Q5

1949
rev 1953

The Estimation of Mutation Rates

Genetics 107
Univ. Wis.

Adapted from Luria and Delbruck, Genetics 28:491; 1943, Armitage, Jour. Hygiene 51:162, 1953, and Lea and Coulson, Jour. Genetics 49:264, 1949.

1. Assumptions
   a. Growth is synchronous and uniform for mutant and non mutant cells.
   b. Mutations occur at cell division, yielding one mutant, one non mutant offspring.
   c. All mutant cells are effectively counted precisely at time of assay.

2. Definitions

N Cell number per culture.

t Generation number. Defined by the law of geometric growth $N = 2^t$ or $t = \log_2 N$.

r Number of mutant cells.

m Number of mutational events. d average mutant clone size = r/m.

a Probability that a division will yield one mutant offspring.

At each generation, N increases to 2N by means of N divisions. By summing the total divisions, it is evident that $N-N_o$ divisions are involved in the increase from an inoculum N_o to a final value N. As N_o is usually negligible in comparison to N, we may usually write that N divisions have occurred in the growth of a culture of size N. Therefore:

(1) $m = aN$

Null fraction method

The probability that a mutation will not occur at a given division is 1-a. Therefore the probability that no mutations will have occurred in a culture is

(2) $p_o = (1-a)^N$ which can be shown to be closely approximated by

(2a) $p_o = e^{-aN}$ or (2b) $\log_e p_o = -aN$, $a = 1/N \log_e 1/p_o = \frac{2.3}{N} \log_{10} \frac{1}{p_o}$

where p_o is that fraction of a series of cultures which contains no mutants.

Average number of mutants

(Warning; Carefully distinguish r from m !!) It will be shown that each generation contributes an equal number of mutants r to the final crop. At the i'th generation, N_i cells are produced from $N_i/2$ by means of $N_i/2 = 2^i/2$ divisions. On the average, there will then be $a.2^i/2$ mutations at this generation. Each of the cells of the i'th generation will increase by a factor $N/N_i = 2^t/2^i$ by the time of assay. The total crop of mutants from the i'th generation will therefore be $a.2^i/2.2^t/2^i$ or $a.2^t/2$. Summing over all n generations we find the total mutant crop

(3) $r = at2^t/2 = atN/2 = a\ N \log_2 N/2$

$a = 2r/N \log_2 N = .602\ r/N \log_{10} N.$

(4) and $d = r/m = t/2.$

Luria and Delbruck's likely average correction

L and D point out that (3) displays the average value of r, including the contribution of a great many mutants from rare, very early mutations. They set up another expression which they consider will give the "likely average" in any given experiment of C cultures, each size N.

It is assumed that no mutations are likely to have occurred prior to t = i. i is selected arbitrarily, as a function of mutation rate, so that in the entire experiment with C cultures there will have been one premature mutation, that is, so that $a.C.N_i = 1$, or $2^i = 1/aC$ and $i = -\log_2 aC$. Thus it is assumed that the mutations in each culture have occurred throughout t-i, rather than all t, generations. (3) then becomes

(5) $r' = 1/2.(t-i)\ aN = (aN/2)(t-i) = (aN/2)\ (\log_2 N + \log_2 aC) = (aN/2) \log_2 aCN.$

This cannot be solved explicitly for a, but may be handled numerically or with the help of a chart provided by L and D. This treatment has been criticized by Armitage; it has also been abused by workers who have pooled estimates of a from different experiments, rather than summing the pooled data. At best this technique does not mitigate the very high variance of r, which makes feasible estimates of its mean very difficult.

Use of the median and upper quartile

The (limited) solution of the theoretical distribution of r by Lea and Coulson allowed the development of two other measures, the median r_2 and the upper quartile r_3 (i.e. the values standing at the positions $(n+1)/2$ and $3(n+1)/4$ in a series of n observations ranked by size).

They have shown that $r_2/m - \log_e m = 1.24$ and $r_3/m - \log_e m = 4.09$, respectively. Tables to assist the calculation of m (=aN, and hence of a) are given by these authors. These methods do not make full use of all the numerical data, but provide more stable estimates of r than methods based on the experimental mean.