

## **S1 Extended methods. Derivation of a formula to calculate repeat loss rates using the mutation accumulation model**

Traditional marker-loss assays employed colony sectoring analyses to estimate the fraction of cells that had lost the selectable marker, followed by classical fluctuation test-based models (Luria–Delbrück method [1], extended by others [2-4]) to estimate marker loss rates. However, this method requires multiple parallel cultures to estimate marker loss rates in a single sample. Additionally, the relatively small population sizes (a few cells to a few hundred cells at most) increases the likelihood of “jackpot events” [1], which are random mutations that occur early during the course of establishment of the population, and predominate the population. Therefore, we chose to use a mutation accumulation model which allows the estimation of rDNA repeat loss rates for any sample from a single culture using 2 measurements, i) fraction of GFP-positive cells, and ii) OD<sub>600</sub>, from 2 time points, t=0 and t=24h, the start and end of the experiment respectively. The use of this model requires starting with, and analyzing population sizes large enough so that the probability of occurrence of the mutation of interest (here, the loss of the *MATα* containing rDNA repeat, which generates a GFP-positive cell) at every generation is nearly 1. The use of flow cytometry to measure the fraction of GFP-positive cells allows us to measure the large numbers of cells required for this type of analysis. Our starting population and the populations of cells analyzed typically contain at least 200,000 cells, which is not only much larger than the population size used in fluctuation tests, but also minimizes the likelihood of jackpot events. Briefly, consider the reporter strain in which the *MATα*-containing rDNA repeat is lost at a rate of  $m$  per cell division, where  $m$  represents the fraction of cells, on average, that lose the *MATα*-containing rDNA repeat and become fluorescent, or GFP-positive, at every division. If the number of GFP-positive cells (cells that have lost the *MATα*-containing

rDNA repeat) and GFP-negative cells (cells that contain a  $MAT\alpha$ -containing rDNA repeat) in the population at any given generation,  $n$  are  $N_n^+$  and  $N_n^-$  respectively, then the total number of cells at any generation  $n$  will be

$$N_n = N_n^+ + N_n^- \quad (1)$$

At every cell division, the number of GFP-negative cells doubles, but also decreases by a fraction of  $m$  as some cells lose  $MAT\alpha$  and become fluorescent. Therefore, at any generation  $n$ , the number of GFP-negative cells can be represented by the equation

$$N_n^- = 2^n(1 - m)^n N_0^- \quad (2)$$

where  $N_0^-$  is the number of GFP-negative cells at  $t=0$ , and  $m$  is the rate of loss of  $MAT\alpha$  per cell division. Further, the number of GFP-positive cells at any generation  $n$  can be obtained from the relation

$$N_n^+ = N_n - N_n^- \quad (3)$$

Therefore, the ratio of GFP-positive cells to GFP-negative cells at any generation  $n$  can be represented by  $R_n$ , and,

$$R_n = \frac{N_n^+}{N_n^-} \quad (4)$$

Substituting the expressions for  $N_n^+$  and  $N_n^-$  from equations (1-3),  $R_n$  may be re-written as

$$R_n = \frac{[N_n - N_n^-]}{[2^n(1-m)^n N_0^-]}$$

$$R_n = \frac{[2^n N_0 - 2^n(1-m)^n N_0^-]}{[2^n(1-m)^n N_0^-]}$$

$$R_n = \frac{[N_0^- + N_0^+ - (1-m)^n N_0^-]}{[(1-m)^n N_0^-]}$$

$$R_n = \frac{[1 + (N_0^+/N_0^-)]}{[(1-m)^n]} - 1$$

$$R_n = \frac{[1 + R_0]}{[(1-m)^n]} - 1$$

where  $R_0$  is the ratio of GFP-positive cells to GFP-negative cells at t=0. Further,

$$(1-m)^n = \frac{[1 + R_0]}{[1 + R_n]}$$

$$m = [(1 + R_0)/(1 + R_n)]^{1/n}$$

For cases where repeat loss rates are very small,  $m \ll 1$ , this equation can be further simplified to

$$1 - mn = (1 + R_0)/(1 + R_n)$$

$$m = 1 - \frac{[1 - \{(1 + R_0)/(1 + R_n)\}]}{n}$$

$$m = \frac{(R_n - R_0)}{n(1 + R_n)} \quad (5)$$

The values for  $R_0$  and  $R_n$  can be obtained simply from the measurements of fractions or numbers of GFP-positive and GFP-negative cells at t=0 and t=24h respectively. The number of generations,  $n$ , can be obtained from the change in the total number of cells over  $n$  generations as

$$n = \log_2 \frac{N_n}{N_0}$$

where  $N_0$  is the total number of cells at t=0. Since  $OD_{600}$  at any generation  $n$ ,  $OD_{600-n}$ , is directly proportionate to the total number of cells at that generation, this relationship can be re-written as

$$n = \log_2 \frac{OD_{600-n}}{OD_{600-0}} \quad (6)$$

where  $OD_{600-0}$  and  $OD_{600-n}$  are the  $OD_{600}$  measurements at t=0 and t=24h respectively. Therefore, the rate of loss of the  $MAT\alpha$ -containing rDNA repeat unit per cell division can be obtained from equations (5) and (6) with measurements of  $OD_{600}$  values and numbers of GFP-positive and GFP-negative cells at the start (t=0) and end (t=24h) of the experiment.

These calculations make the following key assumptions:

1. The rate of loss of the  $MAT\alpha$ -containing rDNA repeat,  $m$ , is constant.
2. GFP-positive cells are generated predominantly by a complete loss of the  $MAT\alpha$ -containing rDNA repeat unit.
3. GFP-positive cells generated by loss of the  $MAT\alpha$ -containing rDNA repeat unit have similar growth rate and doubling time as the GFP-negative cells that retain  $MAT\alpha$ .

4. Loss of GFP-positive or GFP-negative cells from the population, by cell death, for example, is negligible over the course of the experiment.

## **References**

1. Luria SE, Delbruck M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*. 1943;28(6):491-511. Epub 1943/11/01. PubMed PMID: 17247100; PubMed Central PMCID: PMC1209226.
2. Lea DE, Coulson CA. The distribution of the numbers of mutants in bacterial populations. *J Genet*. 1949;49(3):264-85. Epub 1949/12/01. doi: 10.1007/bf02986080. PubMed PMID: 24536673.
3. Luria SE. The Frequency Distribution of Spontaneous Bacteriophage Mutants as Evidence for the Exponential Rate of Phage Reproduction. *Cold Spring Harbor Symposia on Quantitative Biology*. 1951;16(0):463-70. doi: 10.1101/SQB.1951.016.01.033
4. Capizzi RL, Jameson JW. A table for the estimation of the spontaneous mutations rate of cells in culture. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 1973;17(1):147-8. doi: 10.1016/0027-5107(73)90265-0.