LA-4573-MS

c. 3

# LOS ALAMOS SCIENTIFIC LABORATORY
## of the
## University of California

LOS ALAMOS • NEW MEXICO

LOS ALAMOS NATIONAL LABORATORY
3 9338 00387 4400

# Some Elementary Attempts at Numerical Modeling of Problems Concerning Rates of Evolutionary Processes

UNITED STATES
ATOMIC ENERGY COMMISSION
CONTRACT W-7405-ENG. 36

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

This report expresses the opinions of the author or authors and does not necessarily reflect the opinions or views of the Los Alamos Scientific Laboratory.

Printed in the United States of America. Available from
National Technical Information Service
U. S. Department of Commerce
Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

Written: December 1970

Distributed: January 1971

LA-4573-MS
UC-48, BIOLOGY
AND MEDICINE
TID-4500

LOS ALAMOS SCIENTIFIC LABORATORY
of the
University of California
LOS ALAMOS • NEW MEXICO

LOS ALAMOS NATL. LAB. LIBS.
3 9338 00387 4400

# Some Elementary Attempts at Numerical Modeling of Problems Concerning Rates of Evolutionary Processes

by

R. Schrandt and S. Ulam*

*Consultant: Present address: Chairman, Mathematics Department, University of Colorado, Boulder, Colorado 80304

# SOME ELEMENTARY ATTEMPTS AT NUMERICAL MODELING OF PROBLEMS CONCERNING RATES OF EVOLUTIONARY PROCESSES

by

R. Schrandt and S. Ulam

ABSTRACT

An account of numerical work prepared on electronic computers -- the problems concerned the ratio at which favorable mutations spread throughout the population subject to a "survival of the fittest" mechanism. Models of asexual reproduction showed the expected linear growth in the number of improvements. The bisexual process accelerated greatly the average acquisition rate.

## I. INTRODUCTION

In this report, we shall present an abbreviated account of calculations performed by us several years ago. These calculations were preliminary and intended merely as the zeroth approximation to the problem concerning the rates of evolution -- a process which, we here, severely stylized and enormously over simplified. A mention of the results of such calculations in progress at that time was made at a meeting in 1966 at the Wistar Institute in Philadelphia by one of us. The discussion there, as reported in the proceedings of the meeting was rather frequently misunderstood and impression might have been left that the results somehow make it extremely improbable that the standard version of the survival of the fittest mechanism leads to much too slow a progress. What was really intended was indications from our computations -- simple minded as they were -- that a process involving only mitosis, in absence of sexual reproduction or some transfer process, would be indeed much too slow. However, as most biologists realize it anyway, the Darwinian mechanism together with mixing of genes accelerate enormously the rate of acquiring new "favorable" characteristics and leave the possibilities of sufficiency of the orthodox ideas quite open. Numerous requests addressed to us for the elucidations and details of the numerical set up made us decide to give this account of our computations. A more extensive set of numerical experiments with a broader collection of possibilities is planned in the future.

Perhaps the greatest uncertainties -- the strongest objections to any calculation of the sort described in the pages that follow -- must concern the values of the constants which are assumed to start with or should indeed concern even their meaning in the interpretation we have chosen. We have tried to interpret the survivability of individuals by changes in the number of offsprings which carry the species in time measured by a discrete succession of generations. The value of "favorable" mutations was mirrored in the increased proportion of offsprings. The same, needless to say, is true of the frequency of favorable mutations. We have disregarded the lethal and the unfavorable mutations. We assumed a special form of advantages which an individual holds relative to the rest of the

population of comparing the number of his "improvements" with the average number present at that time in the population. We assumed a proportionality law, again arbitrarily. In some problems we have penalized the individuals whose score in the improvements was less than the average; in some of the problems we considered only the positive excess as leading to a greater number of offsprings. Another debatable procedure is the way we have handled the growth of the population by normalizing periodically the total number of individuals to a constant figure. If the number of individuals holding a certain number of "advantages" after normalization dropped below 1 we summarily dismissed such representatives. It should be stressed here very strongly that this procedure makes it very hard to find an analytic model equivalent to our numerical work. We do not have any clear idea of the necessary scaling laws concerning the effect of changing the constants alpha, kappa and the numbers of population. All this is true throughout all the problems. In the calculations involving combinations of genes from both parents, further assumptions were made of independent inheritance of the "improved" genes, etc. etc. As will be seen in the description of the individual problems, we have chosen successively less unrealistic assumptions. Clearly in the counting of the new improved genes coming from "father", and from the "mother" one has to take care not to count the same "improvements" from each twice. As will be seen, this precaution has the effect of slowing down the at first seemingly exponential growth into something more like quadratic function of time (we have studied throughout the calculations the number of "improvements" present in the population as a function of time, that is to say the generation number).

In order to get a feeling for the dependence of the results on the values of the constants, more such computations must be tried in the future and additional variables have to be considered -- certainly the "kind" of the improved or favorable new gene has to be tabulated. A most important question concerns the existence of new genetic instructions involving perhaps logical prescriptions, that is to say recipes for operations and actions of the components rather than merely their chemical composition. An improvement in programming or interpretation of action by a gene or group of genes may be equivalent to a very large number of "favorable changes" with which our computations have dealt so far.

Our first problem (with the code name Adam) concerned asexual reproduction, and we feel that the time scale to acquire a characteristic in an orgaism, such as the development of an eye, by a sequence of consecutive favorable mutations is extremely long if one does not resort to something like a sexual mating in the population. (In the following rough and elementary estimate), the constants assumed are very crude, but err toward "faster" evolution than what is to be expected.

Definitions: Let

$T$ = time of existing life ($\sim 10^9$ yrs.).

$\tau$ = time for one generation say ($\sim$ 3 days).

$G$ = the number of generations $= T/\tau = 10^{11}$.

$N$ = the existing population ($\sim 10^{11}$).

$K$ = the total number of "favorable mutations" necessary to produce the desired characteristic ($\sim 10^6$).

$\alpha$ = the change of a favorable mutation per individual per generation ($\sim 10^{-10}$).

$\gamma$ = the "value" of a single favorable mutation expressed as a survival rate ($\sim 10^{-6}$). That is, an individual having this mutation would have $(k + \gamma)$ descendants, versus k descendants for an individual not having this mutation.

Therefore, in the first generation, the expected value of the population that would have one mutation is $N \cdot \alpha = 10$. In $1/\gamma = 10^6$ generations, a sizeable portion (approx. 1/e) of the population would have this mutation, and in $\sim 10^7$ generations, most of the population would have it. But the time to acquire all the mutations would be $\sim K \cdot 10^7$, or $10^{13}$ generations, which is like the age of the universe.

## II. ADAM

In order to study, on a computer, the rate at which a population can acquire a sequence of mutations we needed a set of more amenable parameters, which, it is hoped, could eventually be scaled down to the set given above. For the first problem, called ADAM, we used the following set:

$N = 100$, $K = 100$, $\alpha = .02$, and $\gamma = .1$.

The method was as follows: In any one generation, each member of the population N had a probability $\alpha$ of acquiring one new independent mutation. Each individual then had one descendant, with a probability of extra descendants given by $\gamma(K_g - K_o)$, where

$K_g$ = the total number of different mutations possessed by this individual, and

$K_o$ = the minimum number of mutations possessed by any individual in the population.

(If $n/\gamma < K_g - K_o \leq (n + 1)/\gamma$, $n = 1, 2, 3, \ldots$, the individual had n extra descendants, and a probability of $\gamma(K_g - K_o - n/\gamma)$ for an $(n+1)^{th}$ descendant.)

The descendants were then given the number of mutations possessed by their parent. The parent population then was assumed to have died, and the descendants formed the new population. The number of mutations in the population were recorded in categories by count $n_i$, $i = 0, 1, 2, 3, \ldots$, where $n_i$ = the number of individuals having a total of i mutations, and $\Sigma\, n_i = N'$, the total new population. In the next generation, each member of this new population had the same probability $\alpha$ of acquiring another new mutation, and had descendants according to the above recipe. These descendants with their number of mutations recorded then became the population for the succeeding generation, etc.

It was necessary to renorm the population periodically, since the number of descendants increased in each generation. This was done by reducing the count $n_i$ in each category by 1/2, to the nearest smaller integer, when the population reached 200, or double the initial one.

The weighted average number of mutations possessed by the population, $\mu$, was computed each generation from the categories $n_i$. This average was then plotted as ordinate against the generation time as abscissa. The slope of this curve is then the rate at which the population can acquire a sequence of mutations, as a function of the parameters $\alpha$ and $\gamma$.

This rate of acquiring mutations turned out indeed to be linear. (For the parameters $\alpha = .02$ and $\gamma = .1$, the slope was about .1; or a majority of the population acquired an additional mutation every 10 generations.) Several problems were run with smaller values of $\gamma$, i.e., $\gamma' = f \cdot \gamma$, f = a fraction. The graphs were all linear, with decreasing slopes, which decreased more with $\sqrt{f}$ than with f itself. There was no appreciable change in the slope by doubling the initial population to 200. Figure 1 shows a plot of 3 cases: $N = 100$, $\alpha = .01$, and $\gamma = .1$, .05 and .01 respectively.

A second version of ADAM was run with the $K_o$ in the probability recipe using the <u>average</u> number of mutations in the population, instead of the minimum number. Those individuals which had less than the average number of mutations, $(K_g < K_o)$, had a probability of $\gamma(K_o - K_g)$ of no children; (if $K_o - K_g \geq 1/\gamma$, they had no children deterministically), and a probability of $1 - \gamma(K_o - K_g)$ of 1 child. The individuals with more than the average number of mutations had the same probability for extra children as the one defined above. This version required fewer renormings of the population and it led to a somewhat greater slope than the $K_o$ = min. recipe. The graph was still linear. Figure 2 illustrates the two versions with the same parameters $N = 100$, $\alpha = .02$ and $\gamma = .1$. The $K_o$ = average recipe was used in all subsequent problems.

## III. EVE

In the second class of problems, we introduced reproduction in the population, and also, what seems important, fluctuations in the manner in which the offspring received mutations from their parents. This problem was naturally called EVE. The initial parameters used were the same as in ADAM. The population acquired new mutations according to the probability $\alpha$, as before. A random mating of the population N was then defined, resulting in N/2 pairs of individuals. (For N odd, the population was arbitrarily reduced by one to obtain $(N - 1)/2$ pairs.) Each pair then constituted a set of parents. The number of children from each set was again determined by the probability function $\alpha(K_g - K_o)$. Here

$K_g$ = the total number of mutations possessed by the set of parents, and

$K_o$ = the average total number of mutations of all the pairs of this mating.

The children were produced in pairs. Thus if a set of parents had $K_g > K_o$, they had two children

for certain, and with a probability given by $\gamma(K_g - K_o - n/\gamma)$, $n = 1, 2, 3, \ldots$, where $n/\gamma < K_g - K_o < (n+1)/\gamma)$, 2n extra children. If for the set of parents, $K_g < K_o$, they had, with the probability $\gamma(K_o - K_g)$ no children, and with probability, $1 - \gamma(K_o - K_g)$ two. Again, if $K_o - K_g \geq 1/\gamma$, they had no children.

The number of mutations acquired by each child was obtained under a binomial distribution centered about the average number of mutations of the parents. If the parents had a count of $(x + y)$ mutations, the number for each child was obtained under the distribution centered at $(x + y)/2$, with minimum at zero and maximum at $(x + y)$. The number of mutations for each child, separately, was determined under this distribution. Thus a child could possibly obtain as much as the sum of the numbers of mutations of its parents. This recipe involving fluctuations in the inheritance would, we thought, speed up the rate of acquisition of mutations (compared to just giving the offspring always $(x + y)/2$ favorable mutations).

The parents having died, the children became the new population. As before, they were classified in terms of counts of the number of individuals having a certain number of mutations. As before, the individuals had the chance $\alpha$ of new mutations, and were mated at random in N/2 pairs. These pairs then had children whose mutations were again determined from the binomial distribution, and these children constituted the population for the following generation, etc.

In our random mating, the sex of the individual was not distinguishable. No attempt was made to keep members of the same "family" from mating. Their number of mutations was not necessarily the same, since it was determined separately for each child under the probability distribution. (The identity of the family was lost once the members were classified according to their mutations.)

The norming of the increasing population was done in the same way as in ADAM: All categories were halved when the population doubled.

The rate of acquiring mutations turned out to be much faster than in ADAM, and appeared to be exponential. In Fig. 3 we have plotted this rate on a semi-log scale for four problems; $N = 100$, $\alpha = .02$, and $\gamma = .1$, .05, .025, and .01 respectively. The reduction in acquisition rate with $\gamma$ was somewhat similar to that in ADAM.

In these problems no attempt was made to keep the histories of the different mutations. We define $\alpha$ to be the rate of acquiring new mutations, but we divide the population only in terms of numbers of individuals having a fixed number of mutations. The children too acquire mutations only under the distribution of the total count of the parent's mutations. Thus one might suspect that the exponential rate of acquisition could be due to a doubling of the identical mutations possessed by both parents.

A. EVE PQ

To correct for this, we computed an expected number of mutations, $\nu$, that the parents should have in common. This is given by

$\nu = (T_1 \cdot T_2)/S$ where

$T_1$ = the total number of mutations possessed by one parent,

$T_2$ = the total number of mutations possessed by the other parent, and

S = the total number acquired by the entire population.

(The total S is accumulated as each individual in each generation has the probability $\alpha$ of acquiring a new mutation. If $\overline{N}$ is the average population in each generation, after k generations, S is approximately $\overline{N} \cdot k \cdot \alpha$.)

We allotted this number $\nu$ to the children for certain and then played a game of chance for additional mutations using the reduced binomial distribution centered at the midpoint of the total count of independent mutations possessed by the parents. In this manner, we count more correctly, the number in common only once.

This method leads to a slower rate of acquisition; it is still exponential in the beginning but tails off to something like a quadratic function, as more mutations must of necessity, be held in common. Results from four problems are plotted in Fig. 4 on a semi-log scale. The cases PQ1 and PQ2 have $N = 100$, $\alpha = .02$ and $\gamma = .1$ and .05 respectively. The case PQ3 is the same as PQ2 except that the sample size was doubled, $N = 200$. It had a somewhat higher value for $\mu$, but the slopes are like those of PQ2.

The last case, PQ4, had the same parameters as PQ3 except that $\alpha$ was cut to .01. This problem was run to 290 generations and showed a definite bending over of the curve to almost a linear rate of acquisition after 230 generations as the population had more and more mutations in common.

B. EVE-PM

In the two versions of EVE already discussed, the population mated at random uniformly. We next considered a version where the mating was non uniform, i.e., preferential in the following sense: We arbitrarily divided the population equally into three groups ranked according to their number of mutation, i.e., the first had the individuals with the greatest number of mutations. We specified that 3/4 of the population in each group mate at random uniformly within their own group. The remaining 1/4 would mate at random with equal probability from either of the two remaining groups. For example, if we name the groups A, B, and C, an individual from group A would have a 3/4 chance of acquiring a mate from group A, and a 1/8 chance of a mate from group B and group C. We called this problem EVE-PM. The EVE-PQ method was used to estimate the mutations in common and to count them only once.

The rate of acquisition of mutations should reflect this preferential mating. A comparison of the curve PM1 of Fig. 5 with PQ1 of Fig. 4, (both with the same parameters), shows indeed that the initial acquisition rate of mutations under preferential mating is much higher than under random mating. But the curve PM1 tails off very rapidly after 100 generations, and at 200 generations the mutation rate is almost the same for the two problems. This indicates that at this point most individuals have acquired most of the mutation available in the total population, so that preferential mating has the same effect as the uniform mating. Our relatively small sample size of 100 in part causes this phenonemon to occur so soon.

The computing time goes up quite rapidly with the initial population; one problem of preferential mating with an initial sample of 400 was run to 96 generations. The result is shown in the curve PM2 of Fig. 5, with the same parameters as PM1 (except for the population size). The preferential mating has a greater initial effect in the larger population, but the slope of this curve too is beginning to decrease.

The mutation acquisition rate in all the EVE problems of sexual reproduction can apparently be divided into three stages:

1.) An initial exponential rate, as few mutations are held in common (compared to the size of the population).
2.) A rate, probably quadratic, as more mutations are held in common by the parents. This number in common is approximated by computing the expected intersection of the number possessed by each parent, assuming the parents had acquired them independently.
3.) A terminal rate, only almost linear, as most of the mutations in the population are in common. If all the mutations were in common, the subsequent rate of acquisition must be linear, since new mutations are obtained only by the $\alpha$ rate of new acquisitions, which is a linear function.

C. EVE-POS

In order to check somewhat, these assumptions, a problem was run in which histories were kept of each new mutation. This was done by representing each new mutation as a bit position in a matrix of words in the computer. Each individual in the population and its children had their mutations recorded in such a matrix, that is to say each mutation was specifically identified.

We called this problem EVE-POS. The sexual reproduction scheme was the same as before. The population was mated at random, and each set of parents had a probability of having extra children given by $\gamma(K_g - K_o)$. In both $K_g$ and $K_o$ the actual mutations in common were counted only once. This number was known for each mating, so no approximation for the expected intersection was needed.

If both parents possessed the i-th mutation, it was then given to each of the children. If neither parent possessed this mutation, it was not acquired by any of their offspring. If only one parent possessed this i-th mutation, each child had a probability of 1/2 of acquiring it.

The norming for this problem was different than before. When the population doubled, each individual was given a 1/2 chance of surviving.

With this recipe for receiving mutations, any individual mutation can be lost to the population, since the parents die off in each generation. For example, the k-th mutation is initially acquired in the $\alpha$ recipe by one individual. If this individual and its mate have no children, the mutation is lost. If they have n children, there is a probability of $(1/2)^n$ that none of the children get it, in which case it is lost in the next generation. There is a chance that this particular mutation will be lost in each subsequent generation, although these probabilities are getting smaller. The k-th mutation is initially acquired only once, and by only one individual. The mutations that survived were "packed" in the bit positions of the matrices. This relieved the space limitation in the memory and allowed the problem to be run much further in time.

It was discovered that approximately 80% of the total of new mutations acquired by individuals were "lost" after matings in subsequent generations. Thus the parameter $\alpha$ in problem EVE-POS has a different meaning than in EVE-PQ. In EVE-POS it denotes the probability of an individual acquiring a new mutation. In EVE-PQ it denotes the probability of acquiring a mutation that survives and will eventually be acquired by the entire population.

Figure 6 shows the mutation rate for two cases of EVE-POS, plotted on a log scale. The parameters are $\alpha = .02$, $\gamma = .1$, and N = 100 and 400 respectively. The case POS1 (N = 100) was run to 251 generations, and POS2 (N = 400) to 90 generations.

Note that the acquisition rate and the values of $\mu$ are considerably smaller in POS1 than in PQ1, which has the same parameters. This is due to the different interpretation of the parameter $\alpha$. A reasonable comparison with POS1 would be to run PQ1 and compute our $\nu = (T_1 \cdot T_2)/S'$, where $S' = .2S'$, since about 80% of the mutations are lost.

The acquisition rate for POS1 becomes linear after about 100 generations. The relatively small sample size is a contributing factor. Some statistics were compiled on the distribution of the available mutations. They were given in Table I.

An estimate was made of the expected number of mutations held in common compared to the actual number held by an average set of parents. The expected number was computed assuming that each parent had the average number of mutations $\mu$. Then $\nu = \mu^2/S$, where S = the total number of surviving mutations. The actual number held in common was about 1.3 times this expected number, after 251 generations.

In the problem with the larger sample size of 400 (POS2), the mutation acquisition rate remains quadratic through 90 generations. At this point 18% of the mutations were held by at least 50% of the population, versus 36% for the population of 100. For this larger sample size, at 90 generations, separate distributions were kept of mutations acquired in the first 45 generations, versus those acquired in the last 45 generations. This data and some statistics on those mutations held in common are given in Table II.

The above problems give some indication of the rate at which a population can acquire mutations in terms of a finite and rather small sample size, and in terms of the procedures which we adopted for acquiring and transmitting of the mutations. These methods involved using rather large values of the parameters alpha and gamma. The problem remains to find scaling factors. The computing time on the old IBM-7094 for the problem POS2 was over one hour, for 90 generations of growth. The computing time for the PQ code was much faster, in the order of minutes, but the parameter $\alpha$ was in effect much larger than the one in the POS code.

## IV. SUMMARY

The problem ADAM with asexual reproduction gave a linear rate of acquisition of mutations. In reducing the parameter $\gamma$ to $f \cdot \gamma$, f a fraction, the acquisition rate was reduced more than $\sqrt{f}$ than like f itself.

In problem EVE with sexual reproduction, the acquisition rate appears to be exponential if the initial population were large enough. But with a small population, more of the <u>same</u> mutations were held in common by the parents. This caused the rate to change from an exponential to a "quadratic" and eventually to a linear one when most mutations were common to the majority of the population. The problem EVE-PQ involved approximating this number (our formula for $\nu$).

The advantage of preferential mating over random

mating gave an initial pronounced increase in the acquisition rate, but this was soon offset by the smallness of the population. In effect, as more mutations were held in common, the range of the distribution of mutations became narrow. After that the preferential mating was not much different from the uniform one.

The EVE-POS problem (where we kept a history of the mutations) gave us a measure of the distribution of mutations as a function of their age. It showed that most of the mutations initially acquired by one individual were lost in subsequent matings. This caused a re-definition of the probability $\alpha$ in computing the expected number of mutations held in common. This problem also showed that the actual number held in common was greater than the estimated number $\nu$, by a factor of about 1.3 for the sample size of 100, and 2.4 for the sample size of 400. This is not too surprising, since the expected size of the intersection assumes an independent sampling, whereas the mutations are acquired by something more like a Markov process.

TABLE I

Problem -- EVE POS1 --- N = 100, $\alpha$ = .02, $\gamma$ = .1.
After 90 generations, there were 55 surviving mutations out of 266 acquired mutations: 79% were lost.
Distribution of the 55 Mutations:
Min.-15 (Least number held by any individual).
Aver.-21 (Average number held--This is the number $\mu$ that is plotted).
Max.-27 (Greatest number held by an individual).
36% of the mutations were held by at least 50% of the population.
1.8% of the mutations were held by the entire population.
After 251 generations, there were 130 surviving mutations out of 744 acquired: 82% were lost.
Distribution of the 130 Mutations:
Min.--87
Aver.--94
Max.--101
74% of the mutations were held by at least 50% of the population.
41% of the mutations were held by the entire population.

(Table I cont.)
The actual number of mutations held in common was approximately 1.3 times the expected number.

Table II

Problem -- EVE POS2 -- N = 400, $\alpha$ = .02, $\gamma$ = .1.
After 90 generations, there were 239 surviving mutations out of 1067 acquired: 78% were lost.
Distribution of the 239 Total Mutations:
Min.--42
Aver.--55
Max.--71
Distribution of Mutations Initially Acquired in the First 45 Generations (115):
Min.--35
Aver.--49
Max.--63
Distribution of Mutations Acquired in the Last 45 Generations (124):
Min.--1
Aver.--6.4
Max.--13
There were 18% of the mutations held by 50% of the population, 27% held by 40% of the population, and 32% held by at least 30% of the population.
At 90 generations, the actual number of mutations held in common was approximately 2.4 times the expected number.

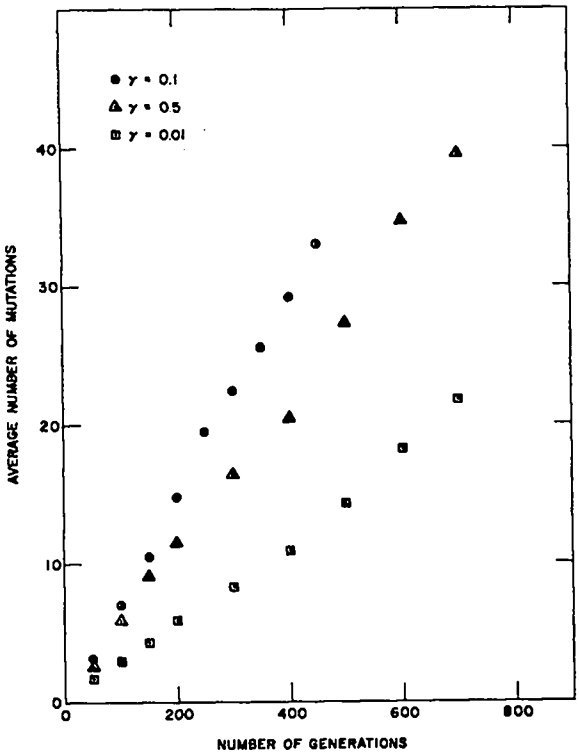


Fig. 1

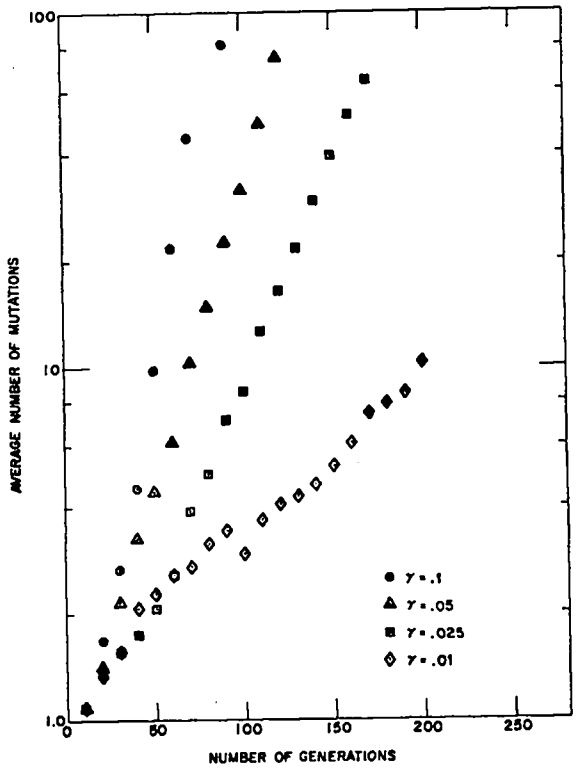


Fig. 3

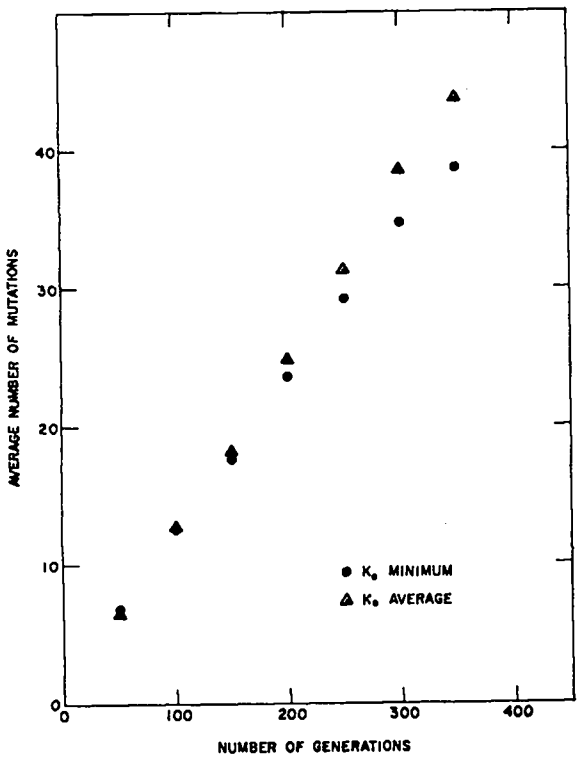


Fig. 2

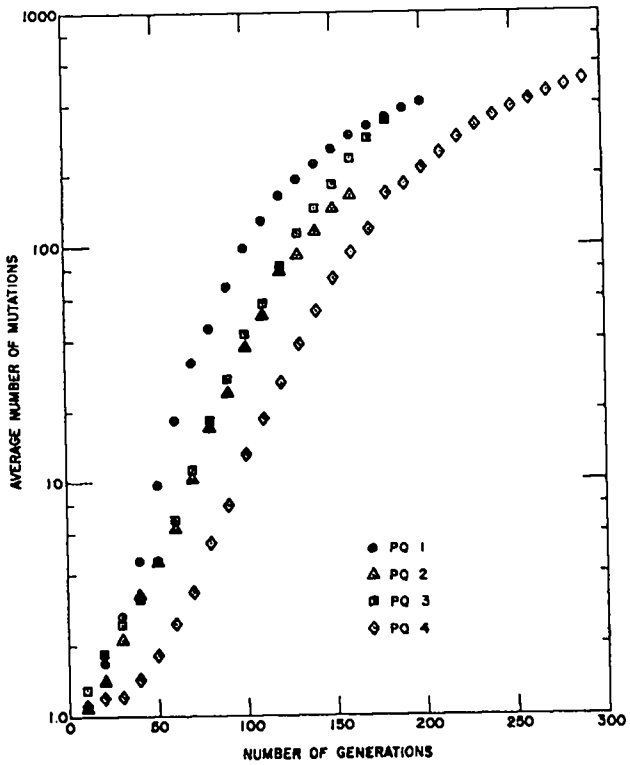


Fig. 4

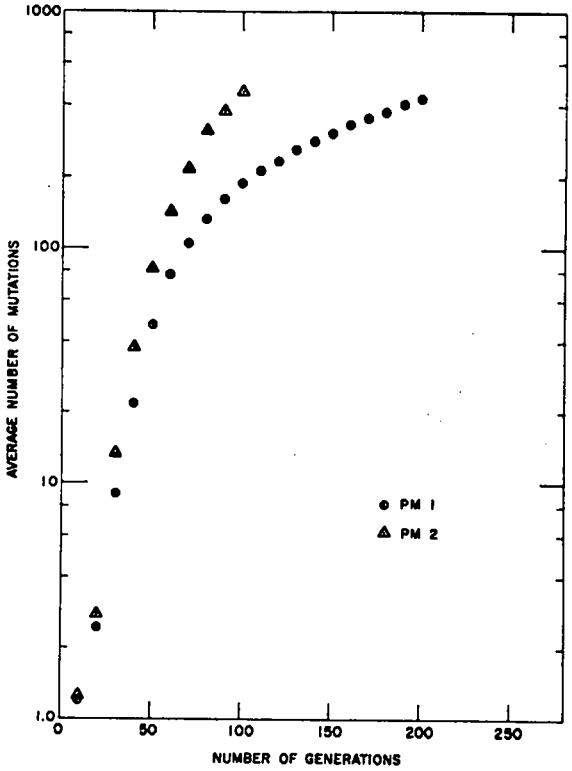
1000
100
10
1.0
0
50
100
150
200
250
AVERAGE NUMBER OF MUTATIONS
NUMBER OF GENERATIONS
PM 1
PM 2

Fig. 5

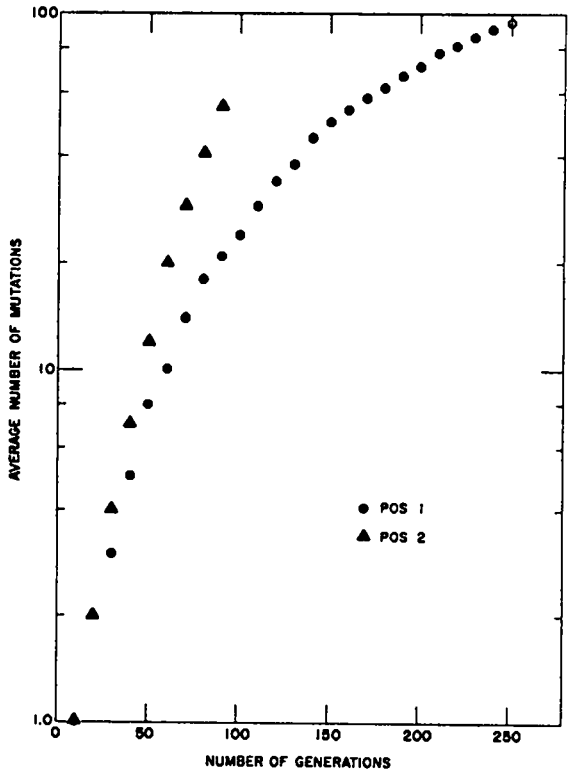
100
10
1.0
0
50
100
150
200
250
AVERAGE NUMBER OF MUTATIONS
NUMBER OF GENERATIONS
POS 1
POS 2

Fig. 6