**Title**
A note on the mating scheme used by the Mutagenesis Project



**Author**
Sen, Saunak

**Publication Date**
2001-06-01

# A Note on the Mating Scheme Used by the Mutagenesis Project

Śaunak Sen*

March 11, 2001


## Abstract

This is a short note on the mating scheme used by the Mutagenesis project. We present an probabilistic analysis of the distribution of the number of mutants in the G3 generation. It is show to be a function of the number of G2 mothers and litter sizes. A computer program is provided to make the calculation. We quantify the odds of a G2 mother being a mutation carrier given that none of its progeny are mutants. Finally we analyze some data from the project; we find the data to be consistent with theory.


## Outline of the mating scheme

A mouse from a specific background strain is exposed to mutagens. This induces mutations in the germ (sperm) cells of the mouse. The progenitor generation is called the G0 generation. The first generation (G1) will carry a single copy of any given mutation. Denote by `X` a specific mutated allele and by `O` the wild-type or normal allele of the background strain at the same locus. The G1 individuals are then `XO`. The G1 females are mated to the background strain which are all `OO` at the specific locus. So the G2 generation is either `XO` or `OO` with equal probability. The G2 progeny are now mated to their parent G1 individual to create the G3 generation. The question is how many G3 progeny have to be collected to get at least two or more mutants in the G3 generation? A mouse is a mutant if it is `XX` at the locus of interest.

## Theoretical analysis

**Finding two or more mutants in a G3 generation** As mentioned above, a G2 mouse is `XO` or `OO` with probability $\frac{1}{2}$. If a G2 mouse is `OO` then no mice in its litters can be `XX`. If a G2 mouse has genotype `XO`, then any individual mouse in its litter will be `OO`, `XO`, or `XX` with probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ respectively. Assuming independent assortment of gametes, in a litter of size $m$, the number of `XX` mice follows a binomial distribution with parameters $m$ and $\frac{1}{4}$. Thus, unconditionally, the distribution of the number of mice in a litter of size $m$ is an equal mixture of the binomial distribution and the degenerate distribution with mass at 0.

Let $X_1, X_2, \ldots, X_k$ be the number of `XX` individuals in litters of size $n_1, n_2, \ldots, n_k$ from $k$ *different* G2 parents. Let $X = \sum_{i=1}^{k} X_i$ be the number of mutants in the G3 generation where a total of $n = \sum_{i=1}^{k} n_i$ G3 potential mutants have been screened. We are interested in the probability distribution of $X$. Of special interest is the probability of getting two or more mutants in the G3 generation that carry the same recessive mutation, $P(X \geq 2)$.

Using numerical convolution methods implemented in Matlab (`www.mathworks.com`) a function for calculating the probability distribution of $X$ has been written. It is reproduced in the last section. It turns out that it is advantageous to screen progeny from as many *different* G2 individuals as possible.

Table 1 lists the probability of finding two or more mutants when 20 G3 progeny are screened from a varying number of G2 mothers. The function `g3prob` reproduced at the end of this paper was used to calculate the probabilities.

*The Jackson Laboratory, `ssen@jax.org`

Table 1: Table showing the probability of finding two or more mutants when 20 G3 progeny are screened from a varying number of G2 mothers. It is assumed that the litter size from each of the mothers is equal. This is not realistic, but is done for approximation. We see that the probability of detecting more than two mutants increases with the number of different G2 mothers used to produce the progeny.

| Number of G2 mothers | Probability of two or more mutants |
|---|---|
| 1 | 0.4878 |
| 2 | 0.6219 |
| 4 | 0.6662 |
| 5 | 0.6785 |
| 10 | 0.7120 |
| 20 | 0.7331 |

**Making bets about the mother mouse** As mentioned above, *a priori* a G2 mouse has 1:1 odds of being a mutant. If its litter shows even one mutant, then we *know* that it is a carrier and that the probability that any one of its children is a mutant is 0.25. We are not so sure if a litter produces no mutants. There is still a possibility that the mouse is a carrier. But larger the litter size with no mutants, smaller the probability that the mother G2 mouse is a carrier. These beliefs can be quantified in terms of odds ratios. In fact, the *a posteriori* odds that the G2 mother is a carrier given that none of its $m$ children are mutants is $\left(\frac{3}{4}\right)^n$:1. As we can see from the formula, the odds get slimmer, larger the litter size. Table 2 shows the posterior odds and posterior probability as a function of litter size. The table can be used as a guide whether or not to wait for a second litter from a mouse.

The above probabilistic analysis suggests that a sequential breeding scheme may be used. After testing the progeny from a mouse, if we find that none are mutants, then chances are that the mother mouse is not a carrier. It may be better to breed a different mouse. On the other hand if even one of the mice in the litter is mutant, we know that the mother mouse is a carrier and its progeny are mutants with probability 0.25.

## A LOOK AT SOME DATA

Kevin Seburn supplied me with some preliminary data from the Mutagenesis Project which are reproduced in Table 3.

The first thing we checked for is the proportion of G2 mothers who showed mutants in their litters. As suggested in the discussion above, this proportion should be 0.5. In the data below, the number of G2 mothers who did not have any phenotypically aberrant progeny is 17 out of a total of 29. This works out to a proportion of 0.59. The two-sided p-value for testing that that this proportion is not 0.5 is 0.46. This suggests that the proportion of G2 mothers who are carriers is consistent with theory.

The second test we did was to find out the proportion of mutants in litters that showed at least one mutant. If a litter has a mutant, then the mother is definitely a carrier. If that is the case then the probability of an individual being a mutant is 0.25. In our data, the number of mutants in litters with at least one mutant is 33. The total number of progeny in litters with at least one mutant is 141. Thus the proportion of mutants is 0.2340. The two-sided p-value for testing that the proportion is 0.25 is 0.66. Agreement with

Table 2: Table showing the probability that a G2 mother is a carrier given that *none* of her children are mutants. This probability decreases with litter size. Note that even if one child in her litter is a mutant, she is a definitely a carrier. These calculations assume that the determination of mutants is done perfectly, without error.

| Litter size | Posterior odds | Posterior probability |
|---|---|---|
| 0 | 1.0000 | 0.5000 |
| 1 | 0.7500 | 0.4286 |
| 2 | 0.5625 | 0.3600 |
| 3 | 0.4219 | 0.2967 |
| 4 | 0.3164 | 0.2404 |
| 5 | 0.2373 | 0.1918 |
| 6 | 0.1780 | 0.1511 |
| 7 | 0.1335 | 0.1178 |
| 8 | 0.1001 | 0.0910 |
| 9 | 0.0751 | 0.0698 |
| 10 | 0.0563 | 0.0533 |

Table 3: Table showing the types of mutants screened and found in different litters. The rows correspond to different G1 parents.

| Mutant ID | Phenotype | Total Deviant Ratio | Deviant | | | |
|---|---|---|---|---|---|---|
| NMF-1 | Unilateral fore and hind limb paralysis, shakes | 1:17 | 1:4 | 0:4 | 0:9 | |
| NMF-2 | Bilateral hindlimb paralysis | 5:29 | 5:13 | 0:16 | | |
| NMF-3 | Shaking, head-tilt, circling | 7:31 | 3:12 | 1:5 | 3:10 | 0:4 |
| NMF-4 | Head tilt, shaking | 7:30 | 1:13 | 3:4 | 3:13 | |
| NMF-5 | Bilateral hindlimb paralysis | 6:36 | 1:11 | 1:11 | 4:10 | 0:4 |
| NMF-6 | Bilateral hindlimb paralysis | 1:24 | 1:4 | 0:12 | 0:8 | |
| NMF-7 | Seizures | 4:20 | 2:9 | 1:3 | 1:5 | 0:3 |
| NMF-8 | Seizure resistant | 1:20 | 1:2 | 0:5 | 0:10 | 0:3 |
| NMF-9 | Head nodding | 1:20 | 1:12 | 0: 8 | | |

theory is comforting. We have not tested for other kinds of distortions that can occur in the collection of the data, for example, the phenotyping could be in error.

## DISCUSSION

The statistical analysis of the data above shows that assuming that the mutants being detected from the mutagenesis mating scheme follow a pattern consistent with theory. A practical implication of the probabilistic analysis is that if a G2 mouse has a moderately large litter (say 8) and has produced no mutants, then chances are that it is not a carrier. This belief can be quantified in terms of Bayesian odds and can be used to decide whether or not to wait for further litters from the mother mouse.

## COMPUTER PROGRAM

The data analysis was performed in Matlab. Reproduced below is the program that was used to calculate the probabilities in Table 1. The function `g3prob` calculates the probability distribution of the number of mutant progeny we may get given a vector of litter sizes (corresponding to different G2 mothers).

```
function p = g3prob( n )
% G3PROB Function to calculate the probabilities of getting recessive
% mutants in the G3 generation of the mutagenesis scheme
%
% P=G3PROB(N)
%
%   P = vector of probabilities for the probability distribution of the
%       number of mutants
%   N = vector of the litter sizes
%
% Example:
% To find out the probability distribution of the number of mutants in
% five litters (of sizes 6 4 9 5 12 respectively ) from five separate G2
% parents use the command
% p = g3prob( [ 6 4 9 5 12 ] )
%
  ng2 = length( n ); % number of different G2 parents
  nprogeny = sum( n ); % total number of G3 progeny to be tested

  % calculate the probability distribution of the number of G2 parents
  % who will be mutation carriers
  pcarriers = binopdf( 0:ng2, ng2, 0.5 );

  % for each of the litters we have to calculate the probability
  % distribution of mutant progeny

  plitter = cell( ng2 );
  for( i=1:ng2 )
    plitter(i) = { [ binopdf( 0:n(i), n(i), 1/4 ); [ 1 zeros( 1, n(i) ) ] ] };
  end

  p = mean( plitter{1} );
  for( i=2:ng2 )
    p = conv( p, mean( plitter{i} ) );
  end
```