

## Inbreeding in artificial selection programmes

BY ALAN ROBERTSON

*Agricultural Research Council Unit of Animal Genetics,  
Institute of Animal Genetics, Edinburgh 9*

(Received 22 October 1960)

The restriction of population size, in terms of the number of individuals allowed to breed, leads to an increase of homozygosity within the population and also to random changes in gene frequency from generation to generation. The magnitude of these effects can be described by introducing the concept of the idealized random breeding population (see Falconer, 1960, p. 49). In Crow's words (1954), 'the offspring come from pairs of gametes which are drawn at random from an infinite pool to which each parent had contributed equally'. The number of progeny contributed by individual parents then has a Poisson distribution.

In practice, parents do not have equal probabilities of contributing to the next generation. Under normal conditions, such inequalities will arise from differences between individuals in inherent fertility and fecundity and between family groups in viability. Under artificial selection, even if based solely on the individual's own performance, there will be a further effect because of differences between progeny groups in the character under selection. In a population composed of full-sib groups, for instance, all pairs of parents may have equal probabilities of contributing progeny to be measured but, if there is any genetic variance in the character under selection, they will not have an equal chance of having the progeny selected as parents of the subsequent generation. This paper presents a theoretical treatment of this problem which was brought to my notice by Dr G. McBride.

A brief reference to the phenomenon was made by Morley (1954). He was primarily interested in the fact that 'in a flock exposed to selection, the genetically superior individuals will tend to be most inbred. As a corollary, selection increases the approach to homozygosity, not only at loci carrying genes determining the character in question but at all loci'.

The problem can be most easily approached in terms of the random changes in the frequency of genes which affect neither natural fitness nor the character under selection. If such a gene is carried by an animal whose progeny mean for the selected character is above the population mean, then its frequency will tend to increase until its association with genes increasing the selected character is broken down by segregation and recombination. The reverse will apply to genes carried by individuals whose progeny mean is below the population mean for the selected character. As a consequence, the random changes in the frequency of such genes

from generation to generation are greater than that expected from the actual number of parents and we can say that the effective population size is below the actual size.

We may take as the simplest example a population split into  $N$  full-sib groups. If the population behaved in the idealized manner, then the expected variance of gene frequency change from generation to generation would be equal to

$$E(\delta q)^2 = \frac{q(1-q)}{4N}.$$

We now assume a large 'gene pool' from which we sample the parents of the full-sib groups. We form the 'gene pool' in the next generation by weighting the genes in each group by a factor depending on the mean of the full-sib group for the selected character. Suppose that the probable contribution of a family is  $f_1$  and that the frequency of the gene in question in that family is  $q_1$ . Then the mean gene frequency in the newly formed gene pool is given by summing over families as  $\sum f_1 q_1 / \sum f_1$ , and the variance of this will be  $q(1-q) \sum f^2 / 4(\sum f)^2$  since each full-sib family is a sample of four alleles. Thus, using the above formula in reverse, we may say that the effective population size  $N_e$  is given by

$$N_e = \frac{(\sum f)^2}{\sum f^2}.$$

Now the families to which these  $f$  factors are ascribed can be considered as members of an infinite population of such families, in which the mean and standard deviation of  $f$  are  $m$  and  $\sigma$ . Then

$$E(\sum f^2) = Nm^2 + N\sigma^2$$

and

$$E(\sum f)^2 = N^2 m^2 + N\sigma^2,$$

so that

$$N_e = \frac{N + C^2}{1 + C^2}, \quad (1)$$

where  $C$  is the coefficient of variation of  $f$  between families. If we speak in the usual way about the relative selective advantages  $s_1, s_2$  of the separate families, it will be seen that  $C^2$  is the variance of  $s$  between families.

It may be of value to contrast this approach with that of Crow (1954) and Crow & Morton (1955) to similar problems. In this paper the approach is *a priori* in predicting effective population size from a knowledge of the variation of selective advantage between families. They were concerned to measure  $N_e$  from the observation of the variation of actual family size.

The two approaches may be connected by noting that Crow's expression for the variance of family size after selection,  $V_K$ , has the expectation  $2(N-1)(1+2C^2)/(N+C^2)$  in a population stable in number.

It remains to determine the variance of  $s$ . We assume that selection is based only on the individual's own performance. It is known that if the selection differential in artificial selection is  $\bar{i}$  standard deviations, then individuals in a group with deviation  $x$  from the mean have a selective advantage  $\bar{i}x/\sigma_w$ , where  $\sigma_w$  is the

standard deviation within groups. This holds only for small values of  $x$ . By trial and error, it has been found that the expression

$$\sigma_s^2 = \frac{\bar{i}^2 \sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

(where  $\sigma_b^2$  is the variance of the selected character between families) is an excellent approximation for the variance of  $s$  up to  $\sigma_b = 2\sigma_w$ , though it is less valuable at high selection intensities.

We have then an expression for the variance of gene frequency changes in the first generation. In the next generation, a new set of full-sib groups will be formed, superimposed on which is a further relationship pattern deriving from the full-sib structure in the first generation. Full-sibs chosen as parents of the second generation will have progeny which are related as single first cousins. So in the selection of individuals in the second generation, those having parents from superior full-sib groups in the first will have a slightly higher chance of being selected. Because of the linear relationship between selected advantage and the deviation of a group from the population mean, the expected change in gene frequency in the second generation will be half that in the first and in the same direction. The expected total change will then accumulate as  $1, \frac{3}{2}, \frac{7}{4} \dots$  to a limiting value of 2. But the effective population size is dependent on the variance of random changes in gene frequency, and  $C^2$  in equation (1) will increase in successive generations as  $1, \frac{9}{4}, \frac{49}{16}$ , up to a limiting factor of 4. Thus, for individual selection, we may finally put

$$\frac{N}{N_e} = 1 + \frac{4\bar{i}^2 \sigma_b^2}{\sigma_b^2 + \sigma_w^2} \tag{2}$$

The parents of the full-sib groups are not a random sample of their generation. As a selected group, they might be expected to show less genetic variation than a random sample. In fact it can be shown that the genetic variation will be reduced by a proportion  $h^2 \bar{i} (\bar{i} - x)$ , where  $h^2$  is the heritability of the character under selection and  $x$  is the abscissa of the unit normal curve when the area cut off is  $p$ . If we are then dealing with full-sib groups between which there is no non-genetic variation, then we may write

$$\frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} = \frac{1}{2} h^2 [1 - h^2 \bar{i} (\bar{i} - x)]$$

and

$$\frac{N}{N_e} = 1 + 2\bar{i}^2 h^2 [1 - h^2 \bar{i} (\bar{i} - x)]$$

In most laboratory selection experiments,  $p$  lies between 20% and 40%, and  $\bar{i}$  then lies between 1 and 1.4. In this range,  $\bar{i} (\bar{i} - x)$  varies little between 0.7 and 0.8. Selection for a character with a high heritability may thus cause an important reduction in the effective population size.

The model implies that animals from each progeny group have an equal chance of being measured, so that at measurement the distribution of family sizes will be Poisson. In laboratory experiments, one generally tries to measure an equal

number of progeny in each mating. This must lead to a reduction of the variation in family size after selection. Using Crow & Morton's formula (14), it can be shown that if at measurement there are  $n$  individuals of each sex in all families, then the variance of group size after selection is reduced by a factor  $1/n$ . For a large number of families the effective population size is then given by

$$\frac{N}{N_e} = \left(1 - \frac{1}{2n}\right) + 4C^2 \left(1 - \frac{1}{n}\right).$$

If  $n = 1$ , implying no selection and equal representation of all families in the next generation, this reduces to the known result  $N_e = 2N$ .

INDEX SELECTION BASED ON FAMILY AND INDIVIDUAL PERFORMANCE

We have been discussing so far selection based solely on the individual's own performance. It is well known that it is more efficient to use a weighted combination of the measurement on the individual and the average measurement of the family it belongs to (Lush, 1947). Writing  $P$ ,  $F$ , and  $\bar{P}$ , for the individual's own measurement, its family mean, and the population mean respectively, we then select on an index

$$I = P - F + k(F - \bar{P}),$$

where  $k$  is chosen to maximize the correlation between  $I$  and the individual's breeding value. In individual selection, which we discussed earlier, we pay no attention to family mean and select only on  $P$ , i.e. we put  $k = 1$ . The inbreeding effect then depended on  $\sigma_b^2/(\sigma_b^2 + \sigma_w^2)$ . In this context, it will be seen that the effect of index selection is to multiply  $\sigma_b^2/\sigma_w^2$  by a factor  $k^2$ .

Using Lush's terminology of  $r$  and  $t$  for the genetic and phenotypic intra-family correlations, we see that  $\sigma_b^2/(\sigma_b^2 + \sigma_w^2)$  is equal to  $t$ . Lush showed that if family size is large, the optimum value of  $k$  is  $r(1-t)/t(1-r)$ , which for full-sib families becomes  $(1-t)/t$ . In index selection, we then have

$$\begin{aligned} \frac{k^2 \sigma_b^2}{k^2 \sigma_b^2 + \sigma_w^2} &= \frac{1}{1 + \sigma_w^2/k^2 \sigma_b^2} \\ &= \frac{1}{1 + \frac{(1-t)}{t} \frac{t^2}{(1-t)^2}} \\ &= \frac{1}{1 + \frac{t}{1-t}} \\ &= 1 - t. \end{aligned}$$

Thus if  $t$  is low, implying a low heritability and considerable attention paid in index selection to family average rather than to the individuals own performance, the effective population size may be much reduced.

On the other hand there may be much non-genetic variation between families leading to a value of  $t$  greater than  $r$ . In individual selection, selection may then be concentrated on only a few families. Index selection, which in this case will put more weight on the individual's deviation from the family mean, will then spread the selection over the families and increase the effective population. In the extreme case, when selection is based entirely on the individual's deviation from family average, the effective size will equal the actual size.

#### DISCUSSION

The problem that has been discussed here has proved difficult to solve exactly and many side-issues have had to be examined. However, the main factors controlling effective population size in a selection programme have been clearly isolated. We may then expect the inbreeding under individual selection to be greater than that calculated from the actual number of parents when both the intensity of selection and the heritability of the character are high. Selection on combined individual and family measurements would be expected to increase this effect except when there is much non-genetic variation between families.

Perhaps the most interesting aspect of the results is in the time-lag in the effect. In the ideal random breeding population, the inbreeding due to a given generation of ancestors is found after two generations (in the absence of self-fertilization) and thereafter remains constant. Here, due to correlated changes in gene frequency in successive generations, the increase in inbreeding from the second to the third generation and from the third to the fourth are similar in magnitude to that in the second generation. It is commonly observed in pedigree analyses of livestock populations that an animal does not contribute greatly to inbreeding until he appears four or five generations back in the pedigree. The effect discussed here may be partly a cause of this but, in addition, in pedigree breeding, selection still operates on animals as grandparents or great-grandparents. The present effect is due to an increase in the contribution of ancestors over several generations because descendants continue to be selected on their own individual merit and not because they had a particular grandfather or great-grandfather.

The earlier discussion was entirely in terms of full-sib groups. In the larger animals, it is customary to keep fewer males than females. The males will be represented by a half-sib group of progeny and will themselves have been intensely selected. In the females, the present effect will be reduced because of the lower intensity of selection but increased because the progeny form a full-sib rather than a half-sib group.

In a recent series of experiments in this laboratory, Dr McBride carried out a selection experiment with *Drosophila melanogaster* in which full pedigrees were kept. In each of two lines, he had 10 full-sib groups and in each sex selected the best 10 individuals on their own score out of 100 measured in each generation. In the early generations, while the response to the selection was still high, the effective population size was 4 and 5 groups in the two cases. The heritability of the

selected character was 50%. On the basis of the present theory, the expected value of  $N/N_e$  is 2.54, in fair agreement with the observed value. In a further experiment, selection was based on a combination of individual and family measurements with the same intensity of selection. The value of  $k$  used was 2. In the early generations, the effective number of parents was 8 pairs. Here the expected value of  $N/N_e$  is 4.31, which is not in such good agreement. However, we have in these experiments a clear demonstration of the actual reduction in effective population size under intense selection.

#### SUMMARY

In a population under artificial selection, the effective population size may be less than the actual number of parents selected because there will be variation between families in the character under selection and consequently in the probability of selection. Expressions are developed for the magnitude of the effect, which will be greater the more intense the selection and the higher the heritability of the selected character. The inbreeding due to outstanding individuals may rise for several generations after their use.

#### REFERENCES

- CROW, J. F. (1954). *Breeding Structure of Populations. II: Effective Population Number in Statistics and Biology*. Iowa State College Press. Pp. 543–556.
- CROW, J. F. & MORTON, N. E. (1955). Measurement of gene frequency drift in small populations. *Evolution*, **9**, 202–214
- FALCONER, D. S. (1960). *Introduction to Quantitative Genetics*. Oliver & Boyd, Edinburgh.
- LUSH, J. L. (1947). Family merit and individual merit as bases for selection. *Amer. Nat.* **81**, 241–261, 362–379.
- MORLEY, F. H. W. (1954). Selection for economic characters in Australian Merino sheep. IV: The effect of inbreeding. *Aust. J. agric. Res.* **5**, 305–316.