The origin of mutants

John Cairns, Julie Overbaugh & Stephan Miller

Department of Cancer Biology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

Nucleic acids are replicated with conspicuous fidelity. Infrequently, however, they undergo changes in sequence, and this process of change (mutation) generates the variability that allows evolution. As the result of studies of bacterial variation, it is now widely believed that mutations arise continuously and without any consideration for their utility. In this paper, we briefly review the source of this idea and then describe some experiments suggesting that cells may have mechanisms for choosing which mutations will occur.

WHEN populations of single cells are subject to certain forms of strong selection pressure, variants emerge bearing changes in DNA sequence that bring about an appropriate change in phenotype. If the selection is being applied to cells growing in liquid, the population with the original genotype will sooner or later be supplanted by the variants; if the cells are immobilized on a plate, the variants will form colonies or papillae, rising above the rest of the population. Our problem is to determine how many of these variants are arising as a direct and specific response to the selection pressure (would not have occurred in its absence) and how many are 'spontaneous' (would have arisen even in the absence of selection).

Luria and Delbrück were the first to attempt this distinction[1]. They studied the origin of the phage-resistant mutants that are found when cultures of *Escherichia coli* are plated out in the presence of bacteriophage T1. They considered two extreme models. In the first model, the mutants are assumed to arise after plating, in response to the attack by the bacteriophage; the mutations occur in a small fraction of the bacteria exposed to the virus, and the number of mutants found in each of a large group of independent cultures should form a Poisson distribution whose mean is simply the product of the number of bacteria plated and the probability that a bacterium can become resistant before it is killed by the virus. In the second model, the mutants are assumed to arise spontaneously, as the result of a rare error in replication that has a constant probability of occurring in the life of each bacterium; in this case, the mutational events will be distributed randomly throughout each culture's previous history (scattered at random through its family tree), and the final number of mutants should be highly variable because occasional cultures will, by chance, have suffered a mutation early in their growth and will therefore end up containing a very large number of mutants. When put to the test, the number of T1-resistant mutants was observed to vary greatly from culture to culture, with the occasional culture containing up to 100 times the median value for the group as a whole. Luria and Delbrück concluded that mutation to phage resistance is the result of rare spontaneous events, occurring during the prior growth of cultures before they had been exposed to any selection.

By various ingenious sampling procedures such as 'replica plating', which tested cells' siblings rather than the cells themselves[2,3], it subsequently became possible to prepare pure cultures of mutants (for example, streptomycin-resistance mutants) without at any time having exposed the cells or their ancestors to any selection pressure. These experiments constituted the final proof that at least some forms of bacterial mutation are occurring spontaneously, before the bacteria can have any indication of each mutation's possible utility. They had, however, shown only that some mutants do arise spontaneously, in the absence of selection. In fact, we now know that mutations to phage and streptomycin resistance are not expressed until several generations after the change in DNA sequence has occurred[4]; thus the resistant mutants one isolates by spreading cultures on selective plates are the result of events that must have occurred many generations before one applies the selection that demonstrates their existence. So these classical experiments could not have detected (and certainly did not exclude) the existence of a non-random, possibly product-oriented form of mutation.

To look for such a process, we have studied mutations that are immediately expressed, and where the selection pressure rewards mutants by letting them multiply but allows all the other, non-mutant cells to survive so that they can at least have the opportunity to perform directed mutation. For this analysis, it was necessary to calculate the expected composite distribution, where some mutants are arising spontaneously (during the prior growth of each culture) and some are arising later, after the cells are put on selective plates.

Mutant distribution in independent cultures

We have to consider two distributions and their combination. The first distribution is created when mutants arise at random while the population in each culture is increasing by binary fission; at this stage, before any selection has been applied, mutants and non-mutants are postulated to multiply at the same rate; when the cultures are put out on selective plates, the mutants are able to form colonies. In these circumstances, the distribution of mutants among a set of independent cultures will be determined by a single parameter, the constant probability of mutation per cell per generation. Lea and Coulson[5] were the first to derive an exact solution. The distribution has certain awkward properties[6,7] but the following argument leads to a useful approximation for its upper end. For a culture undergoing binary fission, the number of cells (n) present at any particular moment plus the number present in all previous generations (that is, n plus $n/2+n/4+n/8+\cdots$) is twice as great as the number present in all previous generations ($n/2+n/4+n/8+\cdots$); so the chance that any particular mutation occurred at least g generations ago is twice the chance that it occurred at least $g+1$ generations ago. If, therefore, we just consider those very rare cultures that suffered a mutation so early in their history that in each culture one single clone (that is, one single mutational event) accounts for virtually all the mutants finally present, we see that the proportion of such cultures that contain at least 2^g cells will be twice the proportion with at least 2^{g+1}; this describes exactly the upper end of the distribution. In fact, if the cultures finally contain N cells and have suffered an average of m mutational events per culture (the mutation rate is $m/2N$ per cell), the probability (P_j) of finding a 'jackpot' of at least j mutants approaches m/j. This approximation had previously been derived for the case where m is very small[8], but actually it holds for the upper end of the distribution irrespective of the value of m. Figure 1*a* shows the exact relation between log x and log P_x (the logarithm of the proportion of cultures with x or more mutants) for several values of m.

A very different distribution will be created if, instead, all the

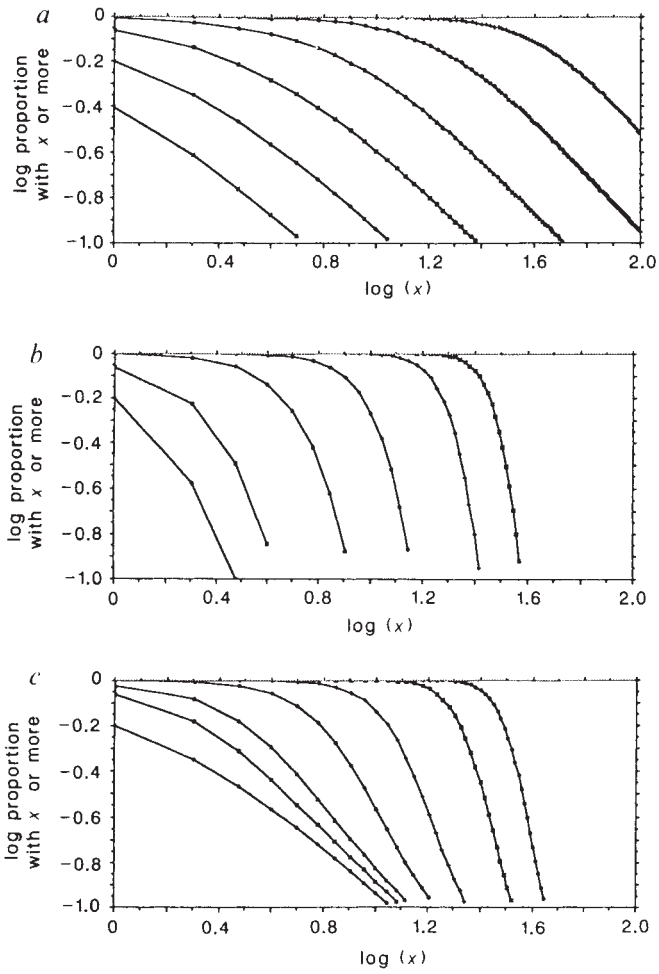

Fig. 1 The expected distributions of mutants among sister cultures that have experienced various kinds of mutational process. Each shows how the logarithm of the probability that a culture has at least x mutants ($\log_{10} P_x$) declines as $\log_{10} x$ increases. *a*, The case where the mutational events are distributed randomly among the entire lineage of each culture (calculated from Lea and Coulson Eq-15, ref. 5). The six examples shown are where the mean number of events per culture (m) is 0.5, 1, 2, 4, 8 and 16. *b*, The case where the mutational events are distributed randomly among the last generation of cells (that is they form a Poisson distribution). The six examples shown are where the mean number of events per culture (μ) is 1, 2, 5, 10, 20 and 30. *c*, Some combinations of the two previous distributions (where m events occur at random throughout the previous lineage of each culture and μ additional events occur in the final generation, among the cells that are being tested for the presence of the mutation). The examples show the expected distribution when $m = 1$ and $\mu = 0$, 1, 2, 5, 10, 20 and 30.

mutants arise after plating, among the bacteria exposed to the selection pressure. The numbers in independent cultures should form a Poisson distribution, and some examples are shown, for comparison, in Fig. 1*b*. The distributions are much narrower and log P_x falls rapidly as log x increases, especially when the average number of mutations per culture is high.

If both forms of mutation are occurring, the distribution of log P_x will be a composite of the two previous distributions. The lower end of the distribution will tend to be dominated by the mutations arising after selection, so that log P_x will initially fall steeply with increasing x; at high values, the distribution will be dominated by the rare jackpots produced by the mutations that occurred early during prior growth, and P_x will be inversely proportional to x. Various composite distributions are shown in Fig. 1*c*.

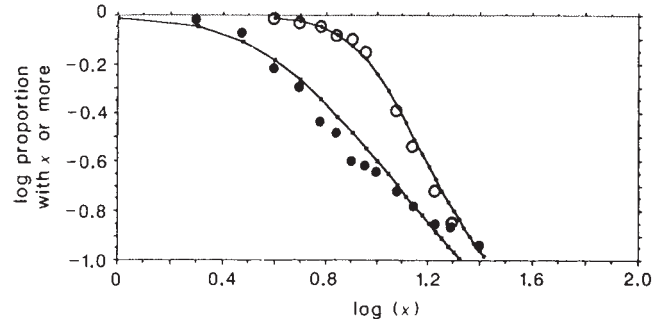

Fig. 2 The observed distribution of Lac$^+$ revertants for *E. coli* Δ *lac pro F' lacZ*$_{am}$ (filled circles) and for a derivative bearing a deletion of *uvrB* and *bio* (open circles). Sixty 3-ml cultures of each strain, in M9 medium plus 0.05% yeast extract, were grown overnight with aeration at 37 °C, and then each culture was centrifuged, resuspended in M9 and plated on M9-lactose plates. The colonies were counted after 48 h. The curves show the expected distributions for $m = 1.6$, $\mu = 2$ (on the left), and $m = 1.6$, $\mu = 7$ (on the right). (Roughly a third of the revertants, produced by these strains, are amber suppressors, and these form slightly smaller colonies at 48 h.)

Mutations affecting lactose fermentation

Bacterial variation in colony morphology was observed soon after the invention of plating procedures. One of the earliest studied examples was the formation of papillae of lactose-fermenting mutants, seen when non-fermenting strains of *E. coli* are plated out on rich media containing lactose, and it was attributed simply to a process of random variation and survival of the fittest[9]. This is a mutation that presumably is expressed as soon as it occurs, and so it is suitable for our present purposes.

Following the publication of tables for the Luria–Delbrück distribution[5], Ryan studied the distribution of Lac$^+$ variants in replicate cultures of Lac$^-$ strains spread on lactose-minimal plates and, surprisingly, he found that in some strains it was narrower than expected[10]. We have confirmed this. The shape of the distribution of Lac$^+$ revertants of strains bearing an amber mutation in *lacZ* depends on the genotype of the strain and on the medium in which it was growing before being plated; in certain circumstances, the distribution looks rather like some of those shown in Fig. 1*c*, as if there are two periods when mutations are occurring—first during the period of growth (producing the kind of distribution shown in Fig. 1*a*) and then in stationary phase (producing the kind of distribution shown in Fig. 1*b*). It is the second of these that is determined by the medium and genotype. In separate experiments, not described here, we have shown that such stationary phase mutation seems to be under the control of genes in the *uvrB-bio* region of the chromosome, but we have so far not been able to derive any simple picture of the regulatory process.

When multiple cultures of *E. coli lacZ*$_{am}$ and of the same strain bearing a deletion in the *uvrB-bio* region are grown in yeast extract or in M9-glycerol and then plated on minimal-lactose plates, the two strains appear to have the same mutation rate during growth (they are indistinguishable at the upper ends of their distributions), but the *uvrB* strain adds a large number of extra mutants that are apparently part of a Poisson distribution (Fig. 2). Interestingly, these extra mutants tend to be slower to produce colonies than the mutants that are members of jackpots (that is, slower than the mutants already in existence at the time of plating). This suggests that the extra mutant colonies are the result of late events, occurring on the plates when the bacteria are in the presence of lactose and under strong selection pressure to become Lac$^+$. The *uvrB* strain produces, each day about 3–5 times as many late mutants as the *uvr*$^+$ strain and it is therefore the strain of choice, when testing for the possible existence of directed mutation.

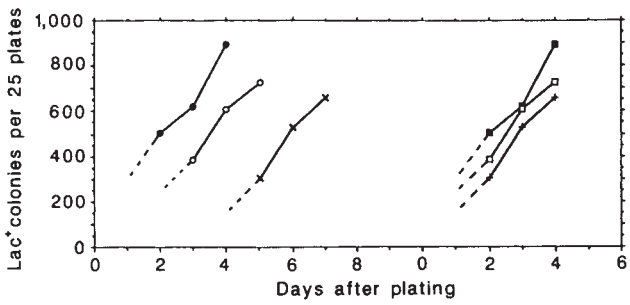

Fig. 3 The accumulation of Lac$^+$ revertants of *E. coli* Δ *uvrB-bio* Δ*lac pro F'lacZ*$_{am}$ after exposure to lactose on M9 plates. Six large cultures were grown overnight in M9 plus 0.1% glycerol. Fifteen 1-ml aliquots from each culture were plated with 2.5 ml top-layer agar on to M9-agar plates; the 90 plates were then overlaid with top-layer agar to ensure that all bacteria were buried in the agar; five plates from each culture were overlaid immediately with top-layer agar containing lactose, five were overlaid at 24 h, and five at 72 h. One of the six cultures proved to have a small jackpot of mutants, and so its plates were discarded; the plates from the remaining five cultures were counted each day, and the total counts are shown in real time on the left and in relation to time after the addition of lactose on the right.

To determine the role of lactose in the production of late Lac$^+$ mutants, aliquots of several replicate cultures of the *uvrB* strain were plated on M9 plates without lactose. When the plates were overlaid immediately with top-layer agar containing lactose, they showed an average of 20 colonies by 48 h and then gained an extra 8 colonies a day thereafter. If, however, the addition of lactose was delayed one or three days, the whole time course of appearance of colonies was delayed one or three days; this delay was seen even when the plates contained the non-hydrolysable inducer of the *lac* operon, isopropyl-thiogalactoside. In short, the accumulation of late Lac$^+$ mutants occurred only in the presence of lactose (Fig. 3).

One trivial explanation could be that this Lac$^-$ strain can grow slowly in lactose (either because its amber mutation in *lacZ* is leaky or because the lactose is impure), and that is why it can continue to produce mutants on plates containing lactose. We have therefore tested for the presence of another rapidly expressed mutation, valine-resistance (Val^R), by overlaying the lactose plates at various times with agar containing glucose and valine, and we find that the number of Val^R colonies countable two days after overlaying with valine actually goes slightly downwards with time (just as did the number of Lac$^+$ colonies in the absence of lactose). In other words, a population of bacteria that is accumulating Lac$^+$ revertants on a lactose-minimal plate is not, at the same time, accumulating Val^R mutants.

This experiment suggests that populations of bacteria, in stationary phase, have some way of producing (or selectively retaining) only the most appropriate mutations. For *E. coli*, under pressure to revert one particular amber codon, 'appropriate mutation' does not add much to the high background mutation rate arising from errors during replication. The next two examples, however, suggest that it plays a much larger role when the selection pressure is for complex changes in genotype that must seldom if ever arise from chance errors in replication.

The Shapiro deletion

In order to study the process of spontaneous deletion, Shapiro[11] constructed a strain in which *araC* (the positive regulatory element controlling the arabinose operon) is placed upstream of *lacZ* but separated from it by a short segment of Mu bacteriophage DNA that contains transcription terminating signals. Both the arabinose and the lactose operons are missing in this strain, so it is Lac$^-$ and Ara$^-$; if, however, it can delete the intervening Mu segment it will be able to grow on lactose, provided arabinose is present—a phenotype we will refer to as Lac(Ara)$^+$. Shapiro has shown that the deletion occurs in populations of cells plated out on lactose-arabinose-minimal plates. The Lac(Ara)$^+$ mutants are not, however, the result of spontaneous deletions occurring during the prior growth of the cultures but are apparently arising on the plates, because all the colonies appear much later than the colonies that are formed when cultures are seeded with a few pre-existing Lac(Ara)$^+$ cells and then grown up and plated. Indeed, the frequency of spontaneous deletion during growth is probably less than 10^{-11}, because immediate-colony-formers have never been found in cultures grown in the absence of selection.

Because this appears to be potentially a case of directed mutation, we have investigated it a little further. The delay in the formation of Lac(Ara)$^+$ colonies on selective plates is not simply some problem, in stationary phase, with the expression of *araC* and *lacZ* following their fusion, for we have prepared fused strains that are Lac(Ara)$^-$ because they have point mutations in *lacZ* or *araC*, and these strains show no such delay in the appearance of their Lac(Ara)$^+$ revertant colonies. Thus the act of gene fusion is indeed a late event, occurring long after the cells are put on lactose-arabinose-minimal plates. The event is influenced by some special property of the segment of Mu DNA that is being deleted, because the deletion does not occur when there is Mu DNA elsewhere in the chromosome[11]. Even so, we would like to know if the deletion is actually dependent on the presence of lactose and arabinose.

We have therefore studied the apperance of Lac(Ara)$^+$ cells (cells that are immediate-colony-formers on lactose-arabinose-minimal plates) in liquid cultures that are kept gently aerated after reaching stationary phase. The results can be summarized as follows. (1) In rich media containing both lactose and arabinose, large numbers of Lac(Ara)$^+$ cells start to appear after 3-4 days at 30 °C. (2) Under these conditions, even minority populations (marked with rifamycin-resistance) contribute to the final population of Lac(Ara)$^+$ cells, and from this we have calculated that the proportion of bacteria that are becoming Lac(Ara)$^+$ must be at least 10^{-8}. (3) In rich media without lactose (or without arabinose), Lac(Ara)$^+$ cells do not accumulate at a detectable rate. (4) When lactose is added to a stationary culture that has arabinose but not lactose, Lac(Ara)$^+$ cells promptly start to accumulate; judging from their rate of increase, the first Lac(Ara)$^+$ cell must be formed within 1-2 h of the addition of lactose.

This therefore seems to be another example of the production of appropriate mutations in response to selection. In one sense, it is a rather clear case, because this particular mutation does not occur at a measurable rate in the absence of selection; so bacteria, in stationary phase, do apparently have access to some process that either can prevent useless mutations from occurring or can destroy unsuccessful mutants soon after they arise. At the same time, it is perhaps a rather special case, because it involves the movement (deletion) of a piece of Mu DNA and does not occur if there is Mu elsewhere in the chromosome.

So far we have only considered examples of spontaneous mutation where bacteria are under pressure to reverse defects that have been previously introduced into their genes by mutagenesis or by genetic engineering. For our final example, we will briefly consider the response of *E. coli* to more natural forms of selection pressure.

Activation of the cryptic genes in *E. coli*

The usual way of distinguishing the various pathogenic and nonpathogenic *Enterobacteriaceae* is to observe which sugars they can use as a source of energy; for example, *E. coli* ferments lactose, whereas the various members of the *Shigella* and *Salmonella* families do not. The standard test is to inoculate a small volume of peptone plus sugar, and observe turbidity, *p*H and

CO_2 production after 1-2 days. When this system of classification was being prepared, early this century, certain bacterial species were observed to be 'late' fermenters of some sugars. Thus it is a regular feature of some bacteria that it takes a week or more before the population of cells, produced by growth on the peptone, are able to generate a variant that can hydrolyse the sugar; for example, *Sh. sonnei* is classified as a 'late' fermenter of lactose, and many strains of *E. coli* are 'late' fermenters of salicin (an aromatic β-glucoside). The genes coding for the appropriate enzymes are there, but they are not readily accessible.

Bacteria apparently have an extensive armoury of such 'cryptic' genes that can be called upon for the metabolism of unusual substrates. The mechanism of activation varies. In some cases it is simply by the movement of an insertion sequence into a position upstream of the cryptic gene[12], but in others it may require several changes in base sequence. It is these latter examples that are of special interest. For instance, *E. coli* turns out to have a cryptic gene (*ebgA*^{0}) that it can call upon to hydrolyse lactose, if the usual gene for this purpose (*lacZ*) has been deleted[13]. The activation of *ebg* requires at least two mutations, one in the repressor (*ebgR*) and one at a particular site in the gene coding for the enzyme (*ebgA*) to make the enzyme capable of hydrolysing lactose[14]. During growth, each of these point mutations occurs at a frequency of less than 10^{-8}; neither on its own will allow a *lacZ* deletion strain to use lactose[15], yet colonies of such a strain, that have grown to stationary phase on a plate containing lactose, will produce Lac$^+$ papillae in about two weeks. That such events ever occur seems almost unbelievable, but we have also to realize that what we are seeing probably gives us only a minimum estimate of the efficiency of the process, since in these cases the stimulus for change must fairly quickly disappear once a few mutant clones have been formed that can exploit the novel sugar. It is difficult to imagine how bacteria are able to solve complex problems like these—and do so without, at the same time, accumulting a large number of neutral and deleterious mutations—unless they have access to some reversible process of trial and error.

Discussion

The main purpose of this paper is to show how insecure is our belief in the spontaneity (randomness) of most mutations. It seems to be a doctrine that has never been properly put to the test. We describe here a few experiments and some circumstantial evidence suggesting that bacteria can choose which mutations they should produce. But we realize that this is too important an issue to be settled by three or four rather ambiguous experiments.

The origin of genetic variation has been the subject of bitter controversy, throughout the nineteenth century and into the first half of the twentieth. Is all variation essentially random, like thermal noise? Or can the genome of an individual cell profit by experience? At its extremes, it was an argument between reductionists and romantics—between those who sought to explain the evolution and behaviour of the biosphere in terms of the laws of physics, and those who wished to make the success of evolution just another manifestation of the mysteriousness of living things.

The early triumphs of molecular biology strongly supported the reductionists. The DNA double helix is an extraordinarily stable structure, but it is subject to spontaneous degradation and to errors in replication, and these changes are, in a sense, due to thermal noise. Furthermore, the discovery of all the elements that lie between DNA sequence and protein structure gave rise to the central dogma of molecular biology, and this doctrine denies any possible effect of a cell's experience upon the sequence of bases in its DNA.

Curiously, when we come to consider what mechanism might be the basis for the forms of mutation described in this paper we find that molecular biology has, in the interim, deserted the reductionist[16]. Now, almost anything seems possible. In certain systems, information freely flows back from RNA into DNA; genomic instability can be switched on under conditions of stress, and switched off when the stress is over; and instances exist where cells are able to generate extreme variability in localized regions of their genome. The only major category of informational transfer that has not been described is between proteins and the messenger RNA (mRNA) molecules that made them. If a cell discovered how to make that connection, it might be able to exercise some choice over which mutations to accept and which to reject.

Since this is the kind of versatility and adaptability we seem to be seeing in these experiments with *E. coli*, it is worth considering briefly how such a connection might be made. In a very direct way, the cell could produce a highly variable set of mRNA molecules and then reverse-transcribe the one that made the best protein; for this, it would have to have a special organelle (perhaps like the gag-pol complex of retroviruses) which contains reverse transcriptase plus some element that somehow monitors the protein product and determines whether the mRNA should go on being translated or should be transcribed into DNA; this could be an efficient process and makes an attractive hypothesis because it gives us an unusual explanation for the origin of retroviruses. Less efficient would be the production of reverse transcripts at random, but even that could achieve the desired end result, if reverse transcription were switched on temporarily, just at the time the cell resumed growth; any cell that started growing would acquire a reverse transcript of the variant sequence that made it able to grow, and so the sequence would eventually be incorporated by recombination into the genome of one of the cell's descendents (indeed, the same result could perhaps be achieved starting with DNA copies of segments of the genome rather than RNA transcripts). Each of these processes would allow individual cells to subject a subset of their informational macromolecules to the forces of natural selection. Each could, in effect, provide a mechanism for the inheritance of acquired characteristics.

We thank Victoria Cairns and Rainer Riester for writing a program for calculating the composite distributions of mutants, Patricia Foster and Richard Peto for comments on the manuscript, and Eric Eisenstadt for the hospitality of his laboratory. The work was supported by the National Science Foundation.

Received 2 May; accepted 11 July 1988.

1. Luria, S. E. & Delbrück, M. *Genetics* **28**, 491-511 (1943).
2. Lederberg, J. & Lederberg, E. M. *J. Bact.* **63**, 399-406 (1952).
3. Cavalli-Sforza, L. L. & Lederberg, J. *Genetics* **41**, 367-381 (1956).
4. Hayes, W. *The Genetics of Bacteria* (Blackwell, Oxford, 1964).
5. Lea, D. E. & Coulson, C. A. *J. Genet.* **49**, 264-285 (1949).
6. Armitage, P. *J. R. statist. Soc. B.* **14**, 1-40 (1952).
7. Mandelbrot, B. *J. appl. Prob.* **11**, 437-444 (1974).
8. Luria, S. E. *Cold Spring Harb. Symp. quant. Biol.* **16**, 463-470 (1951).
9. Hadley, P. *J. infect. Dis.* **40**, 1-312 (1927).
10. Ryan, F. J. *Nature* **169**, 882-883 (1952).
11. Shapiro, J. A. *Molec. Gen. Genet.* **194**, 79-90 (1984).
12. Reynolds, A. E., Felton, J. & Wright, A. *Nature* **293**, 625-629 (1981).
13. Campbell, J. H., Lengyel, J. A. & Langridge, J. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1841-1845 (1973).
14. Hall, B. G. *J. Bact.* **129**, 540-543 (1977).
15. Hall, B. G. & Zuzel, T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3529-3533 (1980).
16. Reanney, D. C. *Nature* **307**, 318-319 (1984).