Published in final edited form as:

Methods Enzymol. 2006 ; 409: 195–213.

Methods for Determining Spontaneous Mutation Rates

Patricia L. Foster

Department of Biology, Indiana University, 1001 E. Third Street, Bloomington, IN 47405, USA, E-mail plfoster@indiana.edu, Tel. (+1) 812 855 4084, Fax (+1) 812 855 6705

Abstract

Spontaneous mutations arise as a result of cellular processes that act upon or damage DNA. Accurate determination of spontaneous mutation rates can contribute to our understanding of these processes and the enzymatic pathways that deal with them. The methods that are used to calculate mutation rates are based on the model for the expansion of mutant clones originally described by Luria and Delbrück and extended by Lea and Coulson. The accurate determination of mutation rates depends on understanding the strengths and limitations of these methods and how to optimize a fluctuation assay for a given method. This chapter describes the proper design of a fluctuation assay, several of the methods used to calculate mutation rates, and ways to evaluate the results statistically.

Introduction

Spontaneous mutations are mutations that occur in the absence of exogenous agents. They may be due to errors made by DNA polymerases during replication or repair, errors made during recombination, the movement of genetic elements, or spontaneously occurring DNA damage. The rate at which spontaneous mutations occur can yield useful information about cellular processes. For example, the occurrence of specific classes of mutations in different mutant backgrounds has been used to deduce the importance of various DNA repair pathways (Miller, 1996).

The mutation rate is the expected number of mutations that a cell will sustain during its lifetime. The mutant fraction or frequency is the proportion of cells in a population that are mutant.¹ Although mutant frequencies can be adequate indicators of the rate at which mutations are induced by DNA damaging agents, they are inadequate indicators of spontaneous mutation rates. This is because the population of mutants is composed of clones, each of which arose from a cell that sustained a mutation. The size of a given mutant clone will depend on when during the growth of the population the mutation occurred. This is the fundamental property of spontaneous mutation that was exploited in the famous Luria and Delbrück fluctuation test (Luria and Delbrück, 1943). Among replicate cultures, the distribution of the numbers of mutations that were sustained is Poisson, but the distribution of the numbers of mutants that result is far from Poisson, and is usually referred to as the Luria-Delbrück distribution.

There are two basic methods to determine mutation rates: mutant accumulation and fluctuation analysis. These two methods are described below, with emphasis on fluctuation analysis.

Terminology

The definitions of the terms used in this chapter are given in Table I. It is important to distinguish between m , the mean number of mutations that occur during the growth of a culture, and μ = the mutation rate, which is the mean number of mutations that occur during the lifetime of a cell. Almost all methods to calculate mutation rates start by determining m and then obtain

¹A mutation is a heritable change in the genetic material; a mutant is an individual that carries a mutation.

μ by dividing m by some measure of the number of cell-lifetimes at risk for mutation, usually N_t . It is also important to distinguish between the number of mutations per culture, m , and the number of mutants per culture, r . r divided by N_t is the mutant fraction or mutant frequency, f .

The Lea-Coulson Model

The methods for calculating mutation rates discussed below are dependent on the model of expansion of mutant clones originally described by Luria and Delbrück (Luria and Delbrück, 1943) and extended by Lea and Coulson (Lea and Coulson, 1949). It is usually called the Lea-Coulson model and has the following assumptions.

- 1) the cells are growing exponentially
- 2) the probability of mutation is not influenced by previous mutational events
- 3) the probability of mutation is constant per cell-lifetime
- 4) the growth rates of mutants and nonmutants are the same
- 5) the proportion of mutants is always small
- 6) reverse mutations are negligible
- 7) cell death is negligible
- 8) all mutants are detected
- 9) no mutants arise after selection is imposed

In addition, for assays using batch cultures, the following assumptions apply

- 10) the initial number of cells is negligible compared to the final number of cells
- 11) the probability of mutation per cell-lifetime does not vary during the growth of the culture

Departures from the Lea-Coulson model affect the distribution of mutant numbers and impact the mutation rate calculation. In general, most of the model's requirements can be met with proper experimental protocols, but some departures reflect real biological phenomenon.

Mutant Accumulation

When a population is growing exponentially the appearance of new mutants plus the proliferation of preexisting mutants results in a constant increase in the mutant fraction each generation. The mutation rate is this increase, and can be determined by measuring the change in the mutant fraction over time (Figure 1). However, an important caveat is that the population must be of sufficient size so that the probability that mutations occur each generation is essentially unity; otherwise, the chance occurrence of mutations will dominate the population (Luria, 1951). For batch cultures, the population must be large enough so that the average number of mutations per culture, m , is much greater than one. For a continuously dividing cell population, μ can be calculated by Eq. (1) (Drake, 1970):

$$\mu = \left(\frac{r_2}{N_2} - \frac{r_1}{N_1} \right) \div \left[\ln \left(\frac{N_2}{N_1} \right) \right] = (f_2 - f_1) \div (\ln N_2 - \ln N_1) \quad \text{Eq. (1)}$$

Although conceptually simple, there are important technical difficulties that limit the use of this method. In general, by the time the population reaches a sufficient size, some mutations have already occurred, and these produce clones of mutants that make it impossible to accurately measure the accumulation of new mutants. Thus, to measure mutant accumulation

a large population with few mutants must be generated. One way to do this is by generating and testing a number of populations and using the ones with a low mutant fraction. This is convenient if the population can be stabilized and used repeatedly for experiments, for example stocks of bacteriophage or viruses. Alternatively, a population can be purged of pre-existing mutants if the mutational target gives a phenotype that can be selected both for and against (*i.e.* a “counterselectable marker”). Selection against the mutants can be used to purge the population, and then selection for the mutants can be used to measure mutation rates. For example, Lac⁺ bacteria with a temperature sensitive *galE* mutation are selected against by lactose at the nonpermissive temperature, but selected for by lactose at the permissive temperature (Reddy and Gowrishankar, 1997). In mammalian cells, mutations in the hypoxanthine-guanine phosphoribosyl transferase (*hprt*) locus make cells sensitive to HAT medium (which contains hypoxanthine, aminopterin, and thymidine) but resistant to 6-thioguanine (Glaab and Tindall, 1997). A few other counterselectable markers are available (Reyrat et al., 1998). Another possibility is to use a mutational target that allows mutants to be eliminated by cell sorting. For example, mutants that allow green fluorescent protein (GFP) to be produced can be eliminated by fluorescence-activated cell sorting (FACS); new mutants can then be detected by flow cytometry (Bachl et al., 1999).

Mutant accumulation has been extensively used to measure mutation rates in chemostats (Kubitschek and Bendigkeit, 1964; Novick and Szilard, 1950). Cell number, N , is constant in a chemostat, so mutant accumulation is a function of the growth rate, λ , and the mutation rate per cell per generation μ . Thus:

$$\mu = \frac{1}{N\lambda} \frac{(r_2 - r_1)}{(t_2 - t_1)} \quad \text{Eq. (2)}$$

Where t_1 and t_2 are the times at which the numbers of mutants, r_1 and r_2 , are measured.

Fluctuation Analysis

Experimental Design

A normal fluctuation test begins by inoculating a small number of cells into a large number of parallel cultures. The cultures are allowed to grow, usually to saturation, and then each culture is plated on a selective medium that allows the mutants to produce colonies. The total number of cells is determined by plating appropriate dilutions of a few cultures on nonselective medium. The distribution of the numbers of mutants among the parallel cultures is used to calculate the mutation rate. This basic design, invented by Luria and Delbrück (Luria and Delbrück, 1943), can be used for single-celled microorganisms, cultured cells, bacteriophage, and viruses. So that individual mutants can be counted, a solid medium is usually used for selection, but the p_0 method (see below) also can be used with liquid medium.

The goal of designing a fluctuation assay is to maximize the precision with which the mutation rate is estimated. Precision is a measure of reproducibility, not accuracy (accuracy is how well the resulting estimate reflects the actual mutation rate, and that will depend on how well the underlying assumptions reflect reality). The important design parameters are: m , the number of mutations per culture; r , the number of mutants per culture; N_0 , the initial number of cells; N_t , the final number of cells; V , the culture volume; and, C , the number of parallel cultures. The first step is to determine a preliminary r and m by plating aliquots from a few parallel cultures on the selective medium. A preliminary m is calculated from the mutant numbers obtained (*eg.*, using Method 2 or 3) and then the other parameters adjusted so that the final m is within a useful range. The value of m will determine which methods can be used to calculate the mutation rate. None of the methods are reliable if m is less than 0.3 unless a prohibitive number of cultures are used. However, if m is above 15 some of the methods are not valid

(Rosche and Foster, 2000). Obviously, the final m also has to be small enough so that the number of colonies on the selective plates is countable. However, if there are a few outlier high counts, they can be truncated at 150 with little loss of precision (Asteris and Sarkar, 1996; Jones et al., 1999).

The desired m is achieved by adjusting N_t , the final number of cells, either by manipulating the cell density or the culture volume. When using defined medium, cell density can be adjusted by limiting the carbon source. However, using other required growth factors, such as vitamins or amino acids, is not recommended because non-requiring mutants will be selected and cell physiology may change. For example, in tryptophan-limited chemostats, bacterial mutation rates become time-dependent instead of generation-dependent (Kubitschek and Bendigkeit, 1964).

The desired N_t can be achieved by adjusting the culture volume. It is usually considered necessary to plate all the cells from each culture on the selective medium because sampling, or low plating efficiency (which is the same thing), increases the proportion of cultures with small numbers of mutants and narrows the distribution (Crane et al., 1996; Stewart et al., 1990). But this requirement restricts the volume of culture that can be used without concentrating the cells (which can be tedious with many cultures). However, if several mutant phenotypes are to be assayed in the same cultures, sampling is unavoidable. In addition, because a large culture contains more “information” than a small culture, it is better to plate a small aliquot from a large culture than all of a small culture if a proper correction can be applied (Jones et al., 1999). Some, but not all, of the methods for calculate mutation rates discussed below are amenable to such corrections.

The validity of the mutation rate calculation requires that N_t be the same in each culture. Usually, but not always, this can be accomplished by growing cells to saturation. If achieving an uniform N_t is a problem, the cell number in each culture can be monitored before mutant selection by measuring the optical density or by counting cells microscopically (*e.g.* using a Petroff-Hausser chamber). Because there is currently no valid method to correct for different N_t 's, deviant cultures must be eliminated from the analysis.

The initial inoculum, N_0 , must contain no preexisting mutants and must be small relative to N_t . Most of the methods to calculate the mutation rate are valid if N_0 is at least 1/1000 of N_t (Sarkar et al., 1992), but this may not insure that N_0 contains no mutants. A reasonable rule of thumb is that N_0 roughly equals $(N_t/m) \times 10^{-5}$. the best way to insure uniformity is to grow a starter culture in the same medium that will be used for the fluctuation assay, dilute these cells to the appropriate density in a large volume of fresh medium, and then distribute aliquots into individual cultures tubes for nonselective growth.

The precision of the estimate of m depends on C , the number of parallel cultures. Most experiments have 10 to 100 parallel cultures, with about 40 being most common. There is little gain in precision if C is larger unless m is less than about 0.3 (Jones et al., 1999; Rosche and Foster, 2000).

The fluctuation test was originally designed to test whether mutations occur before or after selection is imposed (Luria and Delbrück, 1943). If the selection is lethal, then the only mutants that appear must have pre-existed. However, if the selection is not lethal, for example reversion of an auxotrophy or utilization of a carbon source, mutants can arise both before and after selection has been applied. Post-plating mutants can arise because the cells are proliferating on the selective medium (*i.e.* the selection is not stringent), or they can result from mutations that occur in non-growing cells (adaptive mutations). In either case, the distribution of mutant numbers will be a combination of the Luria-Delbrück and Poisson, and the m estimated for pre-plating mutations will be inflated by the post-plating mutations (Cairns et al., 1988). If

non-mutant cells grow on the selective medium because of contaminating nutrients, one solution is to add an excess of scavenger cells that cannot mutate (because, for example, they have a deletion of the relevant gene) to consume the contaminants (Cairns and Foster, 1991). If the non-mutant cells grow because the selection is not stringent, the time it takes for a mutant to form a colony on the selective medium can be determined and then mutant colonies can be counted at the earliest possible time after plating.

Analyzing the Results of Fluctuation Assays

Fluctuation assays give the distribution of the numbers of mutants per culture, r , which is used to calculate m , the mean or most likely number of mutations per culture. m is not itself a particularly interesting parameter since it depends on the cell density and the volume of the culture. However, it is mathematically tractable and yields the mutation rate when divided by some measure of the number of cells. Although the distribution of mutants is not Poisson, the distribution of mutations is, so m is a Poisson parameter. There are many methods to calculate m (often called estimators) but they are all based on the theoretical distribution of mutant clone-sizes described by Luria and Delbrück (Luria and Delbrück, 1943) and Lea and Coulson (Lea and Coulson, 1949). Each method has its advantages and disadvantages, and the choice of method depends on the particular conditions of the experiment and the mathematical sophistication and persistence of the user. The MSS maximum likelihood method (Method 5) is the gold standard because it utilizes all of the results of an experiment and is valid over the entire range of mutation rates. Of the less complicated methods, the Lea-Coulson method of the median (Method 2) and the Jones median estimator (Method 3) are reliable when mutation rates are low to moderate, and the p_0 method (Method 1) can be used when mutation rates are very low ($m \leq 1$). Drake's Formula (Method 4) is particularly useful when comparing data reported as mutant frequencies instead of mutation rates. Methods 6 and 7 can be useful when not all the requirements of the clone-size distribution are met. No method using the mean number of mutants is valid, and none are given here. To see how these various methods behave with real data, see Rosche and Foster, 2000 (Rosche and Foster, 2000).

Method 1: The p_0 method The distribution of the number of mutations that occur during the growth of parallel cultures has a Poisson distribution. If there are no mutants, there were no mutations, and so the mean number of mutations can be calculated from p_0 , the proportion of cultures with no mutants (Luria and Delbrück, 1943). The zeroth term of the Poisson distribution is:

$$p_0 = e^{-m} \quad \text{Eq. (3)}$$

So m is:

$$m = -\ln p_0 \quad \text{Eq. (4)}$$

Although simple, the p_0 method is limited. Its range of usefulness is $0.7 \geq p_0 \geq 0.1$ ($0.3 \leq m \leq 2.3$) and it performs best when p_0 is about 0.3. The p_0 method is inefficient (*i.e.*, requires more cultures for the same precision) compared to other methods (Koziol, 1991; Rosche and Foster, 2000). In addition, p_0 is sensitive to several biologically relevant factors that complicate fluctuation analysis. Phenotypic lag (the delay in expression of a phenotype), poor plating efficiency, and selection against mutants all inflate p_0 and result in an erroneously low m . However, if all cells (not just mutants) have a plating efficiency of less than one, a correction factor can be applied to m . The same correction can be applied if only a fraction, of each culture is plated (Jones, 1993; Stewart et al., 1990). The actual m is calculated from the observed m using Eq. (5) where z is either the plating efficiency or the fraction plated²:

$$m_{act} = m_{obs} \frac{z-1}{z \ln(z)} \quad \text{Eq. (5)}$$

Method 2: Lea-Coulson median estimator This method is based on the observation that for $4 \leq m \leq 15$, a plot of the probability that a culture contains r or fewer mutants, versus $\frac{r}{m} - \ln(m)$ gives a skewed curve about a median of 1.24 (Lea and Coulson, 1949). Rearranging gives the following transcendental equation relating m to the median, $\tilde{r}$:

$$\frac{\tilde{r}}{m} - \ln(m) - 1.24 = 0 \quad \text{Eq. (6)}$$

that can be solved easily by iteration (an example of how to use a spread sheet to solve this equation is given in Figure 2). The Lea and Coulson method of the median is easy to apply and remarkably accurate with in computer simulations (Asteris and Sarkar, 1996; Stewart, 1994) and with real data (Rosche and Foster, 2000). It performs well over the range

$2.5 \leq \tilde{r} \leq 60$ ($1.5 \leq m \leq 15$) (Rosche and Foster, 2000). Because it uses the median, it is relatively insensitive to deviations that affect either end of the distribution, especially if $\tilde{r}$ is relatively large. However, because little of the information obtained from the fluctuation test is used, the method is relatively inefficient (Rosche and Foster, 2000).

Method 3: The Jones median estimator This estimator is based on the theoretical dilution of the experimental cultures that would be necessary to produce a distribution with a median of 0.5 (Jones et al., 1994). The basic equation is:

$$m = \frac{\tilde{r} - \ln(2)}{\ln(\tilde{r}) - \ln[\ln(2)]} = \frac{\tilde{r} - 0.693}{\ln(\tilde{r}) + 0.367} \quad \text{Eq. (7)}$$

Two advantages of the Jones estimator are that it is explicit and that it accommodates dilutions.

If z = the fraction of the culture that is plated or the plating efficiency, and $\tilde{r}_{obs}$ is the observed median, then (Jones et al., 1994):

$$m = \frac{\left(\tilde{r}_{obs} \middle| z \right) - 0.693}{\ln \left(\tilde{r}_{obs} \middle| z \right) + 0.367} \quad \text{Eq. (8)}$$

In computer simulations over the range $3 \leq \tilde{r} \leq 40$ ($1.5 \leq m \leq 10$) the Jones estimator proved to be as reliable and more efficient than the Lea and Coulson median estimator (Jones et al., 1994). The Jones estimator also performs well with real data (Rosche and Foster, 2000).

Method 4: Drake's formula Drake's formula (Drake, 1991) is an easy way to calculate mutation rates from mutant frequencies, and is especially useful in comparing data from different sources. Because it uses frequencies, Drake's formula gives the mutation rate, μ , instead of m (with $\mu = \frac{m}{N_t}$). Starting from Eq. (1) above, Drake sets N_1 to be $1/\mu$, the population size at which the probability of mutation approaches unity. Assuming that no

²When $z = 1$, $\frac{z-1}{z \ln(z)} = 1$ from l'Hôpital's rule, and $m_{act} = m_{obs}$

mutations occur before the population reached this size, f_1 is zero; f_2 is the final mutant frequency, f ; and, N_2 is the final population size, N_t . This gives:

$$u = f \div \ln(uN_t) \quad \text{Eq. (9)}$$

that can be solved for μ by iteration. Drakes' formula is based on the same assumption discussed above for Eq. (1), *i.e.* that mutations occur only during the deterministic period of mutant accumulation. Using the median frequency (if available) instead of the mean minimizes the influence of jackpots (Drake, 1991). Since $\mu = \frac{m}{N_t}$, Drake's formula can be rearranged into the same form as Lea and Coulson's formula, Eq. (6):

$$\frac{\tilde{r}}{m} - \ln(m) = 0 \quad \text{Eq. (10)}$$

When $m < 4$, estimates of m obtained with Eq. (10) are significantly higher than those obtained with Eq. (6), but asymptotically approach those obtained with Eq. (6) as m becomes larger. If $m \geq 30$, the differences are trivial (Rosche and Foster, 2000).

Method 5: The MSS-maximum likelihood method (Sarkar et al., 1992) described a recursive algorithm based on the Lea-Coulson generating function (Lea and Coulson, 1949) that efficiently computes the Luria-Delbrück distribution for a given value of m^3 . Known as the MSS algorithm, it is:

$$p_0 = e^{-m}; \quad p_r = \frac{m}{r} \sum_{i=0}^{r-1} \frac{p_i}{(r-i+1)} \quad \text{Eq. (11)}$$

Note that the equation to calculate p_0 is the same as Eq. (3) and the proportion of cultures with each of the other possible values of r is given by the equation on the right. The algorithm is recursive, meaning that at a given m , the proportions of cultures with 0, 1, 2, 3, *etc.* mutants are $p_0 = e^{-m}$; $p_1 = m \times \frac{p_0}{2}$; $p_2 = \frac{m}{2} \times \left(\frac{p_0}{3} + \frac{p_1}{2} \right)$;

$p_3 = \frac{m}{3} \times \left(\frac{p_0}{4} + \frac{p_1}{3} + \frac{p_2}{2} \right)$; *etc.* for all possible values of r .

This algorithm can be used to estimate m from experimental results using a maximum likelihood function, the formula for which is (Ma et al., 1992):

$$f(r | m) = \prod_{i=1}^C f(r_i | m) \quad \text{Eq. (12)}$$

where $f(r | m) = p_r$ from Eq. (11) and C is the number of cultures. The procedure is to start with a trial m (obtained from Eq. (6), for example) and use Eq. (11) to calculate the probability, p_r , of obtaining each possible r from 0 to the maximum value obtained (even if a given r was not obtained in the experiment it has to be included in the recursive equation). As mentioned above, for most experiments, values of r greater than 150 can be lumped into one category (Asteris and Sarkar, 1996). The likelihood function, Eq. (12), is the product of these calculated p_r 's for r obtained in the experiment. The easiest way to do this calculation is to arrange the

³The algorithm itself was described earlier (Gurland, 1958; Gurland, 1963) but was independently derived by Sarkar and coworkers and applied to fluctuation analysis (Ma et al., 1992; Sarkar et al., 1992)

mutant counts in order and count the number of cultures that had each r mutants, c_r . Then the product of the p_r 's is:

$$(p_0)^{c_0} \times (p_1)^{c_1} \times (p_2)^{c_2} \times (p_3)^{c_3} \times (p_4)^{c_4} \dots \quad \text{Eq. (13)}$$

where each p_r is from Eq. (11) (alternatively $c_r \ln(p_r)$ for each r can be added). Note that values of r that were not obtained in the experiment give a value of 1 in Eq. (13) and so do not have to be included. The procedure is repeated with different m 's over a small range until a m that maximizes Eq. (13) is found.

As mentioned above, the MSS-maximum likelihood method is the best method currently available to estimate m . It uses all the results from a fluctuation experiment and is valid over the entire range of values of m . In addition, computer simulations have shown it to behave in a manner that allows statistical evaluation (Stewart, 1994). A comparison with other methods using real data can be found in Rosche and Foster (Rosche and Foster, 2000)

Method 6: Accumulation of clones Luria (Luria, 1951) pointed out that P_r , the proportion of cultures with r or more mutants, approaches $2m/r$ during the deterministic portion of growth (*i.e.* when m is 1 or greater). Formally:

$$P_r = \sum_{i=r}^{i=N_t} p_i = \frac{2m}{r} \quad \text{Eq. (14)}$$

Taking logarithms gives

$$\ln(P_r) = -\ln(r) + \ln(2m) \quad \text{Eq. (15)}$$

A plot of $\ln(P_r)$ versus $\ln(r)$ will yield a straight line with a slope of -1 and an intercept (where $\ln(r) = 0$) equal to $\ln(2m)$. Dividing P_r at the intercept by 2 gives m .

Method 7: The quartiles method The median is the central (50%) quartile of a distribution. More of fluctuation assay can be incorporated in the calculation of m if the upper (75%) and lower (25%) quartiles are also used. By regressing m versus the theoretical values of r at the quartiles, Koch (Koch, 1982) derived the following empirical equations:

$$m_1 = 1.7335 + 0.4474 \quad Q_1 - 0.002755(Q_1)^2 \quad \text{Eq. (16)}$$

$$m_2 = 1.1580 + 0.2730 \quad Q_2 - 0.000761(Q_2)^2 \quad \text{Eq. (17)}$$

$$m_3 = 0.6658 + 0.1497 \quad Q_3 - 0.0001387(Q_3)^2 \quad \text{Eq. (18)}$$

where Q_1 , Q_2 and Q_3 are the values of r at 25%, 50%, and 75% of the ranked series of observed r 's. For a perfect Luria-Delbrück distribution, the three m 's should be equal. These equations are valid over the range $2 \leq m \leq 14$; Koch also gives graphs that can be used up to values of $m = 120$ (Koch, 1982).

Calculating the Mutation Rate

The mutation rate, μ , is the mean number of mutations, m , normalized to some measure of the number of cells at risk for mutation. Three such measures are used, each of which is based on different assumptions about the underlying mutational process. If the probability of mutation is constant over the cell cycle, then m should be divided by the number of cell divisions that

have taken place. Since the final number of cells in a culture, N_t , arose from N_t-1 divisions, the mutation rate is (Luria and Delbrück, 1943):

$$u = \frac{m}{(N_t - 1)} \approx \frac{m}{N_t} \quad \text{Eq. (18)}$$

The same calculation applies if mutations are assumed to occur at or during division, (Armitage, 1952). If mutations are assumed to occur at the beginning of the cell cycle (*i.e.* shortly after division), then m should be divided by the total number of cells that ever existed in the culture, which is $2 N_t$ (because N_t cells had $N_t/2$ parents, $N_t/4$ grandparents, etc., and the sum of the series is $2 N_t$). Thus, the mutation rate is (Armitage, 1952):

$$u = \frac{m}{2 N_t} \quad \text{Eq. (19)}$$

However, cells usually are not growing synchronously, and in an asynchronous population there are an average of $N/\ln(2)$ cells during one generation period. Thus, the total number of divisions during the growth of a culture is $N_t/\ln(2)$ and the mutation rate is (Armitage, 1952):

$$u = \frac{m \ln(2)}{N_t} = 0.6932 \frac{m}{N_t} \quad \text{Eq. (21)}$$

For the same m , these three equations will give mutation rates that differ by 1: 0.5: 0.693. It is best to use one consistently, and to describe which one was used so that readers can compare results obtained with different methods.

Statistical methods to evaluate mutation rates

The estimates of m or μ obtained from fluctuation tests are neither normally distributed nor unbiased; therefore, no matter how many times a fluctuation experiment is repeated, it is not valid to take the mean and standard deviation of the results (Asteris and Sarkar, 1996; Jones et al., 1994; Stewart, 1994). There are two approaches that allow reasonable confidence limits to be placed around estimates of mutation rates. The first approach is to put confidence limits around the parameter used to calculate m and calculate new m 's using these values; the new m 's will be estimates for the confidence limits of m (Wierdl et al., 1996). This approach is valid only for parameters that have defined distributions, such as p_0 and the median. The second approach is to find a transforming function that gives m a normal distribution; this has been successful only for the MSS maximum likelihood method (Method 5). Once confidence limits are obtained for m , these can be divided by N_t (or $2 N_t$ or $N_t/\ln(2)$) to estimate the confidence limits for μ , the mutation rate. Of course, this procedure ignores the variance of the determination of N_t , (which is approximately N_t). Although nontrivial, it is probably justifiable to ignore this variance as long as N_t is determined accurately.⁴

Confidence Limits for p_0 —Because a culture either has mutants or it does not, p_0 can be considered a binomial parameter with $p_0 = p$ and $(1 - p_0) = q$ (Lea and Coulson, 1949). For a sample population of n , the standard deviation of p is

$$\sigma_p = \sqrt{\frac{pq}{n-1}} \quad \text{Eq. (22)}$$

⁴It is a little-appreciated fact that the expected value of the ratio of two variables is not the ratio of their expected values (*i.e.* not the ratio of their means). Furthermore, the calculation of the variance of a ratio is fairly complicated (*e.g.* see (Rice, 1995). However, if the denominator is larger than the numerator, the variance of the ratio will be smaller than the variance of the numerator, and thus no great harm should be done by ignoring the variance of the denominator.

But for most fluctuation assays, using the binomial is inappropriate and gives meaningless intervals. There are several more widely applicable methods to calculate the CLs for a proportion; the following one uses the F statistic (Zar, 1984):

$$CL_{Upper} = \frac{(np + 1)F_{df1}}{nq + (np + 1)F_{df1}} \quad \text{Eq. (23)}$$

$$CL_{Lower} = \frac{np}{np + (nq + 1)F_{df2}} \quad \text{Eq. (24)}$$

where F is evaluated at the desired α (α = level of significance) and the following degrees of freedom:

$$df1 : \nu_1 = 2(nq + 1); \nu_2 = 2np \quad \text{Eq. (25)}$$

$$df2 : \nu_1 = 2(np + 1); \nu_2 = 2nq \quad \text{Eq. (26)}$$

Confidence limits for the median—The median is, by definition, the value at which the cumulative binomial probability is equal to 0.5 (*i.e.* 50% of the values are above and 50% of the values are below the median). Therefore, the 95% CLs for the median are the values above and below which less than 5% of the values are expected to fall, given n trials and a probability of 0.5. These values can be found by calculating the binomial probabilities for each possible rank-value of a given sample population and then finding the upper and lower rank-values that symmetrically include 95% probability (or any other desired probability). The CLs for the median are the actual sample values that correspond to these rank-values. Conveniently, the binomial probabilities have already been calculated in tables found in many statistic books (*e.g.*, (Zar, 1984). Thus, to calculate CLs for the median

1. Use a table or calculate the binomial probability for each possible rank-value (including 0) for the given population size, $n = C$, and $p = q = 0.5$.
2. Pick i , the highest rank-value that has a probability equal to or less than $\alpha/2$.
3. Pick the $j = n - (i + 1)$ rank-value.
4. Order the samples by increasing value
5. Pick the $(i + 1)^{th}$ and the j^{th} values. These are the CLs for the median value.

Confidence limits for m obtained from the MSS maximum likelihood method—

Using simulated fluctuation tests, Stewart (Stewart, 1994) evaluated the distributions of m 's obtained using several of the common methods. He found that the natural logarithms of m 's obtained using the MSS maximum likelihood method (Method 5) are approximately normally distributed. From this, Stewart calculated the standard deviation of $\ln(m)$:

$$\sigma_{\ln(m)} \approx \frac{1.225m^{-0.315}}{\sqrt{C}} \quad \text{Eq. (27)}$$

where C is the number of cultures. Since $\ln(m)$ is normally distributed, the 95% confidence limits for $\ln(m)$ should be

$$\ln(m) \pm 1.96\sigma_{\ln(m)} \quad \text{Eq. (28)}$$

While this is a reasonable approximation, the true confidence limits must be calculated from the actual m and σ of the population, not the experimentally determined m and σ of the sample.

Methods to calculate or estimate the correct confidence limits are given in Stewart (Stewart, 1994). A close approximation can be obtained from the following equations (Rosche and Foster, 2000)

$$CL_{+95\%} = \ln(m) + 1.96\sigma(e^{1.96\sigma})^{-0.315} \quad \text{Eq. (29)}$$

$$CL_{-95\%} = \ln(m) - 1.96\sigma(e^{1.96\sigma})^{+0.315} \quad \text{Eq. (30)}$$

Once the upper and lower limits for $\ln(m)$ are obtained, the upper and lower limits for m are simply the antilogs.

Departures from the Lea-Coulson Model

All the methods for calculating mutation rates discussed above depend on the Lea-Coulson model of expansion of mutant clones (Lea and Coulson, 1949); therefore, calculated mutation rates will be wrong if the assumptions of the model are violated. However, with some care, several biological meaningful departures can be accommodated, and meaningful mutation rates derived.

Sampling or low plating efficiency—If only a sample of a culture is plated, or if the cells (not just the mutants) have a plating efficiency less than 100%, all clones will be reduced in size by the same relative amount and m will be too small. The observed m can be corrected using Eq. (5) if Method 1 is used, or Eq. (8) if Method 3 is used.

Phenotype lag—If the expression of a mutant phenotype is delayed for several generations, mutants that arise in the last few generations of growth will result in few mutant progeny, whereas mutants that arise early will contribute a normal number. Thus, the lower end of the distribution will be affected, but, depending on the length of the lag, the upper end will not, resulting in an inflated m . The actual m can be estimated graphically with Method 6 by using only the upper part of the curve and eliminating any obvious jackpots (Rosche and Foster, 2000). If the length of the phenotypic lag is known, Koch (Koch, 1982) gives a method for estimating m from the quartiles (Method 7).

Selection against mutants—If during non-selective growth mutants grow more slowly than nonmutants, the result is the opposite of what happens if there is phenotypic lag: mutants that arise in the last few generations of growth will contribute a normal number of mutant progeny, but mutants that arise early will contribute few. This shifts the distribution of mutant numbers from the Luria-Delbrück toward the Poisson (Koch, 1982; Stewart et al., 1990). Koch (Koch, 1982) gives graphs that can be used to estimate m from the quartiles when the growth rate of mutants ranges from 60 to 90% that of nonmutants. If there is more than one type of mutant and each type has a different growth rate, the distribution can be approximated by assuming there is only one type whose growth rate is the average of the two (Stewart et al., 1990).

Adaptive mutation—If mutations occur after the cells are plated on selective medium, the distribution of mutant numbers will have a Poisson component and a plot of $\ln(P_r)$ versus $\ln(r)$ (Eq. (15)) will give a curve that is a combination of the Luria-Delbrück and Poisson (Cairns et al., 1988). The two components can be estimated by fitting the experimental values to the combined distributions (Cairns and Foster, 1991).

Other curve fitting—Using simulated data, Stewart *et al.* (Stewart et al., 1990) have determined the effects of several deviations from the Lea-Coulson model on the shape of the

$\ln(P_r)$ versus $\ln(r)$ curve. Experimental data can be fit to these curves to test whether a given factor is operative. However, all of the deviant curves are rather similar, so any conclusion that a given factor is distorting the distribution would have to be tested experimentally.

Acknowledgements

Research in the author's laboratory is supported by USPHS grant NIH-NIGMS G65175.

References

- Armitage P. The statistical theory of bacterial populations subject to mutation. *J R Statist Soc B* 1952;14:1–40.
- Asteris G, Sarkar S. Bayesian procedures for the estimation of mutation rates from fluctuation experiments. *Genetics* 1996;142:313–326. [PubMed: 8770608]
- Bachl J, Dessing M, Olsson C, von Borstel RC, Steinberg C. An experimental solution for the Luria Delbrück fluctuation problem in measuring hypermutation rates. *Proc Natl Acad Sci USA* 1999;96:6847–6849. [PubMed: 10359801]
- Cairns J, Foster PL. Adaptive reversion of a frameshift mutation in *Escherichia coli*. *Genetics* 1991;128:695–701. [PubMed: 1916241]
- Cairns J, Overbaugh J, Miller S. The origin of mutants. *Nature* 1988;335:142–145. [PubMed: 3045565]
- Crane BJ, Thomas SM, Jones ME. A modified Luria-Delbrück fluctuation assay for estimating and comparing mutation rates. *Mutat Res* 1996;354:171–182. [PubMed: 8764946]
- Drake JW. The Molecular Basis of Mutation. Holden-Day, Inc.; San Francisco, CA: 1970.
- Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 1991;88:7160–7164. [PubMed: 1831267]
- Glaab WE, Tindall KR. Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects. *Carcinogenesis* 1997;18:1–8. [PubMed: 9054582]
- Gurland J. A generalized class of contagious distributions. *Biometrics* 1958;14:229–249.
- Gurland J. A method of estimation for some generalized Poisson distributions. *Int Symp on Classical and Contagious Discrete Distributions* 1963:141–158.
- Jones ME. Accounting for plating efficiency when estimating spontaneous mutation rates. *Mutat Res* 1993;292:187–189. [PubMed: 7692255]
- Jones ME, Thomas SM, Clarke K. The application of a linear algebra to the analysis of mutation rates. *J theor Biol* 1999;199:11–23. [PubMed: 10419757]
- Jones ME, Thomas SM, Rogers A. Luria-Delbrück fluctuation experiments: design and analysis. *Genetics* 1994;136:1209–1216. [PubMed: 8005425]
- Koch AL. Mutation and growth rates from Luria-Delbrück fluctuation tests. *Mutat Res* 1982;95:129–143.
- Kozioł JA. A note of efficient estimation of mutation rates using Luria-Delbrück fluctuation analysis. *Mutat Res* 1991;249:275–280. [PubMed: 2067540]
- Kubitschek HE, Bendigkeit HE. Mutation in continuous cultures. I. Dependence of mutational response upon growth-limiting factors. *Mutat Res* 1964;1:113–120. [PubMed: 14237062]
- Lea DE, Coulson CA. The distribution of the numbers of mutants in bacterial populations. *J Genetics* 1949;49:264–285.
- Luria SE. The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. *Cold Spring Harbor Symp Quant Biol* 1951;16:463–470. [PubMed: 14942756]
- Luria SE, Delbrück M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 1943;28:491–511. [PubMed: 17247100]
- Ma WT, Sandri GvH, Sarkar S. Analysis of the Luria-Delbrück distribution using discrete convolution powers. *J Appl Prob* 1992;29:255–267.
- Miller JH. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu Rev Microbiol* 1996;50:625–643. [PubMed: 8905093]

- Novick A, Szilard L. Experiments with the chemostat on spontaneous mutations of bacteria. *Proc Natl Acad Sci USA* 1950;36:708–719. [PubMed: 14808160]
- Reddy M, Gowrishankar J. A genetic strategy to demonstrate the occurrence of spontaneous mutations in non-dividing cells within colonies of *Escherichia coli*. *Genetics* 1997;147:991–1001. [PubMed: 9383047]
- Reyrat JM, Pelicic V, Gicquel B, Rappuoli R. Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis. *Infect Immun* 1998;66:4011–4017. [PubMed: 9712740]
- Rice, JA. *Mathematical Statistics and Data Analysis*. Wadsworth Publishing Company; Belmont, CA: 1995.
- Rosche WA, Foster PL. Determining mutation rates in bacterial populations. *Methods* 2000;20:4–17. [PubMed: 10610800]
- Sarkar S, Ma WT, Sandri GvH. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* 1992;85:173–179. [PubMed: 1624139]
- Stewart FM. Fluctuation tests: how reliable are the estimates of mutation rates? *Genetics* 1994;137:1139–1146. [PubMed: 7982567]
- 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]
- Wierdl M, Greene CN, Datta A, Jinks-Robertson S, Petes TD. Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics* 1996;143:713–721. [PubMed: 8725221]
- Zar, JH. *Biostatistical Analysis*. Prentice Hall; Englewood Cliffs, New Jersey: 1984.

Generation = k	Cells = N	New Mutants	Growth of Pre-existing Mutants	Total Mutants = r	Mutant Fraction = r/N	Change per Generation = Δ
4	$N_t/16$	0		0	0	
3	$N_t/8$	1		1	$8/N_t$	$8/N_t$
2	$N_t/4$	2		4	$16/N_t$	$8/N_t$
1	$N_t/2$	4		12	$24/N_t$	$8/N_t$
0	N_t	8		32	$32/N_t$	$8/N_t$

Fig 1.

An illustration of the constant increase in the mutant fraction after a population reaches a size sufficiently large so that the accumulation of mutants is simply a function of population size. Luria's conventions are followed {Luria, 1951 632 /id}): k = the generation numbered backwards from 0; N = the number of cells present at each generation; N_t = the final number of cells in the population; μ = the mutation rate per cell (assuming a synchronous population). At each generation there are $N_t/2^k$ individuals that produce $\mu N_t/2^k$ new mutations, which will produce a total of μN_t mutant progeny by the last generation.

	A
1	[insert median here]
2	1.24
3	[try various values of m here]
4	$=(A1/A3)-LN(A3)-(A2)$

Fig 2. Spreadsheet method to solve transcendental equations by iteration. The example is the Lea-Coulson median estimator. Method: insert the experimentally determined median in A1; try various values of m at A3 to get the value at A4 close to 0.

TABLE I

Terms Used

Term	Definition
m	Number of mutations per culture
μ	Mutation rate; probability of mutation per cell per division or generation
N	Number of cells
N_0	Initial number of cells in a culture = the inoculum
N_t	Final number of cells in a culture
r	Observed number of mutants in a culture
$\bar{r}$	Median number of mutants in a culture
f	Mutant fraction or frequency = r/N
V	Volume of a culture
C	Number of cultures in experiment
p_0	Proportion of cultures without mutants
z	Dilution factor or fraction of a culture plated
p_r	Proportion of cultures with r mutants
c_r	Number of cultures with r mutants
Pr	Proportion of cultures with r or more mutants
Q_1	Value of r at 25% of the ranked series of r
Q_2	Value of r at 50% of the ranked series of r = the median
Q_3	Value of r at 75% of the ranked series of r
σ	Standard deviation
CL	Confidence limit = $(1-\alpha) \times 100$
α	Level of statistical significance, usually 0.05 or 0.01