Genetica **96**: 217–223, 1995.

On the distribution of bacterial mutants: the effects of differential fitness of mutants and non-mutants

Gregg Jaeger & Sahotra Sarkar*
Dibner Institute, MIT E56-100, 38 Memorial Drive, Cambridge, MA 02139, USA
**Author for correspondence*

Received 18 May 1994 Accepted 18 November 1994

Abstract

This paper provides an analytic treatment of the effect of differential fitness of mutants and non-mutants on the Luria-Delbrück distribution, which is used to describe the number of mutant cells obtained prior to selection during a fluctuation test experiment. It also systematizes the treatment of the case when the cultures are seeded with multiple cells. One surprising result is that differential fitness of mutants and non-mutants does not affect the mean of the distribution (though, as expected, it decreases the variance). This treatment completes the analysis of the influence of factors that affect the Luria-Delbrück distribution through mechanisms acting prior to plating. All the results that have been obtained to date are collected in a table for easy reference.

Introduction

Ever since Cairns, Overbaugh and Miller (1988) published the results of experiments that apparently showed that a $Lac^- \rightarrow Lac^+$ mutation in *Escherichia coli* occurred more frequently in the presence of lactose than in its absence, the possibility of such 'directed' mutagenesis has been a subject of intense controversy (see Foster, 1992, 1993; Sarkar, 1991, 1992; Lenski & Mittler, 1993; Cairns, 1993; MacPhee, 1993). The argument for the existence of directed mutations is partly based on experimental deviations of bacterial mutant distributions, as obtained from a fluctuation test, from the Luria-Delbrück distribution, as calculated by Lea and Coulson (1949), which is expected to hold if all mutations are random (non-directed). Critics concede that these deviations occur but argue that they can be accounted for by subsidiary factors (other than directed mutations) such as a finite number of cell-divisions in the fluctuation test-tubes (that is, during clonal growth prior to selection through plating), multiple cells (seeds) from which the clones originated, differential fitness of normal and mutant cells (during clonal growth), phenotypic lag, and incomplete plating efficiency, none of which were taken into account in the Lea-Coulson analysis (see references in Sarkar, 1991 and Lenski & Mittler, 1993). In general, it is argued that each of these factors would decrease the variance of the distribution, which is the type of deviation routinely observed.

In order to judge the relevance of these factors, it is necessary to compute the expected distribution of the mutants in their presence. Stewart, Gordon and Levin (1990) developed a framework for incorporating most of these factors; the expected distribution can then be numerically computed using their program (see also Stewart, 1991). Ma, Sandri and Sarkar (1992) devised a new algorithm that led to a six-fold increase in computational efficiency of these calculations (and a further ten-fold increase if massively parallel machines, such as the CM-2 Connection Machine, are used). This algorithm has been incorporated in the computer program available from Stewart following the procedure listed in Stewart, Gordon and Levin (1990) and Stewart (1991). This approach is obviously fruitful in the comparison of experiment to theory. However, since it is not fully analytic, it does not easily permit certain kinds of theoretical analyses, such as the derivation of the probability of observing jackpots (or a very large number of mutants) on the plates.

Ma, Sandri and Sarkar (1992), therefore, presented recursion relations for computing the probability distribution of the expected number of mutants assuming the original Lea-Coulson model. Sarkar, Ma and Sandri

(1992) explicitly incorporated a finite number of cell divisions into the model, and explicitly incorporated the effect of seeding each test-tube with more than one cell. Pakes (1993) extended some of these results. The strategy followed to obtain all of these results is to write down and analyze a set of differential equations which describe the growth of a culture in a test-tube. The growth of normal cells is treated deterministically while mutation and mutant cell growth are treated stochastically. Next, the generating function for the distribution is obtained from these differential equations. Essentially, this part of the treatment goes back to Lea and Coulson's (1949) pioneering analysis. Now, using the algorithm of Ma, Sandri and Sarkar (1992), recursion relations are obtained to evaluate p_r, the probability of finding r mutants when the contents of the fluctuation test-tubes are plated. Finally, an asymptotic analysis of p_r (as $r \to \infty$) is carried out to find a formula for the probability of finding jackpots of size r.

The main purpose of this note is to extend the results of Sarkar, Ma and Sandri (1992) and Pakes (1993) to incorporate the effects of differential fitness of mutants and non-mutants in the fluctuation test-tubes. The effect of seeding each test-tube with more than one cell, which is bound to happen in actual experiments, is also explicitly analyzed. A somewhat surprising result is that the mean of the distribution is not affected by differential fitness of mutants and non-mutants. Of all the factors listed in the first paragraph which can result in deviations from the distribution calculated by Lea and Coulson (1949), only the first three act prior to plating. Therefore, the results reported here complete the analytic treatment of this part of the process. Armitage (1953) noted one other factor acting during clonal growth that could potentially affect the distribution: variation in mutation rates during the growth process. However, without an explicit model of such variation, it is impossible to incorporate this factor into the generating functions. (A useful numerical treatment has been provided by Stewart, Gordon & Levin, 1990.) Finally, fluctuation analysis is routinely used to estimate mutation rates. This paper also looks at – though without considering questions about the 'goodness' of estimation (see, e.g., Jones, Thomas & Rogers, 1994, and Stewart, 1994) – the effect of these three factors on the calculation of mutation rates. One result is surprising: if the mutation rate estimation ultimately relies only on the mean (and no other moment) of the distribution, differential fitness of the mutant and non-mutant cells makes no difference. (Note, however, that the mean does not necessarily provide the best estimate for the mutation rate (see, e.g., Lea & Coulson, 1949, and Stewart, 1994).) All the results are collected in Table 1, which should be useful to experimentalists.

The basic formulae

The method used to obtain the formulae reported here is identical to that used by Sarkar, Ma and Sandri (1992). If p_r is the probability of finding r mutants in a clone of size n, μ is the mutation rate per cell per generation, and b is the relative fitness of a mutant cell (with fitness being interpreted as growth rate) compared to that of a normal cell (with fitness 1), and $0 < b < \infty$, then,

$$p_r + \frac{dp_r}{dn}dn = p_{r-1}\left(\mu dn + \frac{b(r-1)}{n}dn\right) + p_r\left(1 - \mu dn - \frac{br}{n}dn\right) \quad (1)$$

or

$$\frac{dp_r}{dn} = p_{r-1}\left(\mu + \frac{b(r-1)}{n}\right) - p_r\left(\mu + \frac{br}{n}\right) \quad (2)$$

In only slightly different notation ($m = \mu n$), this equation was known to Koch (1982). Note that these equations assume $b \neq 0$ (and $b \neq \infty$). If $g(n, s)$ is the generating function for the distribution (where s is the usual 'dummy variable' used to define a generating function, $0 \leq s \leq 1$), then $g(n, s)$ obeys the partial differential equation:

$$\frac{\partial g}{\partial n} = \mu(s-1)g + \frac{b}{n}s(s-1)\frac{\partial g}{\partial s}. \quad (3)$$

The solution of this equation, with n_0 mutants at the time of seeding and n mutants at the time of plating, is:

$$g(n,s) = \left(1 - s + \left(\frac{n_0}{n}\right)^b s\right)^{(\mu n/b)((1-s)/s)} = \exp\frac{\mu n}{b}\left(\frac{1-s}{s}\right)\ln\left(1 - s + \left(\frac{n_0}{n}\right)^b s\right). \quad (4)$$

To obtain the case considered by Sarkar, Ma and Sandri (1992), set $b = 1$, that is, assume equal fitnesses. The resulting equation was known to Bartlett (1978). To obtain the generating function of Lea and Coulson (1949), now set $n_0/n = 0$. In this equation, since n_0

is the initial number of (normal) cells from which a culture is grown, if the culture is a clone grown from a single cell, $n_0 = 1$.

Following the algorithm of Ma, Sandri and Sarkar (1992), the distribution of mutants is now calculated from equation (4):

$$p_0 = \exp\left[\frac{\mu n}{b}(\alpha^b - 1)\right];$$

$$p_k = \frac{1}{k}\frac{\mu n}{b}\sum_{i=0}^{k-1} p_i(k-i) \times \left(\frac{(1-\alpha^b)^{k-i}}{k-i} - \frac{(1-\alpha^b)^{k-i+1}}{k-i+}\right), \quad (5)$$

where $\alpha = n_0/n$. The limiting procedures indicated in the last paragraph give rise to the recursion relations for the other models. Table 1 summarizes these manipulations. Only the case when the original Lea and Coulson (1949) generating function, which is conceptually somewhat different, and has a radically different asymptotic behavior of p_r as $r \to \infty$, will be explicitly written down here. In that case (from Ma, Sandri & Sarkar, 1992),

$$p_0 = e^{-m}$$

$$p_k = \frac{m}{k}\sum_{i=0}^{k-1}\frac{p_i}{(k-i+1)} \quad (6)$$

where m is the number of mutations that occurred during the growth of the clone. (Formally, one can think of $m = \mu n$ and obtain this equation as the appropriate limit of equation (5). However, since this model effectively assumes infinite population growth, one would then have to assume that, after an infinite number of cell divisions, the clone had grown from size 1 to size n. This was the counterfactual assumption of the original Lea-Coulson analysis that was removed by Bartlett (1978) and Sarkar, Ma and Sandri (1992).).

The probability of obtaining a jackpot of size k ($k \gg 1$) is given by the asymptotic form of p_k as $k \to \infty$. For the model considered by Sarkar, Ma and Sandri (1992), that is, with equal fitnesses but allowing for seeding with multiple cells and a finite number of cell divisions, the generating function

$$g(s) = (1 - s + \alpha s)^{m(1-s)/s}$$

where $\alpha = n_0/n_t$ and $m = \mu n_t$. Note that $g(s)$ has a radius of convergence, $r = 1/(1-\alpha)$. Let $L(\alpha, s) \equiv g(s)$ and $f(s) \equiv g(rs)$. Then

$$f(s) = \frac{L(1,s)}{L(\alpha,s)}(1-s)^{-\mu n_t \alpha}$$

If $f(s) = \sum_{i=0}^{\infty} q_i s^i$, then $q_k = (1-\alpha)^{-k} p_k$. Then, using a Tauberian theorem (see Feller, 1970, p. 423), as $k \to \infty$,

$$q_k \sim \frac{(1-\alpha_k)^{-k}}{\Gamma(\mu n_t)} k^{\mu n_t - 1}$$

where $\Gamma(\cdot)$ is the gamma function.

Therefore,

$$p_k \sim \frac{(1-\alpha)^k}{\Gamma(\mu n_t)} k^{\mu n_t - 1} \quad (7)$$

This argument is due to Pakes and the result was first presented in Pakes (1993). For values of b other than one, this can be generalized to:

$$p_k(\alpha) \sim \frac{(1-\alpha^b)^k k^{(\mu n_0/b)-1}}{\Gamma\left(\frac{\mu n_0}{b}\right)}. \quad (8)$$

For the original Lea and Coulson (1949) model, the asymptotic behavior is entirely different. In that case, Ma, Sandri and Sarkar (1992) and Pakes (1993) have shown that

$$p_k \sim \frac{m}{k^2} \quad (9)$$

There does not appear to be any obvious limiting procedure which allows equation (9) to be obtained from equation (8). Figure 1a plots p_k as obtained from equation (5) (with $b = 1$) and equation (7); Figure 1b plots p_k from equation (5) and equation (8) with $b = 0.9$. In both cases it is seen that convergence to the asymptotic form is quite slow. Therefore, experimentalists would be well-advised to use equation (5) rather than equations (7) or (8). Ma, Sandri and Sarkar (1992) obtained fast convergence (at $n \sim 50$) of p_k evaluated by equation (6) to that obtained by using equation (9). If $n \gg n_0$, therefore, equation (9) can be profitably used.

Note that all these models are applicable only if both the mutants and non-mutants have a finite growth rate, that is, $0 < b < \infty$. If $b = \infty$, $g(n,s) \equiv 0$ whereas if $b = 0$, $g(n,s)$ is infinite. In either case, the formulae reported here – and in the next section –

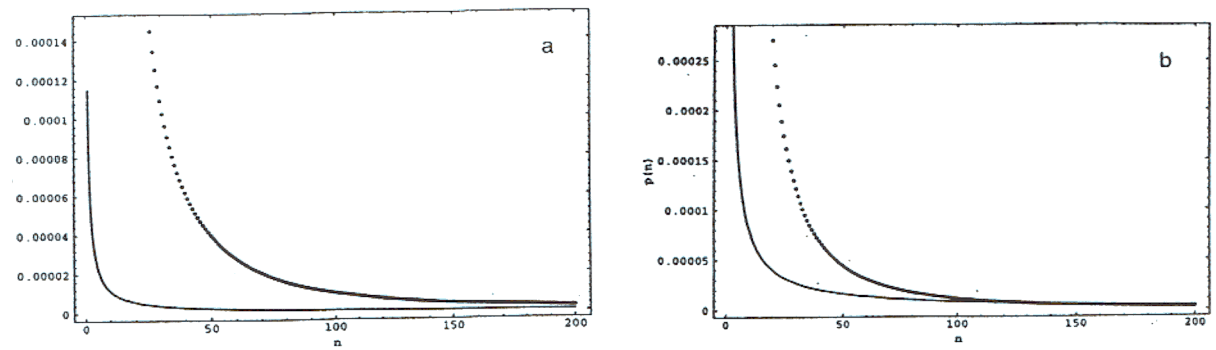

Fig. 1. Probability distributions. The dotted line represents the exact distribution (equation (5)); the solid line represents the asymptotic form (equation (8)). $\mu n = 0.1$; Figure (1a) $b = 1$; Figure (1b) $b = 0.9$.

cannot be used. The results diverge at these limits (in a continuous manner).

Variances, means and mutation rates

The mean, M, and the variance, V, of any discrete probability distribution can be obtained from its generating function using the relations $M = g'(1)$ and $V = g''(1) + g'(1) - [g'(1)]^2$. Using equation (4), this gives:

$$M = \frac{\mu n}{b} \ln\left(\frac{n_0}{n}\right)^{-b} = \mu n \ln\left(\frac{n}{n_0}\right) \qquad (10)$$

and

$$V = -\mu n \ln\left(\frac{n}{n_0}\right) + 2\frac{\mu n}{b}\left(\left(\frac{n}{n_0}\right)^{b} \right) \qquad (11)$$

The usual formula given for V (with $b = 1$) leaves out the factor of 2 in the second term of the equation (11) (see, e.g., Lea & Coulson, 1949). The original Lea and Coulson (1949) generating function gives rise to infinite moments, as is well-known (Bailey, 1964; Bartlett, 1978). Pakes (1993) has shown that the truncated moments have the asymptotic behavior of $m \ln(n)$ for the first moment and $(m/k)n^k$ for the k-th moment, with $k > 1$ (verifying the numerical work of Ma, Sandri & Sarkar, 1992). For the other cases, the appropriate formulae for the mean and the variance can be obtained from equations (10) and (11) using the appropriate limits as indicated in the last section. Explicit formulae are presented in Table 1.

The variance of the distribution of bacterial mutants has been routinely used as an indicator in the controversy over the existence of directed mutations in bacteria. The general argument is that all reasonable factors influencing the original Lea and Coulson (1949) distribution reduce the variance. Figure 2 shows the dependence of V and $\ln(V)$ on b for various values of μn. If $b < 1$ (that is, the mutants are less fit in the test-tube medium), the variance would be decreased. The important observations are that, assuming a constant mutation rate, μ, the larger n is, the greater the effect of the variance and, *especially*, the difference is significant for b close to 1. (In Fig. 2a, for example, $V(1.0) - V(0.95) \gg V(0.95) - V(0.9)$ for each of the curves.) Thus, the usual experiments with $n \gg 10^6$ will feel this effect most seriously. (If $b > 1$, the variance is increased. This is interesting as a theoretical point, but not relevant in experiments designed to explore the possibility of directed mutations where $b \leq 1$). $n_0 > 1$ acts to decrease the variance in either case, verifying the common intuition about the effect of this factor (Sarkar, Ma & Sandri, 1992).

Fluctuation analysis is one of the standard methods used to estimate mutation rates and critics of the hypothesis of directed mutations (see, especially, Lenski & Mittler, 1993) have contended that the factors listed in the second paragraph of the first section invariably lead to an over-estimation of mutation rates from a fluctuation experiment. There are many different ways in which the results of a fluctuation experiment can be used to estimate mutation rates (see Stewart, 1994). However, when the mean of the distribution is used to form (at least) rough estimates as is sometimes still done (though, as will again be emphasized below, the mean is not a reliable estimator for mutation rates), the fact that the mean, M, is independent of b has an unexpected consequence: since equation (10) has no dependence on b, differential fitness of mutants and non-mutants will have no effect on mutation rates estimated

Table 1. Properties of models of modified Luria-Delbrück distributions.

Model	Generating Function	Recursion Relation	Probability of Jackpot of size r	Mean	Variance	Effect on Variance	Effect on Mutation Rate Estimate
Lea–Coulson Model (LC)	$(1-s)^{m_0 s/(1-s)}$	Equation (6)	$\frac{m}{k^2}$	∞	∞	N.A.	N.A.
LC Model with Finite Number of Cell Divisions (LCF)	$\left(1-\frac{1}{n}s\right)^{\mu n_0 s/(1-s)}$	Equation (4), $b=1$, $n_0=1$	$\frac{\left(1-\frac{1}{n}\right)^{k} k^{\mu n-1}}{\Gamma(n)}$	$\mu n \ln n$	$-\mu n \ln n + 2\mu n(n-1)$		0
LCF Model with Multiple Seed	$\left(1-\frac{n_0}{n}s\right)^{\mu n_0 s/(1-s)}$	Equation (4), $b=1$	$\frac{\left(1-\frac{n_0}{n}\right)^{k} k^{\mu n-1}}{\Gamma(\mu n)}$	$\mu n \ln\left(\frac{n}{n_0}\right)$	$-\mu n \ln\left(\frac{n}{n_0}\right) + 2\mu n\left(\frac{n}{n_0}-1\right)$		
LCF Model with Differential Fitness	$\left(1-\frac{1}{n^b}s\right)^{\frac{\mu n_0 s}{b}/(1-s)}$	Equation (4), $n_0=1$	$\frac{\left(1-\left(\frac{1}{n}\right)^{b}\right) k^{\frac{\mu}{b}-1}}{\Gamma\left(\frac{n}{b}\right)}$	$\mu n \ln n$	$-\mu n \ln n + \frac{2\mu n}{b}(n^b-1)$		0
LCF Model with Multiple Seed and Differential Fitness	$\left(1-\left(\frac{n_0}{n}\right)^{b}s\right)^{\frac{\mu n_0 s}{b}/(1-s)}$	Equation (4)	$\frac{\left(1-\left(\frac{n_0}{n}\right)^{b}\right) k^{\frac{\mu n}{b}-1}}{\Gamma\left(\frac{\mu n}{b}\right)}$	$\mu n \ln\left(\frac{n}{n_0}\right)$	$-\mu n \ln\left(\frac{n}{n_0}\right) + \frac{2\mu n}{b}\left(\left(\frac{n}{n_0}\right)^{b}-\right.$		

For an explanation of the models and notation, see text. It is assumed that the mutant cell fitness $b < 1$. The mutation rates are assumed to be estimated by the means. (Key for last two columns: 'N.A.' – not applicable; '0' – there is no difference from the value for the LC model; '–' – there is a decrease from the value for the LC Model.)

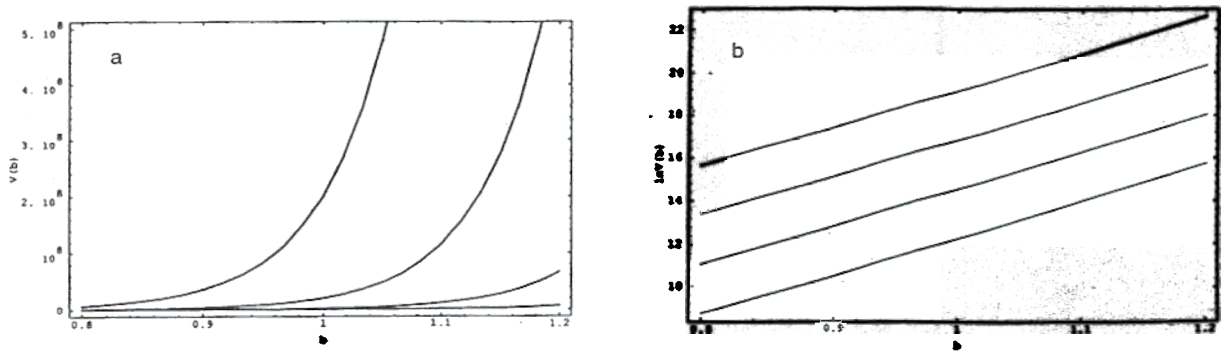

Fig. 2. Variances. Here $n = 10^8$, $n_0 = 1$; in Figure (2a), the curves, from left to right, have $\mu n = 1.0, 0.1, 0.01, 0.001$; in Figure (2b), the lines from top to bottom have $\mu n = 1.0, 0.1, 0.01, 0.001$.

from the mean. However, if the test-tubes are seeded with multiple cells, then equation (10) shows that the mutation rate calculated by $(M/n \ln(n))$ would be an under-estimate of the actual mutation rate.

Discussion

All results for $b \neq 1$ are new. Figure 2 shows that for $0.9 < b < 1.0$, the influence of b in decreasing the variance can be non-negligible for most experimental contexts. The result that the mean, M, is independent of b is surprising. This result is formally due to the cancellation of b in equation (10), which gives the formula for the mean. Intuitively, it can be understood as follows. Compare the situations when $b = 1$ (both mutants and non-mutants grow at the same rate) and when $b < 1$ (the mutants grow more slowly than the non-mutants). In the second case, the p_i for small i (but $i > 0$) are higher than those of the first (because there will be fewer mutants), while the p_i for high i are lower. (The place where the switch takes place depends on b.) However, since $M = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + \cdots k \cdot p_k + \cdots$, the p_i for small i are multiplied by smaller numbers than the p_i for high i. What equation (10) shows is that the two effects cancel each other out. (Alternatively, observe that the mean of the number of mutants should depend only on the mutation rate, μ and the time of clonal growth, measured by n. Thus, b and n should not enter the formula independently. What equation (10) shows is that the b enters both the numerator and the denominator, and then gets cancelled.) In any case, this result has one important consequence: the estimation of mutation rates using the observed mean of the distribution is *not* affected by differential fitness. A contrary position is advocated by Lenski and Mittler (1993) who argue, though present no calculation to show, that all such factors lead to a bias in the estimated mutation rate.

The observation that the mean, M, of the distribution does not depend on b should not, however, be interpreted as an endorsement of the M to estimate mutation rates. As has been known since the pioneering paper of Lea and Coulson (1949) (for a recent treatment see Stewart, 1994), the high variance of the distribution makes the mean an unreliable estimator though it continues to be used for rough estimates (see Foster, 1993). Simulations reported by Stewart (1994) show that the maximum-likelihood procedure is the best known procedure for estimating mutation rates at present.

The inoculation of test-tubes with more than one cell (of presumably identical genotype), which is inevitable in most experimental circumstances, turns out to increase the mean but decrease the variance of the distribution. Therefore, the theoretical analysis reported here confirms the recent experimental results obtained by Dijkmans, Kreps and Mergeay (1994).

As noted before, the results obtained here complete what can be expected from this kind of mathematical analysis of cell growth prior to plating. Whether equally transparent analytic results can be obtained to incorporate post-plating factors – phenotypic lag, plating efficiency less than 1, etc. – remains a subject for further work.

Acknowledgments

We would like to thank A.G. Pakes and F.M. Stewart for discussions, and to an anonymous referee for pointing out an error in the original version of this paper.

References

Armitage, P., 1952. The statistical theory of bacterial populations subject to mutation. Journal of the Royal Society B: 14: 1–40.

Armitage, P., 1953. Statistical concepts in the theory of bacterial mutation. Journal of Hygiene 51: 162–184.

Bailey, N.T.J., 1964. The elements of stochastic processes. Wiley, New York.

Bartlett, M.S., 1978. An introduction to stochastic processes. 3rd edn. Cambridge University Press, Cambridge.

Cairns, J., 1993. Directed mutation. Science 260: 1221–1222.

Cairns, J., J. Overbaugh & S. Miller, 1988. The origin of mutants. Nature 335: 142–145.

Dijkmans, R., S. Kreps & M. Mergeay, 1994. Poisson-like fluctuation patterns of revertants of leucine auxotrophy (leu-500) in *Salmonella typhimurium* caused by delay in mutant cell division. Genetics 137: 353–359.

Feller, W., 1970. Introduction to Probability Theory and Its Applications. Vol. 2. New York: Wiley.

Foster, P.L., 1992. Directed mutation: between unicorns and goats. Journal of Bacteriology 174: 1711–1716.

Foster, P.L., 1993. Adaptive mutation: the uses of adversity. Annual Reviews of Microbiology 47: 469–504.

Jones, M.E., S.M. Thomas & A. Rogers, 1994. Luria-Delbrück fluctuation experiments: design and analysis. Genetics 136: 1209–1216.

Koch, A.L., 1982. Mutation and growth rates from Luria-Delbrück fluctuation tests. Mutation Research 95: 129–143.

Lea, D.E. & C.A. Coulson, 1949. The distribution of the number of mutants in bacterial populations. Journal of Genetics 49: 264–285.

Lenski, R.E. & J.E. Mittler, 1993. The directed mutation controversy and neo-Darwinism. Science 259: 188–194.

Luria, S.E. & M. Delbrück, 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28: 491–511.

Ma, W.T., G.v.H. Sandri & S. Sarkar, 1992. Analysis of the Luria-Delbrück distribution using discrete convolution powers. Journal of Applied Probability 29: 255–267.

MacPhee, D.G., 1993. Directed mutation: paradigm postponed. Mutation Research 285: 109–116.

Pakes, A.G., 1993. Remarks on the Luria-Delbrück distribution. Journal of Applied Probability 30: 991–994.

Sarkar, 1991. Lamarck contre Darwin, reduction versus statistics: conceptual issues in the controversy over directed mutagenesis in bacteria. In A.I. Tauber (Ed.). Organism and the Origin of Self. Kluwer, Dordrecht, pp. 235–271.

Sarkar, S., 1992. Neo-Darwinism and the problem of directed mutations. Evolutionary Trends in Plants 6: 73–79.

Sarkar, S., W.T. Ma & G.v.H. Sandri, 1992. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. Genetica 85: 173–179.

Stewart, F.M., 1991. Fluctuation analysis: the effect of plating efficiency. Genetica 84: 51–55.

Stewart, F.M., 1994. Fluctuation tests: how reliable are the estimates of mutation rates? Genetics 137: 1139–1146.

Stewart, F.M., D.M. Gordon & B.R. Levin, 1990. Fluctuation analysis: the probability distribution of the number of mutants under different conditions. Genetics 124: 175–185.