Published in final edited form as:

J Math Biol. 2009 August ; 59(2): 175–191. doi:10.1007/s00285-008-0225-8.

A Stochastic Model for Estimation of Mutation Rates in Multiple-replication Proliferation Processes

Xiaoping Xiong¹, James M. Boyett², Robert G. Webster³, and Juergen Stech⁴

¹Department of Biostatistics, St. Jude Children's Research Hospital, 332 North Lauderdale Street, Memphis, Tennessee 38105, USA. Email: xiaoping.xiong@stjude.org

²Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

³Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

⁴Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, 17493 Greifswald, Insel Riems, Germany

Abstract

In this paper we propose a stochastic model based on the branching process for estimation and comparison of the mutation rates in proliferation processes of cells or microbes. We assume in this model that cells or microbes (the elements of a population) are reproduced by generations and thus the model is more suitably applicable to situations in which the new elements in a population are produced by older elements from the previous generation rather than by newly created elements from the same current generation. Cells and bacteria proliferate by binary replication, whereas the RNA viruses proliferate by multiple replication. The model is in terms of multiple replications, which includes the special case of binary replication. We propose statistical procedures for estimation and comparison of the mutation rates from data of multiple cultures with divergent culture sizes. The mutation rate is defined as the probability of mutation per replication per genome and thus can be assumed constant in the entire proliferation process. We derive the number of cultures for planning experiments to achieve desired accuracy for estimation or desired statistical power for comparing the mutation rates of two strains of microbes. We establish the efficiency of the proposed method by demonstrating how the estimation of mutation rates would be affected when the culture sizes were assumed similar but actually diverge.

Keywords

Mathematical Modeling; Genetic Mutation; Cell; Virus; Bacteria

1 Introduction

An important parameter in the evolution of populations, consisting of cells, viruses, bacteria, or other microbes, is the speed at which individuals mutate to individuals with altered biological properties. The rate of mutation can be measured with respect to the continuous time or to the generation of pedigree lineage. In this paper we simply call the former the *mutation rate in time* and the latter the *mutation rate in replication*. The mathematical models on mutation rates initiated by Luria and Delbruck [10] and further developed by other authors (e.g., [8], [1], [9], and [14]) were in terms of the mutation with respect to the continuous time. As the mutation occurs during replication of the genome, the mutation rate should be better defined as a probability that an error was made during replication so that the genome of a daughter element differs from that of the parent, rather than as a probability that an element changes its nature in an unit time at any moment of its lifetime. However, mathematically the *mutation rate in*

time and the *mutation rate in replication* are about the same if the unit-time of the former is the mean time-span of life cycle of an element. The *mutation rate in time* can be meaningfully defined and used when the proliferation rate in culture is constant in time and uniform among multiple cultures. The assumption of constant and uniform proliferation rate may sometimes not hold, for example, initially similar cultures could grow into very different colony sizes in the same period of time. If the proliferation rate varies much among cultures, the *mutation rate in time* cannot be estimated reliably and nor can the estimated mutation rate be meaningfully interpreted or applied. The *mutation rate in replication*, defined as the probability of mutation in each replication, remains constant even when the proliferation rate changes, and hence can be estimated and applied reliably in different situations.

In this paper we propose a stochastic model based on a branching process, in which the variable for the time domain of the stochastic process is the log of population size n rather than the real time t , and hence the estimation of mutation rate would not be affected by probable temporal variations of proliferation among cultures during the process. The traditional Luria-Delbruck type models assume that any element, recently born or had been existed for certain time, at any moment are equally likely to give birth to new elements, or equivalently assume that the life time of each element since its birth is exponentially distributed. The model we proposed differs from the Luria-Delbruck type models with the following conditions: 1) elements can be born simultaneously (as compared to being born only one after another); 2) new elements can be produced only from a part of the current population, i.e., the elements having existed for certain time (as compared to that they can be produced from the whole current population); 3) the births and mutations of elements are independent among new born elements (as compared to that the birth and mutation of the n th element depends on the status of the $(n-1)$ th element). For easier mathematical derivation, we simplify the model by assuming equal life span of elements (equivalently the synchronous replications of elements) and assuming a fixed number of offsprings produced by each element. This model is suitable for application situations in which the new elements are unlikely to be produced by recently born elements in the same current generation but by elements that had been existed for certain times (the previous generation). The fixed life span of elements assumed in the model is understood as the mean life spans of elements, and the fixed number of offsprings assumed in the model is understood as the mean numbers of offsprings.

We propose the model in terms of multiple replication (e.g., the RNA viruses proliferate by multiple replication) which includes the binary replication as a special case (e.g., cells and bacteria proliferate by binary replication). We also propose statistical procedures for estimating and testing the mutation rate as probability of mutation per replication per genome, based on data from multiple cultures with divergent culture sizes. The number of cultures is formulated for planning experiments to achieve desired accuracy for estimation or desired statistical power for comparing the mutation rates of two strains of cells. We derive an estimator for the *mutation rate in time* from the estimator for the *mutation rate in replication* based on the proposed model. Finally, we show how the estimation of mutation rate can be affected under the assumption of a common culture size when culture sizes actually diverge.

2 Mutation Rates in Replication

The size of a growing population in an experimental culture increases exponentially under good conditions. As an idealization, the growth process can be imagined as a result of repeated multiplication of individuals from $N_0 = N_0 b^0$ at the 0th generation to $N_g = N_0 b^g$ at the g th generation, assuming each individual produces b daughters. The development of a population can be easily described as a tree starting with N_0 original individuals at generation 0 and branching b -tuple at each node. For easier formulation, it is assumed that an individual with its genome turns into exact b progenies as its next generation; the mutants do not have altered

growth properties, neither advantageous nor disadvantageous. A mutant is an individual that has undergone a mutation and differs from its parent by a change in the genome like a point mutation (nucleotide substitution), insertion, or deletion. A mutation rate, calculated from mutants with known changes in their genomes, is denoted as the *mutation rate at the genomic level*. Any mutation can (but does not need to) lead to altered phenotypic traits of the individual; when distinguishing only by phenotypic traits, the mutation rate is denoted as the *mutation rate at the phenotypic level*. For example, for mutants with altered phenotypes requiring only one point mutation at one of k sites in the genome, the mutation rate at phenotypic level can be converted to the mutation rate at genomic level by a factor of $1/k$ if k is not large. This applies to the example in Section 3 studying influenza viruses.

In this paper, we denote the mutation rate per replication as μ , the number of mutants in the population as M , the population size as N , the frequency of mutants in the population as $f = M/N$. Their initial values at generation 0 are M_0 , N_0 , and $f_0 = M_0/N_0$, respectively. We denote the base for the multiple-replication proliferation as b , and denote a coefficient related to b as $r_b = \ln(b)/(1 - 1/b)$. For cells and bacteria, $b = 2$; for RNA viruses, $b \geq 2$, and b could be different for different types of viruses.

2.1 A Stochastic Model

We propose a stochastic model for the number of mutants in the proliferation process of a culture based on a branching process which can be found in classic mathematical books such as Karlin & Taylor [5] or in recently published books on mathematical biology such as Haccou, Jager & Vatutin [3] or Kimmel & David [7]. Let X_j be the number of unmutated individuals in

the j th generation of the population, then $X_{j+1} = \sum_{i=1}^{X_j} \xi_i$, for $j = 1, \dots, g$, where ξ_i is the number of the $(j+1)$ th generation unmutated descendants by the i th individual of the j th generation. The sequence $X_0, X_1, \dots, X_g, \dots$ forms a branching process. Assume a microbe gives birth to b daughters (b is the base of replication), in which, one daughter definitely does not mutate and for the rest $b - 1$ daughters, each mutates with probability μ . Let ξ be the number of unmutated daughters, then ξ takes possible values of $1, \dots, b$. The number of mutated daughters $b - \xi \sim B(b - 1, \mu)$ the binomial distribution with parameters $b - 1$ and μ . Then $E(\xi) = b - (b - 1)\mu$ and $\text{Var}(\xi) = (b - 1)\mu(1 - \mu)$. Assume a culture (population) starts with N_0 microbes (generation 0) which include M_0 mutants. Then $X_0 = N_0 - M_0$ which counts the unmutated microbes at generation 0. Let $N_j = N_0 b^j$ be the population size and M_j the number of mutants at the j th generation, we have $M_j = N_j - X_j$, in which the M_j 's are random numbers and the N_j 's are deterministic numbers. Assume a culture at the end of the experiment has a size $N = N_g = X_0 b^g$ and a number of mutants $M = M_g = N_g - X_g$ (for the g th generation). The mean and variance of M are, respectively,

$$\begin{aligned} E(M) &= N - E(X_g) = N \left\{ 1 - (1 - f_0)(1 - (1 - 1/b)\mu)^{\log_b(N/N_0)} \right\}, \\ \text{Var}(M) &= \frac{(1 - f_0)\mu N^2}{b N_0} (1 - (1 - 1/b)\mu)^{\log_b(N/N_0) - 1} \left\{ (1 - (1 - 1/b)\mu)^{\log_b(N/N_0)} - \frac{N_0}{N} \right\}, \end{aligned} \quad (1)$$

where $f_0 = M_0/N_0$. The derivation of equations in (1) is sketched in A1 in Appendix. Assume $E(M)/N$ and f_0 are close to 0, by which $\ln(1 - E(M)/N) \approx -E(M)/N$ and $\ln(1 - f_0) \approx -f_0$. Then we have

$\mu \approx \frac{r_b \{ \frac{E(M)}{N} - f_0 \}}{\ln(N/N_0)}$ by solving μ from the equation for $E(M)$ in (1), and the derivation is sketched in A2 in Appendix. Consequently, an estimator for μ is

$$\widehat{\mu} = \frac{r_b \left(\frac{M}{N} - f_0 \right)}{\ln \left(\frac{N}{N_0} \right)}, \quad (2)$$

where $r_b = \frac{\ln(b)}{1-1/b}$. The estimator $\widehat{\mu}$ in (2) is approximately unbiased if $\frac{E(M)}{N}$ and f_0 are close to 0, as shown in A2. Methods for multiple cultures developed later in this paper are based on the $\widehat{\mu}$ in (2). Under this assumption, the variance of $\widehat{\mu}$ is

$$\text{Var}(\widehat{\mu}) \approx \frac{r_b^2 (1 - f_0) \mu}{b N_0 \left\{ \ln \left(\frac{N}{N_0} \right) \right\}^2}, \quad (3)$$

and the derivation of this equation is sketched in A3 in Appendix. Replacing μ by $\widehat{\mu}$ in the equation above, we obtain an estimator for $\text{Var}(\widehat{\mu})$ as

$$\widehat{V}_{\widehat{\mu}} = \frac{r_b^3 (1 - f_0) \left(\frac{M}{N} - f_0 \right)}{b N_0 \left\{ \ln \left(\frac{N}{N_0} \right) \right\}^3}. \quad (4)$$

If the assumption that $E(M)/N$ and f_0 are close to 0 does not hold, then the approximation of μ is, as shown in A2 in Appendix,

$$\mu \approx \frac{r_b \{ \ln(1 - f_0) - \ln(1 - \frac{E(M)}{N}) \}}{\ln \left(\frac{N}{N_0} \right)} \quad (5)$$

and its estimator $\widehat{\mu}$ with improved accuracy can be obtained by replacing $E(M)$ in the equation (5) by M . The variance of this $\widehat{\mu}$ is, as shown in A3 in Appendix,

$$\text{Var}(\widehat{\mu}) \approx \frac{r_b (1 - f_0)}{b N_0} \left[\frac{r_b \mu}{\left\{ \ln \left(\frac{N}{N_0} \right) \right\}^2} - \frac{2\mu^2}{\ln \left(\frac{N}{N_0} \right)} \right], \quad (6)$$

which can be estimated by replacing μ in the equation (6) with this $\widehat{\mu}$.

Taking for example, the special case $b = 2$ is the most common in biology, for which the ξ is distributed with $P(\xi = 1) = \mu$ and $P(\xi = 2) = 1 - \mu$; its expectation $E(\xi) = 2 - \mu$ and variance $\text{Var}(\xi) = \mu(1 - \mu)$. With $r_b = \ln(4)$ for $b = 2$, the $\widehat{\mu}$ in (2) became

$$\widehat{\mu} = \frac{\ln(4) \left(\frac{M}{N} - f_0 \right)}{\ln \left(\frac{N}{N_0} \right)}, \quad (7)$$

and $\text{Var}(\widehat{\mu})$ in (3) became

$$\text{Var}(\widehat{\mu}) \approx \frac{\{\ln(4)\}^2 (1 - f_0) \mu}{2N_0 \left\{ \ln\left(\frac{N}{N_0}\right) \right\}^2}. \quad (8)$$

If $f_0 = 0$, then $\widehat{\mu}$ and $\text{Var}(\widehat{\mu})$ by (7) and (8) are the same as the results by Rossman et al. [12], which were in a different form and obtained through a different approach.

2.2 Mutation Rate at Genome Level

In most of applications, the mutation rate at the phenotypic level of mutation can be estimated more easily than that at the genomic level, because the mutants can be distinguished from nonmutants with much less efforts at the phenotypical level than that at the genomic level. However, in biological studies, most investigators are not interested in mutation rates at phenotypic level but at genomic level, because the comparison of mutation rates at phenotypic level can be misleading if the different changes in the genome lead to the same phenotypic change. For example, in virology, in the case of influenza viruses, virions are resistant to amantadine, an antiviral drug, if there is a point mutation at one of k ($=4$) known specific sites in the nucleotide sequence, shown by Hay et al. [4]. We denote the mutation rate at the phenotypic level as μ_p , the mutation rates at genomic level as μ_{gi} for genomic sites $i = 1, \dots, k$. Within one virus, let X be the number of mutations at the genomic level, e.g., the number of mutated nucleotides at the k sites. If genome mutations are mutually independent, then $\mu_p = P(X > 0) = 1 - \prod_{i=1}^k (1 - \mu_{gi})$ where $\mu_{gi} = P(\text{genome mutation on site } i)$. If $\mu_{gi} = \mu_g$ for $i = 1, \dots, k$, then $\mu_p \approx k\mu_g$ and hence $\mu_g \approx \frac{\mu_p}{k}$. Assuming genotypic mutations causing the *phenotypic* mutation equally likely and independently, we have the estimator for the mutation rate at genomic level

$$\widehat{\mu}_g = \frac{r_b \left(\frac{M}{N} - f_0 \right)}{k \ln\left(\frac{N}{N_0}\right)}, \quad (9)$$

and its variance is

$$\text{Var}(\widehat{\mu}_g) \approx \frac{r_b^2 (1 - f_0) \mu_g}{kbN_0 \left\{ \ln\left(\frac{N}{N_0}\right) \right\}^2}. \quad (10)$$

The derivation of the equation in (10) is sketched in A4 in Appendix. The culture size N and the number of mutants M are obtained at the phenotypic level (e.g., virus level), for example, by counting viruses in the culture before and after adding amantadine, respectively. Replacing μ_g by $\widehat{\mu}_g$ in the equation for $\text{Var}(\widehat{\mu}_g)$ above, we obtain an estimator for $\text{Var}(\widehat{\mu}_g)$,

$$\widehat{V}_{\widehat{\mu}_g} = \frac{r_b^3 (1 - f_0) \left(\frac{M}{N} - f_0 \right)}{k^2 b N_0 \left\{ \ln\left(\frac{N}{N_0}\right) \right\}^3}. \quad (11)$$

The methods for multiple cultures developed later in this paper are based on equations (9) and (11). In any equation for the genotypical level that we will develop below, statistical inference on mutation rate at the phenotypic level can be obtained by setting $\mu_p = \mu_g$ and $k = 1$.

2.3 A Luria-Delbruck Type Model for the Mutation Rate in Replication

Drake [2] proposed a model in which the *mutation rate in replication* was used instead of the *mutation rate in time* and a differential equation $dM = (\mu + \frac{M}{N}) dN$ was used to link the number of mutants in a culture with the population size of the culture rather than with the real time since the start of the culture. Solving this equation, an estimator for the *mutation rate in*

replication μ was obtained as $\hat{\mu} = \frac{\frac{M}{N} - f_0}{\ln(\frac{N}{N_0})}$. The increment of mutants dM actually has two components: 1) increment by proliferation of existing mutants, or $\frac{M}{N} dN$; 2) mutation of the increment of existing nonmutants, or $\mu(1 - \frac{M}{N}) dN$. Taking into account this fact we can modify

the above model to $dM = \{\mu(1 - \frac{M}{N}) + \frac{M}{N}\} dN$, and which yields an estimator $\hat{\mu} = \frac{\ln(1-f_0) - \ln(1-\frac{M}{N})}{\ln(\frac{N}{N_0})}$. The latter estimator improves the former when M is large such that the condition $\frac{M}{N} \approx 0$ barely holds. The two $\hat{\mu}$'s above derived from the model above as functions of M , N , N_0 , and f_0 are similar to the two $\hat{\mu}$'s in Section 2.1 derived from the proposed stochastic model for $b = 2$, but differ with a coefficient $r_b = 1.386$. The similarity suggests that the differences in the assumptions of the proposed stochastic model and the Luria-Delbruck type model do not lead to much different estimators for *mutation rates*. We will show in Section 4.2 that the main difference between the proposed model and the Luria-Delbruck type models is in the variances of the estimators for the mutation rate.

3 Statistical Inferences

For simplicity, from now on we will use μ instead of μ_g to denote the mutation rate at the genotypic level. The μ is very small in most applications, hence for statistical inference we measure the accuracy of $\hat{\mu}$ with the relative scale of $\gamma \equiv \sigma_{\hat{\mu}}/\mu$ rather than with the scale of $\sigma_{\hat{\mu}}$, where $\sigma_{\hat{\mu}}^2 = \text{Var}(\hat{\mu})$. The smaller is the γ , the more accurate is the estimation of μ . The

relationship between γ and the other parameters of the proposed model is $\gamma = \frac{r_b \sqrt{1-f_0}}{\ln(\frac{N}{N_0}) \sqrt{\mu k b N_0}}$, which was obtained easily by joining the definition of γ and the equation for $\text{Var}(\hat{\mu}_g)$ in (10). Using the equation for γ above, we calculated its value for several usual settings of parameters for estimating mutation rates of influenza viruses, and the results indicated that if an estimation is obtained from a single culture, then the γ would be too large for application purposes. Multiple cultures are needed for applications if one wants to obtain estimations with reasonably good confidence levels.

3.1 Estimation from Multiple Cultures

For C cultures, let N_i be the culture size, M_i the number of mutants, and $\hat{\mu}_i$ the estimator for μ for the i th culture, $i = 1, \dots, C$, where μ is the mutation rate common for all cultures. We assume N_i 's are known and constants (for being conditioned on N_i 's), M_i 's are random variables. There are two ways to pool estimators $\hat{\mu}_i$'s: the *Simple Average* and the *Weighted Average*.

Simple Average—Define $\hat{\mu}_a = \frac{1}{C} \sum_{i=1}^C \hat{\mu}_i$, by (9) we have

$$\hat{\mu}_a = \frac{r_b}{kC} \sum_{i=1}^C \frac{\frac{M_i}{N_i} - f_{0i}}{\ln\left(\frac{N_i}{N_{0i}}\right)}, \quad (12)$$

where N_{0i} and f_{0i} are the initial culture size and frequency of mutants in the i th culture. The variance of $\hat{\mu}_a$ is

$$\text{Var}(\widehat{\mu}_a) = \frac{r_b^2 \mu}{kbC^2} \sum_{i=1}^C \frac{1 - f_{0i}}{N_{0i} \left\{ \ln \left(\frac{N_i}{N_{0i}} \right) \right\}^2}, \quad (13)$$

which can be estimated by replacing the μ in (13) by the $\widehat{\mu}_a$ in (12).

Weighted Average—Define $\widehat{\mu}_w = \sum_{i=1}^C w_i \widehat{\mu}_i$, where w_i 's are weights ($\sum_{i=1}^C w_i = 1$ and $w_i \geq 0$). The variance of $\widehat{\mu}_w$ can be minimized by allocating w_i according to the inverse of variance of $\widehat{\mu}_i$ (e.g., see p308 in Rao [11]), by which the optimal weighted-average estimator is

$$\widehat{\mu}_w = \frac{r_b}{kT} \sum_{i=1}^C \frac{N_{0i} \left(\frac{M_i}{N_i} - f_{0i} \right) \ln \left(\frac{N_i}{N_{0i}} \right)}{1 - f_{0i}}, \quad (14)$$

where $T = \sum_{j=1}^C N_{0j} \left\{ \ln \left(\frac{N_j}{N_{0j}} \right) \right\}^2 / (1 - f_{0j})$. The variance of $\widehat{\mu}_w$ is $\text{Var}(\widehat{\mu}_w) = \frac{r_b^2 \mu}{kbT}$, Which can be estimated by

$$\widehat{V}_{\widehat{\mu}_w} = \frac{r_b^3}{b(kT)^2} \sum_{i=1}^C \frac{N_{0i} \left(\frac{M_i}{N_i} - f_{0i} \right) \ln \left(\frac{N_i}{N_{0i}} \right)}{1 - f_{0i}}. \quad (15)$$

In general, the weighted-average estimator $\widehat{\mu}_w$ has a smaller variance and hence should be preferred. If the sizes of all cultures are similar, then $\widehat{\mu}_w$ and $\widehat{\mu}_a$ are approximately the same. Some users may favor the simple-average estimator $\widehat{\mu}_a$ for its intuitiveness and simplicity.

3.2 Comparing Mutation Rates

To compare the mutation rates of two different strains of microbes is to test $H_0: \mu_1 \leq \mu_2$ versus $H_a: \mu_1 > \mu_2$, where μ_1 and μ_2 are the two mutation rates. Let $\widehat{\mu}_1$ and $\widehat{\mu}_2$ be the estimators for μ_1 and μ_2 , and let $\widehat{V}_{\widehat{\mu}_1}$ and $\widehat{V}_{\widehat{\mu}_2}$ be the estimated variances of $\widehat{\mu}_1$ and $\widehat{\mu}_2$. Because $\widehat{\mu}_1$ and $\widehat{\mu}_2$ are linear sums of independent estimators, the test statistic $Z = \frac{\widehat{\mu}_1 - \widehat{\mu}_2}{\sqrt{\widehat{V}_{\widehat{\mu}_1} + \widehat{V}_{\widehat{\mu}_2}}}$ is distributed approximately normal if both C_1 and C_2 are large. The two mutation rates are significantly different if $Z > z_\alpha$, where z_α is the upper α -quantile of standard normal distribution. For smaller C_1 and C_2 , the z_α should be replaced by $t_\alpha (C_1 + C_2 - 2)$ which is the upper α -quantile of t-distribution with degree freedom of $C_1 + C_2 - 2$.

3.3 Examples

In 1979, an H1N1 avian influenza virus crossed the species barrier and established a new lineage in European swine. To determine whether a high mutation rate (in this study a higher rate of point mutations) had contributed to interspecies transmission, Stech et al. [13] compared the mutation rate of A/Swine/Germany/2/81(H1N1), a well-characterized early isolate of the above-mentioned lineage, with that of A/Mallard/New York/6750/78(H2N2), a virus that was well-established in avian hosts. Each culture was started with a single wild-type virus, hence $N_0 = 1$ and $f_0 = 0$. The final culture size N and the number of mutants M were obtained by counting viruses in absence and presence of amantadine. Since a virus survives in the presence

of amantadine if and only if there is at least one mutated nucleotide at any of 4 specific sites in its genome, hence $k = 4$. There were 17 cultures for the Swine virus and 10 cultures for the Mallard virus. Let μ_{sw} be the mutation rate of Swine virus and μ_{mal} be that of the Mallard virus. The interest of the study is to test $H_0 : \mu_{sw} \leq \mu_{mal}$ vs. $H_a : \mu_{sw} > \mu_{mal}$. We applied the method of weighted average for estimations and tests, and did the calculation for two cases: $b = 2$ and $b = 5$. Assuming $b = 2$, for the Swine virus: $\hat{\mu}_{sw} = 6.67 \times 10^{-6}$ and $\hat{\sigma}_{\hat{\mu}_{sw}} = 1.46 \times 10^{-5}$; for the Mallard virus: $\hat{\mu}_{mal} = 3.23 \times 10^{-5}$ and $\hat{\sigma}_{\hat{\mu}_{mal}} = 4.58 \times 10^{-5}$. Assuming $b = 5$, for the Swine virus: $\hat{\mu}_{sw} = 9.68 \times 10^{-6}$ and $\hat{\sigma}_{\hat{\mu}_{sw}} = 1.59 \times 10^{-5}$; for the mallard virus: $\hat{\mu}_{mal} = 4.69 \times 10^{-5}$ and $\hat{\sigma}_{\hat{\mu}_{mal}} = 5.06 \times 10^{-5}$ for $b = 2$, $Z = -0.534$ and p-value = 0.703. for $b = 5$, $Z = -0.701$ and p-value = 0.758. Both results indicate that the mutation rate of Swine virus is not significantly larger than that of Mallard virus.

3.4 Number of Cultures: Sample Size Calculation

To plan an experiment with desired accuracy for estimation of a mutation rate or desired power for detecting a given difference for comparison of two mutation rates, we need a sufficient number of cultures. At the planning stage, we may assume $N_{0i} = N_0$, $f_{0i} = f_0$, and $N_i = N$ for all i 's. Then

$$V_{\hat{\mu}_a} = V_{\hat{\mu}_w} = \frac{r_b^2 (1-f_0) \mu}{CkbN_0 \left\{ \ln \left(\frac{N}{N_0} \right) \right\}^2}$$

Estimation—The number of cultures for estimation with a given γ is

$$C = \frac{r_b^2 (1-f_0)}{\mu kbN_0 \{\gamma \ln(N/N_0)\}^2}, \quad (16)$$

where $\gamma \left(= \frac{\sigma_{\hat{\mu}_a}}{\hat{\mu}_a} = \frac{\sigma_{\hat{\mu}_w}}{\hat{\mu}_w} \right)$ measures the accuracy of estimators $\hat{\mu}_a$ and $\hat{\mu}_w$ relative to the magnitude of μ . A smaller γ leads to a better estimation, but demands a larger sample size C , as shown in Tables 1 and 2.

Comparison—To test the hypotheses $H_0 : \mu_1 \leq \mu_2$ versus $H_a : \mu_1 > \mu_2$ with significance level α and power $1 - \beta$ to detect a given relative difference d , the number of cultures in each strain group is

$$C = \left\{ \frac{r_b (z_\alpha + z_\beta)}{d \ln \left(\frac{N}{N_0} \right)} \right\}^2 \frac{(d+2)(1-f_0)}{\mu_2 kbN_0}, \quad (17)$$

where $d = \frac{\mu_1 - \mu_2}{\mu_2}$ which is the difference of μ_1 and μ_2 relative to the magnitude of μ_2 . For a two-sided test, C is obtained by replacing z_α by $z_{\alpha/2}$ in (17). C_1 and C_2 can be different, but must

satisfy equation $\left(\frac{z_\alpha + z_\beta}{d\mu_2} \right)^2 = \frac{1}{V_{\hat{\mu}_1} + V_{\hat{\mu}_2}}$, where $V_{\hat{\mu}_1} = \frac{r_b^2 (1+d)\mu_2}{kbC_1} \frac{1-f_0}{N_0 (\ln(N/N_0))^2}$ and $V_{\hat{\mu}_2} = \frac{r_b^2 \mu_2}{kbC_2} \frac{1-f_0}{N_0 (\ln(N/N_0))^2}$.

The numbers of cultures are calculated in Tables 1 and 2 for the estimation and the comparison of mutation rates, respectively. For example, to test $H_0 : \mu_1 \leq \mu_2$ versus $H_a : \mu_1 > \mu_2$ with $\alpha = 0.05$ and power $1 - \beta = 0.8$ for detecting a relative difference of $d = 3$. Assume $k = 4$, $N_0 = 1$, $f_0 = 0$; $N_i = 10^{10}$ for all i 's; $\mu_2 = 3 \times 10^{-5}$ and $b = 5$, then we have $C = 43.7$ from Table 2. If the initial number of viruses per culture is increased to $N_0 = 10$ and there are no initial mutants ($f_0 = 0$), then the number of cultures per strain can be reduced to $C = 5.4$.

3.5 Efficiency of Estimation and Comparison

The number of elements in population N increases from N_0 at the beginning to its final value at the end of experiment. During this process the number of mutants M as a function of N can be viewed as a stochastic process with the time domain in terms of N or $\log_b(N)$. In real experiments, however, M and N during this process are unobservable except their last values at the end of the experiment. The μ hence can only be estimated from the last values of N and M . On the other hand, the times at which mutations really happen during the process cause a very large variation in the M at the end of experiment, and hence causes a large variation in $\hat{\mu}$. For example, one mutation at the first generation will produce mutant descendents as many as $1/b$ of population at the end of the experiment, whereas one mutation at the second to last generation produces b mutant descendents. The times of mutations during the process are unobservable and hence unknown. Therefore, the variance of $\hat{\mu}$ is very large which makes the statistical inference with required accuracy very difficult because the number of cultures required could be impractically large. To reduce the variability of estimation, by contending an analogy of hitting-jackpot, Luria and Delbruck [10] proposed an estimator by assuming no mutation until a time when one mutation is expected in the population. This proposition actually is to let the culture size increase from N_0 to N_0^* until the time mentioned above and assume no mutants in the N_0^* elements and take them as new initial starters. The accuracy of estimation by this practice is not controllable because the N_0^* is large and hence there is a good chance that mutants could have existed among the N_0^* pretended new starters. The mutation rate is overestimated by this method with a probability of 0.632, as pointed out by Armitage [1].

We propose another approach for improving the efficiency of statistical inference. As the trends shown in Tables 1 and 2, the number of cultures C can be reduced by increasing initial culture size N_0 and/or final culture size N , for achieving the same accuracy in estimation or power in hypotheses testing. The difficulties are that the N is limited in real experiments, and that if N_0 is greater than 1, then the number of mutants M_0 among the N_0 starters is unknown (f_0 is hence unknown). However, the latter difficulty can be get around in a following way. The number of mutants M consists of two parts, $M = M_1 + M_2$. M_1 is the number of descendents from M_0 initial mutants, or $M_1 = M_0 b^g$ where g is the number for current generation; M_2 is the number of mutated descendents from the $N_0 - M_0$ initial unmutated elements. It is always true that $f \geq f_0$, because $f = M/N \geq M_1/N = M_0 b^g / N_0 b^g = f_0$. In a culture, if $M/N < 1/N_0$ is observed, then it implies $f_0 = 0$ (or $M_0 = 0$). The N_0 initiators for the generation 0 should have been selected before culture starts, and must not be retrospectively claimed after observing M and N . It is incorrect to retrospectively choose an arbitrary N_0 that satisfies $M/N < 1/N_0$ and claim it as the initial population size, because N_0 should be a constant that does not depend on M and N . For multiple parallel cultures, assuming N_{0i} initiators for the i th culture, we may identify cultures that have $M_{0i} = 0$ by checking whether $M_i/N_i < 1/N_{0i}$. We then estimate μ only from the cultures with $M_{0i} = 0$ and using $\hat{\mu}_a$ in (12) or $\hat{\mu}_w$ in (14) and for which setting $f_{0i} = 0$ for all i 's.

4 Mutation Rate in Time

The *mutation rate in time* (denoted as a) is the probability that a microbe mutates in an unit time. Luria and Delbruck [10] proposed two methods for estimating this mutation rate: the P_0 method and the mean method. The P_0 method is simple to use, however, information from data is not fully utilized because only the presence and absence of mutants are counted whereas the numbers of mutants in cultures are ignored, and hence it requires a large number of cultures. The mean method of Luria and Delbruck model was further developed by Lea and Coulson [8] by deriving the probability distribution of the number of mutants and proposing the median method and the maximum likelihood method based on the distribution. A more general model was proposed by Armitage [1], which extends the results by Luria and Delbruck [10] and Lea

and Coulson [8] in the way that mutants and nonmutants in a culture may grow with two different rates exponentially and the model includes the back-mutation. However, the inferences of mutation rates basically remain the same because the parameters of the more delicate models are not estimable from only two available values, M and N , from the data. In this section, we derive an estimator for the *mutation rate in time* based on the proposed stochastic model. The estimator is similar to that by the Luria and Delbruck type models, but with a substantially larger variance.

4.1 A Stochastic Process: Number of Mutants M_t

Let N_t denote the N at time t and assume N_t a deterministic function (e.g., $N_t = N_0 e^{\rho t}$) because the variation of N_t is usually unsubstantial for a given culture. Let M_t denote the M at time t , then M_t is a stochastic process with t in time domain. M_t has a large variability and hence can not be approximated by a deterministic function. Let $M_t^* \equiv E(M_t)$ be the expectation of M_t . In the proposed model, M_t^* is a deterministic function of t for being a function of N_t . Theoretically, the N_t is allowed to be any positive nondecreasing function of t in a certain period of time since the start of the experiment. Assuming $f_0 = 0$ and $k = 1$, then from the equation (1),

$$M_t^* = N_t \left\{ 1 - (1 - (1 - 1/b)\mu)^{\log_b(N_t/N_0)} \right\}, \quad (18)$$

where μ is the *mutation rate in replication*. As a function of t , the M_t^* would be known if N_t is given. From the equation for $\text{Var}(M)$ in (1), the variance of number of mutants M_t at time t is

$$\text{Var}(M_t) = \frac{\mu N_t^2}{b N_0} (1 - (1 - 1/b)\mu)^{\log_b(N_t/N_0)-1} \cdot \left\{ (1 - (1 - 1/b)\mu)^{\log_b(N_t/N_0)} - N_0/N_t \right\}. \quad (19)$$

4.2 Exponential Growth

Assume $N_t = N_0 e^{\rho t}$, where ρ is the proliferation rate, then $\rho = \frac{N'_t}{N_t}$ which implies that ρ is the individual growth rate, or the growth per element per unit time. A culture grows from the generation 0 to the generation g in the time interval $[0, t]$, hence g and t are related by $g = \lambda t$ where $1/\lambda$ is the mean life-span of one generation. Then $N_t = N_0 e^{\rho t}$ can be obtained from $N_g = N_0 b^g$ by letting $\rho = \lambda \ln(b)$. Let a be the *mutation rate in time*, the probability of mutation per element per unit time. Assume a is obtained by allocating probability mass μ (the probability of mutation per element per replication) uniformly on an element's life-span, hence

$a = \frac{\mu}{1/\lambda} = \mu\lambda$. Then $\mu = \frac{a \ln(b)}{\rho}$ and by (1) we have

$$M_t^* = N_0 e^{\rho t} \left[1 - \left\{ 1 - (1 - 1/b) \frac{a \ln(b)}{\rho} \right\}^{\frac{\rho t}{\ln(b)}} \right] \quad (20)$$

and $\text{Var}(M_t) = \frac{a \ln(b)}{\rho b N_0} e^{2\rho t} \left(1 - (1 - 1/b) \frac{a \ln(b)}{\rho} \right)^{\frac{\rho t}{\ln(b)}-1} \left\{ \left(1 - (1 - 1/b) \frac{a \ln(b)}{\rho} \right)^{\frac{\rho t}{\ln(b)}} - e^{-\rho t} \right\}$. If $at \approx 0$ (i.e., $\mu \ln(N) \approx 0$), then we have

$$M_t^* \approx (1 - 1/b) at N_0 e^{\rho t} \quad \text{and} \quad \text{Var}(M_t) \approx \frac{r_b a}{(b-1)\rho N_0} e^{2\rho t}. \quad (21)$$

At the time t , the expected number of mutants in culture, M_t^* in (21), is similar to that of Luria and Delbruck [10], differing with a coefficient $(1-1/b)$. The variance of the number of mutants $\text{Var}(M_t)$ in (21) as a function of t is $e^{2\rho t}$ with a constant coefficient and thus is similar to $a^2 N_0^2 e^{2\rho t}$, the one by Luria and Delbruck [10], which is also $e^{2\rho t}$ with a constant coefficient.

However, the ratio of two coefficients is $r_b/aN_0^3(b-1)\rho$, which implies that the coefficient of former would be substantially larger than that of the latter when a is very small as in most applications. Hence the variance of the number of mutants in a population by the proposed model could be substantially larger than that by the Luria and Delbruck type model, even though the estimates for the number of mutants by the two models are very similar. For multiple parallel cultures, the proliferation rate ρ may vary substantially among cultures because parallel cultures can grow into quite different sizes in a same period of time. For the i th culture with the initial size N_{i0} and the size N_{it} at time t , the proliferation rate ρ_i can be estimated by

$\widehat{\rho}_i = \ln\left(\frac{N_{it}}{N_{i0}}\right)/t$. The mutation rate in time a varies among cultures if ρ does, however, on the contrary the mutation rate in replication μ is a constant across multiple cultures. The mutation rate in time for the i th culture $a_i\left(=\frac{\mu\rho_i}{\ln(b)}\right)$ can be estimated by $\widehat{a}_i = \frac{M_{it}}{(1-1/b)\ln N_{it}}$ from the i th culture alone; or by $\widehat{a}_i = \widehat{\mu} \ln\left(\frac{N_{it}}{N_{i0}}\right)/t \ln(b)$, where $\widehat{\mu}$ is the estimator for μ from all cultures.

5 Divergent Culture Sizes

All cultures in an experiment are assumed to grow into around one common culture size in the traditional fluctuation analysis. In this section, we investigate how the estimation of mutation rate could be affected under the assumption of a common culture size when culture sizes actually diverge. The following analysis is based on the proposed stochastic model, however, its conclusion should be implicative to the fluctuation analysis using the traditional Luria-Delbruck type models.

Let $\bar{N}$ be the sample mean of N_i 's and $\bar{M}$ be the sample mean of M_i 's for C cultures for which

all cultures have the same N^0 and f^0 and assuming $f^0 = 0$. Then $\widehat{\mu}_{cs} = \frac{r_b \bar{M}}{\bar{N} k \ln(\bar{N}/N_0)}$ by (9) is an estimator for γ obtained by assuming equal culture sizes, where the subscript “cs” stands for “common

size”. We can rewrite $\widehat{\mu}_{cs} = \frac{1}{C} \sum_{i=1}^C \widehat{\mu}_i$, where $\widehat{\mu}_i = \frac{r_b M_i}{N_i k \ln(N_i/N_0)}$. Then we have

$$E(\widehat{\mu}_{cs}) = \mu \frac{\frac{1}{C} \sum_{i=1}^C N_i \ln(N_i/N_0)}{\bar{N} \ln(\bar{N}/N_0)} \quad \text{and} \quad \text{Var}(\widehat{\mu}_{cs}) = \frac{r_b^2 \mu}{kbCN_0} \frac{\frac{1}{C} \sum_{i=1}^C N_i^2}{\left\{ \bar{N} \ln(\bar{N}/N_0) \right\}^2}$$

. The first equation above indicates that $\widehat{\mu}_{cs}$ is unbiased ($E(\widehat{\mu}_{cs}) = \mu$) if N_i 's are all equal, and could be biased if otherwise. On the contrary, the $\widehat{\mu}_a$ by (12) is unbiased ($E(\widehat{\mu}_a) = \mu$) even for unequal N_i 's. The culture sizes N_i 's of the parallel cultures can be regarded as samples of a random variable N_t . Now we consider the situation when the number of parallel cultures, C , becoming large. Let $r_{\mu_{cs}} \equiv \lim_{C \rightarrow \infty} E(\widehat{\mu}_{cs})/\mu$, then $r_{\mu_{cs}} = \frac{E[N_t \ln(N_t/N_0)]}{E(N_t) \ln(E(N_t)/N_0)}$. Asymptotically, $\widehat{\mu}_{cs}$ is unbiased to μ if $r_{\mu_{cs}} = 1$, and $\widehat{\mu}_{cs}$ is biased if $r_{\mu_{cs}} \neq 1$. Let $b_{\widehat{\mu}_{cs}} \equiv [E(\widehat{\mu}_{cs}) - \mu]/\mu$, which is the bias of $\widehat{\mu}_{cs}$ relative to μ , then $r_{\mu_{cs}} = 1 + b_{\widehat{\mu}_{cs}}$. Let $r_{\sigma_{cs}^2, \sigma_a^2} \equiv \lim_{C \rightarrow \infty} \text{Var}(\widehat{\mu}_{cs})/\text{Var}(\widehat{\mu}_a)$, then

$r_{\sigma_{cs}^2, \sigma_a^2} = \frac{E(N_t^2)}{[E(N_t)]^2 [\ln(E(N_t/N_0))]^2 E[(\ln(N_t/N_0))^{-2}]}$. Asymptotically, $\widehat{\mu}_{cs}$ is less efficient than $\widehat{\mu}_a$ if $r_{\sigma_{cs}^2, \sigma_a^2} > 1$, and $\widehat{\mu}_{cs}$ is at least as efficient as $\widehat{\mu}_a$ if $r_{\sigma_{cs}^2, \sigma_a^2} \leq 1$.

For the population size $N_t = N_0 e^{\rho t}$, we assume ρ is constant for one culture in the entire proliferation process, but ρ could be different for different cultures and thus can be considered as a random number for the multiple parallel cultures. Let $m(\tau)$ be the moment generating function of ρ , then $m(\tau) \equiv E(e^{\rho \tau})$. Let $\varphi(\tau) \equiv \ln(\ln(m(\tau)))$ and let $\varphi'(\tau)$ be the derivative of $\varphi(\tau)$. Then it is straightforward to show that $r_{\mu_{cs}} = t\varphi'(t)$ for the $r_{\mu_{cs}}$ defined above. If ρ is normally distributed with mean ρ^* and variance σ_ρ^2 , then $m(\tau) = \exp(\rho^* \tau + \sigma_\rho^2 \tau^2 / 2)$, and with some

algebra, $b_{\mu_{cs}} = \left(\frac{2\rho^*}{\sigma_\rho^2} + 1\right)^{-1}$ for the relative bias $b_{\mu_{cs}}$ defined above. This equation implies that the bias of $\widehat{\mu}_{cs}$ equals 0 if $\sigma_\rho^2 = 0$. That is, if the proliferation rates are identical across the parallel cultures, then culture sizes would be the same at the end of the experiment, and consequently the estimation of mutation rate is unbiased. The equation also indicates that if the proliferation rates are not identical ($\sigma_\rho^2 > 0$), then the bias increases in t , that is, the bias is getting larger if the experiment lasts longer time. Fortunately, this relative bias is asymptotically bounded by 1. We may assume $\rho > c$ for some $c > 0$, practically by excluding cultures which almost do not

grow. Then $r_{\sigma_{cs}^2, \sigma_a^2} \geq e^{\sigma_\rho^2 t^2} / \left(\frac{2\rho^*}{c^2 t} + \frac{\sigma_\rho^2}{c^2}\right)$, which implies that the ratio of variances increases in t unboundedly. The loss of efficiency could be substantial if the variance of $\widehat{\mu}_{cs}$ is much larger than that of $\widehat{\mu}_a$, because the sample sizes required are proportional to the variances of the estimators for μ . For the example in Section 3.3, the culture sizes range from 3.4×10^8 to 1.2×10^{10} . We have $\widehat{r}_{\mu_{cs}} = 1.037$ and $\widehat{r}_{\sigma_{cs}^2, \sigma_a^2} = 3.32$. The relative bias of $\widehat{\mu}_{cs}$ for μ is $\widehat{b}_{\mu_{cs}} = 0.037$ and is practically negligible. The ratio of variances for $\widehat{\mu}_{cs}$ to $\widehat{\mu}_a$ is 3.32 which implies that the number of cultures for $\widehat{\mu}_{cs}$ must be 3.32 times as large as that for $\widehat{\mu}_a$ in order to achieve a same accuracy for estimation or a same power for comparison. The analysis above indicates that an estimation assuming equal culture sizes may lose efficiency substantially when culture sizes actually diverge.

6 Conclusion

Estimation and comparison of mutation rates in cells and microbes are important measurements in biological studies. However, the results could be misleading if the variation of the estimation or the errors of the comparison are ignored. Large variability in estimations had been observed by biological researchers (e.g., Kendal and Frost [6]), and could not be adequately explained by the traditional models. In this paper we proposed a model based on the branching process such that some realistic aspects of proliferation and mutation of cells and microbes are better represented in the model. The large variability is inherent to the number of mutants in a culture, and hence is inherent to the estimators for mutation rates. For biological researchers who want to obtain reliable estimations and comparisons of mutation rates, we propose statistical planning of experiments by assuming adequate initial and final culture sizes and calculating number of cultures required to achieve desired accuracy in estimation and statistical power for comparison. The traditional fluctuation analysis assume a common culture size among all cultures. In practice, culture sizes could differ substantially, especially when culture grow into large sizes. We demonstrated that the efficiency of estimation could be much lowered if assumed homogenous culture sizes actually diverge. For the traditional models, to make culture sizes homogenous, the cultures cannot be allowed to grow into very large sizes, which is a restriction that lowers the efficiency of estimation and comparison. On the contrary, the proposed model allows cultures grow into large sizes by taking account of different culture

sizes. By this model, if the cultures can grow into larger sizes, then a smaller number of cultures is required to achieve a same accuracy of estimation or testing. An implication of this fact is that the laboratory work in experiments could be reduced by letting cultures grow for a longer time. The proposed model should be applicable to the mutation in the proliferation of cells or the asexual proliferation of eukaryotic and prokaryotic organisms, in which the new births are unlikely being given by the new members of the population and the mutation rate in replication is constant during the proliferation process.

Acknowledgements

This work was supported in part by Cancer Center Support grant P30 CA-21765 from the National Cancer Institute, Bethesda, MD, and by the American Lebanese Syrian Associated Charities (ALSAC), Memphis, Tennessee, USA. The authors wish to thank the two anonymous referees for their comments, especially the first referee for very helpful suggestions.

7 Appendix

A1

Derivation of equations in (1): Since $E\xi = 1 - (1 - 1/b)\mu$, $N = N_0 b^g$ and $X_0 = N_0(1 - f_0)$, we have $EX_g = X_0(E\xi)^g = X_0[b - (b - 1)\mu]^g = N(1 - f_0)[1 - (1 - 1/b)\mu]^g$, which leads to the equation for $E(M)$ in (1) by joining with $E(M) = N - EX_g$ and $g = \log_b(N/N_0)$. For branching process, $\text{Var}(X_g) = X_0 \text{Var}(\xi)(E\xi)^{g-1} \frac{(E\xi)^g - 1}{E\xi - 1}$, which leads to the equation for $\text{Var}(M)$ in (1) by joining with identities $\text{Var}(M) = \text{Var}(N - M) = \text{Var}(X_g)$, $E\xi = b - (b - 1)\mu$, and $\text{Var}(\xi) = (b - 1)(1 - \mu)\mu$.

A2

Derivation of the equation (2): Solving μ in the equation for $E(M)$ in (1), we have identity $\mu = \frac{b}{b-1} \left[1 - \exp \left\{ \frac{-\ln(1 - E(M)/N) + \ln(1 - f_0)}{\log_b(N/N_0)} \right\} \right] \approx \frac{b}{b-1} \frac{-\ln(1 - E(M)/N) + \ln(1 - f_0)}{\log_b(N/N_0)}$. Assume $E(M)/N$ and f_0 are close to 0, then $\ln(1 - E(M)/N) \approx -E(M)/N$ and $\ln(1 - f_0) \approx -f_0$; consequently, $\mu \approx \frac{b}{1-b} \frac{E(M)/N - f_0}{\log_b(N/N_0)}$ which leads to the equation (2) by joining with $\log_b(N/N_0) = \ln(N/N_0)/\ln(b)$ and $r_b = \ln(b)/(1 - 1/b)$ and replacing $E(M)/N$ by its unbiased estimator M/N .

A3

Derivation of the equation (3): Assume $N_0/N \approx 0$, which is true for most of applications. Let $g = \log_b(N/N_0)$, then the equation for $\text{Var}(M)$ in (1) leads to

$\text{Var}(M) \approx \frac{(1-f_0)\mu N^2}{bN_0} [1 - (1 - 1/b)\mu]^{2g-1} \approx \frac{(1-f_0)\mu N^2}{bN_0} [1 - (2g - 1)(1 - 1/b)\mu + O(\mu^2)]$. Since M is the only random number in the right side of the equation (2), we have

$\text{Var}(\widehat{\mu}) = \frac{r_b^2}{\{\ln(N/N_0)\}^2} N^2 \text{Var}(M)$. Then we get $\text{Var}(\widehat{\mu}) \approx \frac{r_b(1-f_0)}{bN_0} \left[\frac{r_b\mu}{\{\ln(N/N_0)\}^2} - \frac{2\mu^2}{\ln(N/N_0)} \right]$ by joining two equations above. If μ is much smaller than $\frac{1}{\ln(N/N_0)}$, then the second term of $\text{Var}(\widehat{\mu})$ above can be ignored and thus the equation (3) holds.

A4

Derivation of the equation (10): For $\widehat{\mu}_g$ in (9) and $\widehat{\mu}$ in (2), we have $\widehat{\mu}_g = \widehat{\mu}/k$ and thus

$\text{Var}(\widehat{\mu}_g) = \text{Var}(\widehat{\mu})/k^2$, which leads to (10) by joining with the equation (3) and the identity $\mu_g = \mu/k$.

References

- [1]. Armitage PJ. The Statistical Theory of Bacterial Populations Subject to Mutation. *Journal of The Royal Statistics Society* 1952;B14:1–40.
- [2]. Drake, JW. *The Molecular Basis of Mutation*. Holden-Day; San Francisco: 1970.
- [3]. Haccou, P.; Jagers, P.; Vatutin, V. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press; Cambridge, UK: 2005.
- [4]. Hay AJ, Wolstenholme AJ, Skehel JJ, Smith MH. The molecular basis of the specific antiinfluenza action of amantadine. *European Molecular Biology Organization Journal* 1985;4:3021–3024.
- [5]. Karlin, S.; Taylor, HM. *A First Course in Stochastic Process*. Academic Press; New York: 1975.
- [6]. Kendal WS, Frost P. Pitfalls and Practice of Luria-Delbruck Fluctuation Analysis: A Review. *Cancer Research* 1988;48:1060–1065. [PubMed: 3277705]
- [7]. Kimmel, M.; Axelrod, D. *Branching Processes in Biology*. Springer-Verlag; New York: 2002.
- [8]. Lea DE, Coulson CA. The Distribution of the Numbers of Mutants in Bacterial Populations. *Journal of Genetics* 1949;49:264–285.
- [9]. Li IC, Fu J, Hung YT, Chu EHY. Estimation of Mutation Rates in Cultured Mammalian Cells. *Mutation Research* 1983;111:253–262. [PubMed: 6633553]
- [10]. Luria SE, Delbruck M. Mutation of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* 1943;28:492–511.
- [11]. Rao, CR. *Linear Statistical Inference and Its Applications*. Vol. 2nd Ed.. Wiley; New York: 1973.
- [12]. Rossman TG, Goncharova EI, Nadas A. Modeling and Measurement of the Spontaneous Mutation Rate in Mammalian Cells. *Mutation Research* 1995;328:21–30. [PubMed: 7898501]
- [13]. Stech J, Xiong X, Scholtissek C, Webster RG. Independence of Evolutionary and Mutational Rates after Transmission of Avian Influenza Viruses to Swine. *Journal of Virology* 1999;73(3):1878–1884. [PubMed: 9971766]
- [14]. Stewart FM, Gordon DM, Levin BR. Fluctuation Analysis: The Probability Distribution of the Number of Mutants Under Different Conditions. *Genetics* 1990;124:175–185. [PubMed: 2307353]

Table 1

Number of Cultures for Estimating Mutation Rate¹

True μ	Base b	$\frac{\sigma_{\hat{\mu}}}{\mu}$	$N = 10^9$			$N = 10^{10}$			$N = 10^{11}$		
		γ	$N_0 = 1$	$N_0 = 10$		$N_0 = 1$	$N_0 = 10$		$N_0 = 1$	$N_0 = 10$	
$7 \cdot 10^{-6}$	2	0.5	319.6	40.5		258.9	32.0		214.0	25.9	
		1.0	79.9	10.1		64.7	8.0		53.5	6.5	
		1.5	35.5	4.5		28.8	3.6		23.8	2.9	
	5	0.5	269.3	34.1		218.1	26.9		180.3	21.8	
		1.0	67.3	8.5		54.5	6.7		45.1	5.5	
$3 \cdot 10^{-5}$		1.5	29.9	3.8		24.2	3.0		20.0	2.4	
	10	0.5	217.7	27.6		176.4	21.8		145.8	17.6	
		1.0	54.4	6.9		44.1	5.4		36.4	4.4	
		1.5	24.2	3.1		19.6	2.4		16.2	2.0	
	2	0.5	74.6	9.4		60.4	7.5		49.9	6.0	
		1.0	18.6	2.4		15.1	1.9		12.5	1.5	
		1.5	8.3	1.0		6.7	0.8		5.5	0.7	
	5	0.5	62.8	8.0		50.9	6.3		42.1	5.1	
		1.0	15.7	2.0		12.7	1.6		10.5	1.3	
		1.5	7.0	0.9		5.7	0.7		4.7	0.6	
	10	0.5	50.8	6.4		41.2	5.1		34.0	4.1	
		1.0	12.7	1.6		10.3	1.3		8.5	1.0	
		1.5	5.6	0.7		4.6	0.6		3.8	0.5	

¹ Calculated by equation (16), assuming $f(0) = 0$, $k = 4$; $\gamma = \frac{\sigma_{\hat{\mu}}}{\mu}$, where $\hat{\mu}$ is either $\hat{\mu}_a$ or $\hat{\mu}_w$.

Table 2

Number of Cultures for Comparing Mutation Rates^{1,2}

True μ_2	Base b	$\frac{\mu_1 - \mu_2}{\mu_2}$	d	$N=10^9$	$N_0=1$	$N_0=10$	$N=10^{10}$	$N_0=1$	$N_0=10$	$N=10^{11}$	$N_0=1$	$N_0=10$
$7 \cdot 10^{-6}$	2		1	1482.2		187.6	1200.6		148.2	992.2		120.1
			3	274.5		34.7	222.3		27.4	183.7		22.2
			5	138.3		17.5	112.1		13.8	92.6		11.2
	5		1	1248.6		158.0	1011.3		124.9	835.8		101.1
			3	231.2		29.3	187.3		23.1	154.8		18.7
			5	116.5		14.7	94.4		11.7	78.0		9.4
	10		1	1009.6		127.8	817.8		101.0	675.9		81.8
			3	187.0		23.7	151.4		18.7	125.2		15.1
			5	94.2		11.9	76.3		9.4	63.1		7.6
$3 \cdot 10^{-5}$	2		1	345.8		43.8	280.1		34.6	231.5		28.0
			3	64.0		8.1	51.9		6.4	42.9		5.2
			5	32.3		4.1	26.1		3.2	21.6		2.6
	5		1	291.3		36.9	236.0		29.1	195.0		23.6
			3	54.0		6.8	43.7		5.4	36.1		4.4
			5	27.2		3.4	22.0		2.7	18.2		2.2
	10		1	235.6		29.8	190.8		23.6	157.7		19.1
			3	43.6		5.5	35.3		4.4	29.2		3.5
			5	22.0		2.8	17.8		2.2	14.2		1.8

¹ Calculated by equation (17) for testing $H_0 : \mu_1 \leq \mu_2$ vs. $H_a : \mu_1 > \mu_2$.

² Significance level $\alpha = 0.05$, power $1 - \beta = 0.8$; assuming $\beta_0 = 0$, $k = 4$, and $C = CI = C2$.