STATISTICAL CONCEPTS IN THE THEORY OF BACTERIAL MUTATION

By P. ARMITAGE

From the Medical Research Council's Statistical Research Unit, London School of Hygiene and Tropical Medicine

(With 7 Figures in the Text)

1. INTRODUCTION

1·1. The statistical approach to the study of bacterial populations subject to mutation has come into prominence since the publication of Luria & Delbrück's paper (1943), although mathematical concepts had appeared in work published before that date (Deskowitz & Shapiro, 1935; Bunting, 1940). Other mathematical studies on this subject have appeared (Shapiro, 1946; Lea & Coulson, 1949), and I have recently attempted to survey all this work from a mathematical standpoint (Armitage, 1952). Luria & Delbrück's methods for estimating mutation rates are now widely used and advocated (Catcheside, 1951), and it seemed perhaps worth while to discuss some of the statistical aspects of their and other workers' results, avoiding as far as possible complicated mathematics. Many of the results presented here are given without proof; for this, reference may usually be made to the previous paper (Armitage, 1952).

1·2. In the next section I consider the change in the proportion of mutants during the long-term growth of a bacterial population. In §3 the precise meaning of the term 'mutation rate' is discussed. §§4 and 5 are concerned with the distribution of the number of mutants in each of a series of replicate cultures, and with methods of estimating the mutation rate from such a series.

2. DEVELOPMENT OF BACTERIAL POPULATIONS SUBJECT TO MUTATION

2·1. When a culture is in the 'logarithmic' phase of growth (dividing, that is, at approximately equal intervals, with no deaths), the population size increases exponentially. If the generation times were all exactly equal the population size would increase in jumps, taking in turn the values 1, 2, 4, 8, 16, ..., etc. In practice the moments of fission soon get out of step, and the total population increases fairly smoothly. As a mathematical model for this simple growth process, we could assume that, of the N organisms present at some instant t, a proportion $a\,dt$ will divide during a small interval of time of length dt. In mathematical terms, $dN = aN\,dt$. This leads to the differential equation

$$\frac{dN}{dt} = aN,$$

which expresses the rate of change, dN/dt, of the population size, N, as a function of N itself. Since an expression for N as an explicit function of t is required, this differential equation must be solved, the solution being

$$N = N_0 e^{at}. \qquad (1)$$

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

Here e is the base of natural logarithms, 2·7183...., and N_0 is the initial population size when $t=0$.

2·2. We now introduce the concept of mutation. Let the population at time t consist of x members of the wild-type X, and y members of the mutant type Y. Suppose that in a small time interval of length dt a proportion $a\,dt$ of the x wild-type organisms divide, thus producing $ax\,dt$ new organisms, of which a fraction λ are of the mutant type. Similarly, we may suppose that in the same time interval dt a proportion $b\,dt$ of the y mutants divide, producing $by\,dt$ new organisms, of which a fraction μ are of the wild type. The increases dx and dy in the X and Y populations are respectively

$$dx = a(1-\lambda)\,x\,dt + b\mu y\,dt$$

and

$$dy = b(1-\mu)\,y\,dt + a\lambda x\,dt,$$

which lead to the differential equations

$$\left.\begin{aligned}\frac{dx}{dt} &= a(1-\lambda)\,x + b\mu y,\\ \frac{dy}{dt} &= a\lambda x + b(1-\mu)\,y.\end{aligned}\right\} \tag{2}$$

Of the four constants appearing in these equations, a and b represent the growth rate of the two strains, while λ and μ represent the forward and back mutation rates and will usually be very small. It will be convenient to refer to λ and μ as 'relative mutation rates' until the definition of a mutation rate has been discussed below.

2·3. The complete solution of (2) is given by Armitage (1952). The solution has also been discussed by Shapiro (1946); this paper should be read with care as it contains a number of algebraic errors. It will be sufficient here to point out one or two features of particular interest *when the growth rates are equal* (i.e. when $a=b$).

(a) Suppose $a=b$. The proportion of mutants in the total population is $y/(x+y)$. This proportion, which will be denoted here by ψ, changes as the population grows, and is given by a simple formula

$$\psi = \frac{\lambda}{\lambda+\mu} - \frac{(\lambda x_0 - \mu y_0)}{(\lambda+\mu)\,(x_0+y_0)}\,e^{-a(\lambda+\mu)t}, \tag{3}$$

where x_0 and y_0 are the initial numbers of wild-type and mutant organisms respectively.

Putting $t=\infty$ in (3), we find that ψ approaches a limiting value

$$\psi_\infty = \frac{\lambda}{\lambda+\mu}, \tag{4}$$

depending only on the ratio between the two relative mutation rates λ and μ. Putting $t=0$ in (3), we obtain the initial proportion of mutants,

$$\psi_0 = \frac{y_0}{x_0+y_0},$$

as, of course, would be expected.

Figs. 1 and 2 illustrate the way in which ψ changes with time, according as the initial proportion ψ_0 is less or greater than the limiting value ψ_∞.

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

(*b*) In the situation typified by Fig. 1, where $\psi_0 < \psi_\infty$, equation (3) shows that

$$\psi_\infty - \psi = \frac{(\lambda x_0 - \mu y_0)}{(\lambda+\mu)(x_0+y_0)} e^{-a(\lambda+\mu)t}.$$

ψ therefore approaches its limit ψ_∞ exponentially. If $\log_e(\psi_\infty - \psi)$ were plotted against t, a straight line with a slope of $-a(\lambda+\mu)$ should be obtained.

(*c*) If the culture is grown from a pure inoculum of the wild type, $y_0 = 0$, and equation (3) now becomes

$$\psi = \frac{\lambda}{\lambda+\mu}\{1 - e^{-a(\lambda+\mu)t}\}. \qquad (5)$$

A useful property of equation (5) is that at time 0 the initial gradient, or rate of increase, of ψ with respect to t is $a\lambda$. Similarly, in a culture grown purely from the mutant type, the initial rate of decrease of ψ is $a\mu$ (or, more generally, $b\mu$, for these relations hold even when the growth rates a and b are unequal). Fig. 3 illustrates the change in ψ in two cultures, one (A) grown purely from a wild-type inoculum, the other (B) purely from a mutant inoculum. The gradients of the two dotted lines would be respectively $a\lambda$ and $b\mu$.

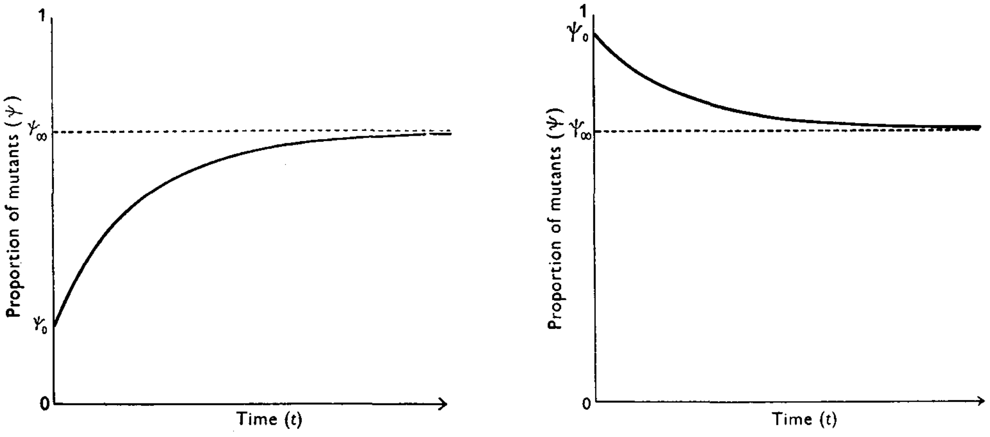

Fig. 1

Fig. 2

Figs. 1, 2. Change in proportion of mutants with growth of culture, according as initial proportion ψ_0 is less or greater than limiting value ψ_∞.

2·4. These considerations suggest at least the approach to the analysis of long-term experiments in which the growth of a bacterial population is observed for a sufficiently long period of time for the proportion of mutants to approach closely its limiting value ψ_∞. The length of time required for such experiments will depend on the growth and mutation rates, but is likely to be a matter of days, or even weeks. In some of his long-term experiments Stocker (1949) grew cultures of *Salmonella typhi-murium* for over 800 generations.

The procedure suggested here is only an outline, as the exact statistical methods appropriate to such experiments have not yet been worked out. We assume that observations of the value of ψ are made at different points of time.

(i) The limiting value ψ_∞ is estimated empirically.

(ii) $\text{Log}_e(\psi_\infty - \psi)$ is plotted against t. If there is no appreciable departure from linearity, equality of growth rates may be tentatively assumed (cf. (b) above). (The situation when a difference in growth rates can be demonstrated is not considered here.) A regression line of $\log_e(\psi_\infty - \psi)$ on t is fitted by the usual statistical technique, and its slope is taken as an estimate of $-a(\lambda+\mu)$. (The natural logarithm of any number may be obtained by multiplying the logarithm to base 10 by 2·3026.)

(iii) The constant a is estimated from the rate of growth of the total population. If a population growing at rate a increases from N_0 to N_1 during an interval of length t, then by (1)

$$N_1 = N_0 e^{at_1},$$

and

$$a = \frac{1}{t_1}\log_e\left(\frac{N_1}{N_0}\right).$$

(iv) Estimates of $\lambda/(\lambda+\mu)$, $a(\lambda+\mu)$ and a having been obtained from (i), (ii) and (iii) respectively, the three constants a, λ and μ can be estimated separately.

2·5. If cultures are grown from pure inocula of the two types, as in Fig. 3, a simple check on the equality of the growth rates is available. If each growth rate is equal to a, the initial rates of increase of ψ in the two curves are $a\lambda$ and $a\mu$, as stated above. Their ratio is therefore λ/μ, and may be compared with $\psi_\infty/(1-\psi_\infty)$, which is also equal to λ/μ (cf. equation (4)). This is quite a sensitive method for detecting slight differences in the growth rates; proportionate differences between a and b even of the order of λ and μ appreciably affect the value of ψ_∞, whereas the initial gradients are effectively unaltered.

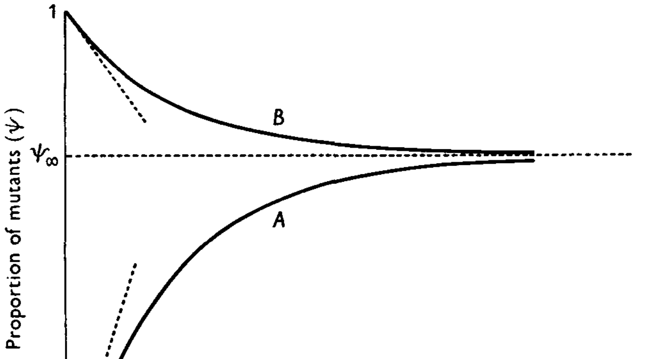

Fig. 3. Change in proportion of mutants in two cultures: (A) grown from a wild-type inoculum, (B) from a mutant inoculum.

2·6. If the growth rate changed during the course of the experiment, as it would if growth were interrupted by refrigeration, it would be necessary to measure t in terms of generations. This is equivalent to giving a fixed value to the growth rate a.

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

For, if t is measured by the number of generations, the population size at time t will be $N_0 2^t$. Comparing this expression with equation (1), we have

$$e^{at}=2^t,$$

whence $$at=t\log_e 2,$$

and $$a=\log_e 2=0\cdot6931.$$

Luria & Delbrück (1943) use as a time unit the generation time divided by $\log_e 2$; in these units $a=1$.

The assumption is made here that the relative mutation rates λ and μ are constant, irrespective of changes in the growth rates a and b. In other words, we assume constant probabilities of mutation per generation time, rather than absolute time. This assumption is usually made, but may not be valid when changes of growth rate depend on the availability of growth factors (Novick & Szilard, 1950).

2·7. Long-term experiments of the type envisaged here are time consuming, and involve some method of continuous culture in order to prevent overcrowding of the bacterial population. It is, therefore, not surprising to find only a small number of such investigations reported in the literature. Deskowitz & Shapiro (1935) studied variants of *Salmonella aertrycke*, and found mutation from the 'rough' to the 'smooth', but not from smooth to rough—the reverse of the usual experience. Bunting (1940, 1946) studied changes in pigmentation of colonies of *Serratia marcescens*. Stocker (1949) investigated changes of phase of flagellar antigen in *Salmonella typhi-murium*; his Text-fig. 4 may be compared to the theoretical model illustrated in Fig. 3 of the present paper.

2·8. The effect of phenotypic delay (which is discussed more fully below) will usually be negligible in long-term experiments. The effect of phenotypic delay of, say, three generations on situations typified by Figs. 1—3 is merely to shift the graph to the right by a time interval of three generation lengths; phenotypic delay of this order of magnitude will usually be small in comparison with the time-span of the whole experiment.

There may, however, be an important bias in the determination of the initial rates of increase and decrease of ψ in pure cultures (§§2·3 (*c*), 2·5), for, in the presence of phenotypic delay, ψ will remain at 0 or 1 for several generations. Such a bias will be avoided if these gradients are determined by the usual method of fitting regression lines of ψ on t, since these lines are not constrained to pass through the origin.

3. THE DEFINITION OF A MUTATION RATE

3·1. The quantities λ and μ, which were introduced as mathematical concepts in §2·2, occur so frequently in the study of this subject that they deserve to be named. It is clear from the definition of λ that this quantity measures in some way the rate of mutation relative to the rate of growth of the wild-type organisms, and μ similarly represents the back mutation rate. It is convenient to describe λ and μ as 'relative mutation rates'. Some consideration must now be given to their interpretation in terms of the life cycle of an individual bacterium.

3·2. Consider a large population of wild-type organisms, growing at a rate a. We define the *generation time*, Λ, to be the length of time required for the population

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

size to increase from N to $2N$. The generation time is clearly a function of the growth rate, a. By (1)

$$2N = N\,e^{a\Lambda},$$

whence

$$\Lambda = (\log_e 2)/a = 0{\cdot}6931/a. \tag{6}$$

(The results of §2·6 may be verified from equation (6). For, putting $a = 0{\cdot}6931$ in (6), we have $\Lambda = 1$ unit; that is, the time is measured in generations. Similarly, putting $a = 1$ in (6), we have $\Lambda = 0{\cdot}6931 = \log_e 2$ units; the unit of time is now the generation time divided by $\log_e 2$.)

If the population is to increase smoothly, the N organisms initially present must be in different phases of their growth cycle, and will divide at different points of time throughout Λ.

3·3 For purposes of illustration it will be convenient to suppose that the mutation rate is small, so that any mutations taking place during a single generation time do not materially affect the size of the wild-type population. In §2·2, λ has been defined in terms of the rate of production of mutant organisms; as the population size increases by an increment $aN\,dt$, the number of mutations occurring is $\lambda aN\,dt$, which can be written $\lambda\,dN$, since $aN\,dt = dN$, the increment in the total population size. It follows that the number of mutations expected to occur as the population increases from N_1 to N_2 is

$$\int_{N_1}^{N_2} \lambda\,dN = \lambda(N_2 - N_1). \tag{7}$$

Equation (7) relates λ to the number of mutations expected during any period of growth.

The initial population size N_1 will frequently be negligible in comparison with the final population size N_2. In such a situation the expected number of mutations is denoted by m, and equation (7) reduces to

$$m = \lambda N_2, \tag{8}$$

a relationship of considerable importance (cf. (8a) below).

3·4. The solution to the differential equations (2), special cases of which have been discussed in §2·3, relates the number of *mutants* y to be expected at time t to the relative mutation rate λ, and hence (from (7)) to the number of *mutations* expected to occur. One of the consequences of the assumptions made at the beginning of §2·2 is that the size of a mutant clone, at a time t after the mutation, is e^{at}, provided the growth rates, a and b, are equal. Now this will be true, *on the average*, only if mutations occur with equal probability at any point during a division cycle. The point is illustrated by Fig. 4, which depicts the growth curves for mutant clones under three different assumptions about the time at which mutations occur:

(a) *Mutations occur at any point during the cycle, all points being equally likely.* The curve marked (a) in Fig. 4 illustrates the *average growth* of mutant clones. The growth curve in individual cases will vary between the limits (b) and (c).

(b) *Mutations occur only at the beginning of a cycle.* The mutant cell will not divide until the end of the first cycle.

(c) *Mutations occur only at the end of a cycle.* The mutant cell will divide immediately after the mutation.

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

For a given mutation rate λ, therefore, the number of mutants to be expected at time t is less or greater than that given by the model of §2, according as mutations tend to occur towards the beginning or towards the end of a life cycle.

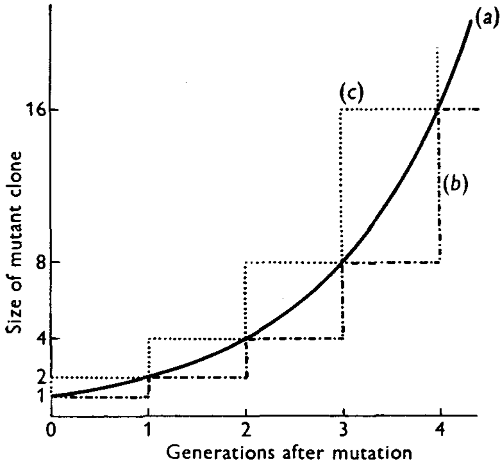

Fig. 4. Development of mutant clones according as mutations occur (*a*) with equal frequency throughout the life-cycle (showing average growth), (*b*) at the beginning of a cycle, (*c*) at the end of a cycle.

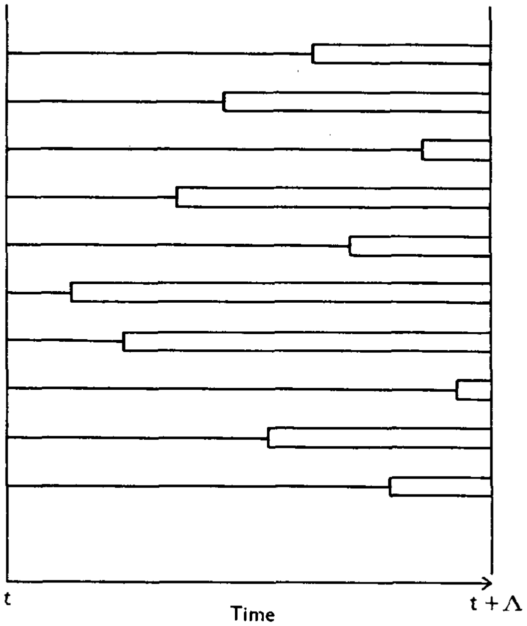

Fig. 5. Development of ten organisms during a generation time period, showing different moments of fission.

3·5. These three hypotheses, (*a*), (*b*) and (*c*), also affect the interpretation of λ in terms of the probability of a mutation during the life cycle of an individual bacterium. That this is so may be seen from the example illustrated in Fig. 5. For purposes of illustration we shall suppose that $\lambda = 0{\cdot}1$, a very much higher value than would occur in practice. Fig. 5 shows the development of ten organisms, dividing

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

at different instants during the generation time period, and giving rise to twenty organisms at the end of the period. From (7), the number of mutations expected during this period is (0·1) (20 − 10) = 1. Consider in turn the three hypotheses about the time at which mutations occur:

(*a*) *Mutations throughout the cycle.* The sum of the lengths of all the segments of life cycles in Fig. 5 is about 14·4 times the generation time. We expect exactly one mutation during this total 'exposure'. The probability of a mutation during one life cycle is therefore 1/14·4 = 0·070.

(*b*) *Mutations at the beginning of a cycle.* Exactly twenty life cycles start during our period of observation, and so there are twenty opportunities for mutation. The probability of a mutation per opportunity is therefore 1/20 = 0·05.

(*c*) *Mutations at the end of a cycle.* Ten life cycles end during the observation period, and the probability of a mutation per opportunity is 1/10 = 0·1.

In general, the probability that a mutation occurs during the life cycle of an individual organism is $0{\cdot}693\lambda$ ($=\lambda \log_e 2$), $0{\cdot}5\lambda$, or λ, according as mutations occur (*a*) with equal probability at all points of the cycle, (*b*) only at the beginning of the cycle, or (*c*) only at the end.

Previous workers have generally used the first value $0{\cdot}693\lambda$, thus implicitly assuming that mutations occur with equal frequency at all points of the cycle. Examples are the 'mutation rate per bacterium per division cycle' of Luria & Delbrück and Newcombe; and the 'mutation rate' used by Stocker. The relative mutation rate, λ, which is appropriate under hypothesis (*c*), is called by Luria & Delbrück the 'mutation rate per bacterium per time unit', and by Newcombe the 'mutation rate per bacterial division'.

3·6. The terms 'mutation' and 'mutation rate' refer essentially to genotypic changes, rather than to the phenotypic changes to which they give rise. If the phenotypic appearance coincides with the mutation, there is no ambiguity, and the results of § 3·5 may be taken to refer to either phenomenon. In general, however, the genotypic mutation may be followed by a 'phenotypic delay' of, perhaps, several generations, and the probability that a phenotypic change occurs during a life cycle is no longer equal for all organisms.

If, however, mutations occur randomly throughout the cycle, as in assumption (*a*) of § 3·4, and if the phenotypic delay is so small that the phenotypic appearance takes place at the end of the cycle, then, in § 3·4, (*a*) could refer to genotypic mutations and (*c*) to phenotypic appearances. The values of λ for genotypes and phenotypes would differ, for, denoting these values by λ_G and λ_P respectively, the probability of a mutation during an individual life cycle is, from § 3·5,

$$0{\cdot}693\lambda_G = \lambda_P.$$

This relationship can be expressed alternatively by the observation that when each phenotypic change takes place the population size is always rather larger than when the mutation occurred; hence λ_P must be smaller than λ_G.

We have here a particular case of the general rule, which has often been discussed in the literature on this subject, that the longer the phenotypic delay, the smaller the apparent mutation rate (as determined from *phenotypic* changes).

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

4. DISTRIBUTION PROBLEMS

4·1. The theory outlined in §2 is of the type described by statisticians as 'deterministic' rather than 'stochastic'; the population sizes x and y at any time t are determined *exactly* in terms of the growth rates, mutation rates and initial population sizes. In the description of physical phenomena, deterministic relationships are adequate for most practical purposes, although the 'stochastic', or random, approach is an essential part of modern theoretical physics. The growth of a bacterial population subject to mutation is an excellent example of a biological phenomenon the main feature of which—the occurrence of a mutation— is a random event, and to which a deterministic theory can provide only a rough approximation.

Consider, for instance, the change in the proportion of mutants, ψ, as illustrated by Figs. 1–3. In any one culture the value of ψ, if it could be measured accurately at each point of time, would be found to fluctuate about the theoretical curve. More precisely, in the situation illustrated by curve A in Fig. 3, the value of ψ in any one culture might depart appreciable from the curve in the very early stages of the growth period, and approach ψ_∞ consistently either below or above A. The theoretical curve may be regarded as an average over a large number of replicate cultures.

Now, in long-term experiments on one culture these fluctuations will probably not obscure the general trend of the curve, and may even be of less importance than the sampling errors involved in the estimation of the population sizes. In short-term experiments, throughout which ψ remains negligibly small, it is of great importance to examine the random fluctuation in the number of mutants from one replicate culture to another.

4·2. The first work in which the random fluctuation between replicate cultures was emphasized was that of Luria & Delbrück (1943) on the resistance of *Escherichia coli*, strain B, to bacteriophage. These authors observed the number of resistant organisms in each of a series of replicate cultures grown for the same length of time. This number was found to vary considerably from one culture to another, a result which, as Luria & Delbrück recognized, would be expected if resistance were due to spontaneous mutation. A few cultures, purely by chance, would experience an early mutation, and would accumulate very many more mutants than cultures in which the first mutation occurred later. As a control experiment a single culture was grown for a certain length of time, at the end of which replicate samples from the culture were exposed to attack by phage. The numbers of resistant organisms in these control replicates varied very much less than those in the separate cultures, and in fact no more than would be expected in a Poisson distribution.

Luria & Delbrück considered that these results supported the mutation theory for the origin of phage resistance, rather than that of adaptation, arguing that on the latter theory replicate cultures would be expected to have almost equal numbers of organisms which were able to adapt themselves to the new environment, as in the control experiment. These conclusions have been criticized by C. N. Hinshelwood and his colleagues (Jackson & Hinshelwood, 1950; Dean & Hinshelwood, 1952) on the grounds that the control experiment is inadequate; on the adaptation theory, samples from cultures grown separately would be expected to show more variation

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

in the proportion of cells able to adapt than samples from the same culture. Ryan (1952*a*), studying the adaptation of *Escherichia coli* to ability to use lactose, rejects these criticisms. He concludes 'that no solid evidence for Hinshelwood's theory exists. This, of course, does not exclude the possibility that in some instances, when the proper experimental tests have been performed, his theory will provide the best interpretation. But, before it is accepted in any instance, modern genetic conceptions must be shown clearly not to apply.'

Refinements of the experimental technique, such as that suggested by Newcombe (1949), will no doubt resolve this controversy. It may prove that certain types of bacterial variants arise by mutation, that others are due to adaptation, and that in some situations both mechanisms are at work. In any event the mathematical prediction of the distribution of number of mutants in replicate cultures, to be expected on the mutation theory, is likely to be an important tool in any such investigations.

4·3. Luria & Delbrück obtained some properties of the theoretical distribution, and an explicit formula was obtained by Lea & Coulson (1949). Lea & Coulson's distribution is expressed in terms of only one parameter, m, the number of mutations expected during the period of growth. If the final population size, N, is large in comparison with the initial size, equation (8) shows that

$$m = \lambda N. \tag{8a}$$

The probability p_y, that y mutants are present at time t, is given as the coefficient of z^y in the expansion of the generating function

$$G(m,z) = e^{-m} \exp\left[m \left\{ \frac{z}{1.2} + \frac{z^2}{2.3} + \dots \right\} \right].$$

In particular

$$\left.\begin{aligned} p_0 &= e^{-m}, \\ p_1 &= \tfrac{1}{2} m e^{-m}, \\ p_2 &= (\tfrac{1}{6} m + \tfrac{1}{8} m^2) e^{-m}. \end{aligned}\right\} \tag{9}$$

and

The distribution is extremely skew, and cannot easily be illustrated graphically. Lea & Coulson provide tables of the distribution for certain values of m up to 15. For higher values of m one can use the fact that a transformed variate χ, which is defined by

$$\chi = \frac{11 \cdot 6}{y/m - \log_e m + 4 \cdot 5} - 2 \cdot 02, \tag{10}$$

is very nearly normally distributed with zero mean and a standard deviation of unity.

4·4. It may be useful to emphasize here certain assumptions on which the theory of Luria & Delbrück and of Lea & Coulson rests, some of which may not always be valid. Assumptions (*a*)–(*f*) appear to be the most important.

(*a*) *No deaths take place.* It is doubtful whether this condition has been achieved in any of the investigations so far reported. A moderate death-rate is not likely to affect the shape of the distribution appreciably, but the mutation rate will be over-estimated.

(*b*) *The growth rates of the two types are equal.* It is not known to what extent the distribution is affected by differential growth rates; estimates of the mutation rate

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

from the number of mutants rather than the number of mutations would clearly be biased.

(c) *There is no delay in phenotypic expression.* There is strong reason to believe that this assumption is frequently invalid (Newcombe, 1948). The effect of phenotypic delay is to reduce the frequencies of cultures with small numbers of mutants much more drastically than the frequencies of cultures with high numbers. For in the latter cultures most of the mutants, being descended from early mutations, will have achieved phenotypic expression; whereas in cultures with less than, say, twenty mutants, a phenotypic delay of only a few generations will cause most of the mutant population to remain undetected. Methods of estimating mutation rates from the lower end of the distribution will therefore produce lower estimates than those based on the upper tail of the distribution (cf. §5 below). This tendency was noticed by Luria & Delbrück.

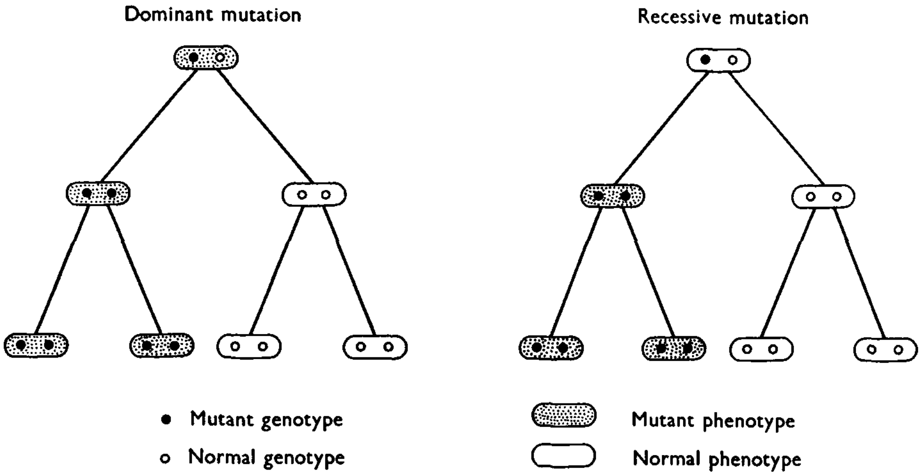

Fig. 6. 'Dominant' and 'recessive' mutation in cells with two nuclei (adapted from Lieb, 1951).

(d) *The bacteria are uninucleate.* Some consequences of the multinucleate arrangement of cells have been discussed by Newcombe and Hawirko (1949) and by Lieb (1951), who distinguish between 'dominant' and 'recessive' mutations. The two situations are illustrated in Fig. 6, which is adapted from Lieb's Fig. 1. A dominant mutation in one of the nuclei causes an immediate phenotypic change (if phenotypic delay is not also present), which continues in half the progeny. For a recessive mutation the phenotypic change does not take place until all the nuclei in one of the progeny are of the mutant type; in the example illustrated in Fig. 6, where the cells have two nuclei, this occurs in the generation following the mutation.

It is convenient to restrict the term 'phenotypic delay' to the uninucleate effect considered by Newcombe. It is not, therefore, applied to recessive mutation, although this phenomenon does involve a delay of phenotypic expression. The essential difference between the two phenomena is that in phenotypic delay in uninucleate organisms there is no delay in the *development* of the mutant clone, but only a temporary failure of the organisms to become phenotypically active. In

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

recessive mutation, on the other hand, the development of the mutant clone does not start for several generations; the effect is exactly the same as if the organisms were uninucleate and the mutation had occurred one or more generations later than it actually did.

For a recessive mutation, then, the theoretical distribution is still valid. The effect of a dominant mutation, however, is to curtail the upper end of the distribution in relation to the lower end—precisely the opposite effect to that of phenotypic delay.

(*e*) *The mutation rate is constant.* As pointed out in § 2·6, this condition may not always be satisfied.

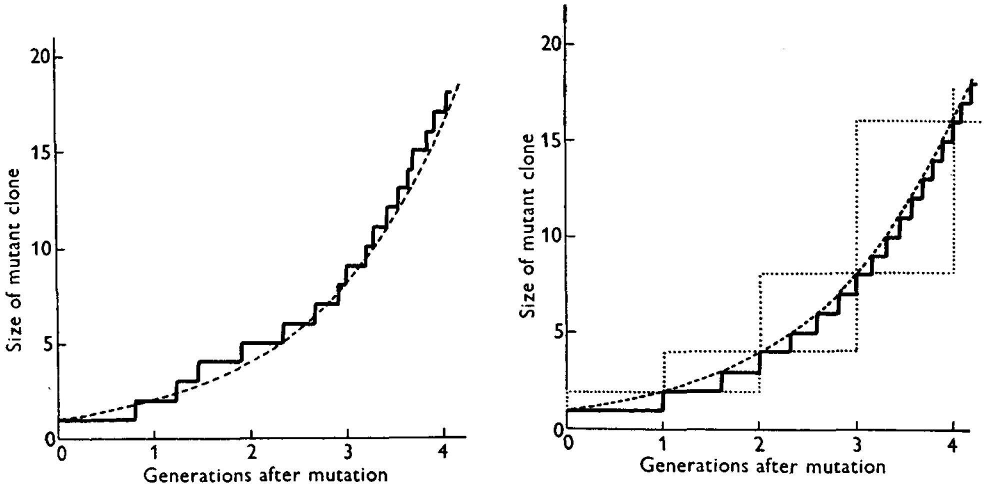

Fig. 7*a*

Fig. 7*b*

Fig. 7. Development of mutant clones in the two derivations of Lea & Coulson's distribution: (*a*) with random growth, (*b*) with deterministic growth.

(*f*) *The growth of mutant clones must approximate to the type shown in Fig.* 4, *curve* (*a*). Lea & Coulson obtained their result under two quite different assumptions about the development of a mutant clone. In their first derivation they regard the growth as a random phenomenon; the times of division would fluctuate from one clone to another, but if a very large number of clones were developed in parallel, the average clone size at a given time after the mutation would be the exponential function illustrated by Fig. 4, curve (*a*). The way in which a typical clone might develop is shown by the thick line in Fig. 7 (*a*); the curve (*a*) of Fig. 4 is given as a dotted line for comparison.

In Lea & Coulson's second derivation, the growth of the mutant clone is deterministic. The assumed type of growth is depicted by the thick line in Fig. 7 (*b*); the size of the clone is always equal to the whole number below the exponential function e^{at} illustrated in Fig. 4, curve (*a*). The curves (*a*), (*b*) and (*c*) of Fig. 4 are shown in dotted lines in Fig. 7 (*b*).

These two different types of mutant growth, assumed in the two derivations of Lea & Coulson's distribution, have, then, one important property in common. They both approximate, in some sense, to the exponential growth of Fig. 4, curve (*a*)—

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

in the first derivation *on the average*; and, in the second derivation, especially during the later stages of growth.

Neither of these two growth processes will be exactly realized in practice, but, provided the growth approximates on the average to the type illustrated in Fig. 4, curve (a), Lea & Coulson's distribution is not likely to be seriously invalidated on this account. The mutant clone will grow on the average as in Fig. 4, curve (a) if the mutations take place with equal probability at any point throughout the cycle (cf. § 3·4). Under other hypotheses as to the time during the cycle at which mutations occur, the lower tail of Lea & Coulson's distribution may be seriously affected. For instance, if hypothesis (c) of § 3·4 is true, the first mutant cell will immediately divide (as in Fig. 4, curve (c)), and there will be a considerable deficit of cells with one mutant organism.

4·5. The remaining restrictions are perhaps less serious:

(g) *N is large, λ and ψ are small.* These conditions can usually be met without any difficulty. If ψ remains small during the experiment the possibility of back-mutations can be neglected.

(h) *The final population sizes in different cultures are equal.* This condition will never be achieved, for even if the growth rate were constant from culture to culture and the cultures were grown from inocula of exactly one organism, these parent cells would be in different phases of the division cycle, and, at a given time, would have generated populations of different sizes. However, these population sizes N, are likely to vary by a factor of less than 2 about the average. A few numerical investigations suggest that variations in N of this order of magnitude are unimportant.

Variation in population size will also be caused by random fluctuations in the generation times, within any one culture, particularly during the early stages of growth. Lea & Coulson assume deterministic, exponential growth for the wild-type organisms. The effect of random fluctuations in the generation times is, however, likely to be considerably less than that, just discussed, of differences in phase of single parent cells, and can therefore be ignored.

(i) *Only one type of mutant is involved.* This restriction can be removed, provided all the other conditions are satisfied—in particular, the different types of mutant must have the same growth rate. The theory would then still hold, λ being interpreted as the sum of the individual mutation rates.

(j) *The inocula must contain no mutants.* This condition may be achieved, at least to a high degree of probability, by selecting small inocula from a population with a very low proportion of mutants.

4·6. In view of this rather formidable list of assumptions, each of questionable validity, it would perhaps be surprising to find in practice any very close agreement with the theory. There are unfortunately not many published results sufficiently extensive to permit a searching comparison.

I have elsewhere (Armitage, 1952) examined the observed distributions of Luria & Delbrück (1943, Exp. 23) and Newcombe (1948, Exps. A–H pooled); these refer to the resistance of *Escherichia coli*, strain B, to T 1. In Table 1 the distributions are compared with Lea & Coulson distributions having the same proportions of cultures with no mutants. In each case the observed distribution has too long an upper tail.

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

A similar discrepancy is found in the results of Demerec & Fano (1945) on the same phenomenon; their distributions are not published in full, but the point is borne out by their summary of the data.

Table 1. *Distribution of Luria & Delbrück* (1943, *Exp.* 23), *and Newcombe* (1948, *Exps. A–H pooled*), *on resistance of* Escherichia coli, *strain B, to phage T* 1. *Observed distributions compared with those expected on theory of Lea & Coulson* (1949)

	No. of cultures with given no. of mutants			
	Luria & Delbrück		Newcombe	
No. of mutants	Observed	Theoretical	Observed	Theoretical
0	29	*29·0*	29	*29·0*
1	17	*15·9*	29	*27·9*
2	4	*9·7*	11	*22·8*
3–4	6	*10·8*	17	*32·2*
5–8	6*	*9·2*	16	*34·0*
9–16	5*	*6·0*	22	*25·5*
17–32	5*	*3·2*	17	*14·6*
33–64	6*	*1·6*	28	*7·2*
> 64	9*	*1·6*	31	*6·6*
	87	*87·0*	200	*199·8*

* The grouping used here is determined by Lea & Coulson's tables, and differs from that given by Luria & Delbrück. The frequencies indicated have therefore been estimated by interpolation, and will probably be slightly inaccurate. The distribution as originally given by Luria & Delbrück is as follows:

No. of mutants	No. of cultures	No. of mutants	No. of cultures
0	29	6–10	5
1	17	11–20	6
2	4	21–50	7
3	3	51–100	5
4	3	101–200	2
5	2	201–500	4
			87

Table 2. *Distributions of Luria & Delbrück and Newcombe, fitted by a model assuming an average phenotypic delay of about four generations*

	No. of cultures with given no. of mutants			
	Luria & Delbrück		Newcombe	
No. of mutants	Observed*	Theoretical	Observed	Theoretical
0–8	62	*62·0*	102	*102·0*
9–24	8	*10·5*	30	*34·4*
25–40	4	*4·4*	20	*17·2*
41–72	4	*3·9*	18	*17·0*
73–136	4	*2·8*	10	*13·0*
137–264	2	*1·6*	15	*8·0*
265–520	3	*0·9*	4	*4·3*
> 520	0	*0·9*	1	*4·3*
	87	*87·0*	200	*200·2*

* Observed frequencies estimated from Luria & Delbrück's original table (see footnote to Table 1).

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

Newcombe suggested that these discrepancies were due to phenotypic delay, which, as stated in §4·4 (*c*), would have this sort of effect. Table 2 shows that the *general shape* of the distribution can be explained by a phenotypic delay, varying for the different descendants of each mutation, but on the average about four generations in length. The fitted distributions have been chosen to have the same proportions in the first group as the observed distributions. The method of fitting, and the rather curious method of grouping adopted, are explained in the earlier paper.

The distributions shown in Table 2 are too heavily grouped to permit any useful examination of the lower end of the distribution. For such an examination a more specific statistical model for phenotypic delay is required. One such model discussed by Armitage (1952) gave an unsatisfactory fit to the data of Luria & Delbrück and Newcombe. In this model it was assumed that each member of a mutant clone which, at the beginning of its life cycle, had not become phenotypically active, had a constant probability, ϵ, of doing so before its moment of fission. The distribution to be expected on this model is difficult to calculate completely,* but the probabilities of obtaining 0, 1, 2 and 3 phenotypically active organisms in a culture were obtained. This 'lower tail' of the theoretical distribution was fitted to each set of data by equating the observed and theoretical frequencies of cultures with no resistant organisms. Three different values of ϵ were tried; for each of the two series the fit was unsatisfactory. In spite of the failure of this particular model for phenotypic delay, other, possibly more realistic, formulations may be more successful.

Ryan (1952*a*, *b*) has reported some experiments on lactose-utilizing mutants in two strains of *E. coli*. For one strain the observed distribution of the number of mutants per culture is fitted quite well by Lea & Coulson's distribution; for the other strain the fit is unsatisfactory. Ryan refers to a number of other experiments of the Luria-Delbrück type, which are reported in the literature. Some of these, mostly with a small number of replicate cultures, provide distributions which agree reasonably well with Lea & Coulson's result; in others the theoretical distribution is inadequate.

An adequate experimental agreement, even on a large series of observations, with Lea & Coulson's or any other mathematical theory does not necessarily confirm the assumptions on which that theory rests. For example, phenotypic delay and a multinucleate structure with dominant mutant exert opposite types of disturbance on Lea & Coulson's distribution; if both these factors were operating together one might get a spurious agreement between theory and practice. Conversely, as Ryan points out, a discrepancy between Lea & Coulson's result and any experimental distribution does not immediately disprove the hypothesis of mutation. It is to be hoped that mathematical workers in this field will continue to explore the consequences of alternative models suggested by the experimenters. Meanwhile it would be of the utmost value if experimental workers could investigate the effect on the distribution of varying some of the experimental factors, such as the length of growth of the replicate cultures.

* An alternative formulation of this model by Kendall (1952, 1953) is likely to prove more flexible.

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

5. ESTIMATION OF MUTATION RATE FROM REPLICATE CULTURES

5·1. In the Luria-Delbrück type of experiment C replicate cultures are grown from small inocula to about the same size, N. In each culture the number of mutants, y, is determined. The theoretical distribution of y has been discussed in §4; on the assumptions stated there, the distribution depends on only one parameter, m, the number of *mutations* expected on the average in cultures growing to a size N. Equation (8a) shows the relation between m, N, and λ (the relative mutation rate).

The problem of estimating λ from an observed distribution of y in a series of C cultures has been much discussed in the literature, particularly by Luria & Delbrück, Lea & Coulson, and Armitage. The procedure in general is to estimate m by one of a number of alternative methods, and then to use equation (8a) to estimate λ. If the 'mutation rate per bacterium per division cycle' is required, on the assumption that mutations occur at all points of the division cycle, the estimate of λ must be multiplied by $\log_e 2 = 0{\cdot}693$ (cf. §3·5).

In some experiments it may be possible to observe directly the number of mutations occurring in each culture (Ryan, 1952a, b). These numbers should follow a Poisson distribution, and their average provides an estimate of m. This estimate will tend to be too low if phenotypic delay is present; if the cells have a multinucleate arrangement with dominant mutation, the mutation rate estimated by (8a) will tend to be too high, since we shall be estimating the mutation rate per cell, rather than per nucleus.

In most experiments, however, the problem is to estimate m, and hence λ, from the distribution of y. The seven principal methods so far proposed are described below. They are arranged not in order of efficiency, but in approximate order of sensitivity to two sorts of departure from the ideal conditions of §§4·4 and 4·5—phenotypic delay, and multinucleate arrangement with dominant mutation. The order, in fact, reflects roughly the relative dependence of each method on the lower or the upper portion of the distribution. The relative efficiencies of some of these methods have been discussed by Lea & Coulson and (more briefly) by Armitage. The results stated by these authors should be regarded with some reserve as they refer (a) to large values of C, and (b) to the ideal conditions of §§4·4 and 4·5. It may be possible, and preferable, to use formulae for sampling errors which are less dependent than those of Lea & Coulson on the exact validity of their distribution (cf. §5·2 below.)

Of the methods described below, no. 1 considerably underestimates m in the presence of phenotypic delay, numbers 2, 3 and 4 do so to a rather less extent, and numbers 5, 6 and 7 are affected only slightly. If the cells are multinucleate with dominant mutation, the estimate of λ from equation (8a) will tend to be too high by method 1, rather less so by methods 2, 3 and 4, and will be hardly affected by methods 5, 6 and 7. Recessive mutation produces in itself no disturbance: the estimate of λ given by any of the methods should be a valid estimate of the relative mutation rate per nucleus.

Method 1. *From the proportion of cultures with no mutants.* This is one of Luria & Delbrück's original methods. From equation (9),

$$m = -\log_e p_0.$$

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

The observed proportion of cultures with no mutants, $\hat{p}_0$, is substituted for p_0 in this equation, to give the estimate $\hat{m}_1$ of m. Under the ideal conditions this method is highly efficient for $m < 0{\cdot}5$, and quite satisfactory for $m < 1$.

Method 2. *Maximum likelihood.* This is the most efficient method under the ideal conditions. It is described fully by Lea & Coulson, who give tables to facilitate its use.

Method 3. *From the equation* $\Sigma\chi = 0$. Details are given by Lea & Coulson. Under the ideal conditions this method is fairly efficient for values of m greater than about 3.

Method 4. *From the median, r.* The C values of y are arranged in order, and the $\frac{1}{2}(C+1)$th observation, r, is called the median. (If C is an even number, the median is taken half-way between the $\frac{1}{2}C$th and $(\frac{1}{2}C+1)$th observations.) If all the y's were transformed into χ's by equation (10), the median of the χ's should, on the average, be 0, since χ is normally distributed with zero mean. Hence an estimate $\hat{m}_4$ of m is obtained by the equation

$$\frac{11{\cdot}6}{r/\hat{m}_4 - \log_e \hat{m}_4 + 4{\cdot}5} - 2{\cdot}02 = 0,$$

whence

$$r/\hat{m}_4 - \log_e \hat{m}_4 = 1{\cdot}24.$$

Lea & Coulson provide a table for the estimation of m by this method, and show that under the ideal conditions it is fairly efficient for values of m greater than about 3, but less so than method 3.

Method 5. *From the upper quartile.* The upper quartile, q, is defined as the $\frac{3}{4}(C+1)$th observation when the y's are arranged in increasing order. If $C+1$ is not divisible by 4, interpolation between adjacent values is necessary (as in the example below).

By a similar argument to that used for the median, the estimate $\hat{m}_5$ of m is obtained by the equation

$$q/\hat{m}_5 - \log_e \hat{m}_5 = 4{\cdot}09. \tag{11}$$

Table 3 below shows the value of $\hat{m}_5$ corresponding to different values of q. Under the ideal conditions this method is less efficient than method 4.

Method 6. *From the mean,* $\bar{y}$. Luria & Delbrück recognized that an estimate of m obtained by equating $\bar{y}$ to the theoretical mean would *usually* be too low; very rarely a culture would experience a particularly early mutation, and the estimate would be very much too high. They proposed a method of estimating m from $\bar{y}$, which was intended to correct for this skewness in the distribution of $\bar{y}$. Their estimate m_6' is obtained by solving the equation

$$\bar{y} = m_6' \log_e (Cm_6').$$

In the previous paper I have shown that this method will over-estimate the mutation rate considerably more than half the time. I therefore suggested a method which is just as likely to under-estimate m as to over-estimate it. This consists in solving the equation

$$\bar{y} = \hat{m}_6 \log_e (3{\cdot}46C\hat{m}_6).$$

However, any method based on the mean is very inaccurate owing to the tremendous sampling fluctuations in the mean, and the suggested method appears to have no advantage over that based on the quartile.

Method 7. *From the maximum value,* y_m. This method was proposed by Newcombe (1948), and uses only the highest value of y in the series, and the mean $\bar{y}$. The estimate $\hat{m}_7$ of m is given by

$$\hat{m}_7 = (y_m - \bar{y})/C.$$

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

There is probably less to be said in favour of this method than for any of the others, as its accuracy does not increase as the number of cultures C is increased.

5·2. Of the seven methods described above, the first four are affected more than the others by phenotypic delay and dominant mutation. Of the last three, which are based predominantly on the upper end of the distribution, method 5 is the most accurate. There would appear, therefore, to be strong reasons for estimating m from the upper quartile q, by equation (11), except in experiments in which m is less than about 3 or 4. (For smaller values of m the theoretical basis of equation (11) is unreliable; furthermore, as m decreases the method becomes increasingly affected by phenotypic delay and dominant mutation.)

Table 3. *Estimate, $\hat{m}$, of expected number of mutations, in terms of upper quartile, q, of observed distribution of numbers of mutants*

q	0	1	2	3	4	5	6	7	8	9
0	—	0·33	0·57	0·78	0·98	1·18	1·36	1·55	1·73	1·90
10	2·07	2·25	2·41	2·58	2·75	2·91	3·07	3·23	3·39	3·55
20	3·70	3·86	4·01	4·17	4·32	4·47	4·63	4·78	4·93	5·07
30	5·22	5·37	5·52	5·67	5·81	5·96	6·10	6·25	6·39	6·54
40	6·68	6·82	6·96	7·11	7·25	7·39	7·53	7·67	7·81	7·95
50	8·09	8·23	8·37	8·51	8·64	8·78	8·92	9·06	9·19	9·33
60	9·47	9·60	9·74	9·87	10·01	10·15	10·28	10·41	10·55	10·68
70	10·82	10·95	11·08	11·22	11·35	11·48	11·62	11·75	11·88	12·01
80	12·15	12·28	12·41	12·54	12·67	12·80	12·93	13·06	13·19	13·32
90	13·45	13·58	13·71	13·84	13·97	14·10	14·23	14·36	14·49	14·62
100	14·75	14·88	15·00	15·13	15·26	15·39	15·52	15·64	15·77	15·90
110	16·03	16·15	16·28	16·41	16·53	16·66	16·79	16·91	17·04	17·16
120	17·29	17·42	17·54	17·67	17·79	17·92	18·04	18·17	18·29	18·42
130	18·54	18·67	18·79	18·92	19·04	19·17	19·29	19·42	19·54	19·66
140	19·79	19·91	20·04	20·16	20·28	20·41	20·53	20·65	20·78	20·90
150	21·02	21·14	21·27	21·39	21·51	21·63	21·76	21·88	22·00	22·12
160	22·25	22·37	22·49	22·61	22·73	22·86	22·98	23·10	23·22	23·34
170	23·46	23·58	23·71	23·83	23·95	24·07	24·19	24·31	24·43	24·55
180	24·67	24·79	24·91	25·03	25·15	25·27	25·39	25·51	25·63	25·75
190	25·87	25·99	26·11	26·23	26·35	26·47	26·59	26·71	26·83	26·95
200	27·07	—	—	—	—	—	—	—	—	—

Table 3 enables the estimate $\hat{m}$ (from which we now drop the suffix) to be obtained readily for any value of q up to 200. As an example, consider Newcombe's Exp. A, in which $N = 3{\cdot}1 \times 10^8$. The numbers of mutants in the twenty-five cultures, when arranged in order, are:

0	3	13	36	60
0	3	14	37	140
0	4	27	43	160
1	8	30	48	231
1	9	35	55	447

The upper quartile is the $(3 \times 26)/4$th, i.e. the $19\frac{1}{2}$th observation, which we take as half-way between the 19th and the 20th. Thus

$$q = \tfrac{1}{2}(48 + 55) = 51{\cdot}5.$$

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

Interpolating in Table 3 for $q=51{\cdot}5$, we find

$$\hat{m}=8{\cdot}30.$$

Finally, our estimate of λ is, from equation (8*a*),

$$\frac{\hat{m}}{N}=\frac{8{\cdot}30}{3{\cdot}1\times10^{8}}=2{\cdot}7\times10^{-8}.$$

(The result given in Table 3 of Armitage (1952) is $2{\cdot}6\times10^{-8}$. The slight discrepancy is due to the use of a different, and less satisfactory, definition of a sample quartile in the calculations summarized in Table 3 of the earlier paper.) On the assumption that mutations take place with equal frequency at all points of the cycle, the 'mutation rate per bacterium per division cycle' is estimated as

$$(2{\cdot}7\times10^{-8})\,(0{\cdot}693)=1{\cdot}9\times10^{-8}.$$

The standard error of the estimate $\hat{m}$ given by (11) may be calculated by the method used by Lea & Coulson for the median (pp. 278–9 of their paper). The result is

$$\text{S.E.}\,(\hat{m})=\frac{8{\cdot}7m}{(5{\cdot}1+\log_e m)\sqrt{C}}.$$

Substituting the estimated value $\hat{m}$ for m in this equation, and making use of (11), this formula becomes

$$\text{S.E.}(\hat{m})=\frac{8{\cdot}7\hat{m}^2}{(\hat{m}+q)\sqrt{C}}. \qquad (12)$$

This derivation depends, however, on the validity of the normal transformation (10). In view of the possibility of disturbing factors like phenotypic delay, and of the doubtful relevance of the standard error unless C is very large, an alternative approach is desirable. It is possible to calculate confidence limits for the expected upper quartile, which do not depend upon any assumptions about the form of the distribution. Table 4 may be used for the 95 % level of confidence (which corresponds, in the earlier method, to taking ± 1·96 times the standard error on either side of $\hat{m}$). The limits for q obtained by using Table 4 will include the true quartile in about 95 % of the cases.

The use of Table 4 may be illustrated on the data (Newcombe's Exp. A) given earlier. The quantities tabulated in Table 4 are the ranks of the members of the sample (when put in ascending order) which are to be used as confidence limits. In our example, $C=25$, and the ranks for the limits are 14·8 and 23·2. Interpolating between the 14th and 15th, and between the 23rd and 24th members of the sample, we find the limits for the quartile to be 34·0 and 174·2. From Table 3 these correspond to $m=5{\cdot}81$ and 23·97, respectively. Finally, the limits for the relative mutation rate, λ, are

$$\frac{5{\cdot}81}{3{\cdot}1\times10^{8}} \quad\text{and}\quad \frac{23{\cdot}97}{3{\cdot}1\times10^{8}}$$

$$=1{\cdot}9\times10^{-8} \quad\text{and}\quad 7{\cdot}7\times10^{-8}.$$

The limits for m, obtained from equation (12) by taking ± 1·96 S.E. $(\hat{m})$ on either side of $\hat{m}$, are 4·4 and 12·2. The fact that these limits are considerably closer together than the values 5·8 and 24·0, which were obtained from Table 4, does not, of course, justify the first method. Indeed, the asymmetry of the limits obtained from Table 4 suggests that the use of symmetrical limits based on (12) involves too crude an approximation.

The theoretical basis of Table 4 is explained in the Appendix.

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

Table 4. *Ranks of members of sample to be used as* 95% *confidence limits for upper quartile*

No. of replicate cultures, C	Rank for Lower limit	Rank for Upper limit
10	5·2	—
11	5·8	—
12	6·4	—
13	7·0	13·0
14	7·7	13·8
15	8·3	14·7
16	8·9	15·6
17	9·6	16·4
18	10·2	17·3
19	10·9	18·1
20	11·5	19·0
21	12·2	19·8
22	12·8	20·7
23	13·5	21·5
24	14·2	22·4
25	14·8	23·2
26	15·5	24·0
27	16·2	24·9
28	16·8	25·7
29	17·5	26·5
30	18·2	27·4
31	18·8	28·2
32	19·5	29·0
33	20·2	29·8
34	20·9	30·7
35	21·5	31·5
36	22·2	32·3
37	22·9	33·1
38	23·6	33·9
39	24·3	34·8
40	24·9	35·6
41	25·6	36·4
42	26·3	37·2
43	27·0	38·0
44	27·7	38·8
45	28·4	39·7
46	29·0	40·5
47	29·7	41·3
48	30·4	42·1
49	31·1	42·9
50	31·8	43·7

For $C > 50$, use the following approximations:
Rank for lower limit $= 0{\cdot}75C - 0{\cdot}849\sqrt{C} + 0{\cdot}5$.
Rank for upper limit $= 0{\cdot}75C + 0{\cdot}849\sqrt{C} + 0{\cdot}5$.

6. NOTATION

The notation used in this paper differs somewhat from that of the previous paper. In particular, it seemed desirable to introduce simpler symbols for the two relative mutation rates, λ and μ, and for the two growth rates, a and b, than were used previously. It also seemed convenient to denote the upper quartile of a series of observations by q, rather than l. The following table shows the relation between the main symbols used here, and those used in my previous paper (A), by Luria & Delbrück (LD), by Shapiro (S), and by Lea & Coulson (LC).

Present	A	LD	S	LC
a	$a+g$	(Unity)	$a+m$	β
b	$b+h$	·	$c+b$	·
C	C	C	·	N
m	m	m	·	m
N	N_t	N_t	$x+y$	n
q	l	·	·	·
r	·	·	·	r_0
x	x	·	x, s, P	·
y	y	·	y, r, R	r
λ	$g/(a+g)$	a	$m/(a+m)$	α/β
μ	$h/(b+h)$	·	$b/(c+b)$	·
χ	χ	·	·	x

SUMMARY

This paper is a short exposition of the mathematical and statistical theory of the growth of bacterial populations subject to mutation.

A mathematical model for the long-term development of a mixed population with two types of organism is proposed. The proportion of organisms which are of the mutant type eventually approaches an asymptotic value, which is independent of the initial composition of the population. A procedure is outlined for estimating the forward and backward mutation rates from a long-term experiment.

The exact interpretation of the constants representing mutation rates requires some assumption about the point of time, during an individual life cycle, at which mutations occur. The usual assumption is that mutations can occur with equal frequency at all instants during the cycle.

In short-term experiments, in which the proportion of mutants is at all times negligible, it is important to consider the variation between the numbers of mutants developing in replicate cultures. The theoretical distribution of Lea & Coulson may be disturbed by the failure of any one of a number of assumptions; the effects of such disturbances are considered in some detail.

Various methods of estimating the mutation rate from an observed series of replicate cultures are examined. Two of the main sources of disturbance of the theoretical distribution may be delay of phenotypic expression, and the existence of multinucleate cells with dominant mutation. These factors affect particularly the lower tail of the distribution, and it is suggested that a fairly safe procedure may be to estimate the mutation rate from the upper quartile of the observed distribution. Tables 3 and 4 enable the estimate of the mutation rate, together with 95 % confidence limits, to be readily calculated.

I am indebted to Dr Forrest Fulton and Mr D. G. Kendall for many valuable suggestions on the presentation of this paper; to Miss Irene Allen for her help in the computation of Tables 3 and 4; and to Mrs M. G. Young for drawing the figures.

APPENDIX: CONFIDENCE LIMITS FOR THE UPPER QUARTILE

Suppose that, in successive samples, each of C observations drawn at random from some population and arranged in ascending order of magnitude, the k_1th and k_2th members of each sample are used as lower and upper confidence limits for the upper quartile of the population.

The probability that the population upper quartile is less than the lower limit, defined in this way, is the probability that less than k_1 observations out of C fall below the upper quartile of the population. This is the sum of the first k_1 terms of the binomial expansion $(0{\cdot}25+0{\cdot}75)^C$

$$=(0{\cdot}25)^C+C(0{\cdot}25)^{C-1}\,(0{\cdot}75)+\ldots+\binom{C}{k_1-1}\,(0{\cdot}25)^{C-k_1+1}\,(0{\cdot}75)^{k_1-1}.$$

This partial sum of the binomial expansion may be written, in terms of the incomplete beta-function, as $$I_{0{\cdot}25}(C-k_1+1,\,k_1).$$

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

Similarly, the probability that the population upper quartile exceeds the upper limit, defined as the k_2th member of the sample, is the probability that at least k_2 observations out of C fall below the population upper quartile. This probability is the sum of all except the first k_2 terms of the binomial expansion $(0{\cdot}25+0{\cdot}75)^C$,

$$=I_{0\cdot 75}(k_2,\, C-k_2+1).$$

Now, if it were possible to choose integers k_1 and k_2 such that

$$I_{0\cdot 25}(C-k_1+1,\, k_1)=I_{0\cdot 75}(k_2,\, C-k_2+1)=0{\cdot}025, \tag{13}$$

these limits would define a central confidence interval for the upper quartile, with confidence coefficient 0·95, independently of any assumption about the functional form of the distribution. The values of k_1 and k_2 satisfying (13) (which are the lower and upper limits tabulated in Table 4) are, however, not integers. A similar difficulty arises in the theory of confidence limits for parameters of discrete distributions, and more than one approach has been suggested. The solution proposed here is that a rank of, say, 13·8, should be interpreted by linear interpolation between the observations whose ranks are 13 and 14. The resulting limits will not have a confidence coefficient of exactly 0·95, but the discrepancy will probably be small.

The existence of distribution-free confidence intervals for population percentiles appears to have been first noted by Thompson (1936). A table for obtaining confidence intervals for the median is given by Nair (1940). It is analogous to the present Table 4, except that the ranks are given as integers, so that the confidence coefficient is always greater than 0·95.

The entries in Table 4 were obtained by inverse interpolation in tables of the cumulative binomial distribution (National Bureau of Standards, 1950), and checked by inverse interpolation in the tables of the incomplete beta-function (Pearson, 1934). For $C<13$, the upper limit cannot be calculated: there is a probability greater than 0·025 that the extreme member of the sample falls below the upper quartile of the distribution.

Neither set of tables referred to above permits the extension of Table 4 beyond $C=50$. For $C>50$, satisfactory approximations to k_1 and k_2 are provided by the following formulae (based on the normal approximation to the binomial distribution):

$$k_1=0{\cdot}75C-0{\cdot}849\,\sqrt{C}+0{\cdot}5,$$

$$k_2=0{\cdot}75C+0{\cdot}849\,\sqrt{C}+0{\cdot}5.$$

For $C=50$, for example, these formulae give the approximate values

$$k_1=32{\cdot}0,\quad k_2=44{\cdot}0,$$

as compared with the correct values (from Table 4),

$$k_1=31{\cdot}8,\quad k_2=43{\cdot}7.$$

Table 4 may also prove of value in other problems, in which confidence intervals for the *lower* quartile are required. For a given sample size, C, the quantities k_1 and k_2 are read from Table 4. Lower and upper 95 % confidence limits for the lower quartile are then provided by the members of the sample whose ranks are respectively $C-k_2+1$ and $C-k_1+1$, the observations being arranged in ascending order. Fractional ranks are, as before, interpreted by linear interpolation between adjacent members of the sample.

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press

REFERENCES

Armitage, P. (1952). The statistical theory of bacterial populations subject to mutation. *J. R. statist. Soc.* B, **14**, 1.

Bunting, M. I. (1940). The production of stable populations of color variants of *Serratia marcescens* no. 274 in rapidly growing cultures. *J. Bact.* **40**, 69.

Bunting, M. I. (1946). The inheritance of color in bacteria, with special reference to *Serratia marcescens*. *Cold Spr. Harb. Symp. quant. Biol.* **11**, 25.

Catcheside, D. G. (1951). *The Genetics of Micro-organisms*. London: Pitman.

Dean, A. C. R. & Hinshelwood, C. N. (1952). The applicability of the statistical fluctuation test. *Proc. Roy. Soc.* B, **139**, 236.

Demerec, M. & Fano, U. (1945). Bacteriophage-resistant mutants in *Escherichia coli*. *Genetics*, **30**, 119.

Deskowitz, M. & Shapiro, A. (1935). Numerical relations of an unstable variant of *Salmonella aerotrycke*. *Proc. Soc. exp. Biol., N.Y.*, **32**, 573.

Jackson, S. & Hinshelwood, C. N. (1950). An investigation of the nature of certain adaptive changes in bacteria. *Proc. Roy. Soc.* B, **136**, 562.

Kendall, D. G. (1952). Les processus stochastiques de croissance en biologie. *Ann. Inst. Poincaré* (to appear).

Kendall, D. G. (1953). Stochastic processes and the growth of bacterial colonies. *Symp. Soc. exp. Biol.* **7** (to appear).

Lea, D. E. & Coulson, C. A. (1949). The distribution of the numbers of mutants in bacterial populations. *J. Genet.* **49**, 264.

Lieb, M. (1951). Forward and reverse mutation in a histidine-requiring strain of *Escherichia coli*. *Genetics*, **36**, 460.

Luria, S. E. & Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, **28**, 491.

Nair, K. R. (1940). Table of confidence interval for the median in samples from any continuous population. *Sankhyā*, **4**, 551.

National Bureau of Standards (1950). *Tables of the Binomial Probability Distribution*. Washington: U.S. Government Printing Office.

Newcombe, H. B. (1948). Delayed phenotypic expression of spontaneous mutations in *Escherichia coli*. *Genetics*, **33**, 447.

Newcombe, H. B. (1949). Origin of bacterial variants. *Nature, Lond.*, **164**, 150.

Newcombe, H. B. & Hawirko, R. (1949). Spontaneous mutation to streptomycin resistance and dependence in *Escherichia coli*. *J. Bact.* **57**, 565.

Novick, A. & Szilard, L. (1950). Experiments with the chemostat on spontaneous mutations of bacteria. *Proc. nat. Acad. Sci., Wash.*, **36**, 708.

Pearson, K. (1934). *Tables of the Incomplete Beta-function*. London: 'Biometrika' Office.

Ryan, F. J. (1952*a*). Adaptation to use lactose in *Escherichia coli*. *J. gen. Microbiol.* **7**, 69.

Ryan, F. J. (1952*b*). Distribution of numbers of mutant bacteria in replicate cultures, *Nature, Lond.*, **169**, 882.

Shapiro, A. (1946). The kinetics of growth and mutations in bacteria. *Cold Spr. Harb. Symp. quant. Biol.* **11**, 228.

Stocker, B. A. D. (1949). Measurements of rate of mutation of flagellar antigenic phase in *Salmonella typhi-murium*. *J. Hyg., Camb.*, **47**, 398.

Thompson, W. R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Ann. math. Statist.* **7**, 122.

(*MS. received for publication* 9. ix. 52)

https://doi.org/10.1017/S0022172400015606 Published online by Cambridge University Press