_n\bar{n}\left(p + \frac{q\mu k_n}{(k_n d_N - k_N(\mu + d_n))}\right) = M(\mu + d_n).$$

Rearranging terms,

$$M(k_n - \mu - d_n) = \bar{n}\left(pk_n + \frac{q\mu k_n^2}{k_n d_N - k_N(\mu + d_n)}\right),$$

and solving for $\bar{n}$, we finally obtain:

$$\bar{n}_3 = M\left[\frac{(k_n - \mu - d_n)(k_n d_N - k_N(\mu + d_n))}{pk_n(k_n d_N - k_N(\mu + d_n)) + q\mu k_n^2}\right]. \tag{1.27}$$

Hence, substituting (1.27) into (1.26), we get:

$$\bar{N}_3 = M\left(\frac{\mu k_n}{(k_n d_N - k_N(\mu + d_n))}\right)\left(\frac{(k_n - \mu - d_n)(k_n d_N - k_N(\mu + d_n))}{pk_n(k_n d_N - k_N(\mu + d_n)) + q\mu k_n^2}\right). \tag{1.28}$$

Let us compute $\bar{N}/(\bar{n}+\bar{N})$. We have

$$\frac{\bar{N}}{\bar{n}+\bar{N}} = \frac{\mu k_n}{k_n d_N - k_N \mu - d_n k_N + \mu k_n}. \tag{1.29}$$

If $k_n = k_N$, then we get the following equality,

$$\frac{\bar{N}}{\bar{n}+\bar{N}} = \frac{\mu}{d_N - d_n}. \tag{1.30}$$

This ratio is the same as in the discrete case and is the same as the continuous case with no competition (when the growth rates were equal). Similarly, we compute $\bar{n}/(\bar{n}+\bar{N})$. We get

$$\frac{\bar{n}}{\bar{n}+\bar{N}} = \frac{k_n d_N - k_N \mu - k_N d_n}{k_n d_N - k_N \mu - k_N d_n + k_n \mu}.$$

If $k_n = k_N$, then we get the following equality:

$$\frac{\bar{n}}{\bar{n}+\bar{N}} = \frac{d_N - d_n - \mu}{d_N - d_n} = 1 - \frac{\mu}{d_N - d_n}. \tag{1.31}$$

Solving (1.30) for the mutation rate constant, we arrive at:

$$\mu = (d_N - d_n)\frac{\bar{N}}{\bar{n}+\bar{N}}. \tag{1.32}$$

We can use (1.32) to solve for the mutation rate constant of a bacterial cell by measuring the ratio of mutated to non-mutated and the difference in the death rate constants.

In Figure 1.5, we show the location of the steady states in the phase plane by plotting the nullclines (1.23).

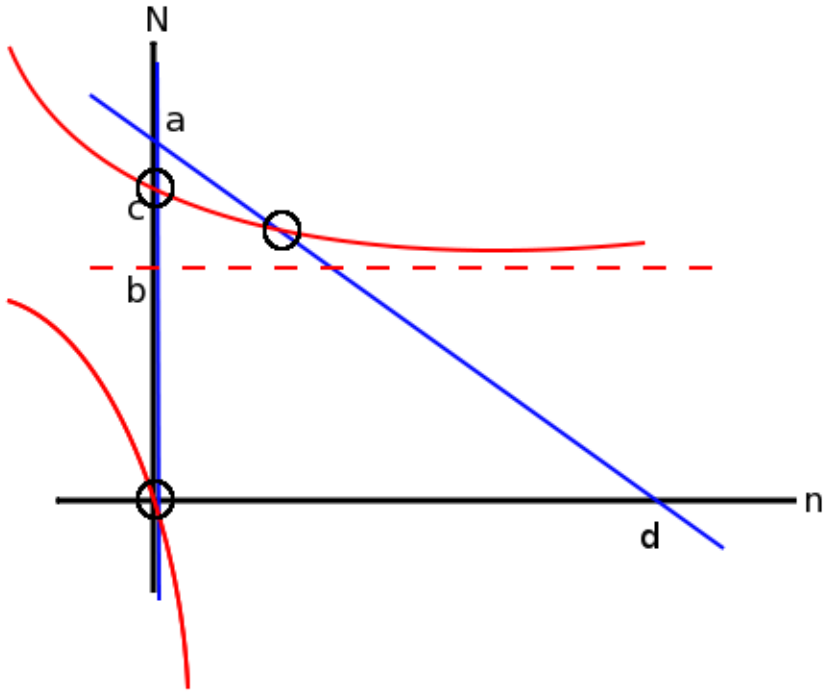

Figure 1.5: The nullclines of the system. Here $a = (M/q)(1 - (\mu + d_n)/k_n)$, $b = (M\mu)/(pk_N)$, $c = (M/(qk_N))\,(k_N - d_N)$, and $d = (M/p)(1 - (\mu + d_n)/k_n)$.

In order for the non-trivial steady state (1.28) and (1.27) to exist, we must have that the non-mutated population is non-zero ($n > 0$) and,

$$k_N < k_n \frac{d_N}{d_n + \mu},$$

or

$$\frac{k_n}{k_N} > \frac{d_n + \mu}{d_N}. \tag{1.33}$$

For the existence of a biologically meaningful steady state, we must also have (see 1.28):

$$k_n \geq d_n + \mu.$$

The above requirement limits how small k_n can be, and since the growth rate constant k_n has to be greater than the combined forces associated with the death rate constant d_n and mutation rate constant μ.

If $k_N = 0$,

$$\frac{\bar{n}}{\bar{n} + \bar{N}} = \frac{d_N}{d_N + \mu}.$$

This ratio approaches one, if μ is close to zero. Since $\frac{\bar{n}}{\bar{n}+\bar{N}} = 1 - \frac{\bar{N}}{\bar{n}+\bar{N}}$, in this case, the ratio, $\frac{\bar{N}}{\bar{n}+\bar{N}}$ is close to zero.

From (1.33), we consider the equality $k_N = k_n \frac{d_N}{d_n+\mu}$ because this is the highest the mutated growth rate constant can become while still allowing the above steady state to exist. Then the ratio $\frac{\bar{N}}{\bar{n}+\bar{N}}$ is,

$$\frac{\bar{N}}{\bar{n} + \bar{N}} = \frac{\mu k_n}{k_n d_N - \frac{k_n d_N \mu}{d_n+\mu} - \frac{k_n d_n d_N}{d_n+\mu} + \mu k_n},$$

$$\frac{\bar{N}}{\bar{n} + \bar{N}} = \frac{\mu k_n (d_n + \mu)}{k_n d_N (d_n + \mu) - k_n d_N \mu - k_n d_n d_N + \mu k_n (d_n + \mu)},$$

or, as $t \to \infty$,

$$\frac{\bar{N}}{\bar{n}+\bar{N}} = \frac{\mu k_n d_n + k_n \mu^2}{\mu k_n d_n + k_n \mu^2} = 1.$$

Since $\frac{\bar{N}}{\bar{n}+\bar{N}} = 1 - \frac{\bar{n}}{\bar{n}+\bar{N}}$, then the ratio, $\frac{\bar{n}}{\bar{n}+\bar{N}}$ is zero.

The summary of the results is as follows:

Criteria	$\frac{\bar{n}}{\bar{n}+\bar{N}}$	$\frac{\bar{N}}{\bar{n}+\bar{N}}$
$k_N = 0$	$\to 1$	$\to 0$
$k_n = k_N \frac{d_n+\mu}{d_N}$	$\to 0$	$\to 1$

These results tell us under what conditions the non-mutated and mutated populations can be driven to saturation or extinction. These conditions will become more important later when we address the analysis of models to minimize the total bacterial load in the lung.

Limiting Ratio in Exponential Phase

We now address the issue of the limiting ratio during exponential phase. This is very important because we will sample the bacterial population during all stages of growth. To find this ratio, we assume the populations are at a stage where the carrying capacity, M, is large in comparison to the bacterial population. Now let $\widehat{M} = \epsilon M$, or $\frac{\widehat{M}}{\epsilon} = M$ in the system (1.22). The model has the form:

$$\begin{aligned} \frac{dn}{dt} &= k_n n\left(1 - \epsilon\frac{pn+qN}{\widehat{M}}\right) - \mu n - d_n n, \\ \frac{dN}{dt} &= k_N N\left(1 - \epsilon\frac{pn+qN}{\widehat{M}}\right) + \mu n - d_N N. \end{aligned} \tag{1.34}$$

We find the zeroth order approximation of the solution of system (1.34) in the finite time interval by

setting $\epsilon = 0$; we arrive at:

$$\begin{aligned} \frac{dn}{dt} &= k_n n - \gamma n - d_n n, \\ \frac{dN}{dt} &= k_N N + \gamma n - d_N N. \end{aligned} \tag{1.35}$$

Now this system (1.35) has the form that is exactly the same as our non competition system (1.11). This means the ratio of mutated cells to the whole population approaches the same ratio constant in the exponential phase as in the steady states. The implications of this mean that we can observe the ratio of mutated cells to the whole population approach a constant by continually regrowing our culture in exponential phase. This is important that we do not have to wait until stationary phase is approached in a batch culture to measure the mutated cell ratio.

1.5 Model with Competition and Constant Inflow of Unevolved Bacteria

Lungs are not isolated from the environment. There is constant small inflow of bacteria into the lungs. Given a population in a lung infection, there might be some influx of very similar bacteria. This assumption is made because if total bacterial extinction is achieved, then the lung will not remain bacteria free for long. To take into account this inflow of bacteria, we will now assume that the non-mutated bacteria are being seeded with a constant rate. Let us consider the extension of the previous model:

$$\begin{aligned} \frac{dn}{dt} &= k_n n\left(1 - \frac{pn + qN}{M}\right) - \mu n - d_n n + \delta, \\ \frac{dN}{dt} &= k_N N\left(1 - \frac{pn + qN}{M}\right) + \mu n - d_N N. \end{aligned} \tag{1.36}$$

Here, n and N are the populations of the non-mutated and mutated bacteria, respectively, μ_n stands for the mutation rate constant, d_n and d_N are the respective death rate constants, and k_n and k_N are the

respective growth rate constants. M/p is the maximum population of n in the absence of N, and M/q is the maximum population of N in the absence of n. Here we add the term δ, which represents the constant inflow rate of non-mutated bacteria.

We solve for the steady states by setting the derivatives in equations (1.36) equal to zero:

$$\begin{aligned} 0 &= (k_n - d_n - \mu)n - \left(\frac{k_n a}{M}\right)n^2 - \left(\frac{k_n b}{M}\right)nN + \delta, \\ 0 &= (k_N - d_N)N - \left(\frac{k_N b}{M}\right)N^2 - \left(\frac{k_N a}{M}\right)nN + \mu n. \end{aligned} \tag{1.37}$$

We can solve these equations and get the following expressions for nullclines (hyperbolas on the phase plane):

$$\begin{aligned} \bar{n} &= \frac{(k_N - d_N)\bar{N} - (k_N q/M)\bar{N}^2}{(k_N p/M)\bar{N} - \mu}, \\ \bar{N} &= \frac{(k_n - \mu - d_n)\bar{n} - (k_n p/M)\bar{n}^2 + \delta}{(k_N q/M)\bar{n}}. \end{aligned} \tag{1.38}$$

From Figure 1.6, the plot of the nullclines, we can see that there is only one steady state.

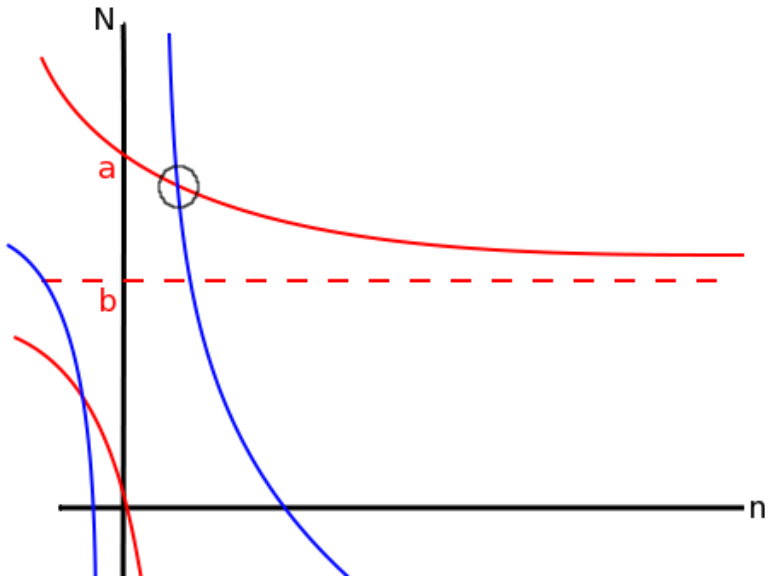

Figure 1.6: The nullclines of the system (1.37). Here $a = (M(k_N - d_N)/(qk_N))$ and $b = (\mu M)/(k_N p)$.

The nontrivial steady state is given by the intersection of the nullclines (1.38).

It is important to note that the ratios are different for different models. The ratios all include the mutation rate constant along with various other constants. In order to use these ratios to estimate the mutation rate, the other constants must be determined. This can be addressed experimentally, e.g., the death rate constant can be figured out with live/dead cell staining or microscopy or similar methods. For our purposes, we will concern ourselves with the ratio expressed in (1.15) with the death rate constants equal to zero for simplicity.

1.6 Conclusions

It is possible to determine the ratio of mutated or non-mutated to the total population based on the models describing the mutated and non-mutated populations previously described. In what follows, we will use this ratio to estimate the mutation rate constant for bacterial populations. In practice,

ratios involving more complicated terms demand more constants that must be known. For example, in the last model, the bacterial influx constant must be known, the mutated and non-mutated birth and death rate constants must be known and also the resource coefficients. This can make mutation rate estimation much more complicated.

Estimating Mutation Rate and Experimental Methods

1.7 Background

Determining the mutation rate of bacterial populations is important in many areas including epidemiology, public health, microbiology, etc [32]. In 1943 Luria and Delbrück designed a test to determine whether or not mutations are the result of a selective force, or if mutations occur naturally with or without selection. This same test is still used today for estimating the mutation rate constants of bacterial populations. H.E. Kubitschek described another way to estimate mutation rate constants of a bacterial strain in a continuous culture such as a chemostat [21], but it relies on expensive and sensitive continuous culture experiments that must be run for long times. The following section reviews the two currently used experimental procedures and the basic analysis that accompanies each of them. Then a new experimental procedure is proposed and discussed.

1.8 Luria-Delbrück Batch Culture Experiment

Overview

In 1943, Luria and Delbrück developed the fluctuation test to determine how mutations occur [24]. This fluctuation test results in mutation frequencies which are used to estimate the mutation rate constant. Mutation frequencies and rates in this experiment are in relation to a presented challenge, which

can be antibiotics, different nutrient sources, a virus or any other measurable challenge.

The Luria and Delbrück experiment consists of inoculating many parallel cultures with the same amount of non-mutated bacteria (which can be experimentally verified by plating on selective media). At the end of the growth phase, the number of mutated cells in each culture is determined by plating samples on selective media. This experiment assumes that mutated cells will be selected on solid medium (plates), and that cultures will be saved to determine the total number of cells by plating on non-selective media.

In order to determine the duration of the growth phase and the number of parallel cultures, some preliminary experiments must be conducted. From these experiments an estimate of the mutational events per culture can be determined and used to design the fluctuation test. Ideally, the culture will be grown up in an appropriate volume so that a mean of 2-5 antibiotic resistant colonies grow on each selective plate.

The analysis in [22] uses the fact that the fraction of mutated cells to the total population approaches a constant. The ratio established is based on the assumption of a discrete population.

The Luria-Delbrück experiment is based on plate counts from selective and non-selective media. The mean is taken and the average of mutational events per culture is computed. A confidence interval is estimated using a standard deviation of the data. The mutation rate constant is estimated by dividing the number of average mutational events per culture by the average number of cell generations in a culture. This analysis lacks mathematical rigor due to the approximations and assumptions made during the calculation.

Formulas

A widely used method for analysis of the data for the Luria-Delbrück experiment is the Lea and Coulson model. The main method for Lea and Coulson model is related to the formation of mutants

via a Yule Process (preferential attachment) [22]. The non-mutated and mutated cells grow with rate constants k_n and k_N respectively. If an non-mutated cell generates an mutated cell at time s, then the distribution of mutated cells resulting from that event is given by $Y(t-s)$. Let $m(t)$ stand for the number of mutational events per culture, which occur at times τ_i, and let μ stand for the mutation rate constant. This Yule process lends itself to the following generating function of total mutants:

$$N(t) = \sum_{i=1}^{m(t)} Y_i(t - \tau_i).$$

Let $G(z;t) = E(z^{N(t)})$ be the probability generating function (p.g.f.) of $N(t)$, the Yule growth process. Since the p.g.f of a Yule process is known, the p.g.f. of $N(t)$ can be shown to be [47]

$$G(z;t) = \exp\left(-m(t) + \sum_{i=1}^{\infty} q_i z^i\right).$$

Here

$$q_i = \mu e^{-k_N t} \int_0^t [1 - e^{-k_N(t-s)}]^{i-1} e^{(k_n+k_N)s} ds,$$

or, using the binomial expansion,

$$q_i = \mu \sum_{j=0}^{i-1} (-1)^j \begin{pmatrix} i-1 \\ j \end{pmatrix} \frac{e^{k_n t} - e^{-(j+1)k_N t}}{k_n + (j+1)k_N}.$$

Now we can use the random variable technique with the cumulant generating functions to arrive at equations for the mean. The random variable technique involves using a Taylor series approximation to solve the system for the resulting generating functions.

Consider the following cumulant generating function of $N(t)$,

$$K(s,t) = \log(E(e^{sN(t)})).$$

The random variable technique [47] gives

$$\frac{\partial K}{\partial t} = k_N(e^s - 1) + \mu e^{k_n t}(e^s - 1).$$

Now we let $K(s,t) = \Sigma_i \kappa_i(t) s^i / i!$, and solve for κ_1 with the condition $\kappa_1(0) = 0$,

$$\kappa_1 = E[N(t)] = \begin{cases} \frac{\mu(\exp(k_n t) - \exp(k_N t))}{k_n - k_N}, & (k_n \neq k_N), \\ \mu t \exp(k_n t), & (k_n = k_N). \end{cases} \tag{1.39}$$

The mutation rate constant, μ, can be determined from (1.39) by computing the limiting ratio and the growth rate constants of the two populations.

$$\mu = \begin{cases} E[N(t)](k_n - k_N)/(\exp(k_n t) - \exp(k_N t)), & (k_n \neq k_N), \\ E[N(t)](t \exp(k_n t)), & (k_n = k_N). \end{cases} \tag{1.40}$$

This analysis is accompanied by the following experimental procedure.

Procedure

A 96-well plate is preferred for this experiment due to the low cost, compact nature and small culture size of each well. Growth media for the bacterial culture should be properly selected to encourage growth.

Materials: For one bacterial strain, the minimum amount of liquid media is around 100 ml. The minimum amount of non-selective and selective plates needed are around 2 each per 96-well culture. A 96-well plate is used, or enough equivalent culture tubes.

Methods: The experiment consists of the following steps:

1. A small bacterial culture is to be grown in a liquid culture in the same growth medium at the same temperature as the rest of the experiment.
2. Before saturation, it should be determined how dense the cell culture is by an optical density reading.
3. A sample of this culture is to be plated out on selective media to assure the presence of no mutants.
4. This sample should be plated out on non-selective media for total cell population counts.
5. Another sample should be sufficiently diluted to a low density, to provide for a consistent low number of cells in each of the 96-well cultures.
6. Of the 96-well cultures, a few wells are kept empty as growth controls, for the optical density.
7. The rest of the 96-well cultures are to be inoculated with the same amount of the previously diluted culture from step 5.
8. After the growth phase and before saturation (as determined by the preliminary experiments), most cultures should be plated out completely on selective media.
9. Some inoculated cultures are to be saved for plating on non-selective media to determine the total population.
10. Analysis is to be completed on the resulting number of mutants to determine the mutation rate via the Lea-Coulson Analysis.

Lea-Coulson Analysis

The following Lea-Coulson Analysis is based on the plate counts from the above protocol. Each culture grown in step 7, in the above procedure, is plated out on q-replicates of both non-selective and selective media. These q-replicates are counted and averaged to obtain an estimate for the number of cells and mutated cells for each culture. The average number of mutants over all cultures is used in Equation (1.39),

$$\kappa_1 = E[N(t)] = \begin{cases} \frac{\mu(\exp(k_n t) - \exp(k_N t))}{k_n - k_N}, & (k_n \neq k_N), \\ \mu t \exp(k_n t), & (k_n = k_N). \end{cases} \tag{1.41}$$

From this, μ can be estimated. This analysis requires knowledge of the growth rate constants of both populations.

Lea-Coulson [35] introduced different methods of analysis that did not require knowledge of the growth rates. The most commonly used one is the method of the median. Given data that represents the estimated number of mutated cells for each culture, define r as the median. Letting m be the number of mutational events per culture, Lea and Coulson noticed that for $4 \leq m \leq 15$, the equation $r/m - \ln m \approx 1.24$, was observed. Then value of m can be used to find the mutation rate by dividing m by the average number of cells at risk in each culture, determined from step 9.

P_0 Method Analysis

The method of analysis used by Luria and Delbrück was the p_0 method [24]. They assumed the number of mutated cells arisen in each culture would follow a Poisson distribution. The simplest way to estimate the number of mutational events per culture, m, was to count the number of cultures with zero mutants. The fraction of cultures with zero mutated cells should be equal to

$$P(x = 0) = e^{-m}.$$

From this, the average number of mutational events per cultures (m), can be estimated. If we know the average number of generations per culture (assuming they are all the same), the mutation rate constant can be estimated.

1.9 Continuous Culture Experiment

Overview

H.E. Kubitschek described a way to estimate mutation rate of a bacterial strain in a continuous culture such as a chemostat [21]. This method is mathematically tractable, but the assumptions are not always met. This method also requires access to expensive equipment and constant maintenance requirements. It should also be noted that evolution can not be stopped in actively dividing populations. There are also some strains that cannot be grown in a continuous culture due to their nature (e.g. mucoidal strains).

This experiment requires one to inoculate a chemostat and continually measure the accumulation of mutants in the system. The rate of mutant accumulation in a chemostat is assumed to depend only on the mutation rate constant and the difference in growth rate constants between the mutated and non-mutated cells. During the chemostat experiment, samples are extracted and plated on non-selective and selective media. From these measurements, the mutation rate can be calculated. It is worth noting that the continuous culture experiment is an improvement in procedure from the Luria-Delbrück batch culture due to the fact that total cell counts are easier to do since the overall cell population stays approximately constant as opposed to exponentially growing population in a batch culture. Also, the age of the culture can be estimated by measuring the outflow of the culture. The analysis also assumes

screening for one class of mutated cells, but there may be multiple classes of mutated cells in the chemostat.

Formulas

The rate at which the mutated cells accumulate in the chemostat is not only defined by the mutation rate but also by the difference in the growth rate constants of the mutated cells to non-mutated cells. We define n to be the constant non-mutated cell population and $N(t)$ to be the mutated cell population. We assume n to be constant due to the large population size compared to N. The corresponding growth rates are k_n and k_N; and the mutation rate is defined by μ.

The rate of change of mutant concentration can then be described as

$$\frac{dN}{dt} = n\mu - (k_n - k_N)N. \tag{1.42}$$

Here, the first term accounts for a change in mutated cell concentration due to mutation rate. The second term is due to the difference in growth rates. If the non-mutated population grows faster than the mutated ($k_n > k_N$), the mutated population will decrease in the chemostat due to displacement.

The solution to (1.42) is

$$N(t) = \frac{(k_n - k_N)N_0 + n\mu}{(k_n - k_N)} e^{(k_N - k_n)t} + \frac{n\mu}{k_n - k_N}. \tag{1.43}$$

Assuming $N_0 = 0$,

$$N(t) = \left(\frac{n\mu}{(k_n - k_N)}\right)\left(e^{(k_N - k_n)t} + 1\right).$$

the limiting ratio of $N(t)$ to the whole population, $N(t) + n$, is given by

$$\lim_{t\to\infty} \frac{N(t)}{N(t)+n} \approx \lim_{t\to\infty} \frac{N(t)}{n} = \frac{\mu}{k_n - k_N}. \tag{1.44}$$

Here we assume that $N(t) + n \approx n$, and $k_n - \mu > k_N$. It is important to note that here $k_n \neq k_N$. If the growth rates are equal, then the ratio can be shown to be

$$\frac{N(t)}{N(t)+n} = \frac{N_0}{n} + \mu t, \tag{1.45}$$

where $N_0 = N(t = 0)$. We see from (1.45) that the mutant population increases linearly with the slope being equal to the mutation rate.

Solving for the mutation rate constant, assuming $k_n > k_N$:

$$\mu = \left(\lim_{t\to\infty} \tfrac{N(t)}{n}\right)(k_n - k_N), \quad (k_n \neq k_N). \tag{1.46}$$

If $k_n = k_N$, then we fit the data using linear regression and the mutation rate constant is equal to the slope of the line. If $k_n < k_N$, then the data must be fit using non-linear regression techniques to equation (1.43).

Procedure

Materials: For one bacterial strain, the liquid media is constantly created to be fed into the chemostat. The experiment is stopped when enough data points have been collected. The minimum amount of non-selective and selective plates needed are determined by how many time points are collected.

Methods: The experiment consists of the following steps:

The experimental procedure for estimating the mutation rate constant in the continuous environment grows a culture over time and records the accumulation of mutated cells. It is important to note that the experiment itself will reveal the relationship between the growth rates. If the mutated cells increase more than linearly (concave up), then there is favorable selection ($k_N > k_n$). If the mutated cell accumulation is less than linear (concave down), then there is unfavorable selection for the mutated cells ($k_N < k_n$). The protocol for this experiment requires the use of chemostats, preferably run in multiple vessels (duplicate, triplicate or more).

1. A chemostat should be prepared and the strain in question will be grown in liquid media in similar temperature, Ph, and composition to the rest of the experiment.
2. The chemostat will be run without a strain for at least 24 hours to check for contamination.
3. A dilution rate is to be chosen according to the average duplication time of the strain and the duration of the experiment.
4. The chemostat should then be inoculated with the strain and a sample taken. This sample should be first taken at time $t = t_0$. Then this sample is to be diluted appropriately and plated on both selective and non-selective media.
5. The population in the chemostat will be sampled frequently enough to collect enough data points to fit mutant accumulation counts. Counts should be fit to a line if $k_n = k_N$, fit to a ratio if $k_n > k_N$ or non-linearly fit to Eq. (1.43) if $k_n < k_N$. (See above analysis).

Analysis

The analysis of the continuous culture will lead to estimates of the mutation rate constant based on the linear fit to the data collected in the above procedure. For the case where $k_n = k_N$, the estimate

for mutation rate constant involves fitting the concentration of mutants (cells per ml) to a straight line. The slope of this line will be the mutation rate constant. The slope will be in dimensional units [(cells per ml)/time], and the mutation rate constant is in units [mutations per generation], so we multiply by a factor of time per generation and divide it by the total cell population per ml in the chemostat. The result is the number of mutants that appear per generation.

The confidence interval for the mutation rate constant is simply the confidence of predicting the slope of a linear equation at a certain confidence level, using standard statistical line fitting techniques.

1.10 Batch Culture Continuous Experiment

Overview

An ideal experiment would combine the benefits of the previous two experiments and have none of their detriments. Like Luria-Delbrück's experiment, the new approach does not rely on complicated and expensive equipment, and similar to the Kubitschek experiment, the analysis would not rely on many approximations and assumptions. We propose a new experiment and mutation rate constant estimation procedure that satisfies these requirements. This method is also versatile and accommodating to different types of systems, e.g., it can be used in cases of back mutations, seeding of mutated or non-mutated cells from the environment, etc.

Formulas

The mutation rate constant will be estimated from (1.15). We can rewrite (1.15) as follows:

$$\lim_{t\longrightarrow\infty} \frac{N}{n+N} = \frac{\mu}{(k_n - d_n) - (k_N - d_N)}. \tag{1.47}$$

If we assume that $d_n = d_N = 0$, then we arrive at

$$\lim_{t\longrightarrow\infty} \frac{N}{n+N} = \frac{\mu}{k_n - k_N}. \tag{1.48}$$

The experiment estimates the ratio in (1.48) by plating samples of the batch culture on selective and non-selective media. The growth rates must also be determined in order to arrive at an estimation for the mutation rate, μ:

$$\mu = \left(\lim_{t\longrightarrow\infty} \frac{N}{n+N}\right)(k_n - k_N) \tag{1.49}$$

Mutation Rate Estimators

From this point forward, $T(t)$ will stand for the total population (non-mutated and mutated). In this experiment, we observe data points representing the numerator ($N(t)$), and denominator ($T(t)$). We will observe the numerator on selective plates and the denominator on non-selective plates.

An estimator for $N(t)/T(t)$ at any time t will be based on observing the numerator, $N(t)$ and the denominator, $T(t)$. In order to reduce the effects of different growth rate constants, the bacterial culture will be re-inoculated in batch cultures and the ratio of mutated, $N(t)$, to the whole population, $T(t)$, will be measured by plating out diluted portions of the batch culture on selective and non-selective plates.

When performing this experiment, samples from the batch culture will be taken out and plated multiple times on both selective and non-selective media. Let p be the number of samples taken over time from the batch culture, let q be the number of replicates of every sample. Also let $N_{i,j}$ stand for the j-th replicate of the i-th sample of the population plated on selective media, and let $T_{i,j}$ stand for the j-th replicate of the i-th sample of the population plated on non-selective media. In order to estimate the

expected value of the i-th sample, we calculate the average over the j replicates.

$$\overline{N}_i = \frac{1}{q}\sum_{j=1}^{q} N_{i,j},$$

$$\overline{T}_i = \frac{1}{q}\sum_{j=1}^{q} T_{i,j}.$$

Then the average is taken again to get an estimate for $\widehat{N}$ and $\widehat{T}$,

$$\widehat{N} = \frac{1}{p}\sum_{i=1}^{p} \overline{N}_i,$$

$$\widehat{T} = \frac{1}{p}\sum_{i=1}^{p} \overline{T}_i.$$

From these estimates we use ratio estimators to arrive at our estimate for the mutated to the population ratio and corresponding confidence intervals. The simple ratio estimator is then

$$\widehat{r} = \frac{\widehat{N}}{\widehat{T}}.$$

Another ratio estimator that reduces the bias, is Quenouille's ratio estimator, [44].

$$\widehat{r_Q} = 2\widehat{r} - \frac{1}{2}(\widehat{r_1} + \widehat{r_2}).$$

Here $\widehat{r_1}$ and $\widehat{r_2}$ are the simple ratios obtained from the first and second half of the data respectively.

There are a number of different ratio estimators [44], but they require the knowledge of how big the population is from which the samples are taken. It has been shown that Quenouille's estimator is less biased than the simple estimator [44]. It is also more efficient when the numerator has a normal distribution [44].

Procedure

The following experiment was performed blind on six *E. coli* strains, and then on one unknown *P. aeruginosa* strain. Both used rifampicin for selection of mutated cells. The *P. aeruginosa* strain results were compared to results from a Luria-Delbrück experiment. The results follow the experimental procedure description.

1. The strain in question was grown in a batch culture (shaken) with the same temperature and pH in the following steps (which were ideal for the strain growth).
2. During step 1, optical densities were taken to identify how long the strain stays in exponential growth mode, and when maximal growth rate was achieved.
3. A new batch culture was re-inoculated with a small, predetermined amount of media containing the strain from the previous batch culture.
4. Optical densities and plate counts were taken. Plating was replicated on 3 of each non-selective and selective plates for each time point.
5. When the strain was approximately growing the fastest, a new batch culture was inoculated as before, and data was collected (platings and optical density).
6. The mutated cell population to whole population ratio was computed and experiment was stopped when the ratio appeared relatively constant.
7. From the last time point, a mutated and non-mutated colony was sampled and grown up in non-selective liquid media separately and the optical densities were taken every hour for 24 hours.

8. The previous optical densities were fit to a logistic growth equation and the difference in growth rate constants were calculated.

9. From the measured mutated cell ratio and the difference in growth rate constants, the mutation rate constant can be determined.

This experiment reduces the number of cultures needed to estimate the mutation rate constant as compared to the Luria-Delbrück procedure. The recommended number of cultures for the Luria-Delbrück experiment is between 20-80 cultures [35]. This new protocol recommends between 15-25 cultures for both experiments (limiting ratio and growth rate determination). Another benefit to this procedure is that there is no complicated, expensive, or hard to maintain equipment involved.

1.11 Results

Media and Strains

To verify this new proposed procedure, a blind experiment with six standard *Escherichia coli* strains were used (labeled A, B, C, D, E, F), of which all mutation rate constants were previously known from Luria-Delbrück analyses. The liquid media used was Davis Broth and the unselective solid media was tryptone agar [7]. Selective media was Rifampicin added to tryptone agar at 300 mg/L. After the mutation rate constants were determined, a *Pseudomonas aeruginosa* strain was used in the new batch culture method to determine the mutation rate constant as well. The liquid media used was Mueller-Hinton Broth and the solid media was Mueller-Hinton agar. The selective media for *Pseudomonas aeruginosa* was Rifampicin added at 300 mg/L to the solid media.

E. coli results

The plate counts for all six *E. coli* strains are in table 1.1 over the course of 6 time points spaced out in 121.5 hours. Again, at each time point ($t > t_0$), the strains were reinoculated into a new batch culture and sampled on selective and non-selective media.

T0								
03:30:00 PM	Culture	OD(550nm)	Non-Selective1	Non-Selective2	Non-Selective3	Selective1	Selective2	Selective3
0 hrs	A	-0.011	872	812	1548	0	0	0
Non-sel Dilution	B	0.014	1536	704	600	0	0	0
2.50E-004	C	0.001	1040	752	1656	0	0	0
	D	-0.003	800	744	1168	0	0	0
Selective Dilution	E	-0.008	760	312	448	0	0	0
1.00E-001	F	0.013	1224	1264	988	0	0	0
T1								
03:30:00 PM	Culture	OD(550nm)	Non-Selective1	Non-Selective2	Non-Selective3	Selective1	Selective2	Selective3
24 hrs	A	0.783	3040	3352	1824	26	23	30
Non-sel Dilution	B	0.713	1472	1416	0	20	31	20
2.50E-006	C	0.78	2448	1792	1944	0	0	1
	D	0.833	1536	2456	1176	11	7	11
Selective Dilution	E	0.708	1368	2224	1744	57	75	46
2.50E-002	F	0.786	3624	2792	1536	15	16	18
T2								
12:00:00 PM	Culture	OD(550nm)	Non-Selective1	Non-Selective2	Non-Selective3	Selective1	Selective2	Selective3
44.5 hrs	A	0.795	2312	1632	1848	36	33	45
Non-sel Dilution	B	0.739	1160	1848	1760	27	44	23
2.50E-006	C	0.805	2776	3432	3184	1	2	2
	D	0.843	1392	2560	2536	3	8	19
Selective Dilution	E	0.763	1312	1600	1688	42	37	35
2.50E-002	F	0.811	2000	2192	1448	41	31	58
T3								
03:00:00 PM	Culture	OD(550nm)	Non-Selective1	Non-Selective2	Non-Selective3	Selective1	Selective2	Selective3
71.5 hrs	A	0.8	377	346	240	28	38	39
Non-sel Dilution	B	0.739	252	288	312	25	17	35
4.00E-007	C	0.838	326	278	328	0	1	0
	D	0.839	512	412	280	5	10	7
Selective Dilution	E	0.759	502	282	356	30	41	43
2.50E-002	F	0.813	390	540	478	14	22	26

Table 1.1: Raw Data (Plate Counts and Dilution Rates). It is worth noting that most of the plate counts were extraordinarily high. Counting was done by griding plate and counting a smaller part.

The ratio, $\widehat{R}$ calculated for each time point is a total average across the three plates for each of the six different strains. The separate data is plotted in the following figures. The standard deviation for the ratio $N/(n+N)$ = Selective/Non-selective, is given by the (1.50). Again, $\widehat{n}$ stands for the average

non-mutated cells, and $\widehat{N}$ stands for the average mutated cells and also $\widehat{R}$ standing for the average ratio, we have the following formula:

$$\sigma_R = \widehat{R} \cdot \sqrt{\left(\frac{\sigma_n}{\widehat{n}}\right)^2 + \left(\frac{\sigma_N}{\widehat{N}}\right)^2} \,. \tag{1.50}$$

The culture ratios through the six time points are shown in Table (1.2).

	T0	T1	T2	T3	T4	T5
A	0.00E+000	9.62E-007	1.97E-006	1.74E-006	1.22E-007	2.99E-008
B	0.00E+000	1.64E-006	1.97E-006	1.45E-006	9.71E-008	2.73E-008
C	0.00E+000	1.62E-008	5.32E-008	1.72E-008	1.81E-007	1.12E-007
D	0.00E+000	5.61E-007	4.62E-007	2.92E-007	0.00E+000	2.16E-008
E	0.00E+000	3.34E-006	2.48E-006	1.60E-006	0.00E+000	0.00E+000
F	0.00E+000	6.16E-007	2.30E-006	7.05E-007	9.96E-007	2.53E-008

Table 1.2: Mutator ratios of all six cultures across the six time points.

Visualizing this in a graph, with the averages plotted, we arrive at Figure 1.7. Figure 1.8 is a log-y plot of the same data.

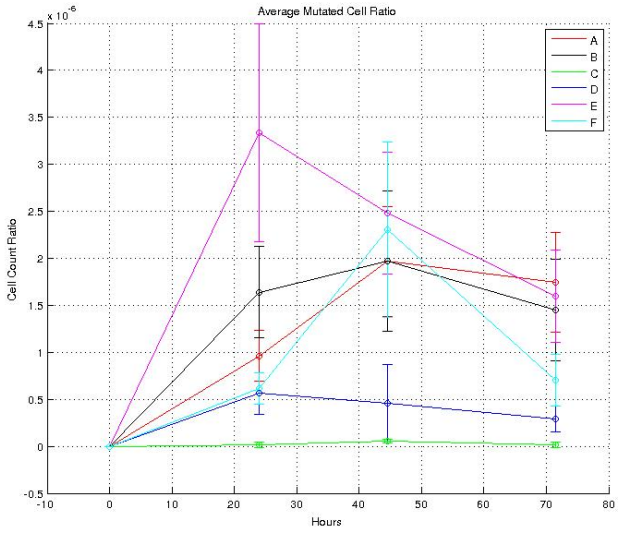

Figure 1.7: Ratio of mutated cells to total population over time.

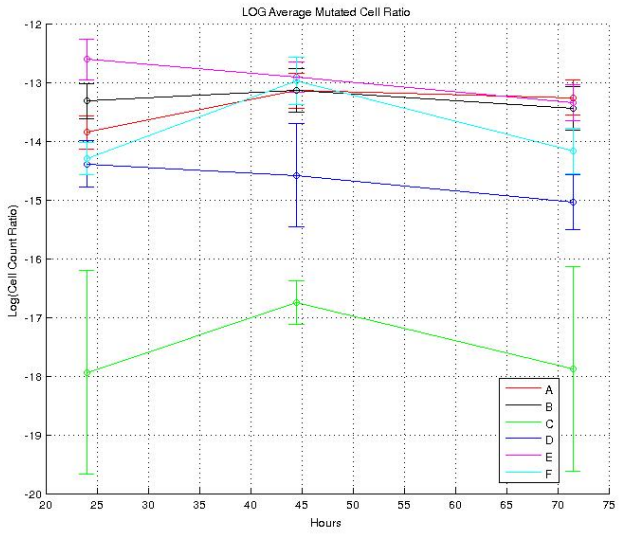

Figure 1.8: Log-Ratio of mutated cells to total population over time.

The ratio, $\widehat{R}$ is computed using only data values collected at the last time point, as the limit of the ratio is exponentially approached over time (see (1.14)). Next, colonies from the last plating are saved and regrown in liquid culture for growth rate determination.

Growth Rate Experiments

In order to figure out the growth rates constants of the mutated cells and non-mutated cells, a selective and non-selective plate was saved from the end of the previous mutated cell ratio experiment. A sample of the non-selective plates and selective plates were grown up from each and optical density samples were taken every hour until stationary phase was achieved.

The data was fit using non-linear least squares (Matlab Statistics Toolbox) to the following logistic equation:

$$P(t) = \frac{K \cdot P_0 e^{rt}}{K + P_0(e^{rt} - 1)}. \tag{1.51}$$

Bootstrapping was done to determine a standard error estimate for the growth rates. Bootstrapping resamples the residuals of the data with replacement, scaled by the leverages. For each sample, the growth rate constants, carrying capacity, and initial population were estimated, giving 5000 estimates for each parameter. A 95 percent confidence interval was then computed. This bootstrapping procedure was done on all three parameters for each population, the initial population N_0, n_0, the corresponding growth rate constants, k_N, k_N and the carrying capacities, M_N, M_n. The results for the growth rate constants are shown in Table 1.3.

Plotting the 5% to 95% ordered growth rate constants from the bootstrapping in Equation (1.51) around the best fit line with the data produces the graphs in Figure 1.9.

Strain Type:	Growth Rate Constant:	SD:
A-non-mutated	1.225	0.171
A-mutated	1.053	0.107
B-non-mutated	0.993	0.054
B-mutated	0.940	0.076
C-non-mutated	0.330	0.130
C-mutated	0.867	0.510
D-non-mutated	1.205	0.122
D-mutated	0.919	0.226
E-non-mutated	0.781	0.055
E-mutated	0.510	0.021
F-non-mutated	0.845	0.068
F-mutated	0.841	0.066

Table 1.3: Growth Rate constants for the six *E. coli* strains.

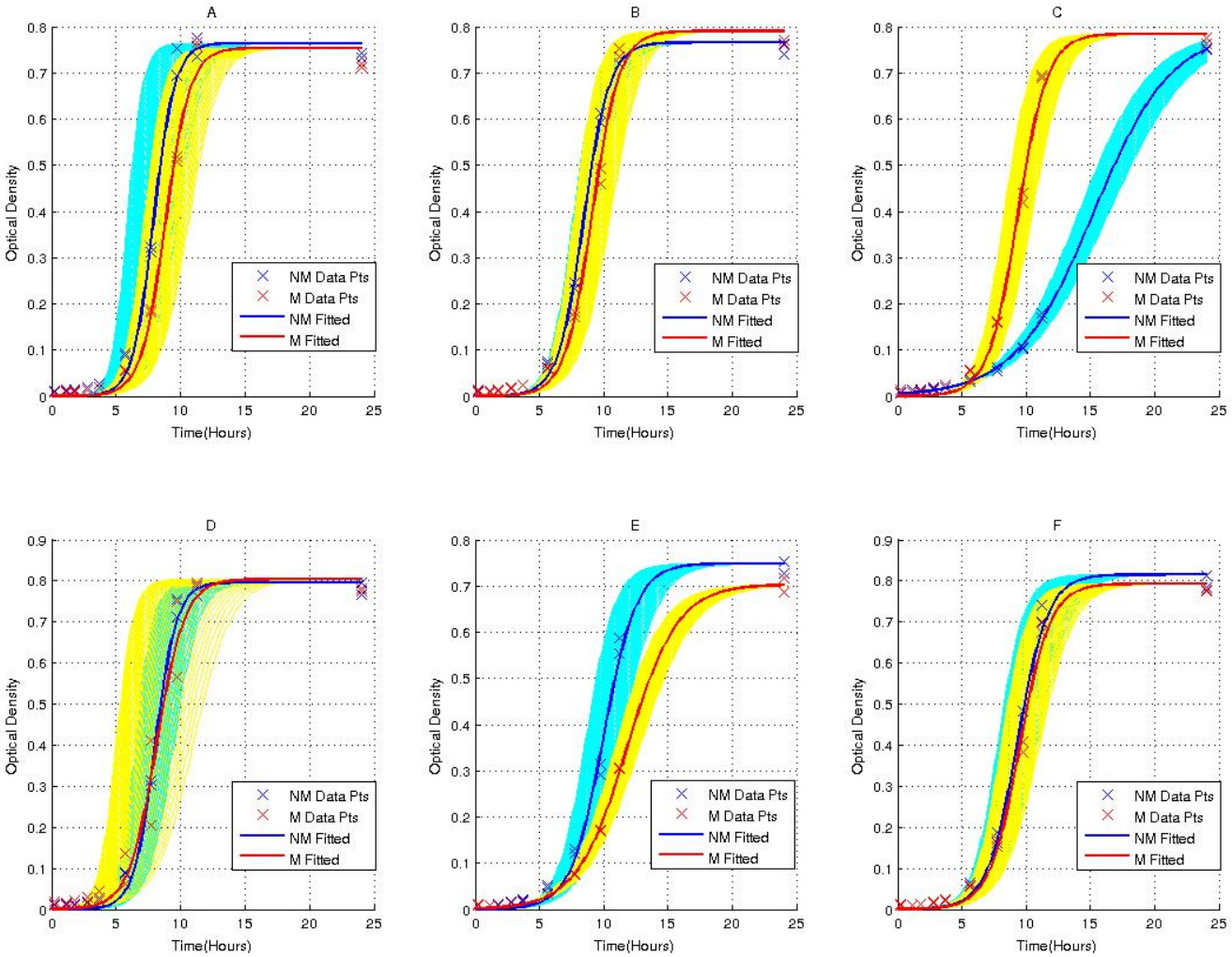

Figure 1.9: Both best fit growth curves for the non-mutated and mutated cells, with the 5% to 95%

We can formally test if the growth rates are equal by storing all the 5000 bootstraps and comparing the growth rates for each bootstrap. We count how many times the number of bootstrap growth rate constant estimates for the mutated strain is greater than the non-mutated strain. This number divided by the number of bootstraps gives an estimate of the p-value of the hypothesis that the mutated strain grows slower than the non-mutated strain. The p-values are listed in Table 1.4.

Strain	p-value
A	0.0826
B	0.1399
C	0.9998
D	0.0600
E	<0.0002
F	0.2400

Table 1.4: P-values for the hypothesis that the non-mutated strain has a higher growth rate constant (Estimated from 5000 bootstrap iterations). Strain C's value of 0.9998 indicates that the mutated strain has a significantly higher growth rate constant.

Using the growth rate constant and limiting ratio data, we can calculate the mutation rate constants.

Mutation Rate

The mutation rate constant can be found using the following formula:

$$\mu = \left(\lim_{t \longrightarrow \infty} \frac{N}{n+N}\right)(k_n - k_N) \tag{1.52}$$

The units of the mutation rate constant are in number of mutated cells to challenge per time measurement. Since we measured the growth rate constants in hours and used rifampicin, the mutation rate constant is number of rifampicin resistant cells per hour.

Solving Eq. 1.52 for μ, the mutation rate constant, we arrive at Table 1.5, which contains the mutation rate constants and standard deviations.

Strain:	Mutation Rate Constant:	SD:
A	2.997E-007	4.253E-007
B	7.714E-008	1.321E-007
C	N/A	N/A
D	8.350E-008	8.387E-008
E	4.338E-007	1.503E-007
F	3.108E-009	5.659E-008

Table 1.5: Table of calculated mutation rate constants using (1.52).

Strain C does not have a mutation rate constant due to the fact that the mutated cells grow faster than the non-mutated cells, (1.52) assumes the opposite. This might be due to the fact that strain C was an environmental strain, not yet acclimated to laboratory conditions.

To address when $k_N > k_n$, let us revisit the original system in (1.13),

$$\begin{aligned} n &= C_1 e^{at}, \\ N &= C_2 e^{gt} + \frac{bC_1}{a-g} e^{at}. \end{aligned} \tag{1.53}$$

We now assume that $a < g$ and proceed to look at the ratio $n/(n+N)$.

$$\frac{n}{n+N} = \frac{C_1 e^{at}}{C_1 e^{at} + \frac{bC_1}{a-g} e^{at} + C_2 e^{gt}}.$$

This simplifies to

$$\frac{n}{n+N} = \frac{C_1 e^{(a-g)t}}{C_1 e^{(a-g)t} + \frac{bC_1}{a-g} e^{(a-g)t} + C_2}.$$

If we assume that $a < g$, then we can write

$$\lim_{t\longrightarrow\infty} \frac{n}{n+N} \mid_{a<g} = 0. \tag{1.54}$$

And similarly,

$$\lim_{t\longrightarrow\infty} \frac{N}{n+N} \mid_{a<g} = 1. \tag{1.55}$$

These ratios in (1.54) and (1.55) do not allow the prediction of the mutation rate μ. To fit μ, we must fit the dynamics to a non-linear fitting procedure using the original system (1.13).

P. aeruginosa results

Similarly the experiment was performed on a *P. aeruginosa* strain, with the results shown in Table 1.6.

The first time point dilutions were inaccurately done resulting in a uncountable amount of colonies on the unselective plates and zero colonies on the selective plates. This leaves us with only 3 time points (time points 2 through 4). The ratio r calculated for each time point is a total average across the two plates and across the three strains. The separate data is plotted in the following figures. The standard deviation for the ratio $N/(n+N)$ = Selective/Non-selective, is given by (1.50).

Visualizing this in a graph, with the average plotted, we arrive at Figure 1.10.

Start:					
06/29/11 08:30 PM		Culture 1	Culture 2	Culture 3	Dilution Rate
Time Point 1	A550	1.913	1.970	1.994	
06/30/11 07:30 PM	Unselective1	X	X	X	500000
T_between = 23hrs	Unselective2	X	X	X	500000
T = 23	Selective1	0	0	0	5000
r = 0	Selective2	0	0	0	5000
Time Point 2	A550	2.016	1.959	2.095	
07/02/11 08:00 AM	Unselective1	1700	1412	980	6250000
T_between = 36.5hrs	Unselective2	1612	908	1616	6250000
T = 59.5	Selective1	2	2	3	50
r = 0.000000015556636	Selective2	2	4	3	50
Time Point 3	A550	1.980	2.112	2.021	
07/03/11 08:00 PM	Unselective1	175	198	196	6250000
T_between = 36hrs	Unselective2	181	201	207	6250000
T = 95.5	Selective1	6	11	6	50
r = 0.000000352331482	Selective2	11	10	7	50
Time Point 4	A550	1.930	1.945	1.873	
07/04/11 08:45 PM	Unselective1	168	248	269	6250000
T_between = 24.75	Unselective2	224	316	290	6250000
T = 120.25	Selective1	40	43	44	5
r = 0.000000143630342	Selective2	45	59	41	5

Table 1.6: Data from the *P. aeruginosa* experiment.

	Culture 1		Culture 2		Culture 3	
	Ratio	SD	Ratio	SD	Ratio	SD
Time Point 2	9.662E-009	3.631E-010	3.820E-007	2.150E-007	1.735E-007	6.010E-008
Time Point 3	2.069E-008	8.620E-009	4.211E-007	2.871E-008	1.447E-007	1.670E-008
Time Point 4	1.849E-008	4.040E-009	2.581E-007	7.220E-008	1.216E-007	8.868E-009

Table 1.7: *P. aeruginosa* ratio of mutated cells over time to rifampicin.

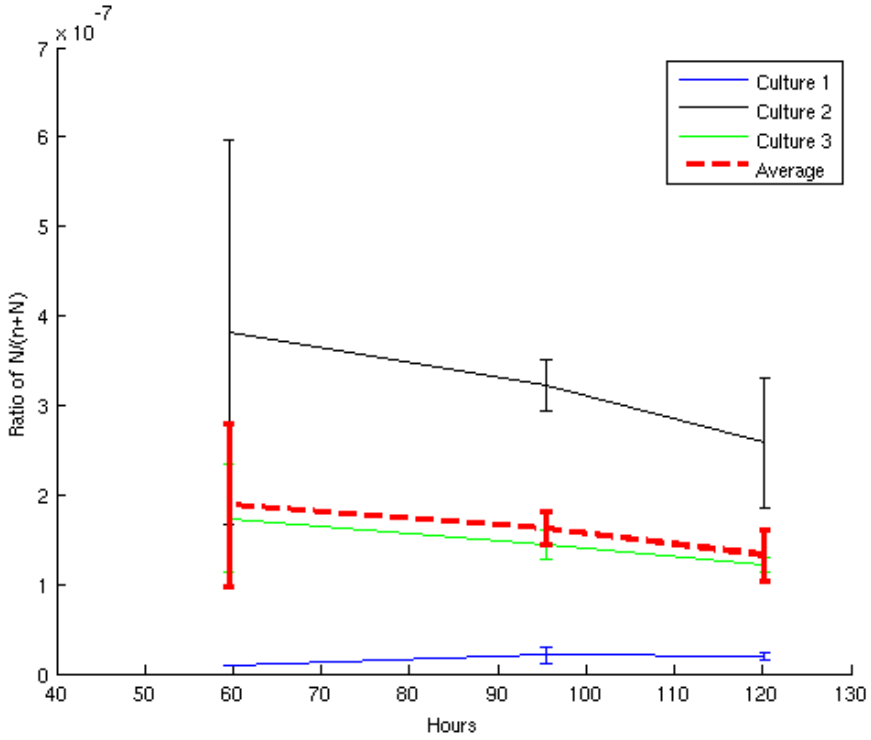

Figure 1.10: Ratio of mutated cell population to total cell population over time. The red dashed line is the average of all three cultures.

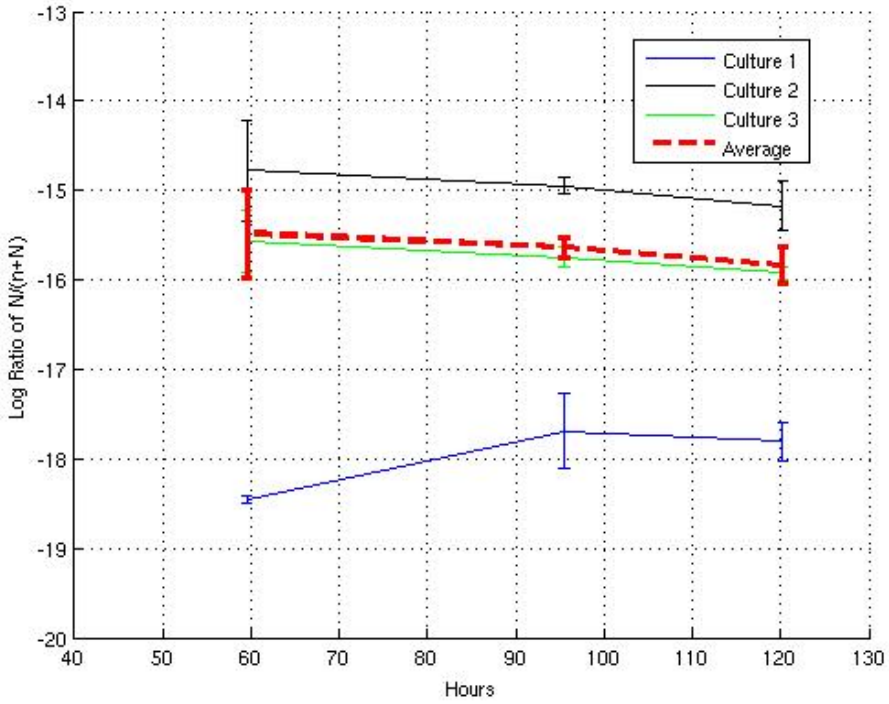

Figure 1.11: Log-Ratio of mutated cell population to total cell population over time. The red dashed line is the average of all three cultures.

Pseudomonas aeruginosa Growth Rate Constant Experiment

In order to figure out the growth rate constants of the mutated cells and non-mutated cells, selective and non-selective plates were saved from the end of the mutated ratio experiment above. A colony was grown up from each and optical density samples were taken every hour until stationary phase was achieved. The experimental protocol was as follows.

1. Two 10 ml tubes of Mueller-Hinton broth were inoculated, one with a colony from the the non-selective plate and one with a colony from the selective plate. The populations in these tubes grew up for 24 hours at 30°C.

2. Then 0.3ml of each of the two culture was inoculated in three separate 50ml beakers (six cultures total) and an optical density of the populations was taken at time point zero hours.

3. Every hour, for 10 hours, optical densities of the cultures were taken. After that point, optical densities were taken every 6 hours until saturation occurred (around 2 O.D., 550nm, for *P. aeruginosa*).

The above mentioned data were fit using non-linear least squares (Matlab Statistics Toolbox) to the following logistic equation:

$$P(t) = \frac{K \cdot P_0 e^{rt}}{K + P_0(e^{rt} - 1)}.$$

After this fit, 5000 bootstraps were done on all three parameters for each population, the initial population N_0, n_0, the corresponding growth rate constant, k_N, k_N and the carrying capacity term, M_N, M_n. The results are in Table 1.8.

The bootstrap histograms for the non-mutated cells and the mutated cells are in Figures 1.12 and 1.13.

Non-mutated Cells				
Parameter	Least-Sq Fit	Bootstrap Mean	Bootstrap SE	95% C.I.
n_0	0.024	0.025	0.008	(0.010-0.041)
k_n	0.578	0.582	0.051	(0.495-0.701)
M_n	1.851	1.853	0.052	(1.733-1.942)
Evolved Cells				
Parameter	Least-Sq Fit	Bootstrap Mean	Bootstrap SE	95% C.I.
N_0	0.016	0.016	0.003	(0.010-0.023)
k_N	0.520	0.522	0.028	(0.472-0.579)
M_N	1.958	1.959	0.020	(1.917-1.999)

Table 1.8: Fitted data, with 5,000 bootstraps.

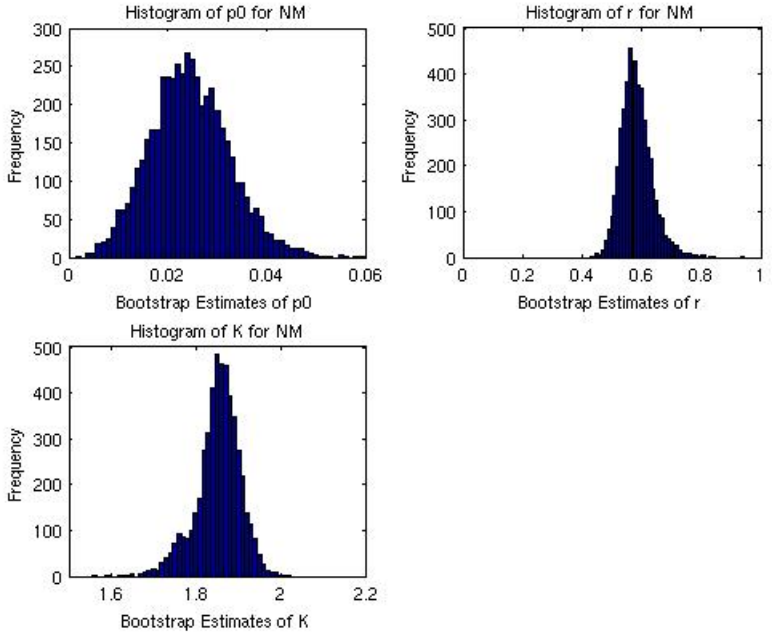

Figure 1.12: Histogram of the 5000 Bootstrap values for the non-mutated cells.

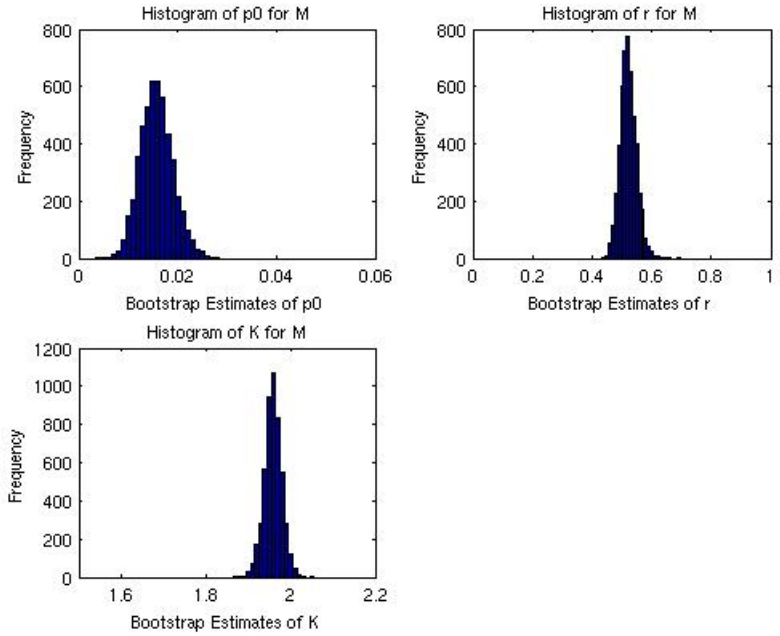

Figure 1.13: Histogram of the 5000 Bootstrap values for the mutated cells.

Plotting the 5% to 95% ordered growth rate constants from the bootstrapping in Equation (1.51) around the best fit line with the data produces the graphs in Figure 1.14. If we repeat the hypothesis test that the mutated cells have a lower growth rate constant using the bootstrap fits, we then arrive at a p-value of 0.0554. This is significant enough to proceed with the mutation rate constant estimation.

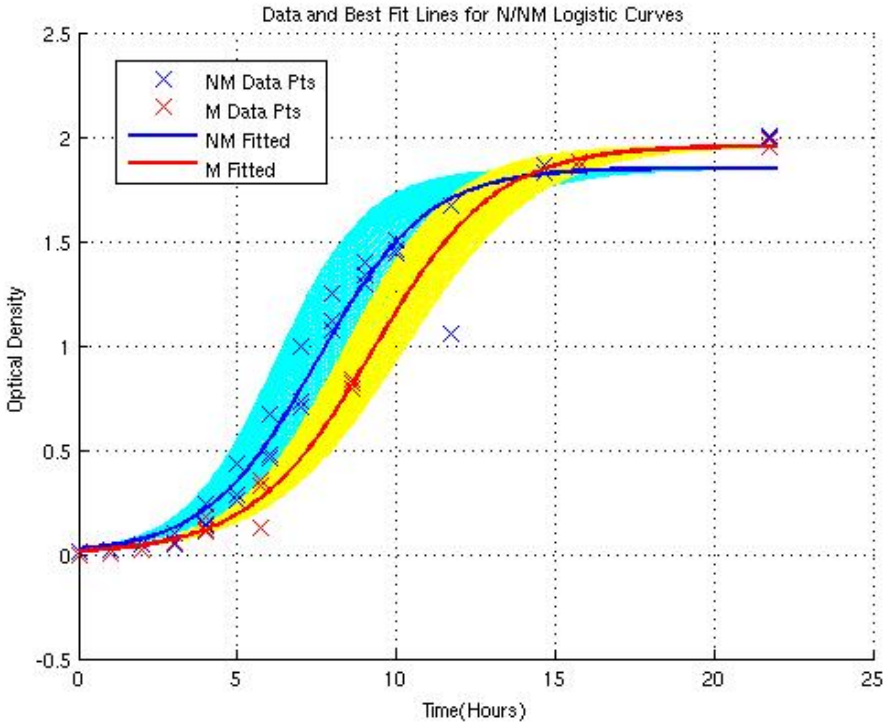

Figure 1.14: Both growth curves for the non-mutated and mutated cells. Also, the 5% to 95% ordered growth rate constants from the bootstrapping in Equation (1.51) plotted around the best fit line with the data.

Observing the data, we see that there are two outliers, one from each population. If we exclude these outliers, we get the results shown in Table 1.9.

The bootstrap histograms for the non-mutated cells and the mutated cells are shown in Figures 1.15 and 1.16.

Non-mutated Cells				
Parameter	Least-Sq Fit	Bootstrap Mean	Bootstrap SE	95% C.I.
n_0	0.021	0.021	0.004	(0.014-0.030)
k_n	0.597	0.598	0.029	(0.543-0.659)
M_n	1.900	1.900	0.032	(1.848-1.968)
Evolved Cells				
Parameter	Least-Sq Fit	Bootstrap Mean	Bootstrap SE	95% C.I.
N_0	0.020	0.020	0.003	(0.015-0.026)
k_N	0.496	0.497	0.017	(0.462-0.530)
M_N	1.964	1.964	0.014	(1.941-1.994)

Table 1.9: Fitted Data, with 5,000 bootstraps, and excluding the two outliers.

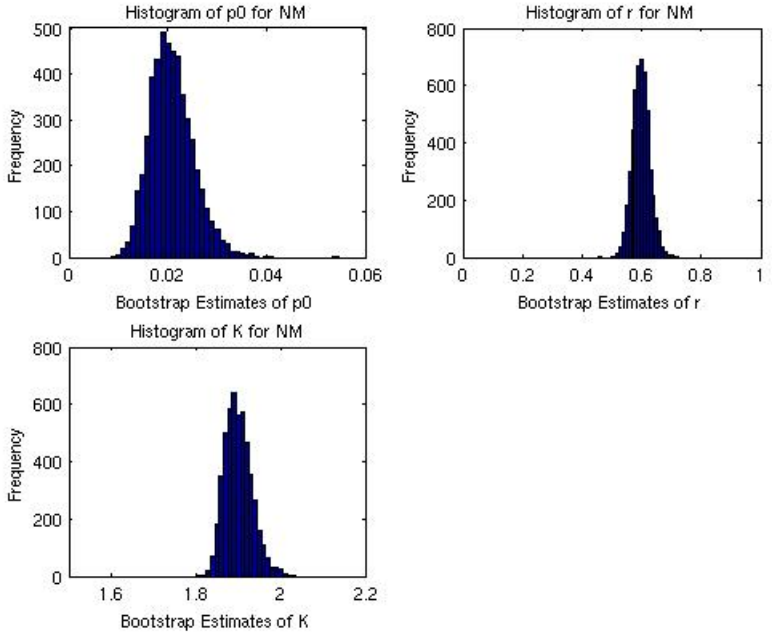

Figure 1.15: Histogram of the 5000 Bootstrap values for the non-mutated cells.

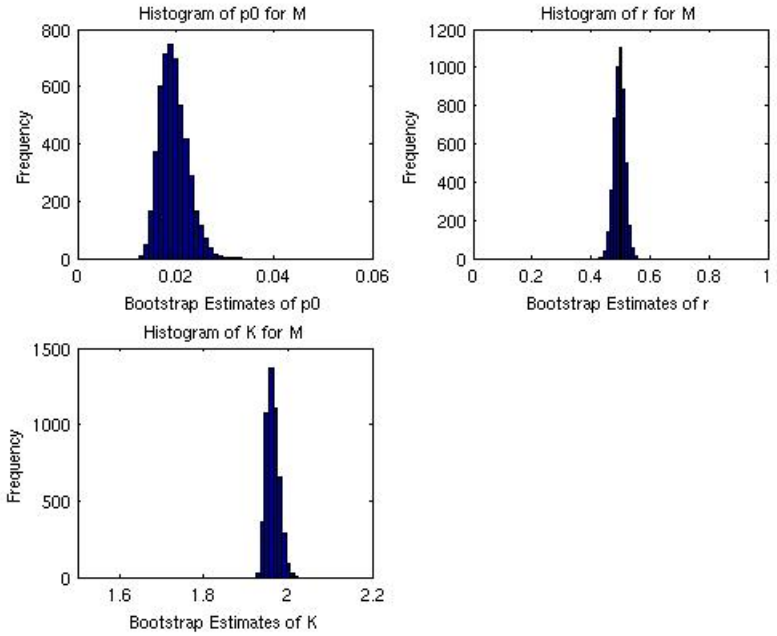

Figure 1.16: Histogram of the 5000 Bootstrap values for the mutated cells.

Plotting the 5% to 95% ordered growth rate constants from the bootstrapping in Equation (1.51) around the best fit line with the data produces the graphs in Figure 1.17.

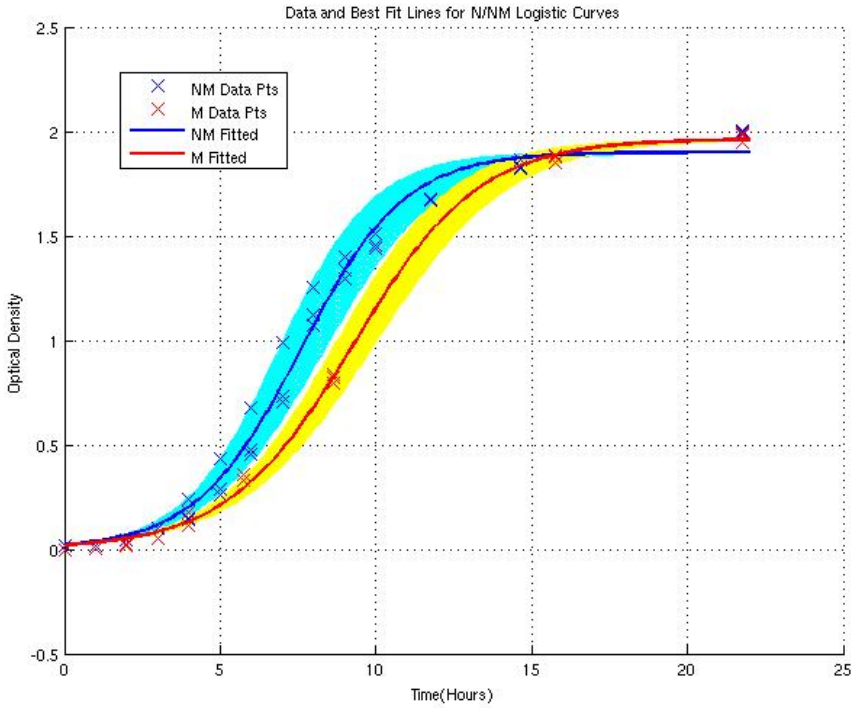

Figure 1.17: Both growth curves for the non-mutated and mutated cells. Also, the 5% to 95% ordered growth rate constants from the bootstrapping in Equation (1.51) plotted around the best fit line with the data.

From the above data, we can estimate the mutation rate constant.

Mutation Rate Constant Estimation

The mutation rate constant can be found using the following formula:

$$\mu = \left(\lim_{t\longrightarrow\infty} \frac{N}{n+N}\right)(k_n - k_N) \tag{1.56}$$

The mutation rate constant of this strain can be found using the averages and the standard deviation formula (1.50):

$$\lim_{t\longrightarrow\infty}\frac{N(t)}{n(t)+N(t)}=\frac{\mu}{0.597-0.496}=1.327\times10^{-7},$$

Solving for μ, the mutation rate constant, we arrive at

$$\mu=13.477\times10^{-9}\pm3.87\times10^{-9}.$$

1.12 Estimating Mutation Rate Constants in a Sample

Up until now in this thesis, all work has been done to estimate the average mutation rate of a colony grown from a single cell. This does not reveal much information as to the environmental make up of mutation rates in a sample. Here the assumption is made that there are two populations of concern, a high mutating subpopulation and a normal mutating population. The tests discussed previously reveal the average mutation rate of any single cell. If we run these tests again with a sample that contains two different mutation rates, it is easy to derive the proportion of mutators and non-mutators in the sample.

Consider the average mutation rate of a sample,

$$\mu_{avg}=\frac{N_1\mu_1+N_2\mu_2}{N_1+N_2}.$$

Here μ_{avg} is the average estimated mutation rate, N_1 is the population of non-mutator cells, N_2 is the population of mutators, and μ_1 , μ_2 are the respective mutation rate constants. By the definition of mutators, we can state that $\mu_1\leq\mu_{avg}\leq\mu_2$.

Solving for N_2/N_1,

$$\frac{N_2}{N_1} = \frac{\mu_{avg} - \mu_1}{\mu_2 - \mu_{avg}}. \tag{1.57}$$

Notice that

$$\lim_{\mu_{avg} \to \mu_2} \frac{N_2}{N_1} = \infty. \tag{1.58}$$

This tells us that if the mutation rate constants are equal, then the ratio of high-mutators to low-mutators goes to infinity. Because of this we consider the ratio of high-mutators to the whole population as in (1.59):

$$\lim_{\mu_{avg} \to \mu_2} \frac{N_2}{N_2 + N_1} = \lim_{\mu_{avg} \to \mu_2} \frac{1}{1 + (N_1/N_2)} = 1. \tag{1.59}$$

This limit approaches 1 in the case that the mutation rates are equal; see Figure 1.18.

Figure 1.18 reveals that to determine the ratio of mutators to non-mutators, the mutation rate of both populations and the average mutation rate of the sample must be determined. With this information, the proportion of mutators in the sample can be computed using (1.57).

Two Evolved Populations

In the following model, we consider the case of two evolved bacterial populations. We must make the assumption of the Law of Mass Action, as there are interacting population densities. The model system has the form:

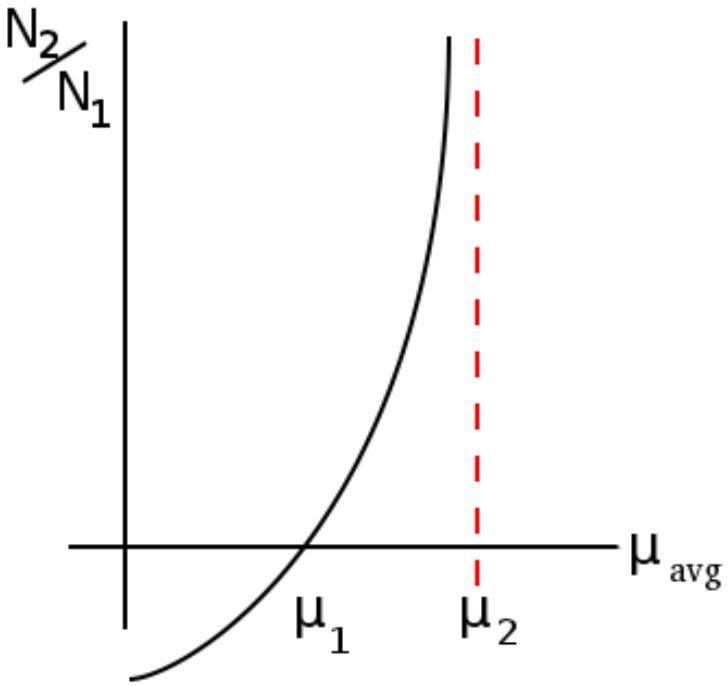

Figure 1.18: The ratio N_2/N_1 vs. μ_{avg}, the black line is the plot of (1.57). As μ_1 approaches μ_2, N_2/N_1 goes to infinity.

$$\begin{aligned} \frac{dn}{dt} &= k_n n - (\mu_1 + \mu_2)n - d_n n, \\ \frac{dN_1}{dt} &= k_{N_1} N_1 + \mu_1 n - d_{N_1} N_1, \\ \frac{dN_2}{dt} &= k_{N_2} N_2 + \mu_2 n - d_{N_2} N_2. \end{aligned} \tag{1.60}$$

Here, n, N_1 and N_2 are the populations of the unevolved and the two different evolved bacteria, respectively, μ_1 and μ_2 stand for the mutation rate constants to the first evolved and second evolved populations respectively, d_n, d_{N_1} and d_{N_2} are the respective death rate constants, and k_n, k_{N_1} and k_{N_2} are the respective growth rate constants.

The exact solution of the system (1.60) with initial conditions, $n(0) = n_0$, $N_1(0) = N_{1_0}$, $N_2(0) = N_{2_0}$, has the form:

$$n(t) = n_0 e^{\lambda t} \text{ , where } \lambda = k_n - \mu_1 - \mu_2 - d_n,$$

$$N_1(t) = \left[N_{1_0} - \frac{\mu_1 n_0}{\lambda + k_{N_1} - d_{N_1}} \right] e^{(k_{N_1} - d_{N_1})t} + \left[\frac{\mu_1 n_0}{\lambda + k_{N_1} - d_{N_1}} \right] e^{\lambda t},$$

$$N_2(t) = \left[N_{2_0} - \frac{\mu_2 n_0}{\lambda + k_{N_2} - d_{N_2}} \right] e^{(k_{N_2} - d_{N_2})t} + \left[\frac{\mu_2 n_0}{\lambda + k_{N_2} - d_{N_2}} \right] e^{\lambda t}.$$

Similar to previous discussion, consider the ratios $N_1/(n+N_1+N_2)$ and $N_2/(n+N_1+N_2)$. If we assume that $\lambda > (k_{N_1} - d_{N_1})$ and $\lambda > (k_{N_2} - d_{N_2})$, then it can be shown that the ratios reduce to the following:

$$\lim_{t\to\infty} \frac{N_1}{n + N_1 + N_2} = \frac{\left(\frac{\mu_1 n_0}{\lambda + k_{N_1} - d_{N_1}}\right)}{\left(\frac{\mu_1 n_0}{\lambda + k_{N_1} - d_{N_1}}\right) + \left(\frac{\mu_{12} n_0}{\lambda + k_{N_2} - d_{N_2}}\right) + n_0},$$

$$\lim_{t\to\infty} \frac{N_2}{n + N_1 + N_2} = \frac{\left(\frac{\mu_2 n_0}{\lambda + k_{N_2} - d_{N_2}}\right)}{\left(\frac{\mu_1 n_0}{\lambda + k_{N_1} - d_{N_1}}\right) + \left(\frac{\mu_2 n_0}{\lambda + k_{N_2} - d_{N_2}}\right) + n_0}.$$

This can be further reduced to:

$$\lim_{t\to\infty} \frac{N_1}{n + N_1 + N_2} = \frac{\mu_1}{\mu_1 + \mu_2 \left(\frac{\lambda + k_{N_1} - d_{N_1}}{\lambda + k_{N_2} - d_{N_2}}\right) + (\lambda + k_{N_1} - d_{N_1})},$$

$$\lim_{t\to\infty} \frac{N_2}{n + N_1 + N_2} = \frac{\mu_2}{\mu_2 + \mu_1 \left(\frac{\lambda + k_{N_2} - d_{N_2}}{\lambda + k_{N_1} - d_{N_1}}\right) + (\lambda + k_{N_2} - d_{N_2})}.$$

We also consider the ratio of the total evolved to the whole population, $(N_1 + N_2)/(n + N_1 + N_2)$,

$$\lim_{t\to\infty} \frac{N_1 + N_2}{n + N_1 + N_2} = \frac{\mu_1 + \mu_2\left(\frac{\lambda+k_{N_1}-d_{N_1}}{\lambda+k_{N_2}-d_{N_2}}\right)}{\mu_1 + \mu_2\left(\frac{\lambda+k_{N_1}-d_{N_1}}{\lambda+k_{N_2}-d_{N_2}}\right) + (\lambda + k_{N_1} - d_{N_1})}.$$

Despite the fact that we now consider two different evolved populations, the ratio of evolved to unevolved approaches a constant. We can generalize this process to multiple evolved populations. This is important for using our model in order to understand the meaning of mutation rate constant values in environmental samples, in which some distributions of bacteria with different mutation rate constants exist.

K Evolved Populations

Now consider generalizing the previous case to k evolved populations. The model will have the form:

$$\frac{dn}{dt} = k_n n - \left(\sum_{i=1}^{k} \mu_i\right) n - d_n n,$$

$$\frac{dN_i}{dt} = k_{N_i} N_i + \mu_i n - d_{N_i} N_i, \text{ for } i = 1, \ldots, k.$$

The initial conditions are $n(0) = n_0$, $N_i(0) = N_{i_0}$. The corresponding explicit solution can be written as:

$$n(t) = n_0 e^{\lambda t} \text{ , where } \lambda = k_n - \left(\sum_{i=1}^{k} \mu_k\right) - d_n,$$

$$N_i(t) = \left[N_{i_0} - \frac{\mu_i n_0}{\lambda + k_{N_i} - d_{N_i}}\right] e^{(k_{N_i} - d_{N_i})t} + \left[\frac{\mu_i n_0}{\lambda + k_{N_i} - d_{N_i}}\right] e^{\lambda t}.$$

Now we consider the ratio of evolved population i to the whole population:

$$\lim_{t\to\infty} \frac{N_j}{n + \left(\sum_{i=1}^{k} N_k\right)} = \frac{\mu_j}{\mu_j + \left(\sum_{i\neq j}^{k} \mu_i \frac{\lambda + k_{N_j} - d_{N_j}}{\lambda + k_{N_i} - d_{N_i}}\right) + (\lambda + k_{N_j} - d_{N_j})}.$$

We can also consider the ratio of all evolved bacteria to the total population in the culture:

$$\lim_{t\to\infty} \frac{\left(\sum_{i=1}^{k} N_i\right)}{n + \left(\sum_{i=1}^{k} N_i\right)} = \frac{\left(\sum_{i=1}^{k} \frac{\mu_i}{\lambda + k_{N_i} - d_{N_i}}\right)}{1 + \left(\sum_{i=1}^{k} \frac{\mu_i}{\lambda + k_{N_i} - d_{N_i}}\right)}.$$

Again, here we note that even though there are multiple equations for evolved bacterial populations, all of the ratios to the whole population approach a constant. This suggests possible experiments to determine the relationships of multiple mutation rate constants present in a sample. Previous knowledge of the growth and death rate constants of the unevolved and each evolved population is needed. These growth and death rate constants can be determined by performing growth rate experiments as was done previously with *E. coli* and *P. aeruginosa*.

Minimizing the Bacterial Load

The mathematical mutation model, described by System (1.22), can be used to design an optimal antibiotic schedule of treatments for patients with chronic lung infections. In order to do so, we must define the effect of the antibiotic on the system. During an antibiotic treatment, we set the growth rate, k_n, of the unevolved bacteria population to zero.

During antibiotic treatment, we can predict the actual steady states approached and the limiting ratio of evolved to the total population by plugging in $k_n = 0$ into the corresponding steady state equations, Eq. (1.27) and (1.28). The limiting ratio is then given by eq. (1.29).

Choosing constants from the *Pseudomonas aeruginosa* experiment, we let $\mu = 13.5\times10^{-9}$, $k_n = 0.597$, $k_N = 0.496$, $n_0 = 0.021$, and $N_0 = 0.02$. We choose to exaggerate the carrying capacity to see the effect of antibiotics more clearly, and set $M = 1000$ We must make the assumption that $p > q$, this assumption means that unevolved bacteria use less resources than the same amount of evolved bacteria. We then apply antibiotics from time point 40 to time point 130 hours alternating between 5 hours of antibiotic treatment and 5 hours of no antibiotic treatment. Here we assume that the unevolved population remains susceptible to the antibiotic. The simulation results are shown in Figure 1.19.

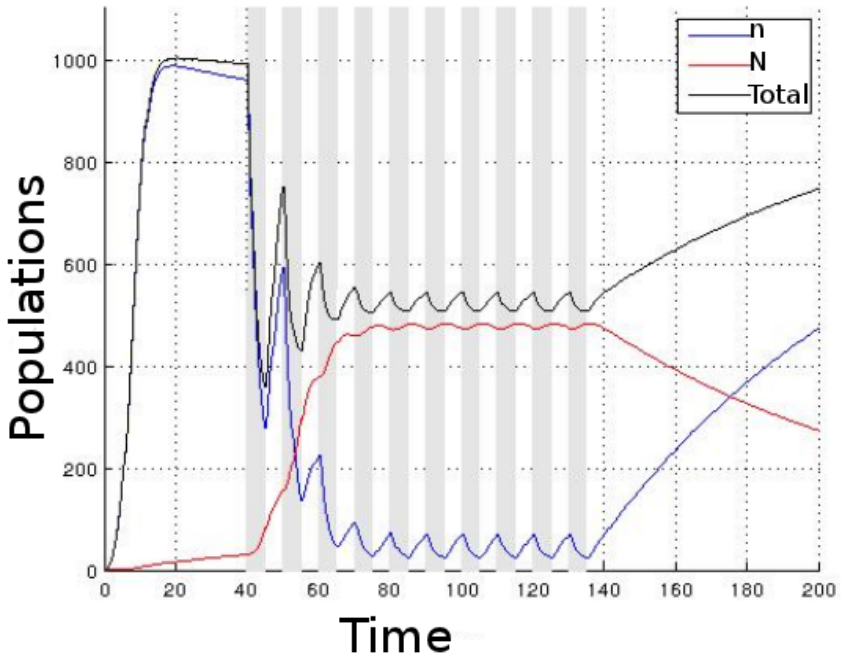

Figure 1.19: A graph of total population (black), evolved population (red), and unevolved population (blue) during an antibiotic schedule that consists of antibiotic treatment (shown in gray) every other 5 hours applied from hour 40 to hour 130.

The total bacterial load (black line in Figure 1.19) goes down significantly during antibiotic treatment if and only if $p > q$. This means that when antibiotics are applied, the population of evolved bacteria

increases (as they are unaffected by the antibiotic) and the total population goes down because evolved bacteria use more resources. Continual applications of the antibiotic are needed to keep the bacterial load at a minimal level. What is also important to notice is that after the antibiotic treatment is over ($t > 130$), the unevolved bacteria take longer to reach the steady state than the initial population due to higher competition with the evolved population.

The implications of this simulation show that antibiotics are useful for decreasing the overall bacterial load, not only immediately, but in the long run as well and providing lasting effects after the antibiotics are removed.

1.13 Conclusions, Pitfalls, and Future Directions

In this chapter, we have introduced a new method for measuring the mutation rate constant of bacterial populations. We have verified the procedure experimentally with multiple strains. This procedure is mathematically simpler than the fluctuation experiment and is experimentally simpler and cheaper than the continuous mutation accumulation assay. It would appear that these methods work for most of the clinical strains tested. The strains that the new method does not work with are strains that break the assumption about the mutated cells growing slower than the non-mutated cells. In order for this new method to work, that assumption must be made.

It is also possible to use more complicated ratios, but measurements of the strains death rate constants must be done. This is possible with live/dead staining or using microscopy. As for comparing the results of the new method to the fluctuation analysis, the latter is the one to be trusted more. This new method must be more rigorously tested. Also given that the plate counts were high for the ratio experiment, the results of the new experiment are not to be taken as final.

It should be mentioned that the new method outlined in this chapter is only relevant when the growth rate constant of the mutated strain is smaller than the growth rate constant of the non-mutated strain.

Even if the growth rate constants are close to each other, this could introduce large errors in the calculations. To get around this, the other two methods may be used in conjunction with this new method to verify results.

There are other methods to measure mutation rate, e.g. sequencing, bottlenecking populations, etc. These methods measure mutation rate constants in different units than the methods addressed in this chapter. They might be more reliable and more useful depending on the question at hand. The new method proposed here only addresses the probability of a strain generating a measurable mutation (by way of measuring phenotype) per generation. It does not address nonsense mutations, nor distinguish between different mutations that result in the same phenotype, or anything similar.

For future directions, the experiment must be repeated to verify results. This research would be furthered by measuring the mutated and non-mutated strain death rates and use the more complicated ratios to see the difference in the outcomes. It would be recommended to compare this method with different measurements of mutation rates, for example, to sequence strains over time and record the mutation rate of the genome and compare it to how the new method measures mutation rate. It is important to note that the new method measures mutation rate as the probability of generating resistance to an antibiotic per generation, and sequencing would measure the probability of changing the genetic code per generation. These are different measurements, and the relationship between them is not fully known.

Chapter 2

Microarray Comparison and Classification

2.1 Introduction

A microarray is a molecular biology tool that can simultaneously measure the amount of cRNA (copy RNA) or cDNA (copy DNA) in a sample, which is linked to gene absence or presence. [8]. A microarray consists of thousands of DNA probes attached to a surface. When a sample is run over the microarray chip, bits of cDNA or cRNA hybridize to the probes and fluoresce. The machine then measures the amount of fluorescence. This measurement is compared to a standard control and the comparison reveals deletions or up-regulations of genes by computing the logarithm of the ratio of the test strain fluorescence to the control strain fluorescence. There are different types of microarray tests with different types of chips and fluorescence. Each test can result in different values for the same gene and strain. The microarray tests in our data sets consists of sheared *Pseudomonas aeruginosa* genomes that were assessed on aCGH (comparative genome hybridization) microarrays [46].

Many protocols call for different microarray tests to be run for confidence in the results [36]. Running

many microarray tests may be expensive and time consuming for experiments. It would be convenient to have a cutoff value determined such that every data point greater than it shares a specified confidence level that another microarray test would give the same result.

With this specific data set, we are interested in gene deletions. Thus a cutoff will be chosen to maximize the correlation of gene presence or absence values between tests below such a specified cutoff value. The basis of this is that a highly deleted genes will show a decreased measurement on all microarray tests. If a too small cutoff is chosen then too much noise is included, if a too extreme cutoff is chosen, important gene values will be excluded and the sample size will be too small.

2.2 Microarray Comparison

In this section, two microarrays will be compared across 5,437 genes with nine strains. Both tests will be compared. The problem of the two different microarray tests having two different means and variances will be addressed below along with maximizing the correlation of gene absence values between tests below a specified cutoff.

Co-Inertia Analysis

There are different ways for comparing two different tests on the same strains. One method is Co-inertia Analysis (C.I.A.) [11]. Co-inertia Analysis uses principle component analyses to transform the two data sets. Specifically it finds axes that minimize the variance. The data are then rotated and shifted about these axes on top of each other.

After this transformation, the RV coefficient is used to compare the two data sets containing only genes with all strain values more extreme than the cutoff [40]. The RV coefficient compares the two matrices, similar to the multivariate Pearson Correlation.

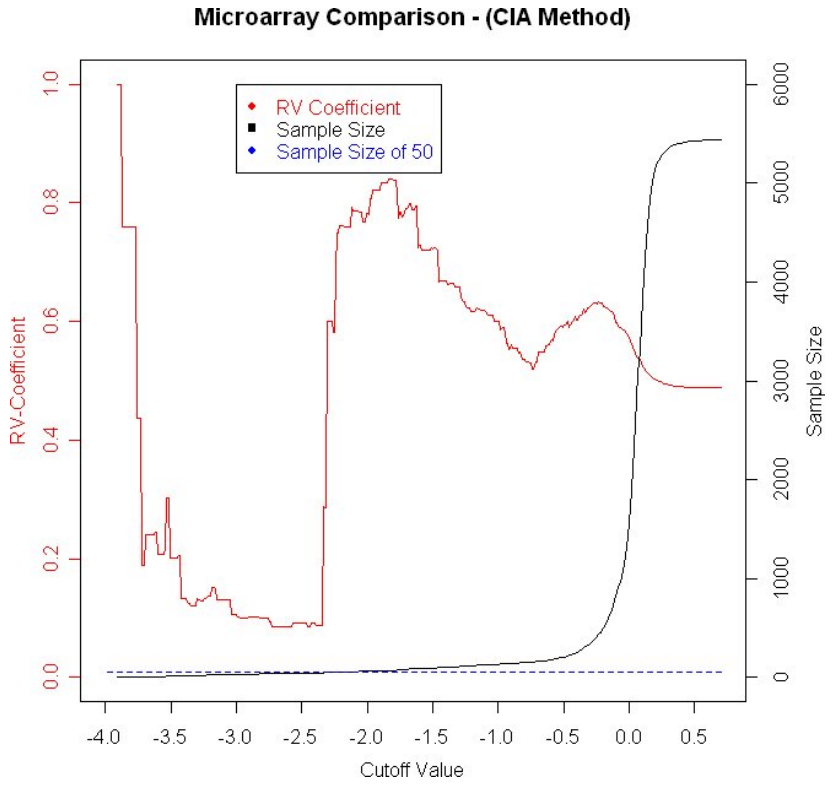

Figure 2.1: Red line shows us the RV-coefficient (left y-axis), only including gene values below a specified cutoff value (x-axis). On the right y-axis, sample size is shown (black line), i.e. how many genes are blow such a cutoff and for reference, a horizontal blue dashed line depicts a sample size of 50 genes, suggesting that the low sample size is a reason for a drop in the correlation.

Given two data matrices, X and Y, define the trace of a matrix as X^T. The RV-coefficient between the two matrices, X and Y, depends on the covariance matrix given by the following:

$$\Sigma_{XY} = E(X^T Y).$$

The RV-coefficient is defined as

$$RV(X, Y) = \frac{Tr(\Sigma_{XY}\Sigma_{YX})}{\sqrt{Tr(\Sigma_{XX}^2) \cdot Tr(\Sigma_{YY}^2)}}.$$

To address the noise in the data set, genes with all 9 strain values above a specified cutoff will be thrown out and not included. The RV coefficient is then computed for the remaining genes between the two microarray tests (See Figure 2.1). Note that as the cutoff value gets more negative, the number of genes included in the sample size get small fast.

In Figure 2.1, a local maximum RV coefficient of 0.84 lies at a cutoff value of -1.98.

Here it is worth nothing that between the values of -2.2 and -3.7 the RV coefficient is low, suggesting those genes below those cutoff values are not correlated. This might be explained by the vanishingly small sample size of the genes. Also extremely negative numbers around -4 suggest those genes have multiple deletions and have a greater chance of appearing that way in other microarray tests, which might explain the high correlation in the very few genes remaining below -3.7.

Using PCA in transforming the data suggests that the variance in the data is important. There is no reason to suspect that variance in the microarray data is important. Also, this method assumes that a transformation is needed to get the data from the two tests on the same scale. Since both tests generally have the same mean (approximately zero) and similar variances, a transformation might not be needed and another method should be used.

Average Pearson Correlation

If a transformation is not needed, then another method might be to maximize the average correlation between the two tests when including only genes whose values are below a specified cutoff. The following steps are followed to find the maximum average correlation.

- A cutoff point is chosen.
- Any genes with all values above the cutoff point are thrown out from both tests.
- The Pearson correlation is computed between corresponding strains from both tests and then the

average is taken.

- The procedure is repeated for different cutoffs and the cutoff that maximizes the average Pearson correlation is chosen.

Following the above procedure results in Figure 2.2.

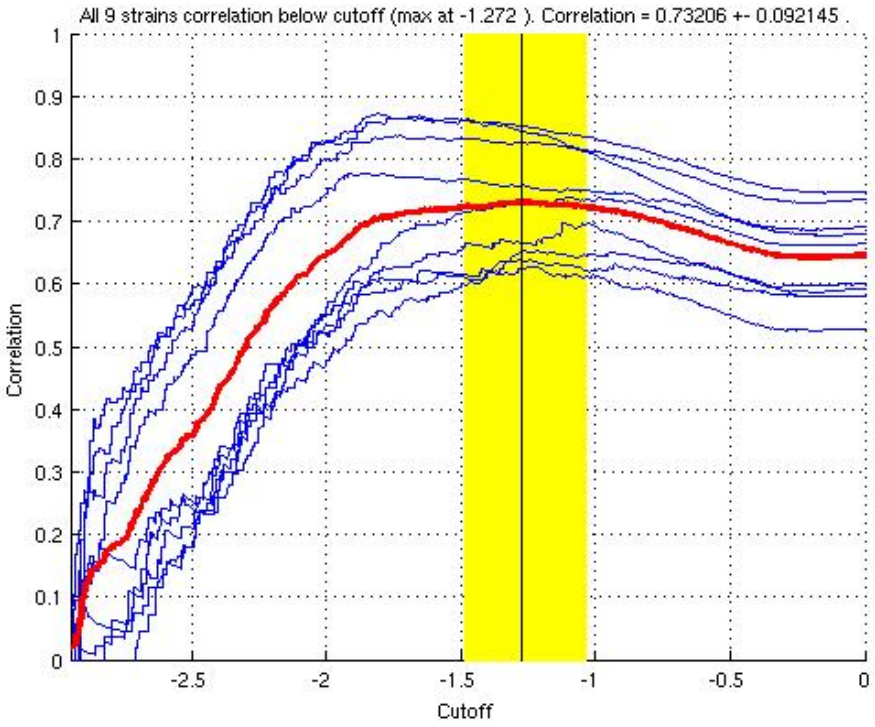

Figure 2.2: Red line shows the average Pearson correlation (left y-axis), only including gene values below a specified cutoff value (x-axis). The blue lines are the individual correlations for the 9 strains. The maximum is shown at -1.272 with a vertical black line. The yellow region depicts a region of 1% deviation from the maximum value (-1.50, -1.15). Bootstrapping was performed at the maximum (90% sampled 10000 times) to obtain a standard deviation on the correlation.

A confidence region is chosen to be all cutoff values that result in a 1% deviation from the maximum value. To minimize the chance of false positives, the lower part of the region may be chosen (-1.5) as the cutoff value.

Discussion

Given the similar results from the Co-Inertia Analysis and the average Pearson correlation, the latter method may be preferred due to less involved calculations with large matrices of data. Also it may be argued that transforming the data may result in the loss of information for correlation and predictions.

2.3 Microarray Classification

Mutator cells and mutated (or evolved) cells are important to distinguish in the next sections. Mutators are cells that have a higher than normal mutation rate. This does not mean they are resistant to an antibiotic, but they have a high probability of developing resistance. Non-mutator cells are cells with a lower mutation rate and can also develop resistance, but have a lower probability of doing so due to a lower mutation rate constant. When either population develops resistance, they are called evolved or mutated cells.

It is desirable to determine the mutator/non-mutator status of a strain without having to determine the mutation rate. To do this, a single microarray data set of 48 known strains will be analyzed for a method of classification. All methods will require the removal of each strain, one at a time, and it will classified under the assumption that the mutator status is unknown. This is known as Leave-One-Out Crossvalidation (LOOCV).

Normal Distribution Classification

In this classification, we assume that all the strain microarray measurements come from a normal distribution. We will remove one strain at a time and label it as an unknown strain. The unknown strain is then compared to the mutator and non-mutator values using a Welch t-test, because the two corresponding strains may have unequal variances. The higher p-value will tell which distribution the

unknown strain is the closest. This was done for all 48 strains in the data set. The results are in Table 2.1.

Linear Discriminant Classification

One assumption made in the above classifications is that all genes are equally important in determining a mutator or non-mutator. Since that is probably not the case biologically, principal component analysis (P.C.A.) will be used to describe the 2 or 3 components (dimensions) that describe the greatest variation in the data. From these components, a linear discriminant analysis can be done to determine if an unknown strain can be classified as mutator or non-mutator. The results are seen in Table 2.1.

Average Projection Classification

If strains are thought as vectors with n dimensions (n genes), an unknown strain can be classified in a group with the largest projection. To accomplish this, the mutators are averaged together and similarly the non-mutators are averaged together. Let $\boldsymbol{U}$ stand for the unknown strain vector, and $\boldsymbol{D}$, $\boldsymbol{N}$ stand for the average mutator and non-mutator vectors respectively. Then the projection of $\boldsymbol{U}$ on $\boldsymbol{D}$ and $\boldsymbol{N}$ is computed ($\boldsymbol{U}_D$, $\boldsymbol{U}_N$).

$$\boldsymbol{U}_D = \frac{\boldsymbol{U} \cdot \boldsymbol{D}}{||\boldsymbol{D}||},$$

$$\boldsymbol{U}_N = \frac{\boldsymbol{U} \cdot \boldsymbol{N}}{||\boldsymbol{N}||}.$$

If $\boldsymbol{U}_D > \boldsymbol{U}_N$ then the unknown strain is classified as a mutator and if $\boldsymbol{U}_D < \boldsymbol{U}_N$, the unknown strain is classified as a non-mutator. In the unlikely chance that $\boldsymbol{U}_D = \boldsymbol{U}_N$, the unknown strain stays unclassi-

fied. The results are seen in table 2.1.

Allowed Error

The results in Table 2.1 do not show how close the classification values are to each other. For the normality classification using the t-test, the p-values might have been very close to each other, similarly for the average projection, the mutator and non-mutator projections might have been very close to each other. If the strain in question is within a specified error tolerance to both classes, we classify it in a third class, *u* for unknown regardless if the strain was correctly or incorrectly classified. The results of different error tolerance are seen below in Figures 2.3 (Normal test allowed error) and 2.4 (Geometric projection allowed error).

Actual Classification	Normal-T-Test	LDA- 2 Axes	LDA- 3 Axes	Avg Projection
Fraction Misclassified:	23/48	30/48	27/48	12/48

Table 2.1: The four classification outcomes, the fraction of the 48 that were misclassified.

Discussion

Table 2.1 suggests that the best is the average projection method. The fraction of misclassified strains obtained by this method is lower by at least 11 more strains classified correctly when compared to the other methods. The other three methods mis-classify strains about 50% of the time, suggesting that these methods are no improvement over a random guess.

An ideal classification would seek to minimize both misclassified strains and unclassified strains. Using an error tolerance in the p-values of the Welch t-test, or an error tolerance in the values of the projections, decreases the amount of misclassified strains but also increases the amount of unclassified strains. An appropriate error tolerance is subjective and depends entirely on the needs of the classification. A suggestion for a proper amount of error is to consider the level of noise in the data. Looking at Figure 2.4, we observe that allowing error in the projection method of 1.95 units (projected gene

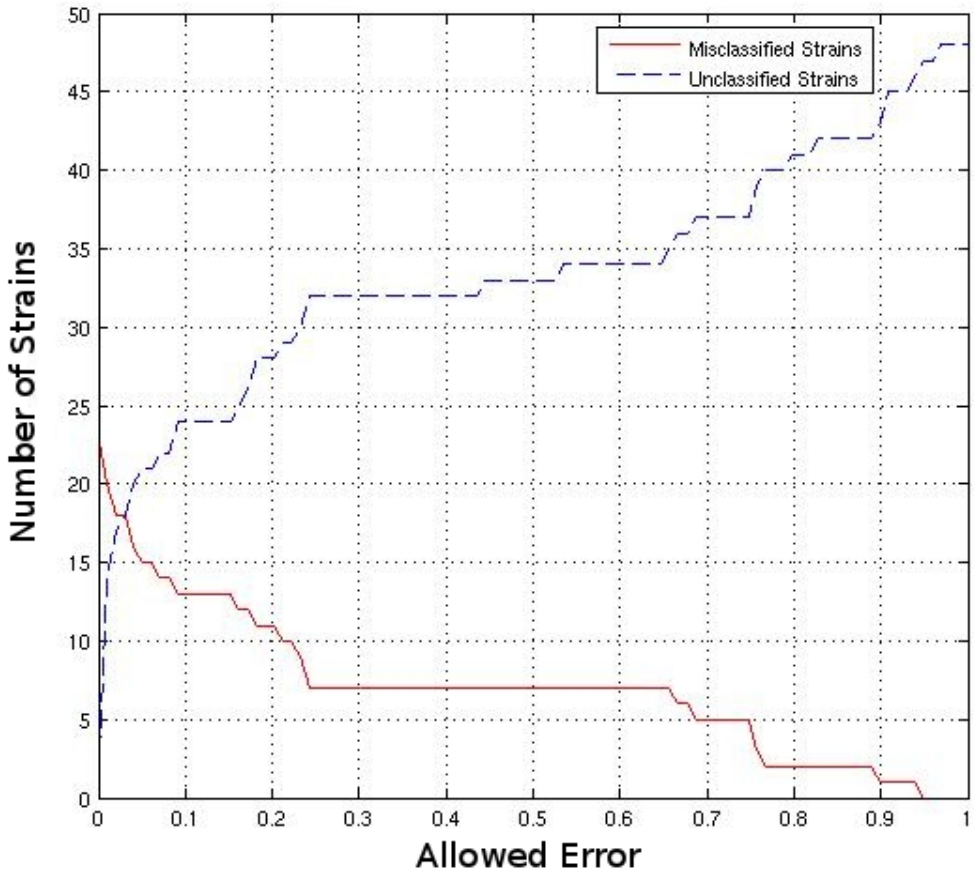

Figure 2.3: Red line shows the number of misclassified strains above a certain error tolerance using the normal classification strategy. The total number of unclassified strains is represented by a blue dashed line.

microarray values), results in no strains misclassified, and under half of the strains unclassified.

Recovering Missing Values

A microarray test can result in missing values due to process error. When this happens, the data value is filled in with a 'NaN' (not a number) value. Since these values remain unknown, the gene values are dropped completely if any NaN values appear. Starting with 5551 genes, and dropping the genes that

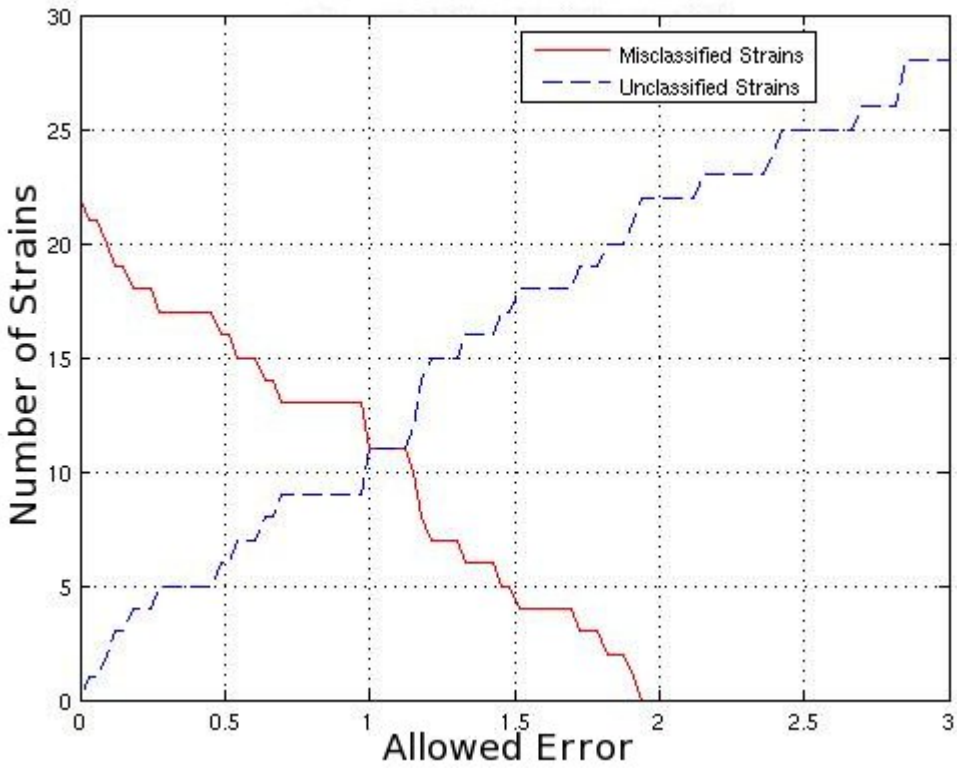

Figure 2.4: Red line shows the number of misclassified strains above a certain error tolerance using the geometric projection classification. The total number of unclassified strains (strains with classification values below a certain error tolerance) is represented by a blue dashed line.

contain any missing data values, the number of genes available to use for classification is reduced to 3695 genes. But the 1856 genes that were dropped might contain information that could be valuable. One way to remedy this is to remove strains instead of the genes. In the data set used for classification, we remove the ten worst strains. Doing this, results in fewer genes that are thrown out due to having NaN values. Instead of starting with 3695 gene values we have 4373 gene measurements. This means that the correlations will have more gene data but less strain data to classify. The results of all the classifications are given in Table 2.2.

Another solution is to try and fill in the NaN values with the average of the gene microarray values. To do this we find the average microarray value of the gene across the strains and fill in the NaN values with that average. This leaves all 5551 genes and all 48 strains in the data set and the classifications

38 Strains, 4373 Genes (worst 10 strains removed)	Misclassified	Misclassified (<1.272)	Misclassified (<1.6)
Normal T-Test	0.526	0.474	0.447
LDA- 2 Axes	0.368	0.368	0.368
LDA- 3 Axes	0.421	0.421	0.421
Avg. Projection	0.395	0.447	0.500

Table 2.2: Table of classification results (fractions of strains available for classification) for the case where the 10 worst strains were omitted. The numbers in the parentheses indicate the usage of genes with mean microarray values below such a number.

are rerun. See Table 2.3 for results.

All Strains, All Genes (NaN's replaced by gene mean)	Misclassified	Misclassified (<1.272)	Misclassified (<1.6)
Normal T-Test	0.521	0.333	0.375
LDA- 2 Axes	0.438	0.438	0.438
LDA- 3 Axes	0.458	0.229	0.229
Avg. Projection	0.208	0.313	0.313

Table 2.3: Table of classification results when filling in NaN values with the corresponding gene data mean.The numbers in the parentheses indicate the usage of genes with mean microarray values below such a number.

Discussion

The results show that the biggest improvement from including more gene microarray data is that the projections classification improve to almost a 20% misclassification rate. The normal t-test and the LDA methods also show some slight improvements.

Highly Informative Genes

Not all genes in a microarray may be indicative of the categories we observe (mutator/non-mutator). Reducing the number of gene microarray values used in classification is important to lower the computational load, and it may also result in knowledge of what specific genes might be responsible for two categories that we want to distinguish. One method to reduce the number of gene values used in classification is to only consider gene values that have high individual LOOCV scores when we are using the average projection method. This requires computing the classification for each gene and only taking the genes with the lowest misclassification percentage. A plot of the number of genes that

individually classify better than a given percentage is shown in Figure 2.5. The top classifying genes are shown in Table 2.4.

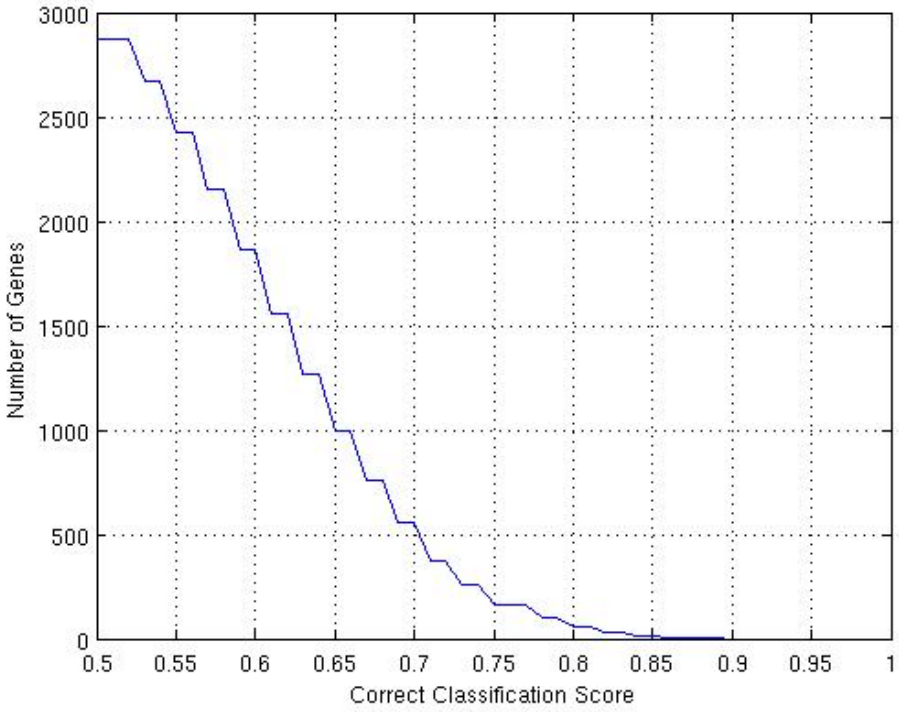

Figure 2.5: Plot of the number of genes (y-axis) that classify better than a given percentage (x-axis).

The gene space can effectively be narrowed down by selecting individual genes that classify correctly above a desired percentage. For illustrative purposes, we use the genes that classify greater than 85% of the time. Such genes can be seen in Table 2.4. Using only these 17 genes, the average projection classification and normal t-test classification improves (Table 2.5).

Individually picking out good classifying genes may introduce bias into our system. In other words, there maybe a different selection of genes that results in less misclassification. It is hard to find these genes as there are many different possible groups that can be used.

Gene Description	> 85%	> 87%	> 89%
Hypothetical Protein	PA356	PA356	
Hypothetical Protein	PA377		
Hypothetical Protein	PA388		
Hypothetical Protein	PA581		
Amino Acid ABC Transporter Membrane Protein	PA1340		
Probable Chemotaxis Transducer	PA1646		
NADH Dehydrogenase I Chain H	PA2643		
Cell Division Topological Specificity Factor MinE	PA3245		
Hypothetical Protein	PA3323	PA3323	
Probable Non-Ribosomal Peptide Synthetase	PA3327	PA3327	
Probable ATP-Binding Component of ABC Transporter	PA3375		
Transport Protein HasD	PA3406	PA3406	
Hypothetical Protein	PA3419		
Hypothetical Protein	PA3421	PA3421	PA3421
Hypothetical Protein	PA3440		
Hypothetical Protein	PA3473		
Probable Permease of ABC Transporter	PA3512	PA3512	

Table 2.4: This table shows the genes that individually classify better than a given percentage of genes shown in the top row. The gene descriptions are on the left, they were taken from the *Pseudomonas* Genome Database.

17 Genes Selected by Individual Classification	Misclassified
Normal T-Test	0.104
LDA- 2 Axes	0.542
LDA- 3 Axes	0.479
Avg. Projection	0.104

Table 2.5: Group classification results using only the 17 genes that individually have a correct classification rate of greater than 85%. Notice that the LDA uses PCA to transform the data, resulting in a loss of information in the gene data, and groups of genes do generally worse, despite individually performing well.

2.4 Genetic Algorithms

To reduce bias, a genetic algorithm might be used to explore the different possible groups of genes used for classification. A genetic algorithm uses concepts found in evolution to explore a parameter space. It uses the principle of the "survival of the fittest" to produce better solutions to a problem. It starts with a population of solutions, evaluates them and ranks them from best to worst. A top percentage of solutions are kept, and they generate new "offspring" by recombining parts of the better solutions (recombination). Then the resulting solutions are perturbed slightly (mutation) and re-evaluated. This procedure is repeated over and over until certain stopping criteria are met. This procedure is schematically demonstrated in Figure 2.6.

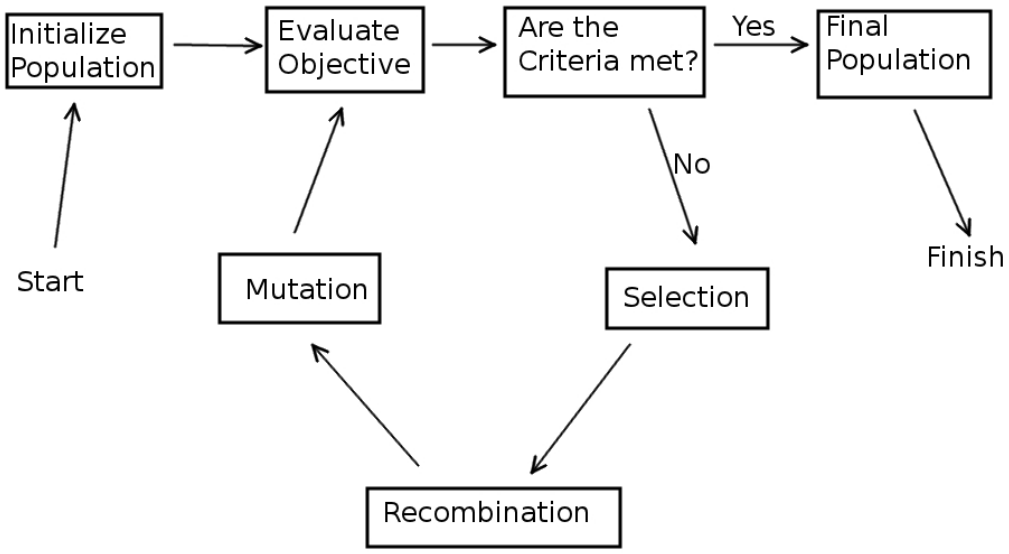

Figure 2.6: Structure and description of the processes that occur during an genetic algorithm implementation.

For our purposes, each individual is a $1 \times n$ binary vector, where n is the number of possible genes. Evaluation of individuals happens by using the genes for classification where the vector has a one-value instead of a zero-value. Each individual in the population is evaluated and ranked based on geometric projection classification performance of the groups of gene values it uses. We also implement a small-number penalty to prevent individuals from using small groups of genes, say only gene PA3412 (see

Table 2.4). We implement this by subtracting a term to the fitness of a gene that has high values for very small number of genes, so a more fit individual has more genes. As a side note, this term can vary and could also have a maximum around the desired amount of genes.

To start, we initialize 100 individuals with a randomly generated binary strings. We evaluate and rank each individual according to how well the gene groups for each classify using LOOCV. Then the top 10% are kept, and the rest are thrown out. Ninety new offspring are generated by randomly selecting two parents. The offspring's binary value at any position is chosen by taking on the value of a randomly picked parent. After we generate the new population, we go through each individual and randomly switch each bit (zero to one or one to zero) based on a mutation probability. For our simulations, the mutation probability was selected to be $1/n$, so that each individual is expected to have one mutation per generation. This mutation probability was found to have the best results. If the mutation probability is too low, then our resulting population may depend too much on the initial population. If the mutation probability is too high, the final population may never converge. For our simulation, we can not implement a stopping procedure based on classification because we do not know how low the misclassification percentage can get with any group of genes. To get around this, the stopping criteria we will use is related to the number of generations. When the number of generations reaches a certain value (500) we stop the simulation. The best individual found used around 1854 genes to get the misclassification of the geometric projection procedure down to 0.1042 (5 out of 48). To reduce the number of genes used even further, we can also take genes that the top 10 individuals all agree on. This means that we only use genes that the top 10 fittest individuals have a '1' in that gene position. For our population of 100 run for 500 generations, we take the genes that the top 10 most fit individuals agree on (all 10 have a '1' in that gene position). This reduces the amount of genes used to 435 for a misclassification of 0.1458 (7 out of 48) using the geometric projection method.

2.5 Conclusions, Pitfalls, and Future Directions

So far, using the methods discussed in this chapter, the lowest LOOCV misclassification percentage that we were able to achieve was around 10%. This suggests that the information in the data is not ideal for predicting mutator/non-mutator status. To understand how much information is contained in the data, as it pertains to mutator/non-mutator status, we plot the variances of the gene microarray values vs. corresponding means. The idea here is that the higher the variance of the gene data, the more information it contains. The plot is shown in Figure 2.7.

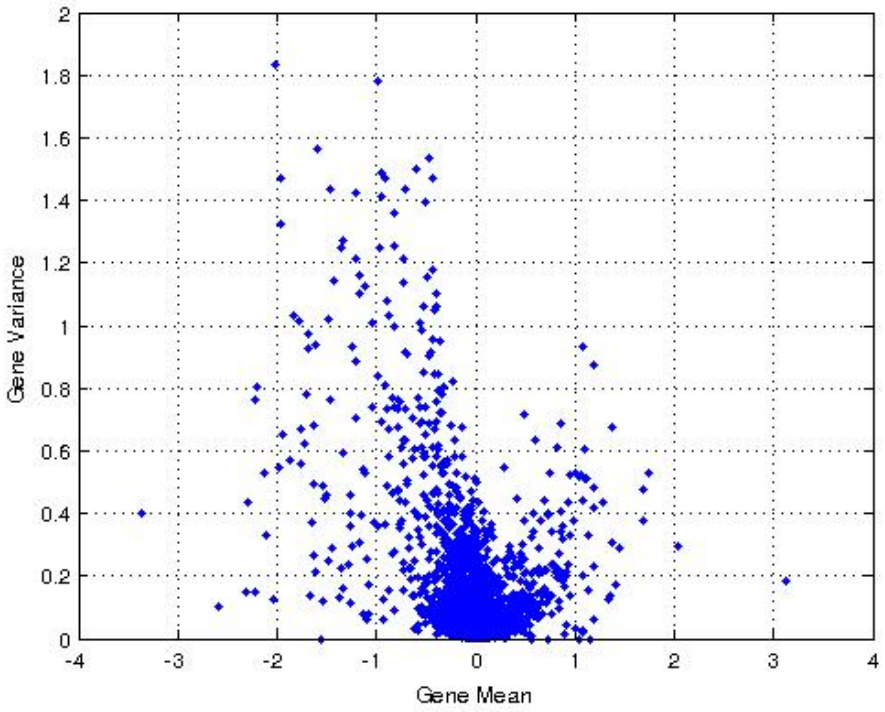

Figure 2.7: Gene microarray value variances vs. corresponding means.

Figure 2.7 indicates that there are gene microarray values with high variances and negative means (implying that there exist important genes that are deleted). If we use gene with negative microarray values for LOOCV classification, then we arrive at the results shown in Figure 2.8. This figure illustrates that despite there being apparent information in the data (high variance), the classification ability

is poor. This means that the mutator/non-mutator classification is dependent on other factors or that the quality of our data is low.

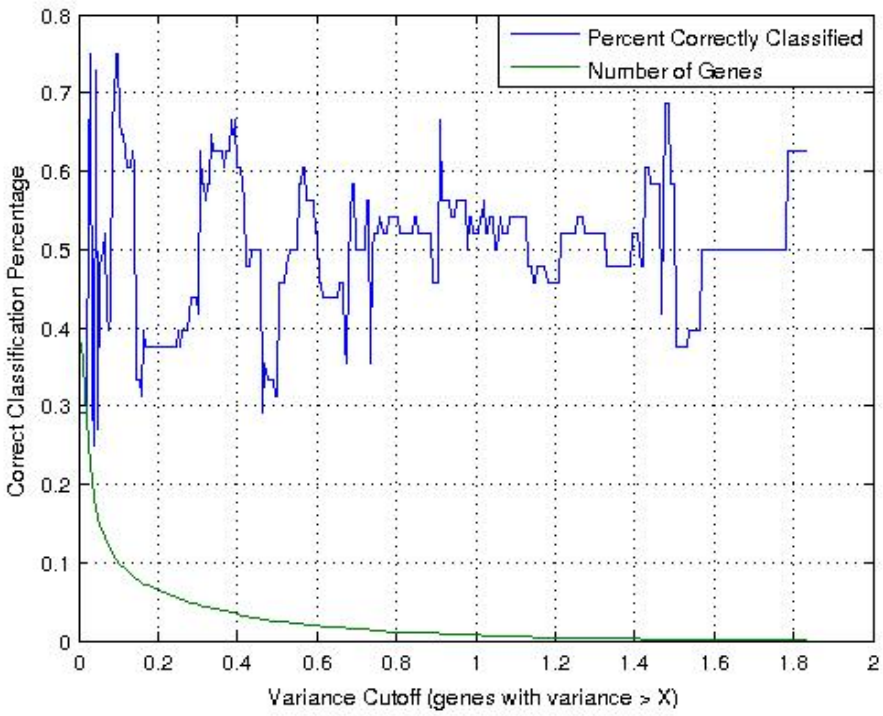

Figure 2.8: Correct classification percentage for the cases where we use only genes with microarray values having variances above a certain value and that also have negative means.

In working with microarray classification, dimension reduction is important for decreasing the computational load and for identifying genes carrying the most useful information for our purposes. Since microarrays generally have a large feature space (number of genes), the quality of the data is very important. With a slight increase in error, the information in the microarray becomes useless. These type of procedures can be used in other applications, wherever generating microarray data are easier than conducting a phenotype test (e.g. classifying cancer cells as being malignant or benign, determining disease status of cells, cellular microarray profiling and others).

For future use and testing of these methods, a data set that had repeated measurements would be very useful for generating an idea of the measurement error in the microarray data set. It could also be

important to take a gene-motivated approach. This approach would consist of identifying ahead of time genes thought to be responsible for controlling or regulating mutation rate, gene repair, or similar genes. These genes could be analyzed with respect to other genes to see if they are indeed affected differently.

Chapter 3

The Effect of Killer Virus on Competition in S. cerevisiae

3.1 Introduction

Saccaromyces cerevisiae (yeast), like all cellular species, is susceptible to predation, competition, and parasitism. Naturally-occurring yeasts frequently harbor so-called Killer viruses [1], double-stranded RNA (dsRNA) molecules that encode proteins that are toxic to non-killer strains, as well as proteins that confer toxin immunity [23]. In the simplest case, encounters between Killers and non-Killers have only two results: mating or death. Like other fungal viruses, yeast Killer viruses can only be transmitted by mating, and propagated only via replication of an infected and immune population [3,23]. Killer viruses have been isolated from yeast that occupy a wide range of habitats across the Earth, including hosts *S. cerevisiae, S. pombe, C. glabrata*, and *P. pastoris*, found in clinical, fermentation, and food isolates.[3, 12, 25, 30]. The wide, but not universal, occurrence of this phenomenon suggests there are a restricted set of ecological conditions that select for the killer phenotype.

The mechanisms of killing and immunity are best understood in *S. cerevisiae* and killer viruses K-1,

K-2, and K-28 [3]. Infection among killer yeast is caused by a member of the cytoplasmic-persisting family Totiviridae. In *Saccharomyces cerevisiae* there are three major killer viruses (ScV-M1, ScV-M2, and ScV-M28) [4], each of which encodes for a specific killer protein (K1, K2 and K28 respectively). Killer virus infection requires the presence of two elements: K-dsRNA and L-dsRNA [39]. K-dsRNA encodes for the killer protein and for cell immunity, while L-dsRNA encodes for the export of the killer protein to the extracellular space [20]. Both must be present for a yeast to be a "killer" [4]. Once the killer protein enters the extra-cellular space, it can not bind to a killer yeast cell, due to the immunity conferred. If, however, the killer protein binds to a non-killer cell, the protein disrupts the cell wall potential (ion leakage) which results in cell death [2].

The killer virus is only transmitted sexually in yeast [5]. When diploid yeast is starved for nitrogen in the absence of a fermentable carbon source, they typically undergo meiosis and produce 4 haploid spores, 2 of each mating type. When infected haploid cells encounter uninfected cells of opposite mating type, transmission of the killer trait follows syngamy and restoration of the diploid states. Diploids may then reproduce mitotically (asexually) or, if poor growth conditions ensue, undergo another round of sporulation and meiosis. The benefits for being a killer are straightforward: in any interaction between killer yeast and sensitive yeast, the killer yeast can outcompete the sensitive yeast in a batch culture by production of killer pre-protoxin. However, if a killer yeast competes with a non-killer resistant yeast, the competitive outcome is not immediately clear [18]. Because the killer yeast expending nutrients and energy towards maintaining the killer virus and killer protein, a resistant yeast may be able to out compete the killer yeast. This chapter studies models of infection in chemostat conditions and dependence of possible biologically meaningful steady states of these models on system parameters. Growth rate is a quantitative measure of yeast cell fitness. We will use the growth rate data of infected and uninfected yeast strains to determine and compare numerical values of quantitative measures of fitness. Our hypothesis is that the fitness of the yeast cell harboring the both K and L dsRNA viruses is less than that of a partially uninfected yeast cell harboring only the L ds-RNA virus. Travisano [18] demonstrates that invasion of the killer yeast is not possible under certain conditions. Analysis of models performed in this paper allows one to specify the conditions on

the system parameters for which washout of the killer yeast is possible.

Below, the following models will be discussed:

- First we analyze the resistant yeast and killer yeast system with no spread of infection.
- Then we study a killer system with susceptible yeast, with no spread of infection.
- Lastly, we address the model with a nonresistant yeast in which infection can occur through sporulation.

3.1.1 Resistant Yeast and Killer Yeast System

Another model to consider is case where uninfected yeast is resistant to "killer" toxin. We arrive at a "pure competition model": the interaction between two types of yeast only occurs through competition for the nutrients. We consider ethanol production conditions in a chemostat where there will be no sporulation, and no transmission of the virus. Model assumptions are listed below:

(a) strains interactions may be described using the law of mass action;

(b) the constant inflow rate equals the outflow rate;

(c) no killing of uninfected yeast by killer toxin and no transmission of virus.

(d) growth rates of the infected and uninfected strains are different.

Under these assumptions, the model has the following form:

$$\begin{aligned} \frac{dG}{dt} &= \frac{F}{V}G_0 - \frac{F}{V}G - \frac{\mu_1}{\gamma_1}YG - \frac{\mu_2}{\gamma_2}IG, \\ \frac{dY}{dt} &= \mu_1 YG - \frac{F}{V}Y, \\ \frac{dI}{dt} &= \mu_2 IG - \frac{F}{V}I. \end{aligned} \tag{3.1}$$

We use the following notation:

G - glucose concentration [mol / liter],

Y - uninfected yeast concentration [cells/liter],

I - infected yeast concentration [cells/liter],

F - flow rate [liter/hour],

V - volume of vessel [liter],

μ_1 - uninfected yeast growth rate constant [liter/hour per mol glucose consumed],

μ_2 - infected yeast growth rate constant [liter/hour per mol glucose consumed],

γ_1 - yield of uninfected yeast from glucose consumption [number of uninfected yeast cells per mol glucose cons

γ_2 - yield of infected yeast from glucose consumption [number of infected yeast cells per mol glucose consume

We rescale the system (1) by introducing new variables, $\widehat{G}$, $\widehat{Y}$, $\widehat{I}$, and $\widehat{t}$:

$$G = \alpha\widehat{G}, \quad Y = \beta\widehat{Y}, \quad I = \delta\widehat{I}, \quad t = \sigma\widehat{t}. \tag{3.2}$$

Here α, β, δ, and σ are constants (in appropriate units of measurement for the respective quantities). Substituting (3.2) into (3.1), we obtain

$$\begin{aligned}
\frac{\alpha}{\sigma}\frac{d\widehat{G}}{dt} &= \frac{F}{V}G_0 - \frac{F}{V}\widehat{G}\alpha - \frac{\mu_1}{\gamma_1}\beta\alpha\widehat{Y}\widehat{G} - \frac{\mu_2}{\gamma_2}\delta\alpha\widehat{I}\widehat{G}, \\
\frac{\beta}{\sigma}\frac{d\widehat{Y}}{dt} &= \mu_1\beta\alpha\widehat{Y}\widehat{G} - \frac{F}{V}\beta\widehat{Y}, \\
\frac{\delta}{\sigma}\frac{d\widehat{I}}{dt} &= \mu_2\delta\alpha\widehat{I}\widehat{G} - \frac{F}{V}\delta\widehat{I}.
\end{aligned} \tag{3.3}$$

We choose the proportionality (scaling) constants as follows:

$$\sigma = \frac{V}{F}, \quad \beta = \frac{\gamma_1}{\mu_1}\frac{F}{V}, \quad \delta = \frac{\gamma_2}{\mu_2}\frac{F}{V} \quad \alpha = G_0.$$

.

To make notations shorter, we omit hats (⌢), and write the nondimensionalized system as follows:

$$\begin{aligned} \frac{dG}{dt} &= 1 - G - YG - IG, \\ \frac{dY}{dt} &= \alpha_1 YG - Y, \\ \frac{dI}{dt} &= \alpha_2 IG - I. \end{aligned} \tag{3.4}$$

Here $\alpha_1 = \mu_1 G_0 \frac{V}{F}$, and $\alpha_2 = \mu_2 G_0 \frac{V}{F}$.

Now, G, Y, and I are the nondimensionalized concentrations of nutrient, yeast, and infected yeast, respectively, in the chemostat. To find the steady states (populations that do not change over time), we set the derivatives in system (3.4) to zero. So, we now have to solve the system:

$$\begin{aligned} 0 &= 1 - \bar{G} - \bar{Y}\bar{G} - \bar{I}\bar{G}, \\ 0 &= \alpha_1 \bar{Y}\bar{G} - \bar{Y}, \\ 0 &= \alpha_2 \bar{I}\bar{G} - \bar{I}, \end{aligned} \tag{3.5}$$

Below we present the steady states, which are solutions of (3.5). We start with the trivial steady state:

$$(\overline{G_1}, \overline{Y_1}, \overline{I_1}) = (1, 0, 0). \tag{3.6}$$

There are no infected or uninfected yeast populations in the chemostat.

Next, we present nontrivial steady states. We only get steady states if either Y or I is zero. It can be shown that since $\mu_1 \neq \mu_2$ (and thus $\alpha_1 \neq \alpha_2$), the two yeast strains cannot coexist.

$$(\overline{G_2}, \overline{Y_2}, \overline{I_2}) = (\frac{1}{\alpha_1}, \alpha_1 - 1, 0). \tag{3.7}$$

For this steady state (3.7) to exist, we must have that $\overline{Y_2} > 0$. If that is true, then the following condition must be satisfied: $\alpha_1 > 1$, or $\mu_1 G_0 > \frac{F}{V}$. In other words, the characteristic growth rate must be greater compared to the characteristic rate of emptying.

We also have another steady state:

$$(\overline{G_3}, \overline{Y_3}, \overline{I_3}) = (\frac{1}{\alpha_2}, 0, \alpha_2 - 1). \tag{3.8}$$

For this steady state (3.8) $\overline{I_3} > 0$ to exist, we must have that $\alpha_2 > 1$, or $\mu_2 G_0 > F/V$. This also says that the characteristic growth rate must be greater than the characteristic emptying rate.

When both conditions, $\mu_1/G_0 > F/V$ and $\mu_2/G_0 > F/V$, are satisfied, we observe a principle of competitive exclusion scenario, where the outcome depends on the initial populations of the two strains. It is important to note that the steady state where all three populations exist, only occurs when $\mu_1 = \mu_2$ and $\gamma_1 = \gamma_2$. If μ_1/γ_1 is bigger than μ_2/γ_2, then the resistant yeast will always win in competition, and this last steady state will not exist. A similar situation happens if μ_1/γ_1 is less than μ_2/γ_2. For fixed values of $G_0 = 1$, $V = 1$, $F = 1$, $\mu_1 = \mu_2 = 1$, and $\gamma_1 = \gamma_2 = 1$, we numerically estimated the domains of attraction for the two strains. The results are shown in Figure 3.1.

Figure 3.1 illustrates an unstable focal point. Letting $\alpha_1 = \alpha_2 = \alpha$, we can find this steady state by letting $U = Y + I$. We then arrive at the system,

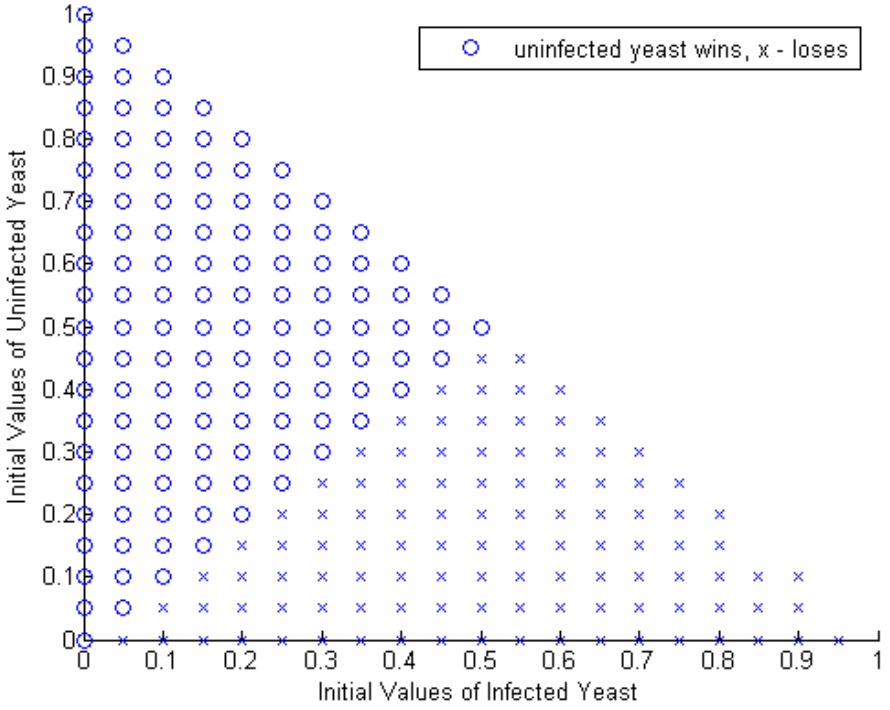

Figure 3.1: Graph illustrating the competitive exclusion principle: initial values belonging to the domain of attraction of the steady state with $\overline{Y_2} \neq 0$, $\overline{I_2} = 0$ are marked be "o", and those belonging to the domain of attraction of the steady state with $\overline{I_3} \neq 0$, $\overline{Y_3} = 0$ are marked by "x". The parameter values used here were $\mu_1 = \mu_2 = 1$, $\gamma_1 = \gamma_2 = 1$, $G_0 = 0.02$, and $F/V = 1$.

$$\begin{aligned} \frac{dG}{dt} &= 1 - G - GU, \\ \frac{dU}{dt} &= \alpha GU - U. \end{aligned} \tag{3.9}$$

The non-zero steady state of System 3.9 is $(\bar{G}, \bar{U}) = (1/\alpha, \alpha - 1)$. To determine the stability of this steady state, we compute the determinant and trace of the Jacobian,

$$J(\bar{G}, \bar{U}) = \begin{pmatrix} -\alpha & -\frac{1}{\alpha} \\ \alpha(\alpha - 1(& 0 \end{pmatrix}. \tag{3.10}$$

This Jacobian has trace equal to $-\alpha$ and determinant equal to $\alpha - 1$. This means this steady state is unstable when $\alpha - 1 = \mu G_0(F/V) - 1 < 0$, or $\mu G_0 < F/V$.

Also, for the same fixed parameter values, and $\mu_1 > \mu_2$, we can also theoretically washout both strains of yeast entirely from the chemostat. This possibility is illustrated in a bifurcation diagram of the flow rate per volume (Figure 3.2).

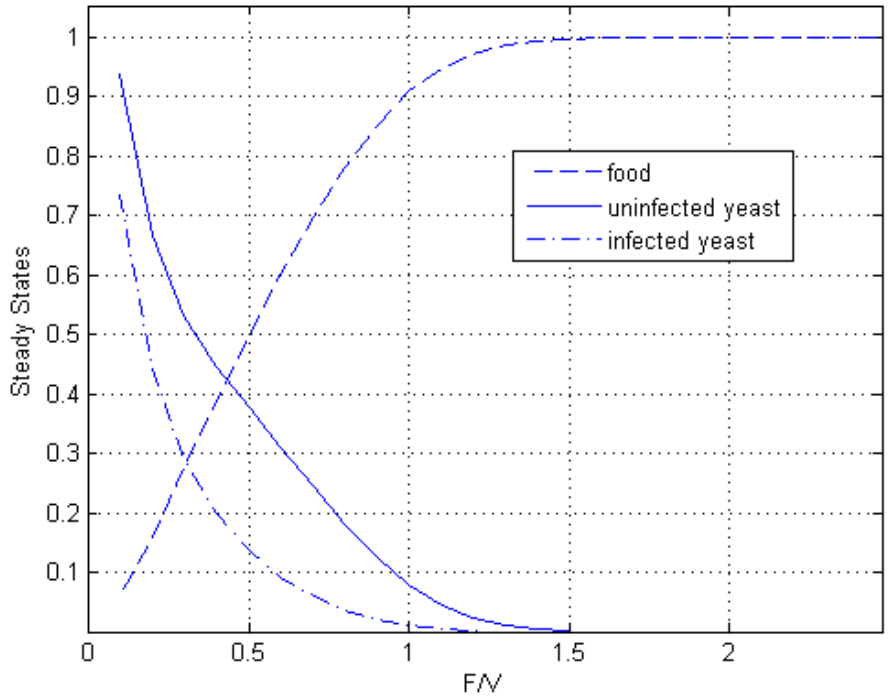

Figure 3.2: Graph showing the possibility of two strains washout. The parameter values used here were $\mu_1 = \mu_2 = 1$, $\gamma_1 = \gamma_2 = 1$, and $G_0 = 0.02$.

3.2 Sensitive Yeast and Killer Yeast Model

For this model, we consider chemostat conditions, where there is no sporulation. If there is no sporulation, there will be no transmission of the virus. In addition to competition for food, the two strains will also interact with each other (i.e., the killer yeast will produce the toxin that will affect the sensitive

yeast). The killing parameter to model can also vary widely, possibly simulating the interaction between killer yeast and partially sensitive yeast. For example, the sensitive yeast can become resistant or partially resistant to the killer toxin by changing the membrane receptors on the cell wall to be not as numerous or not as easy for the killer protein to bind to. We formulate assumptions for this model:

(a) interactions are described using the law of mass action.

(b) the constant inflow rate equals the outflow rate.

(c) only killing of uninfected yeast by toxin produced by killer yeast will occur.

(d) growth rates of the infected and uninfected strains are different.

A "killing" term appears in the uninfected yeast equation. It will depend on the concentrations of both strains, so the interaction will be described by the law of mass action. (Here we also assume that the amount of toxin produced is proportional to the population of infected yeast). The model has the form:

$$\begin{aligned} \frac{dG}{dt} &= \frac{F}{V}G_0 - \frac{F}{V}G - \frac{\mu_1}{\gamma_1}YG - \frac{\mu_2}{\gamma_2}IG, \\ \frac{dY}{dt} &= \mu_1 YG - kYI - \frac{F}{V}Y, \\ \frac{dI}{dt} &= \mu_2 IG - \frac{F}{V}I. \end{aligned} \tag{3.11}$$

Here we use the notation:

G - glucose concentration [mol / liter],

Y - uninfected yeast concentration [cells/liter],

I - infected yeast concentration [cells/liter],

k - death rate of uninfected yeast from toxin produced by infected yeast [liter^2/hour],

F - flow rate [liter/hour],

V - volume of vessel [liter],

μ_1 - uninfected yeast growth rate constant [liter/hour per mol glucose consumed],

μ_2 - infected yeast growth rate constant [liter/hour per mol glucose consumed],

γ_1 - yield of uninfected yeast from glucose consumption [number of uninfected yeast cells per mol glucose cons

γ_2 - yield of infected yeast from glucose consumption [number of infected yeast cells per mol glucose consume

We rescale the system (3.11), by introducing new variables, $\widehat{G}$, $\widehat{Y}$, $\widehat{I}$, and $\widehat{t}$:

$$G = \alpha\widehat{G}, \quad Y = \beta\widehat{Y}, \quad I = \delta\widehat{I}, \quad t = \sigma\widehat{t}. \tag{3.12}$$

Here α, β, δ, and σ are constants. Substituting (3.12) into (3.11):

$$\begin{aligned}
\frac{\alpha}{\sigma}\frac{d\widehat{G}}{dt} &= \frac{F}{V}G_0 - \frac{F}{V}\widehat{G}\alpha - \frac{\mu_1}{\gamma_1}\beta\alpha\widehat{Y}\widehat{G} - \frac{\mu_2}{\gamma_2}\delta\alpha\widehat{I}\widehat{G}, \\
\frac{\beta}{\sigma}\frac{d\widehat{Y}}{dt} &= \mu_1\beta\alpha\widehat{Y}\widehat{G} - k\beta\delta\widehat{Y}\widehat{I} - \frac{F}{V}\beta\widehat{Y}, \\
\frac{\delta}{\sigma}\frac{d\widehat{I}}{dt} &= \mu_2\delta\alpha\widehat{I}\widehat{G} - \frac{F}{V}\delta\widehat{I}.
\end{aligned} \tag{3.13}$$

We choose the proportionality (scaling) constants as follows:

$$\sigma = \frac{V}{F}, \quad \beta = \frac{\gamma_1}{\mu_1}\frac{F}{V}, \quad \delta = \frac{\gamma_2}{\mu_2}\frac{F}{V}, \quad \alpha = \frac{F}{V}\frac{1}{\mu_1}.$$

Now our system becomes:

$$\begin{aligned} \frac{dG}{dt} &= \alpha_1 - G - YG - IG, \\ \frac{dY}{dt} &= YG - \alpha_2 YI - Y, \\ \frac{dI}{dt} &= \alpha_3 IG - I. \end{aligned} \tag{3.14}$$

Here $\alpha_1 = \dfrac{\mu_1 G_0 V}{F}$, $\alpha_2 = k\dfrac{\gamma_2}{\mu_2}$, and $\alpha_3 = \dfrac{\mu_2}{\mu_1}$.

System (3.14) the nondimensionalized equations for the concentration of nutrient, uninfected yeast, and infected yeast in the chemostat. Again, we look for the steady states. We must solve equations that we obtain from setting derivatives of system (3.14) to zero:

$$\begin{aligned} 0 &= \alpha_1 - \bar{G} - \bar{Y}\bar{G} - \bar{I}\bar{G}, \\ 0 &= \bar{Y}\bar{G} - \alpha_2 \bar{Y}\bar{I} - \bar{Y}, \\ 0 &= \alpha_3 \bar{I}\bar{G} - \bar{I}. \end{aligned} \tag{3.15}$$

We obtain four possible steady states. After changing back to original units, we have a trivial steady state:

$$(\overline{G_1}, \overline{Y_1}, \overline{I_1}) = (G_0, 0, 0). \tag{3.16}$$

We also get the following steady state:

$$(\overline{G_2}, \overline{Y_2}, \overline{I_2}) = (\mu_1 G_0 \frac{V}{F}, \gamma_1 G_0 - \frac{\gamma_1}{\mu_1}\frac{F}{V}, 0). \tag{3.17}$$

With $\overline{Y} > 0, \quad \overline{I} = 0$. For existence of the second steady state, the condition $\mu_1 > \frac{F}{V}G_0$ must hold.

We also get a steady state with $\overline{I} > 0$ and $\overline{Y} = 0$:

$$(\overline{G_3}, \overline{Y_3}, \overline{I_3}) = (\mu_2 G_0 \frac{V}{F}, 0, \gamma_2 G_0 - \frac{\gamma_2}{\mu_2}\frac{F}{V}). \tag{3.18}$$

For the third steady state to exist, it can be shown the condition $\mu_2 > \frac{F}{V}G_0$ must hold.

Finally, for coexistence, we want $\overline{Y} > 0$ and $\overline{I} > 0$, the corresponding steady state is

$$(\overline{G_4}, \overline{Y_4}, \overline{I_4}) = (\frac{1}{\mu_2}\frac{F}{V}, \overline{Y_4}, \frac{F}{V}\frac{1}{k}\left(\frac{\mu_1}{\mu_2} - 1\right). \tag{3.19}$$

Here,

$$\overline{Y_4} = \left(\frac{\mu_2}{\mu_1}\gamma_1 G_0 - \frac{F}{V}\frac{\gamma_1}{\mu_1}\right) - \left(\frac{F}{V}\frac{\gamma_1}{\gamma_2}\frac{1}{k\mu_1}(\mu_1 - \mu_2)\right).$$

For existence of the fourth steady state, we have the condition:

$$\mu_2 G_0 > \frac{F}{V} + \frac{F}{V}\frac{1}{k\gamma_2}(\mu_1 - \mu_2). \tag{3.20}$$

To determine the stability properties of the steady states, we consider the Jacobian of the system (3.11):

$$J(G, Y, I) = \begin{pmatrix} -\frac{F}{V} - \frac{\mu_1}{\gamma_1}Y - \frac{\mu_2}{\gamma_2}I & -\frac{\mu_1}{\gamma_1}G & -\frac{\mu_2}{\gamma_2}G \\ \mu_1 Y & \mu_1 G_0 - kI - \frac{F}{V} & -kY \\ \mu_2 I & 0 & \mu_2 G_0 - \frac{F}{V} \end{pmatrix}. \tag{3.21}$$

Evaluating the Jacobian (3.22) at the first steady state, we obtain

$$J(G_0,0,0)=\begin{pmatrix} -\frac{F}{V} & -\frac{\mu_1}{\gamma_1}G_0 & -\frac{\mu_2}{\gamma_2}G_0 \\ 0 & \mu_1 G_0-\frac{F}{V} & 0 \\ 0 & 0 & \mu_2 G_0-\frac{F}{V} \end{pmatrix}. \tag{3.22}$$

Since the matrix is upper-triangular, the resulting real eigenvalues are

$$\lambda_1=-\frac{F}{V},\ \lambda_2=\mu_1 G_0-\frac{F}{V},\ \lambda_3=\mu_2 G_0-\frac{F}{V}.$$

It can be easily seen from conditions on the existence of the second and third steady states that the first steady state becomes unstable if any other steady state exists. If no other steady state exists, then the first steady state is stable.

Next, consider the Jacobian evaluated at the second steady state,

$$J(\frac{1}{\mu_1}\frac{F}{V},\gamma_1(G_0-\frac{F}{V}\frac{1}{\mu_1}),0)=\begin{pmatrix} -\mu_1 G_0 & -\frac{1}{\gamma_1}\frac{F}{V} & -\frac{\mu_2}{\gamma_2}\frac{1}{\gamma_1}\frac{F}{V} \\ \mu_1\gamma_1 G_0-\gamma_1\frac{F}{V} & \mu_1\gamma_1 G_0-\gamma_1\frac{F}{V} & -k\gamma_1 G_0+k\frac{\gamma_1}{\mu_1}\frac{F}{V} \\ 0 & 0 & \frac{\mu_2}{\mu_1}\frac{F}{V}-\frac{F}{V} \end{pmatrix}. \tag{3.23}$$

The corresponding eigenvalues of the Jacobian (3.23) are

$$\lambda_1=-\frac{F}{V},\ \lambda_2=-\left(\mu_1 G_0-\frac{F}{V}\right),\ \lambda_3=\frac{F}{V}\left(\frac{\mu_2}{\mu_1}-1\right).$$

All these eigenvalues are negative, according to the conditions for the second steady state to exist. This implies that this steady state is always stable when it exists.

We now consider the Jacobian evaluated at the third steady state.

$$J(\frac{1}{\mu_2}\frac{F}{V}, 0, \gamma_2(G_0 - \frac{F}{V}\frac{1}{\mu_2})) = \begin{pmatrix} -\mu_1 G_0 & -\frac{\mu_1}{\mu_2}\frac{1}{\gamma_1}\frac{F}{V} & -\frac{1}{\gamma_2}\frac{F}{V} \\ 0 & \frac{\mu_1}{\mu_2}\frac{F}{V} - k\gamma_2 G_0 - \frac{\gamma_2}{\mu_2}\frac{F}{V} - \frac{F}{V} & 0 \\ \gamma_2\left(\mu_2 G_0 - \frac{F}{V}\right) & 0 & 0 \end{pmatrix}. \tag{3.24}$$

The eigenvalues are

$$\lambda_{1,2} = -\frac{1}{2}\mu_1 G_0 \pm \frac{1}{2}\sqrt{(\mu_1 G_0)^2 - 4\frac{F}{V}\left(\mu_2 G_0 - \frac{F}{V}\right)},$$

$$\lambda_3 = \frac{1}{\mu_2}\frac{F}{V}(\mu_1 - \mu_2) - \gamma_2\left(kG_0 - \frac{F}{V}\frac{1}{\mu_2}\right).$$

Here we can see that $\lambda_3 > 0$ for some small value of k and $\mu_1 > \mu_2$. If k is larger than $(F/V)(1/G_0\mu_2)$, then $\lambda_3 > 0$, making the steady state stable. It is worth noting that if we change the flow rate above a certain point, making $(\mu_2 G_0 - F/V)$ negative, the steady state will no longer exist.

We now consider the fourth steady state in the Jacobian,

$$\begin{aligned} J\left(\frac{1}{\mu_2}\frac{F}{V}, \overline{Y_4}, \frac{F}{V}\frac{1}{k}(\mu_1/\mu_2 - 1)\right) &= \begin{pmatrix} -\frac{F}{V} - \frac{\mu_1}{\gamma_1}\overline{Y_4} - \frac{1}{k\gamma_2}(\mu_1 - \mu_2) & -\frac{\mu_1 F}{\gamma_1\mu_2 V} & -\frac{1}{\gamma_2}\frac{F}{V} \\ \mu_1\overline{Y_4} & 0 & -k\overline{Y_4} \\ \frac{1}{k}\frac{F}{V}(\mu_1 - \mu_2) & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} a & b & c \\ d & 0 & e \\ f & 0 & 0 \end{pmatrix}. \end{aligned} \tag{3.25}$$

We write the characteristic equation in the form:

$$\text{Det}(J - \lambda I) = \lambda^3 + \lambda^2(-a) + \lambda(-cf - bd) + (-bef) = 0. \tag{3.26}$$

Here a, b, c, d, e, and f are given by the previous matrix. It can be shown that for the characteristic equation (3.26), the Routh-Hurwitz Criteria does not hold:

$$(-bef) = -\left(\frac{\mu_1 F}{\gamma_1 \mu_2 V}\right)\left(\frac{F}{V}(\mu_1 - \mu_2)\overline{Y_4}\right) < 0.$$

Thus, when this non-trivial steady state exists, it is unstable. In order to the fourth steady state to exist, we must have the following condition, following from (Eq. 3.20):

$$\mu_2 G_0 > \frac{F}{V}\left(1 + \frac{1}{\gamma_2}(\mu_1 - \mu_2)\right). \tag{3.27}$$

The boundary of the region in the parameter plane $(F/V, k)$, where all four steady states may be observed, is given by the equality of the condition (3.27),

$$\mu_2 G_0 - \frac{F}{V}\left(1 + \frac{1}{\gamma_2}(\mu_1 - \mu_2)\right) = 0. \tag{3.28}$$

Solving (3.28) for k as a function of F/V, we arrive at the equation,

$$k = \frac{\frac{\mu_2}{\gamma_2}\left(\frac{\mu_1}{\mu_2} - 1\right)\frac{F}{V}}{\mu_2 G_0 - \frac{F}{V}}. \tag{3.29}$$

A graph of Equation (3.29) is seen in Figure 3.3. The data used for this graph come from the results section.

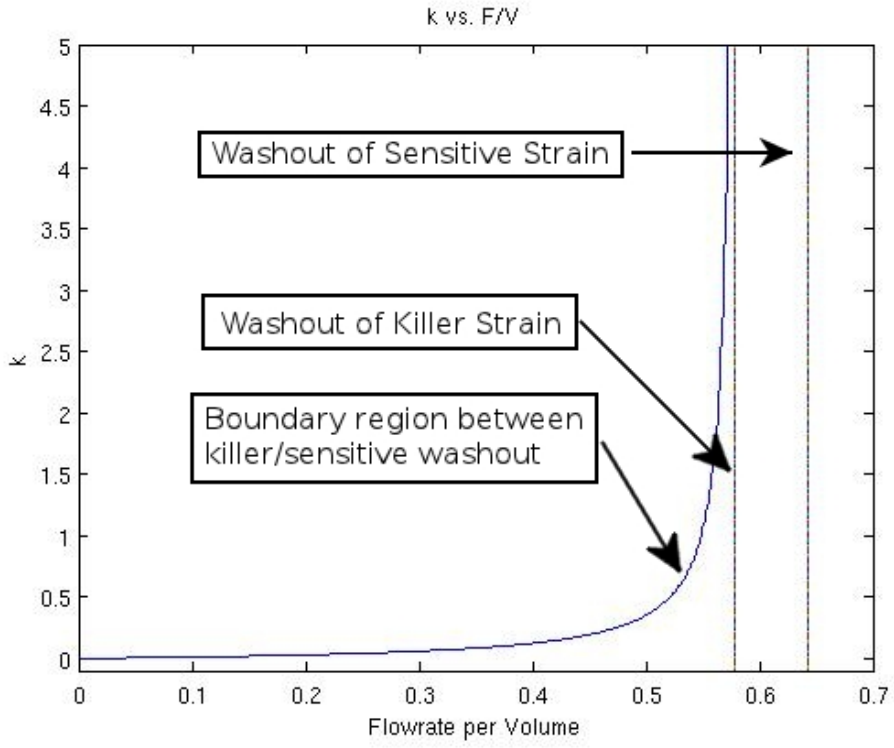

Figure 3.3: A graph of Equation (3.29), with constants determined from the results (Table 3.3).

3.3 Materials and Methods

Yeast strains and medium. The diploid *Saccharomyces cerevisiae* K1 killer yeast used in this experiment was T158c/S14a (ATCC 46427) [23]. The wild-type strain was *Saccharomyces cerevisiae* (ATCC 28684). The sensitive strain was a cycloheximide-cured killer strain as described below.

Batch and Solid Media: Growth medium was YEPD, 1% yeast extract, 2% peptone, and 2% glucose [23]. YPD agar was used for cultivation of the yeast, 1% yeast extract, 2% peptone, and 1% glucose. MB Agar was used [43], which was YPD agar with 0.6% methylene blue buffered to pH 4.7 with a sodium phosphate buffer, which was determined to be the ideal pH for killer protein activity [42]. CYH Agar was made with a sub-lethal concentration of cycloheximide of $400\mu g/mL$ mixed in with YPD agar [15].

Chemostat Media: Chemostat media was prepared to have a final concentration of 2% glucose [45].

Curing of killer strains. The K1 and K2 strains were seeded onto the CYH agar, and incubated for four days at 30°C, then plated onto YPD agar plates for cultivation and phenotypic/genotypic tests for killer activity [31].

In order to determine the impact of the killer virus on the fitness of yeast, we created a genotypically similar sensitive yeast from the killer yeast (K1-T158c/S14a). To pick out which strains to test for genotypic loss of the dsRNA (double stranded RNA) K1 virus, the sensitive lawns on the MB (Methylene-Blue) plates, that were streaked with 12 cycloheximide strains (1-12), we used the plating assessment to visually determine the four strains that had the least killing activity. Strains NMCY-2, NMCY-4, NMCY-8, and NMCY-11 were picked to be genetically tested for sensitivity. To genotypically asses that the K-dsRNA was no longer present, RT-PCR (Reverse Transcriptase Polymerase Chain Reaction) was run (see Figure 5).

Phenotypic assay of killer activity. A plate assay on MB agar was used for a phenotypic test [12]. The possible cured strains, K1 yeast, K2 yeast, and a wild type yeast were spotted onto a lawn of sensitive *S. cerevisiae* grown on MB plates, and incubated at 20°C for four days. Strains that retained the killer activity appeared to have a "kill-zone" of clear dead cells around them, whereas the cured strains did not have such a zone [2].

Genotypic assay of killer activity. Total RNA was prepared for RT-PCR amplification [16], then concentrated with ethanol precipitation [13]. Primers were designed based on the known K-dsRNA sequence [39]. This sequence (see Table 3.1) amplified a 900bp sequence of the K1 viral genome. RT-PCR was then used to screen clones for the presence/absence of the K1 killer virus [38]. Then the dsRNA was prepared from the cells that were washed with 50 mM Na_2EDTA, incubated for 15 minutes in 50 mM Tris-H_2SO_4 containing 2.5% 2-mercaptoethanol, and then mixed for one hour with 0.1 M NaCl/10 mM Tris-HCI, pH 7.5/10 mM Na_2EDTA/0.2% sodium dodecyl sulfate and redistilled phenol. Nucleic acid was recovered by ethanol precipitation, dissolved in 2.0 M LiCl/0.15 M NaCI/0.015 M

Na_3 citrate and incubated at 8 hr at 4C. The LiCl was removed with centrifugation and the nucleic acid was recovered with ethanol [38].

Electrophoresis RT-PCR products were electrophoresed in a 2% TAE agarox gel for 45 minutes at 75 volts, then visualized under UV illumination after staining with ethidium bromide. [33]. From the RT-PCR we can determine which of these strains have lost the dsRNA, and we can take them and do an accelerostat experiment to determine the growth rates.

Growth rate constant determination. To assess strain fitness we evaluated growth rate and yield. To estimate these parameters, we cultured yeast in an accelerostat, a continuous culture system where cell growth is evaluated at multiple steady states that arise in response to incremental increases in dilution rate of nutrient-limiting media. The strains $K1^+R^+$ and $K1^-R^-$strains were run as monocultures, with only the flow rate varied. Accelerostats were fed with nitrogen-limited, glucose sufficient Delft defined media amended with minerals and vitamins added, but no amino acids (glucose concentration of 1%). The flow rate of the vessels was changed from 0.05 vessel volumes per hour and increased by 0.05 vessel volumes per hour after a steady state had been achieved, until a washout occurred. From these data, it is possible to determine the growth rate constants and yields of the strains.

Cell enumeration and biomass determination. Cell counts were estimated using optical density (OD) counts at 600nm. Cell viability counts were also done, at various dilution rates (200X or 100X, determined by OD) as needed onto YPD agar plates.

Linear Regression Data were fitted to a linear regression line. Data were entered into MatLab, and fit to a regression line. Then a 95% confidence region, consisting of sample regression lines obtained for parameter values lying within 2 standard deviations from the least squares determined set of parameters was plotted around the line.

Integration of KanMX-ABr plasmid. To test predictions of competition in the model, we integrated the pFA6-kanMX plasmid into the flo8 gene as described by Geitz and Schiestl [17]. Primers used for integration of antibiotic cassette are described in Table 3.1. The resulting strains, are described in

Table 3.2.

K-dsRNA Primer		Amplicon Size
Forward Primer	3'-AAGCCAACCCAAGAAGAGACC-5'	0.9 - Kilobases
Reverse Primer	3'-TGTGTCCGCCTTCGTGTTTT-5'	

Table 3.1: The two primers used to amplify the K-dsRNA plasmid for the K1 strain killer verification.

Strains	Genotype	Phenotype	Provenance
S. cerevisiae	K-ABr-	K-, AB Sensitive	ATCC#28684
T158c/S14a	K+ABr-	K1-Killer, AB Sensitive	ATCC#46427
T158c/s14a sensitive	K-ABr-	K-, AB Sensitive	CYH-Treatment
Killer-KanMX	K+ABr+	K1-Killer, AB Resistant	kan-MX integration
Sensitive- KanMX	K-ABr+	K-, AB Resistant	kan-MX integration

Table 3.2: The five strains used in this experiment.

		Growth Rate	Yield
1st Run	K+R+	5.5521E+008	11.658E+008
	Sensitive	6.4267E+008	12.2274E+008
2nd Run	K+R+	5.1073E+008	12.1816E+008
	Sensitive	6.4276E+008	12.7098E+008
Combined	K+R+	5.7883E+008	11.4628E+008
	Sensitive	6.4214E+008	12.4708E+008
	Units	liter/hour per mol glucose	number of cells per mol glucose

Table 3.3: The growth rate constants and yields for the killer strain and the sensitive strain for both experiments and for all the data combined.

3.4 Results

Strain development. In order to compare the fitness effect of harboring the K1-killer virus, we need to have two strains with similar genomic backgrounds, one with the killer system and one without the killer system. To cure the killer virus, we subjected the killer strain to a sublethal dosage of cyclohexamide, which resulted in RNA plasmid free cells. To verify that the resulting strains are dsRNA free, we performed an RT-PCR analysis, screening for the 900bp K-dsRNA plasmid, as described before. A gel was run on all resulting strains to verify loss of killer K-plasmid (see Figure 3.4). Strains NMCY-2, NMCY-4, and NMCY-11 showed no 900bp K-dsRNA fragment.

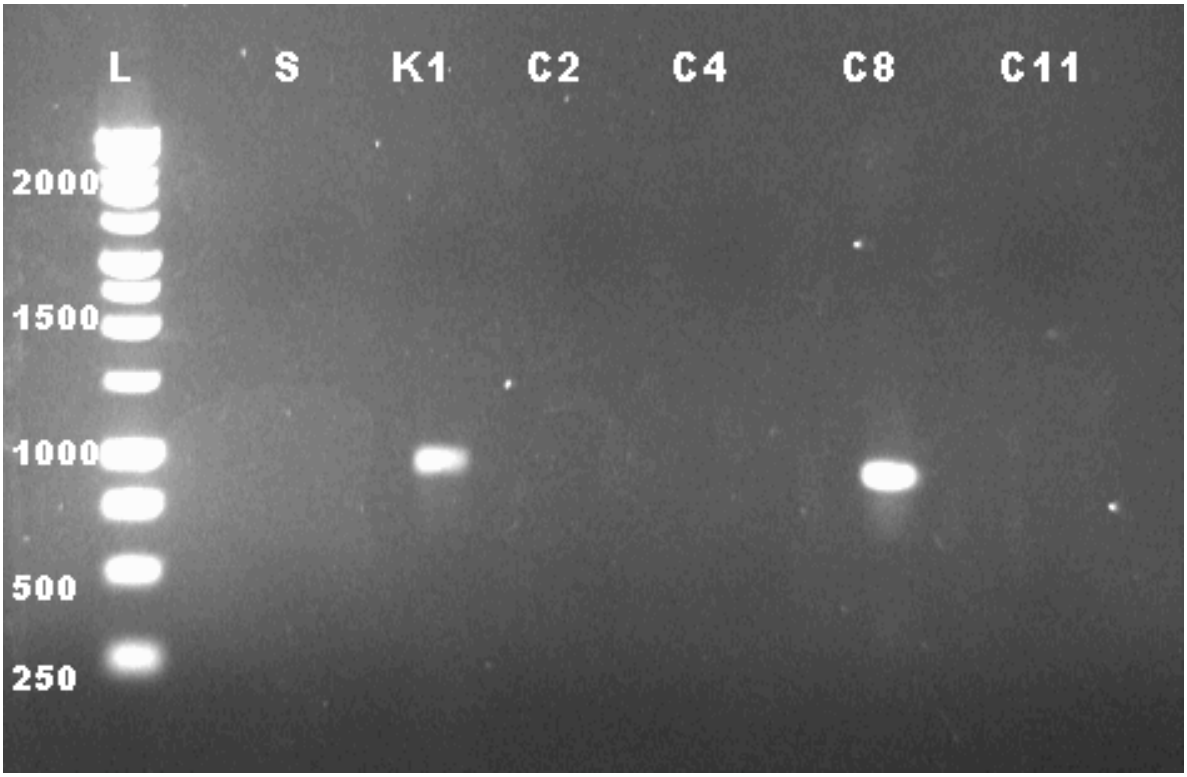

Figure 3.4: RT-PCR, using primers specific to a 900bp fragment of the K_1 ds-RNA virus of the 6 strains: S- Sensitive Non-Resistant Yeast (Negative control); K- K1 T158c/S14a Killer Yeast (Positive control); NMCY-2, NMCY-4, NMCY-8, NMCY-11 - CYH treated strains.

Accelerostat experiment. Running the accelerostat experiment for the T158c/S14a K1 strain, we get a graph of the optical density vs. the flow rate per volume. Two experiments were performed

and replicated. Regression lines were fit to the data (see Figure 3.5). Confidence bands were also constructed around the fitted regression line for the mean of the data. It has been shown that the cell count of *S. cerevisiae* increases linearly with optical density. We can convert these optical density values to solve for the growth rates and the yield. The slopes and intercepts were compared between the two replicate experiments are it was found that they were not significantly different ($p < 0.05$). The results were significantly different between the killer and sensitive ($p < 0.05$) (Table 3.3).

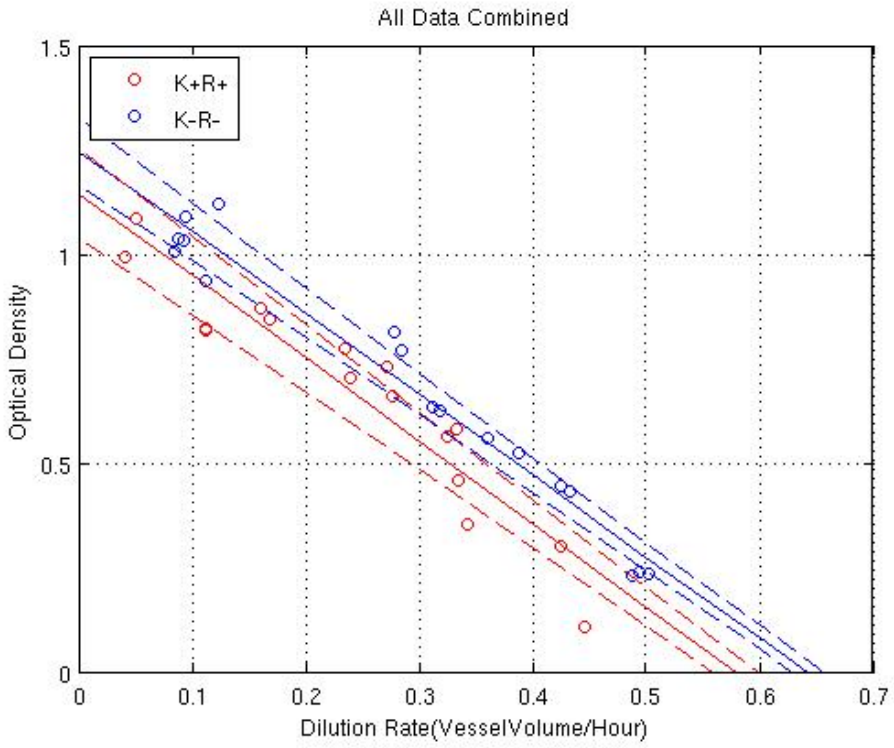

Figure 3.5: Optical Density of Killer (Red lines) and Sensitive (Blue lines) vs. Flow Rate (vessel volumes/hr), with a 95% confidence region plotted in dashed lines. The two lines have significantly different x-intercepts (washout values) ($p < 0.05$).

Growth Rate and Yield Determination. In order to determine the growth rate and yield, we must determine the steady state of a chemostat with one yeast. The steady states are determined from the system:

$$\begin{aligned} \frac{dG}{dt} &= \frac{F}{V}G_0 - \frac{F}{V}G - \frac{\mu_1}{\gamma_1}YG, \\ \frac{dY}{dt} &= \mu_1 YG - \frac{F}{V}Y. \end{aligned} \tag{3.30}$$

To find the steady states, we set the system in (3.30) equal to zero. From the second equation of (3.30) set to zero, we have:

$$0 = \mu_1 \overline{G} - \frac{F}{V},$$

or

$$\mu_1 \overline{G} = \frac{F}{V}. \tag{3.31}$$

Substituting equation (3.31) into the first equation of (3.30) in equilibrium, we obtain

$$\frac{F}{V}(G_0 - \overline{G} - \frac{1}{\gamma_1}\overline{Y}) = 0,$$

or

$$G_0 - \frac{F}{V\mu_1} - \frac{1}{\gamma_1}\overline{Y} = 0. \tag{3.32}$$

From (3.32), when $\overline{Y} = 0$ (uninfected yeast disappears), we obtain $\mu_1^* = F/(VG_0)$, which indicates the parameter value, such that for $\mu > \mu_1^*$, the uninfected yeast is going to be washed out from the system. Since we want to know how $\overline{Y}$ depends on F/V, we rewrite equation (3.32) as

$$\overline{Y} = \gamma_1 G_0 - \frac{\gamma_1}{\mu_1}\left(\frac{F}{V}\right). \tag{3.33}$$

Here, parameter F/V is in units of 1/hr and $\overline{Y}$ is number/vessel volume. One optical density unit of both yeast strains in YPD corresponds to approximately 1mg/mL of dry cell weight. For both yeast strains, 1mg of dry cells is about 1×10^7 cells per ml. Using this fact we can see that 1 unit of optical density corresponds to 2×10^9 cells/vessel volume.

From the RT-PCR results (see Figure 3.4), we see that strains C2, C4, and C11 appear to be K-dsRNA free after the cycloheximide treatment, we chose C2 and C11 to do a duplicate run from the RNA extraction. The accelerostat data for the cycloheximide treated yeast strains derived from the K1 (T158c/S14a) yeast strain is in Figure 3.6, while the accelerostat data for the two killer experiments is in Figure 3.7.

Now we fit the sensitive data to a line ($y = mx + b$) and find $m_s = -1.94207$ with standard error $\sigma_{m_s} = 0.17468$ and $b_s = 1.24708$ with standard error $\sigma_{b_s} = 0.05685$. Computing this for the killer data, we arrive at $m_k = -1.98034$ with standard error $\sigma_{m_k} = 0.32099$ and $b_k = 1.14628$ with standard error $\sigma_{b_k} = 0.08683$.

The above parameters were used in generating the competition-washout hyperbola seen in Figure 3.29. We can see that since the killer strain washes out at a flow rate per vessel volume of 0.555 vv/hr and the sensitive washes out at 0.643 vv/hr, any flowrate in between will washout the killer and not the sensitive strain. This implies that competition in a continuous culture at that flowrate, the sensitive strain can outcompete the killer strain.

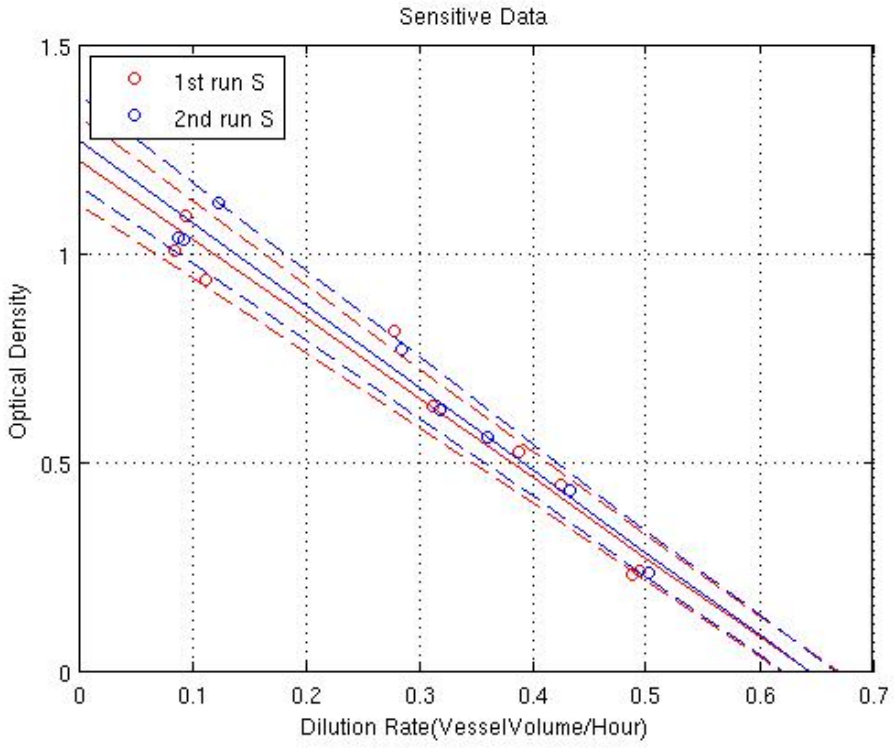

Figure 3.6: Optical Density of the two sensitive accelerostat experiments vs. Flow Rate (vessel volumes/hr) with 95% confidence lines (dashed). The two experiments were not significantly different ($p < 0.05$).

3.5 Discussion, Results, Future Directions and Pitfalls

When harboring a killer virus, yeast can kill competing sensitive yeast. The trade off is that they expend energy to produce killer toxin, when they could use it to maintain themselves or reproduce. Because of this, hosting an infection should result in a lower growth rate.

One of the interesting results following from the analysis is the prediction that, if only the killing of uninfected yeast is present and no transmission of the virus (no sporulation), then we can still drive out the infected yeast from the chemostat: increasing the flow rate puts more stress on reproduction enough to overcome the killing of the uninfected yeast.

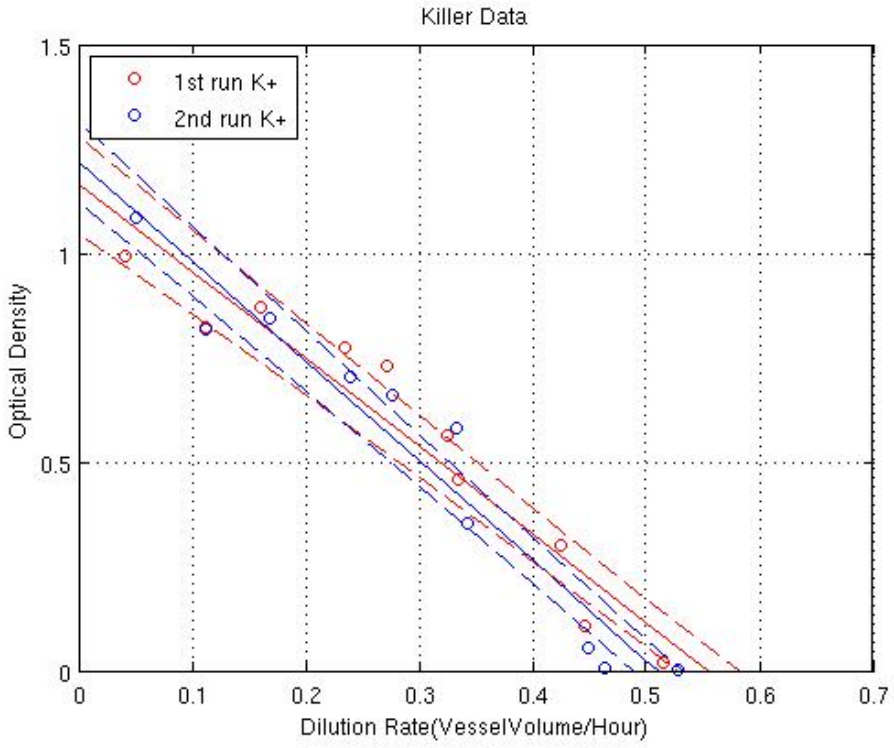

Figure 3.7: Optical Density of two experimental runs of Killer vs. Flow Rate (vessel volumes/hr), with a 95% confidence region plotted in dashed lines. The two lines are not significantly different ($p < 0.05$).

Although transmission of the virus is rare, as yeast only sporulate when nitrogen starved, sporulation promotes viral transmission. Since sporulation may occur more frequently in unfavorable environments, it increases the numbers of the infectious yeast, resulting in a persistence of the infection as long as any uninfected yeast are present with low enough flow rates. Still washout of the infected yeast can occur with a higher flow rate as has been shown.

From the results of the two accelerostat experiments, we can see that the two growth rate constants of the killer strain verses the sensitive strain are similar, the yields are considerably different. For the sensitive strain, the yield is over twice as much. It is quite clear that the killer virus prohibits the full use of the nutrient towards growth and hampers the cells ability to reproduce. An interesting prediction this data can make is the outcome of competition between sensitive and killer yeast. Looking at the

first system we proposed, we have an additional term of killing by the infected strain, with an unknown parameter, k. This term can be predicted by running competition experiments at different flow rates, and measuring the resulting ratio of infected verses uninfected yeast. Even when both killer and sensitive yeast strains are in the chemostat, there is a region in the k and (F/V) parameter space where flow rate is high enough to washout the infection if the infected (killer) yeast has a lower growth rate.

It is important to note that the models used here do not distinguish between relative growth rate, absolute growth rate and the death rate. To determine the growth rate constant separately from the death rate constant, it is important to take into account the self-toxicity of the killer virus. The killer virus assures replication faithfulness because of the killer protein in the extra-cellular space. If a daughter cell fails to acquire the killer virus, it is also not immune to the killing effect of the protein and will die with exposure to the killer protein. It would be interesting to take this further and measure the self-toxicity of the killer virus by live/dead cell staining, microscopy, or similar methods.

Bibliography

[1] I. Alves-Rodrigues. Saccharomyces cerevisiae: A useful model host to study fundamental biology of viral replication. *Research*, 120:49–56, 2006.

[2] F. Breinig. Dissecting toxin immunity in virus-infected killer yeast uncovers an intrinsic strategy of self-protection. *Angewandte Molekularbiologie*, pages 3810–3815, 2006.

[3] F. Breinig and M. Schmitt. The viral killer system in yeast: from milecular biology to application. *Fems Microbio*, 26:257–276, 2002.

[4] F. Breinig and M. Schmitt. Yeast viral killer toxins: Lethality and self-protection. *Nature*, 4:212–221, 2006.

[5] F. Brenig. Kre1p, the plasma membrane receptor for the yeast k1 viral toxin. *Cell.*, 108:395–405, 2002.

[6] Evans M.D. Dizdaroglu M. Lunec J. Cooke, M.S. Oxidative dna damage: Mechanisms, mutation, and disease. *FASEB Journal*, 17:1195–1214, 2003.

[7] Davis and Mingoli. Mutants of *Escherichia coli* equiring methionine or vitamin b12. *J. Bact.*, 60:17–28, 2008.

[8] Fry J.C. Dancer B.N. Dawson, S.L. A comparitive evaluation of five typing techniques for determining the diversity of flourescent pseudomonads. *J. Microbial Methods*, 50:9–22, 2002.

[9] Taylor J.W. Dettman, J.R. Mutation and evolution of microsatellite loci in neurospora. *Genetics*, 168(3):1231–1248, 2004.

[10] Charlesworth B. Charlesworth D. Crow J. Drake, J.W. Rates of spontaneous mutation. *Genetics*, 148:1667–1686, 1998.

[11] S. Dray. Co-inertia analysis and the linking of ecological data tables. *Ecology*, 84:3078–3089, 2003.

[12] A. Eiden-Plach. Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by *Candida glabrata, Pichia pastoris, Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*. *Microbio*, pages 961–966, 2003.

[13] L. Fujimura Esteban, R. Vega. and T. Launching of the yeast 20s rna narnavirus by expressing the genomic or antigenomic viral rna in vivo. *J. of Bio*, 280:33725–33734, 2005.

[14] Keightly P.D. Eyre-Walker, A. The distribution of fitness effects of new mutations. *Nat Rev Genet.*, 8(8):610–618, 2007.

[15] G. R. Fink and C. A. Styles. Curing of a killer factor in *Saccaromyces cerevisiae*. *Proc. Nat. Acad. Sci.*, 69(10):2846–2869, 1972.

[16] D. B. Franken. Genetic and fermentation properties of the k2 killer yeast, *Saccharomyces cerevisiae* t206. *Antonie Van Leewenhoek*, 73:263–369, 1998.

[17] Schiestl R. Gietz, R. Transforming yeast with dna. *Methods in Mol. Cell. Biol.*, 5:255–269, 1995.

[18] Travisano M. Greig, D. Density-dependent effects on allelopathic interactions in yeast. *Evolution*, 62(3):521–527, 2008.

[19] Maruyama T. Kimura, M. The mutational load with epistatic gene interactions in fitness. *Genetics*, 54:1337–1351, 1966.

[20] Teichert S. Meinhardt F. Klassen, R. Novel yeast killer toxins provoke s-phase arrest and dna damage checkpoint activation. *Mol Microbio*, 53:342–351, 2004.

[21] H.E. Kubitschek. *Introduction to Research with Continuous Cultures*. Prentice-Hall, Englewood Cliffs, New Jersey, 1970.

[22] Coulson C. A. Lea, D. E. The distribution of the numbers of mutants in bacterial populations. *Genetics*, 49:264–285, 1949.

[23] Bussey G. Lolle, S. In vivo evidence for posttranslational translocation and signal cleavage of the killer preprotoxin of *Saccharomyces cerevisiae*. *Cell*, pages 4274–4280, 1986.

[24] Delbrck M. Luria, S. E. Mutations of bacteria from virus sensitivity to virus resistance. *The American Statistician*, 28(6):491–511, 1943.

[25] D. Marquina. The biology of killer yeasts. *Microbiol*, 5:65–71, 2002.

[26] J. Neel. Mutation and disease in man. *Canad. J. Gen. and Cytology*, 20(3):295–306, 1978.

[27] M. Nei. Genetic polymorphism and the role of mutation in evolution. *Evolution of Genes and Proteins*, 71:165–190, 1983.

[28] Canton R. Pilar C. Fernando B. Blazquez J. Oliver, A. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science*, 288:1251–1253, 2000.

[29] Lorentzen D. Excoffon K. Zabner J. McCray P. B. Nauseef W. M. Dupuy C. Patryk, M. and B. Bnfi. A novel host defense system of airways is defective in cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 1:26, 2006.

[30] Ramírez M. Regodón J. Pérez, F. A simple and reliable method for hybridization of homothallic wine strains of *Saccharomyces cerevisiae*. *Microbio*, pages 5039–5041, 1998.

[31] Ramírez M. Regodón J. Pérez, F. Influence of killer strains of *Saccharomyces cerevisiae* on wine fermentation. *Antonie Van Leeuwenhoek*, 79:392–399, 2001.

[32] L. Pray. Antibiotic resistance, mutation rates and mrsa. *Nature Education*, 1(1), 2008.

[33] F. Riffer. Mutational analysis of k28 preprotoxin processing in the yeast *Saccharomyces cerevisiae*. *Microbio*, 148:1317–1328, 2002.

[34] Foster P. Rosche, W. Determining mutation rates in bacterial populations. *Methods*, 20:4–17, 2000.

[35] Foster P. Rosche, W. A. Determining mutation rates in bacterial populations. *Methods*, 20:4–17, 2000.

[36] Tsuchiya S. Terasawa K. Tsujimoto G. Sato, F. Intra-platform repeatability and inter-platform comparability of microrna microarray technology. *PLoS ONE*, 4(5):1–12 (e5540), 2009.

[37] D. Schluter. Evidence for ecological speciation and its alternative. *Science*, 323(5915):737–741, 2009.

[38] N. Skipper. Synthesis of a double stranded cdna transcript of the killer toxin-coding region of the yeast m1 double-stranded rna. *Biochem. Biophy. R. Comm.*, 114(2):518–525, 1983.

[39] N. Skipper. Cloning and sequencing of the preprotoxin-coding region of the yeast m1 double-stranded rna. *EMBO J.*, 3(1):107–111, 1984.

[40] A. Smilde. Matrix correlations for high-dimensional data: the modified rv-coefficient. *Bioinf*, 25(3):401–405, 2008.

[41] P. D. Sniegowski. The evolution of mutation rates: Separating causes from consequences. *Bioessays*, 22(12):1057–1066, 2000.

[42] G. Soares and H. Sato. Characterization of the *Saccharomyces cerevisiae* y500-41 killer toxin. *Brazilian J. of Microbio*, 31:291–297, 2000.

[43] Sato. H. Soares, G. Killer toxin of *Saccharomyces cerevisiae* y500-41 active against fleischmann and itaiquara commercial brands of yeast. *Rev. Microbiol*, 30(3):1999, 1999.

[44] M. Tin. Comparison of some ratio estimators. *Am Statistical Assoc*, Mar, 1965.

[45] Postma E. Scheffers W. A. Van Dijken J.P. Verduyn, C. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, 8(7):501–517, 1992.

[46] Boulianne-Larsen C.M. Chandler C. Chiotti K. Kroll E. Miller S. Taddei F. Sermet-Gaudelus S. Ferroni A. McInnerney K. Franklin M.J. Rosenzweig F. Warren, A. Genotypic and phenotypic variation in *Pseudomonas aeruginosa* reveals signatures of seecondary infection and mutator activity in certain cystic fibrosis patients with chronic lung infections. *Infection and Immunity*, 79(12):4802–4818, 2011.

[47] Qi. Zheng. Tprogress of a half century in the st the luria-delbruck distribution. *Math BioSci*, 162:1–32, 1999.