A MODEL TESTING PROBLEM IN GENETICS [*]

W. T. Federer, D. S. Robson, and A. M. Srb

The central question as to the rate of genetic progress attained through selection by recombination of factors free of heterotic effects can be answered by using haploid organisms such as neurospora. This organism is ideally suited for such quantitative inheritance studies. With most diploid organisms, e.g., corn, it is impossible to separate the heterotic from the recombinatory effects as affected by selection. It would be, however, possible to differentiate between these two effects in certain diploid organisms and for certain characters, e.g., drosophila males and sex-linked characters.

In neurospora, some characters which could be used for selection studies to obtain some evidence on the above question are:

(i)   Growth rate (linear and dry weight)

(ii) Spore size (diameter)

(iii) Number of asci per fruiting body

(iv) Number of fruiting bodies.

A series of environments involving varying temperature, pH, nutrition, etc. could be studied to ascertain the size and extent of genotype-environment interaction.

All of the above characters could be considered to be inherited quantitatively. It might be difficult, but not impossible, to obtain data for character number (iv). Selection for rapid growth and large spore size would have considerable practical importance in neurospora experiments since large spores are easier to handle and a faster growth rate could shorten the time for experimentation. Thus, this would represent an actual breeding problem in neurospora.

The experiment could be designed in such a manner that genetic progress through selection for quantitatively inherited characters, could be obtained for intra-specific and inter-specific crosses. Favorable crossable neurospora strains from many parts of the world are available for this. Also,

---

markers could be used in such a way that one could select as follows:

(i)   Select without reference to markers

(ii) Select with reference to markers in <u>each</u> strain

(iii) Select with reference to markers in <u>both</u> strains.

This would require adequately marked strains. One very well marked species is available; the remaining species available to us are sufficiently well marked to conduct such a selection experiment.

Genetic progress through selection would be measured for intra-specific and inter-specific crosses in order to study differences between such crosses. Reciprocal crosses for all crosses would allow checks on cytoplasmic effects.

Several hypotheses may be of interest in this experiment. For example, the hypothesis of equal and additive gene effects for a given character could be postulated. Also, one could hypothesize that no natural selection is operating. Empirical evidence from a properly designed experiment could be obtained for these and other hypotheses.

In addition to the above, population size and sample size from a population could be varied to study their effect on the rate of genetic progress attained through selection by recombination free of heterotic effects. Although this problem can be solved theoretically, as indicated below, empirical evidence on living organisms needs to be obtained.

<u>An example of the effect of family size on the chances of selecting the best genotype in a selfing system.</u>

The selection procedures used in most, if not all breeding programs represent stochastic processes of the Markov type. Individuals selected in one generation are used as the breeding stock for the next generation; consequently, the genotypic distribution and selection probabilities for the n'th stage of the selection process depend only upon the outcome in the (n-1)'th stage. If the population size and selection rate remain constant through successive generations then the corresponding Markov process is stationary in the sense of having constant transition probabilities; in other words, the probability that a particular genotype will be selected from the progeny of a given set of breed stock does not depend upon the number of selection stages required to produce that breed stock. The

construction and analysis of Markovian models for selection processes should

prove useful to breeders as a means of predicting and comparing the properties

of alternative procedures. An example involving selection in a simple 2-factor

diploid selfing system is given here to illustrate the type of information which

such a mathematical analysis will yield.

For purposes of mathematical simplicity in this example we shall assume

that the 9 different genotypes which can appear in this breeding system are

phenotypically distinct, and ordered with respect to the phenotypic character

being selected. The ordering to be considered here is

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ aabb < aaBb < aaBB < Aabb < AaBb < AaBB < AAbb < AABb < AABB \end{array}$$

In practice, particularly in dealing with the so-called quantitative characters,

genotypes will not be phenotypically distinct; the situation we describe with

the present model is therefore better than reality from the breeder's viewpoint.

Our breeding process begins with the cross of two inbred lines which differ

genotypically at 2 independently segregating loci. An F_2 population of r indi-

viduals is obtained, and the individual which is phenotypically best is selected

out and selfed to produce r progeny. Again, the best of these r F_3 individuals

is selected and selfed to produce r F_4 individuals, and so on. We shall now

compute the probability that the individual selected at the n'th stage of this

process will be of the j'th genotype, and we are especially interested in the

probability that this selection will be of genotype number 9 (AABB), the top

ranking genotype.

If at any given stage of the selection process an individual of the i'th

genotype is selected then the probability that the r progeny in the next genera-

tion will include r_1 individuals of genotype number 1, r_2 of type 2, $\cdots$, r_9 of

type 9 is given by the multinomial term

$$\frac{r!}{r_1!r_2!\cdots r_9!} \; a_{i1}^{r_1} a_{i2}^{r_2} \cdots a_{i9}^{r_9}$$

where $a_{i1}, a_{i2}, \cdots, a_{i9}$ are the segregation probabilities for the i'th genotype under selfing. For example, for i=4 these segregation probabilities are $\frac{1}{4}, 0, 0, \frac{1}{2}, 0, 0, \frac{1}{4}, 0, 0$, respectively. The probability that an individual of the j'th genotype will be selected from the r progeny of an individual of the i'th genotype is then

$$p_{ij} = (a_{i1} + \cdots + a_{ij})^r - (a_{i1} + \cdots + a_{ij-1})^r$$

These probabilities p_{ij} are called one-step transition probabilities in the terminology of Markov processes; p_{ij} is the conditional probability that the selected individual in any generation will be of genotype j given that the individual selected in the preceding generation was of genotype i. If segregation follows the Mendelian law then the p_{ij} are easily evaluated; for example,

$$p_{41} = (\tfrac{1}{4})^r, p_{42} = 0, p_{43} = 0, p_{44} = (\tfrac{3}{4})^r - (\tfrac{1}{4})^r, p_{45} = 0, p_{46} = 0, p_{47} = 1 - (\tfrac{3}{4})^r, p_{48} = 0, p_{49} = 0.$$

The probability of a transition from genotype i to genotype j in n steps, called the n-step transition probability $p_{ij}^{(n)}$ is also easily computed for this system. Clearly, for the homozygous genotypes 1,3,7, and 9 we have $p_{11}^{(n)} = p_{33}^{(n)}$ $= p_{77}^{(n)} = p_{99}^{(n)} = 1$ while for a genotype heterozygous at one locus, as genotype 4, we have $p_{4j}^{(n)} = 0$ for j=2,3,5,6,8,9 and

$$p_{41}^{(n)} = p_{41}p_{11}^{(n-1)} + p_{44}p_{41}^{(n-1)}$$

$$= p_{41} + p_{44}p_{41}^{(n-1)}$$

$$= p_{41}\left(\frac{1-p_{44}^{n}}{1-p_{44}}\right)$$

$$= \left(\tfrac{1}{4}\right)^{r}\left\{\frac{1-[(\tfrac{3}{4})^{r}-(\tfrac{1}{4})^{r}]^{n}}{1-(\tfrac{3}{4})^{r}+(\tfrac{1}{4})^{r}}\right\}$$

$$p_{44}^{(n)} = p_{44}p_{44}^{(n-1)}$$

$$= p_{44}^{n}$$

$$= [(\tfrac{3}{4})^{r}-(\tfrac{1}{4})^{r}]^{n}$$

$$p_{47}^{(n)} = p_{47}p_{77}^{(n-1)}+p_{44}p_{47}^{(n-1)}$$

$$= p_{47}\left(\frac{1-p_{44}^{n}}{1-p_{44}}\right)$$

$$= [1-(\tfrac{3}{4})^{r}]\left\{\frac{1-[(\tfrac{3}{4})^{r}-(\tfrac{1}{4})^{r}]^{n}}{1-(\tfrac{3}{4})^{r}+(\tfrac{1}{4})^{r}}\right\}$$

The n-step transition probabilities $p_{5j}^{(n)}$ are of primary interest since the system begins with an F_1 population of individuals of genotype number 5, AaBb. These are obtained in the same manner as above:

$$p_{51}^{(n)} = \frac{1}{1-p_{55}}\left\{p_{51}+\frac{p_{52}p_{21}}{1-p_{21}}+\frac{p_{54}p_{41}}{1-p_{41}}\right\} + \frac{p_{55}^{n}}{1-p_{55}}\left\{\frac{p_{52}p_{21}}{p_{22}-p_{55}}+\frac{p_{54}p_{41}}{p_{44}-p_{55}}-p_{51}\right\}$$

$$- \frac{p_{22}^{n}}{1-p_{22}} \left\{ \frac{p_{52}p_{21}}{p_{22}-p_{55}} \right\} - \frac{p_{44}^{n}}{1-p_{44}} \left\{ \frac{p_{54}p_{41}}{p_{44}-p_{55}} \right\}$$

$$= \frac{1}{1-p_{55}} \left\{ p_{51} + \frac{a(p_{52}+p_{54})}{1-b} \right\} + \frac{p_{55}^{n}}{1-p_{55}} \left\{ \frac{a(p_{52}+p_{54})}{b-p_{55}} - p_{51} \right\}$$

$$- \frac{b^{n}}{1-b} \left\{ \frac{a(p_{52}+p_{54})}{b-p_{55}} \right\}$$

where

$$a = p_{21} = p_{41} = p_{63} = p_{87} = \left(\frac{1}{4}\right)^{r}$$

$$b = p_{22} = p_{44} = p_{66} = p_{88} = \left(\frac{3}{4}\right)^{r} - \left(\frac{1}{4}\right)^{r}$$

$$c = p_{23} = p_{47} = p_{69} = p_{89} = 1 - \left(\frac{3}{4}\right)^{r}$$

and

$$p_{51} = \frac{1}{16^{r}}, \quad p_{52} = \frac{3^{r}-1}{16^{r}}, \quad p_{53} = \frac{4^{r}-3^{r}}{16^{r}}, \quad p_{54} = \frac{6^{r}-4^{r}}{16^{r}}, \quad p_{55} = \frac{10^{r}-6^{r}}{16^{r}}$$

$$p_{56} = \frac{12^{r}-10^{r}}{16^{r}}, \quad p_{57} = \frac{13^{r}-12^{r}}{16^{r}}, \quad p_{58} = \frac{15^{r}-13^{r}}{16^{r}}, \quad p_{59} = \frac{16^{r}-15^{r}}{16^{r}}$$

$$p_{52}^{(n)} = p_{52} \left\{ \frac{b^{n}-p_{55}^{n}}{b-p_{55}} \right\}$$

$$p_{53}^{(n)} = \frac{1}{1-p_{55}} \left\{ p_{53} + \frac{ap_{56}+cp_{52}}{1-b} \right\} + \frac{p_{55}^{n}}{1-p_{55}} \left\{ \frac{ap_{56}+cp_{52}}{b-p_{55}} - p_{53} \right\}$$

$$- \frac{b^{n}}{1-b} \left\{ \frac{ap_{56}+cp_{52}}{b-p_{55}} \right\}$$

$$p_{54}^{(n)} = p_{54} \left\{ \frac{b^n - p_{55}^n}{b - p_{55}} \right\}$$

$$p_{55}^{(n)} = p_{55}^n$$

$$p_{56}^{(n)} = p_{56} \left\{ \frac{b^n - p_{55}^n}{b - p_{55}} \right\}$$

$$p_{57}^{(n)} = \frac{1}{1 - p_{55}} \left\{ p_{57} + \frac{a p_{58} + c p_{54}}{1 - b} \right\} + \frac{p_{55}^n}{1 - p_{55}} \left\{ \frac{a p_{58} + c p_{54}}{b - p_{55}} - p_{57} \right\}$$

$$- \frac{b^n}{1 - b} \left\{ \frac{a p_{58} + c p_{54}}{b - p_{55}} \right\}$$

$$p_{58}^{(n)} = p_{58} \left\{ \frac{b^n - p_{55}^n}{b - p_{nn}} \right\}$$

$$p_{59}^{(n)} = \frac{1}{1 - p_{55}} \left\{ p_{59} + \frac{c(p_{56} + p_{58})}{1 - b} \right\} + \frac{p_{55}^n}{1 - p_{55}} \left\{ \frac{c(p_{56} + p_{58})}{b - p_{55}} - p_{59} \right\}$$

$$- \frac{b^n}{1 - b} \left\{ \frac{c(p_{56} + p_{58})}{b - p_{55}} \right\}$$

The one-step transition probabilities p_{5j} represent the selection probabilities for the F_2 generation; for example, with an F_2 population of size $r=10$ the probability of selecting genotype 9 is

$$p_{59} = 1 - \left(\frac{15}{16}\right)^{10} = .4755 .$$

The n-step transition probabilities $p_{5j}^{(n)}$ approach limiting values of 0 for

$j=2,4,5,6$, and 8 and non-zero values given by the first term in the expressions for $p_{51}^{(n)}$, $p_{53}^{(n)}$, $p_{57}^{(n)}$ and $p_{59}^{(n)}$. For example,

$$p_{59}^{(\infty)} = \frac{1}{1-p_{55}} \left\{ p_{59} + \frac{c(p_{56}+p_{58})}{1-b} \right\}$$

Figure 1 shows the rate of approach to this limiting value for small r; with $r=5$, for example, the chances that the breeder will have selected the best genotype in the F_4 generation are 82 in 100, and the chances of the ultimate selection of genotype 9 if the selection process continues indefinitely are 86.3 in 100.

The primary value of this stochastic model approach to selection would lie in the comparison of alternative selection procedures. Absolute values of the selection probabilities may be substantially altered by changing the segregation ratios or the ordering of the genotypes, but the relative values obtained by two different selection procedures operating on the same inheritance system should correctly reflect the merits of the two procedures.
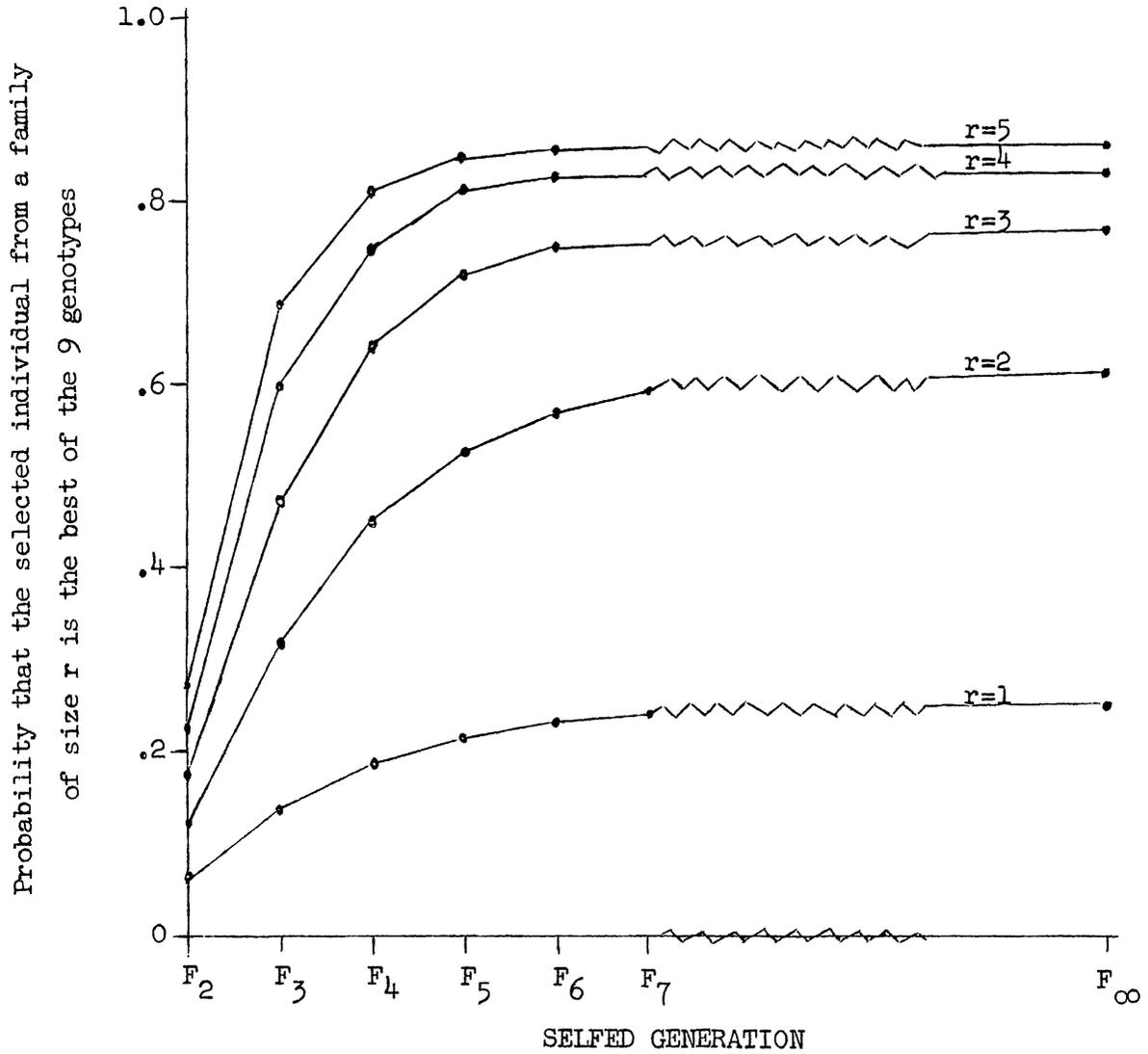
Figure 1. A graph showing the probability of selecting the best of the 9
genotypes in a 2-factor selfing system aabb   aaBb   ••• AABB
when the selection process consists of picking the best genotype
from a family of size r, r=1,2,3,4,5.