Genet. Res., Camb. (1961), **2**, *pp.* 177–188
With 1 *text figure*
Printed in Great Britain

Effects of dominance and size of population on response to mass selection*

By KEN-ICHI KOJIMA

Department of Genetics, North Carolina State College, U.S.A.

(*Received* 14 *September* 1960)

A theory of mass selection has been developed for a large genetic sample and its application for predicting the actual change of certain genetic parameters, such as the mean value of a character in the offspring population from selected parents, is widely made by plant and animal breeders. In this theory the amount of change in mean yield depends upon the relative proportion of the additive genetic variance to the total variance on the phenotypic measurements in the parental populations and upon the intensity of selection, usually, in terms of selection differentials.

In practice of mass selection, however, the evaluation of genotypes is always conducted with a finite number of individuals. In some cases the number is so small that the application of the theory formulated for a large population may become very erroneous. In this paper a few aspects of mass selection conduted in a small genetic sample will be considered. More specifically, the purpose of this paper is threefold: (1) to obtain the expression of the expected change in gene frequency, (2) to derive an approximate variance of the gene frequency change, and (3) to evaluate the effects of dominance on responses to mass selection, when a small number of organisms are tested. Thus, a few concepts hitherto discussed intuitively and qualitatively in regard to consequences of selection in a small population can be presented more quantitatively in terms of selection intensities, sample sizes and gene actions.

Although a single locus with two alleles is primarily considered, the basic framework and method of the approach employed in this paper are not necessarily limited to the situations with one locus. The results, however, may be subjected to considerable changes, if joint effects of different genes and linkages among loci are marked. Generalization of the findings from a single locus situation to many-loci situations will be discussed only when genes do not exhibit epistasis and they are in equilibrium with respect to linkage.

SCHEME OF SELECTION

Since the term mass selection is often used in a very broad sense, it appears necessary to describe explicitly the procedure of selection to be discussed in the present study, along with an introduction of notations to be used.

* Contribution No. 1225 of the Journal Series, North Carolina Agricultural Experiment Station, Raleigh, North Carolina. This work was supported by a grant from National Science Foundation.

Let N individuals be a random sample of diploid organisms taken from a large genetic population G. The G can be a potential or conceptual population which is sometimes called a gene pool. The gene frequency at a locus, say A-a, in G is denoted by q for allele A, and the zygotic frequencies of types, AA, Aa and aa, are represented by U_1, U_2 and U_3, respectively.

The genetic sample consisting of the N random individuals is evaluated in a performance test with respect to a certain quantitative character, of which the value is denoted by Y. The Y's are expressed in terms of the phenotypic standard deviation, σ_t, of the original population G. The density distribution of the character in G is denoted by $\phi(Y)$ with the mean $\overline{Y}$. Furthermore, let $\phi_1(Y)$, $\phi_2(Y)$ or $\phi_3(Y)$ be the density distributions of Y, when the genotype at the locus in question is given as AA, Aa or aa, respectively. Their means are symbolized as $\overline{Y}_1$, $\overline{Y}_2$ and $\overline{Y}_3$ and the deviations of the respective means from $\overline{Y}$ are written as d_1, d_2 and d_3. All of these distributions are assumed to be similar bell-shaped ones. The genotypic variance of the character contributed from the locus in question is $U_1d_1^2+U_2d_2^2+U_3d_3^2$, which in turn becomes a heritability component in broad sense. This component is usually a small fraction for a given locus.

Now the measurements on N individuals are recorded from the performance test. The n best performing individuals are selected according to the rank of Y's, and will be used as parents of the following generation. The intensity of selection in this procedure is then expressed by a fixed proportion, n/N, while the point of truncation is not fixed but varies as a random variate. Let Y_0 be the value of this random variate in a particular test. Then the Y_0 is an order statistic and its density distribution is

$$f(Y_0) = \frac{N!}{n!\,(N-n-1)!}P^n(1-P)^{N-n-1}\phi(Y_0) \tag{1}$$

where $$P = \int_{Y_0}^{\infty}\phi(Y)\,dY = U_1P_1+U_2P_2+U_3P_3$$

and $$P_1 = \int_{Y_0}^{\infty}\phi_1(Y)\,dY, \qquad P_2 = \int_{Y_0}^{\infty}\phi_2(Y)\,dY \qquad \text{and} \qquad P_3 = \int_{Y_0}^{\infty}\phi_3(Y)\,dY$$

EXPECTED CHANGE OF GENE FREQUENCY

The change of gene frequency after one trial of selection depends upon the relative numbers of selected individuals having AA, Aa and aa genotypes. Let n_1, n_2 and n_3 (i.e. $n-n_1-n_2$) be the numbers of AA, Aa and aa individuals in the selected group, respectively. The joint distribution of n_1, n_2 and n_3 for a given value of Y_0 is multinominal with

Means: $$\overline{n}_1' = nU_1P_1/P; \qquad \overline{n}_2' = nU_2P_2/P; \qquad \overline{n}_3' = nU_3P_3/P \tag{2}$$

Variances: $$\left.\begin{aligned} V'_{n_1} &= nU_1P_1(P-U_1P_1)/P^2; \qquad V'_{n_2} = nU_2P_2(P-U_2P_2)/P^2; \\ V'_{n_3} &= nU_3P_3(P-U_3P_3)/P^2 \end{aligned}\right\} \tag{3}$$

https://doi.org/10.1017/S0016672300000689 Published online by Cambridge University Press

Covariances:

$$\mathrm{Cov}'(n_1.n_2) = -nU_1P_1U_2P_2/P^2; \qquad \mathrm{Cov}'(n_1.n_3) = -nU_1P_1U_3P_3/P^2;$$
$$\mathrm{Cov}'(n_2.n_3) = -nU_2P_2U_3P_3/P^2 \tag{4}$$

where prime (′) stands for 'conditional on Y_0'.

In order to obtain the means of n_1 and n_2 for all possible Y_0, $\bar{n}_1'$ and $\bar{n}_2'$ are to be integrated over the distribution of Y_0 given in (1). Then

$$\left.\begin{aligned}\bar{n}_1 &= nU_1\int_{-\infty}^{\infty}\frac{P_1}{P}f(Y_0)\,\mathrm{d}Y_0\\ \bar{n}_2 &= nU_2\int_{-\infty}^{\infty}\frac{P_2}{P}f(Y_0)\,\mathrm{d}Y_0\end{aligned}\right\} \tag{5}$$

which can be approximately written as

$$\left.\begin{aligned}\bar{n}_1 &= nU_1\int_{-\infty}^{\infty}\frac{P+d_1\phi(Y_0)}{P}f(Y_0)\,\mathrm{d}Y_0\\ \bar{n}_2 &= nU_2\int_{-\infty}^{\infty}\frac{P+d_2\phi(Y_0)}{P}f(Y_0)\,\mathrm{d}Y_0\end{aligned}\right\} \tag{6}$$

The approximation used in (6) is not good unless d_1 and d_2 are small enough so that d_1^2 and d_2^2 are negligibly small compared with d_1 and d_2, respectively.

A rigorous mathematical principle to obtain the values of the integral form in (6) is discussed by Ruben (1954) in connexion with the moment of order statistics, but the application of his principle seems to be hopelessly complex for the present case where n takes an arbitrary number. A rather simple argument to obtain approximate solutions for (6) is used (see Appendix), and the results are

$$\bar{n}_1 = nU_1 + NU_1d_1\left[\phi(\bar{Y}_0)+\frac{1}{2}\phi''(\bar{Y}_0)u_2(Y_0)+\frac{1}{3!}\phi'''(\bar{Y}_0)u_3(Y_0)+\ldots\right] \tag{7}$$

$$\bar{n}_2 = nU_2 + NU_2d_2\left[\phi(\bar{Y}_0)+\frac{1}{2}\phi''(\bar{Y}_0)u_2(Y_0)+\frac{1}{3!}\phi'''(\bar{Y}_0)u_3(Y_0)+\ldots\right] \tag{8}$$

where $\bar{Y}_0$ is the mean of the nth largest observation in a random sample of size $(N-1)$; $\phi(\bar{Y}_0)$, $\phi''(\bar{Y}_0)$, $\phi'''(\bar{Y}_0)$, etc. are the density function, $\phi(Y)$, its second, third, etc. derivatives, all evaluated at $\bar{Y}_0$; and $u_2(Y_0)$, $u_3(Y_0)$, etc. are the second, third, etc. central moments of the nth largest observation in a sample of size $(N-1)$ taken from the distribution, $\phi(Y)$.

Now the change in gene frequency from a single trial of selection is

$$\Delta q = \frac{2n_1+n_2}{2n}-q \tag{9}$$

and its expectation is

$$E(\Delta q) = \frac{2\bar{n}_1+\bar{n}_2}{2n}-q \tag{9}'$$

https://doi.org/10.1017/S0016672300000689 Published online by Cambridge University Press

Substituting (7) and (8) into (9)′, it is found that

$$E(\Delta q) = k(U_1 d_1 + \tfrac{1}{2} U_2 d_2) \tag{10}$$

where

$$k = \frac{N}{n}\left[\phi(\overline{Y}_0) + \frac{1}{2}\phi''(\overline{Y}_0)\, u_2(Y_0) + \frac{1}{3!}\phi'''(\overline{Y}_0)\, u_3(Y_0) + \ldots\right] \tag{11}$$

In the equation (10) the values of U_1 and U_2 are arbitrary except that $U_1 + U_2 + U_3 = 1$. When the parental population is in Hardy-Weinberg equilibrium, the expression (10) becomes

$$E(\Delta q) = kq(1-q)\{q(\overline{Y}_1 - \overline{Y}_2) + (1-q)(\overline{Y}_2 - \overline{Y}_3)\} \tag{12}$$

This form is very similar to the well-known expression for the change of gene frequency derived from large sample theory; i.e. the form being the additive genetic comparison multiplied by $kq(1-q)$. Furthermore, if N becomes large without changing n/N and if $\phi(Y)$ is normal, then k is equal to $N/n\phi(\overline{Y}_0)$, because $u_2(Y_0)$, $u_3(Y_0)$, etc. diminish. The k value in such a case is often written as i, and called selection differential. The expression of k given in (11) may be called a generalized selection differential, since it is not restricted by the size of sample and the form of phenotypic distribution. It has been known that the selection differential for a finite population can be computed by using the table of ranked normal deviates (e.g. Fisher & Yates, 1953), when the phenotypic distribution is normal. An evaluation of the expression for k in formula (11) will be made in a later section.

VARIANCE OF CHANGE IN GENE FREQUENCY

From the formula (9) the variance of Δq is written as

$$V_{\Delta q} = \frac{1}{4n^2}\{4V_{n_1} + V_{n_2} + 4\,\mathrm{Cov}\,(n_1, n_2)\} \tag{13}$$

In order to spell out (13) in terms of genetic parameters, the variances and covariance of n_1 and n_2 conditional on Y_0 given in (3) and (4) must be integrated over the distribution (1). They are

$$V_{n_1} = nU_1(1-U_1) + nd_1 U_1(1-2U_1)\,k$$
$$V_{n_2} = nU_2(1-U_2) + nd_2 U_2(1-2U_2)\,k$$
$$\mathrm{Cov}\,(n_1, n_2) = -nU_1 U_2 - nU_1 U_2(d_1 + d_2)\,k$$

when the same order of approximation is permitted as in the case of $\overline{n}_1$ and $\overline{n}_2$. Substituting these variances and covariance into equation (13), the variance of Δq becomes

$$V_{\Delta q} = \frac{1}{n}q(1-q) - \frac{1}{4n}U_2 + \frac{1}{n}k(U_1 d_1 + \tfrac{1}{2}U_2 d_2)(1-2q) - \frac{1}{4n}kU_2 d_2 \tag{14}$$

When the parental population is in Hardy-Weinberg's equilibrium, i.e.,

$$U_1 = q^2 \quad \text{and} \quad U_2 = 2q(1-q),$$

https://doi.org/10.1017/S0016672300000689 Published online by Cambridge University Press

the variance, (14), becomes

$$V_{\Delta q} = \frac{q(1-q)}{2n}\{1+2q(1-q)(d_1+d_3-2d_2)k+d_2k\}$$

or

$$V_{\Delta q} = \frac{q(1-q)}{2n}\{1+q(1-q)\beta k+(1-2q)\alpha k\} \tag{15}$$

where α is the additive genetic comparison defined in (12) and β is the dominance comparison defined by $(\overline{Y}_1+\overline{Y}_3-2\overline{Y}_2)$.

A few special cases are considered:

1. N $= n$; when k is defined to be zero for $N = n$ (because of the approximation used in deriving k, the value of k is not exactly zero at $N = n$), then

$$V_{\Delta q} = \frac{q(1-q)}{2n}$$

2. No dominance; $\overline{Y}_1+\overline{Y}_3-2\overline{Y}_2 = 0$ or $\overline{Y}_1-\overline{Y}_2 = \overline{Y}_2-\overline{Y}_3$, then

$$V_{\Delta q} = \frac{q(1-q)}{2n}\{1+k(\overline{Y}_1-\overline{Y}_2)(1-2q)\}$$

3. Complete dominance; $\overline{Y}_1 = \overline{Y}_2$, then

$$V_{\Delta q} = \frac{q(1-q)}{2n}\{1+k(\overline{Y}_1-\overline{Y}_3)(1-q)(1-3q)\}$$

The variance in (15) can be partitioned into two factors; one from pure sampling, without selection, that is, $q(1-q)/2n$, and the other which reflects the effect of selection on the variance of Δq. The latter effect can be positive or negative depending upon the interrelation between the gene frequency, q, in the parental population and the level of dominance measured by $h = \beta/(\overline{Y}_3-\overline{Y}_1)$. Figure 1 presents this interrelation, in which any point falling in the region surrounded by heavily drawn boundaries (S in Fig. 1) results in $V_{\Delta q}$ being less than $q(1-q)/2n$, and $V_{\Delta q}$ is larger than $q(1-q)/2n$, when the point falls outside of the region S. With no dominance ($h = 0$), the variance with selection for allele, A, is smaller than that expected from the pure random drift in a population of the same size when q is higher than $\frac{1}{2}$. As the level of dominance increases, the region S widens. With complete dominance ($h = 1$) this relation between the two variances holds for $q > \frac{1}{3}$. As h becomes larger, the upper and lower bounds for S tend to approximately $q = 0{\cdot}79$ and $q = 0{\cdot}21$, respectively. Thus one can expect that selection for higher performance reduces the size of variance of Δq in comparison with the variance due to the genetic random drift when the gene frequency ranges from intermediate to high.

When the value of gene frequency and the intensity of selection are given, the variance of change in gene frequency is a function of only the additive and dominance comparisons. As it is easily seen in (15), the higher the degree of dominance the larger the reduction in the variance. As to the additive effect, a larger additive comparison results in a greater reduction in the variance, provided $q > \frac{1}{2}$. It is important to note that the magnitude of these comparisons is expressed relative

to the size of the total phenotypic standard deviation, σ_t. Hence, a locus with larger comparisons in a population is subjected to less random variation than a locus with smaller comparisons in the same population, when the gene frequencies at the two loci are comparable in size.

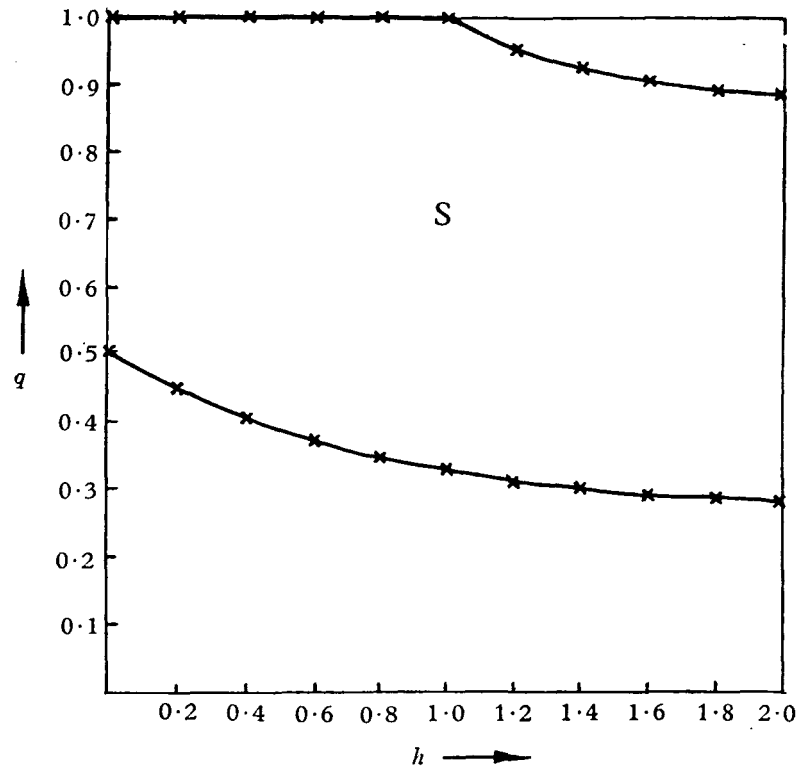

Fig. 1. The region, S, in which the variance, $V_{\Delta q}$, from mass selection is smaller than the variance from random genetic drift when the same number of individuals are taken from a population (see text).

q: Gene frequency
h: Dominance level

EXPECTED GAIN FROM SELECTION

The predicted gain can be defined as the expected difference between the mean of an offspring population produced by selected individuals and the mean of a reference population. In this paper the population in Hardy-Weinberg's equilibrium with the same q as in G is considered as the reference population. Random mating among n selected parents is assumed. Let $\overline{Y}'$ stand for the mean of an offspring population, while $\overline{Y}$ is the mean of the reference population. Both populations can be assumed to be infinitely large.

In an offspring population the gene frequency is $(q+\Delta q)$, then the difference, $\overline{Y}'-\overline{Y}$, from a single trial of selection is equal to

$$\Delta\overline{Y} = 2\Delta q\{q(\overline{Y}_1-\overline{Y}_2)+(1-q)(\overline{Y}_2-\overline{Y}_3)\}+(\Delta q)^2\{\overline{Y}_1+\overline{Y}_3-2\overline{Y}_2\} \qquad (16)$$

https://doi.org/10.1017/S0016672300000689 Published online by Cambridge University Press

when $\Delta\bar{Y}$ is measured in terms of σ_t. The first term in (16) is proportional to the additive comparison, $\{q(\bar{Y}_1-\bar{Y}_2)+(1-q)(\bar{Y}_2-\bar{Y}_3)\}$, and the second to the dominance comparison, $\{\bar{Y}_1+\bar{Y}_3-2\bar{Y}_2\}$. Then, the expectation of $\Delta\bar{Y}$ or the predicted gain is equal to

$$2E(\Delta q)\alpha+\{E(\Delta q)\}^2\beta+V_{\Delta q}\beta \tag{17}$$

Now each term in (17) is evaluated stepwise. The first term is equal to $2q(1-q)\alpha^2 k$, which is equal, in turn, to $k\sigma_\alpha^2$, where σ_α^2 is the additive genetic variance (Kojima, 1959) or the heritability since the character is measured by σ_t. This term is equivalent to the expression of the conventional prediction equation, $i\sigma_\alpha^2$, except that k is a generalized value of i. The term, therefore, represents the expected change in the mean of population associated with a linear (or additive) effect of genes.

The second term can be written as

$$\frac{k\sigma_\alpha^2}{2}kq(1-q)\beta.$$

According to Kojima (1959), the quantity, $q^2(1-q)^2\beta^2$, is equal to the dominance variance. Then the second term becomes

$$\pm\tfrac{1}{2}(k\sigma_\alpha^2)\,k\sigma_\beta$$

where the sign depends on the direction of dominance. When the allele A is dominant over the allele a, i.e. $\bar{Y}_1-\bar{Y}_2<\bar{Y}_2-\bar{Y}_3$, the sign is negative, and positive otherwise. Then this term represents an increment, negative or positive, onto the value of $E(\Delta\bar{Y})$ when there is dominance. The magnitude of this contribution is proportional to the gain due to the linear effect of genes, $(k\sigma_\alpha^2)$, and to the degree of dominance measured by σ_β.

The nature of the third term $V_{\Delta q}\beta$, is quite different from that of the previous two terms. While the previous two were directly proportional to the linear effect of genes, the third term does not depend on the linear effect but is proportional to the variance in the change of gene frequency. As given in formula (15), this variance is composed of two factors. One is the sampling variance of n parents taken from the population without selection (i.e. the variance due to the genetic random drift), and the other is a factor which modifies the sampling variance as a function of the genetic effects and the selection differential.

The sign of the contribution from the third term to $E(\Delta q)$ depends again on the direction of dominance as it was in case of the second term. The third term can be written as $\pm V_{\Delta q}\sigma_\beta/q(1-q)$, where $+$ is used when $\bar{Y}_1-\bar{Y}_2<\bar{Y}_2-\bar{Y}_3$ and $-$ when $\bar{Y}_1-\bar{Y}_2>\bar{Y}_2-\bar{Y}_3$.

Summing up all three terms in (17), the expected gain is equal to

$$k\sigma_\alpha^2-\tfrac{1}{2}(k\sigma_\alpha^2)\,k\sigma_\beta-V_{\Delta q}\sigma_\beta/q(1-q) \tag{18}$$

when the heterozygote is better than the mean of the two homozygotes, and otherwise

$$k\sigma_\alpha^2+\tfrac{1}{2}(k\sigma_\alpha^2)\,k\sigma_\beta+V_{\Delta q}\sigma_\beta/q(1-q) \tag{18}'$$

Roughly speaking, the first term represents an increment due to the additive effect, the second modifies the first term according to the direction and degree of dominance, and the third term represents the joint effect of dominance and inbreeding due to the finiteness of sample size. When n and N become large, the last term in (18) or (18)′ tends to zero, and $E(\Delta\overline{Y})$ becomes a function of σ_α^2 and σ_β. The gross effect of dominance on $E(\Delta\overline{Y})$ is obvious; with $\beta < 0$, $E(\Delta\overline{Y})$ is affected downwards, and with $\beta > 0$, upwards.

When there is no selection, i.e. $k = 0$ and $E(\Delta q) = 0$, the total expected gain is equal to $\pm(\sigma_\beta/2n)$. This is the change of mean due to the random fixation of the locus. When the heterozygote is superior to the mean of the two homozygotes, the change is negative. This value, $-(\sigma_\beta/2n)$, is nothing but the amount of inbreeding depression due to dominance and the finiteness of the sample. On the other hand, with dominance such that $\overline{Y}_1 - \overline{Y}_2 > \overline{Y}_2 - \overline{Y}_3$, the mean of the offspring population tends to increase by $+(\sigma_\beta/2n)$, through the increase of homozygosity.

For the purpose of an illustration of the theory developed in this section, the ratio, $E(\Delta\overline{Y})/k\sigma_\alpha^2$, is examined as an indicator of discrepancy between the expected gain and the gain by the usual prediction equation for one locus with complete dominance. Substituting into (18) the expression for $V_{\Delta q}$ is given in the previous section, the ratio becomes

$$1 - \tfrac{1}{2}q(1-q)(\overline{Y}_1 - \overline{Y}_3) - \frac{1}{4nk(1-q)^2(\overline{Y}_1 - \overline{Y}_3)} - \frac{(1-3q)}{4n(1-q)} \qquad (19)$$

As an example let $q = \frac{1}{2}$, $n = 10$ and $N = 50$. For $(\overline{Y}_1 - \overline{Y}_3)$, which is the difference between the two homozygotes, 0·1 may be assigned. With these specifications, $\sigma_\alpha^2 = 1/800$ and $\sigma_\beta^2 = 1/1600$ for this locus, and $k = 1{\cdot}37$. The ratio in question is 0·28. Since the ratio has to be 1 for complete agreement between the expected gain and the gain predicted by $k\sigma_\alpha^2$, this ratio, 0·28, means 72 % overprediction by the usual prediction equation. When 0·2 is assigned for $(\overline{Y}_1 - \overline{Y}_3)$, then $\sigma_\alpha^2 = 1/200$; $\sigma_\beta^2 = 1/400$ and the ratio becomes 0·63, indicating 37 % overprediction.

DISCUSSION

A numerical evaluation of the selection differential derived in (11) can be made when $\phi(Y)$ is normal. In Table 1 the exact values $(\overline{Y}_s)$ for the selection differential computed from the tabulated values of normal order statistics in Fisher and Yates' Table (1953) are compared with the approximate values computed from the first and second terms in formula (11). For the expression of $u_2(Y_0)$ the asymptotic variance of Y_0 is used; that is,

$$u_2(Y_0) = \frac{n(N-n+1)}{N(N+1)}\{\phi(\overline{Y}_0)\}^{-2}$$

where the $\overline{Y}_0$ is the mean of the nth largest value in a sample of size N. The discrepancies between the corresponding k and $\overline{Y}_s$ are surprisingly small, even though only the first two terms in (11) are used for the computation of k. This

remarkably good agreement may be a peculiarity when $\phi(Y)$ is normal, but does provide a mean for checking adequacy of formula (11). It should be noted that the formula for k possesses only some theoretical interests when the distribution of phenotypes is assumed to be normal. For practical purposes the average of top n values in Fisher and Yates' tabulation will serve a better role for the selection differential if $\phi(Y)$ is normal. When the distribution is non-normal, the formula for k can be used to obtain appropriate values of k by using asymptotic formulae for $\bar{Y}_0$, $u_2(Y_0)$, etc.

Generalization of the findings from one-locus situations to many-loci situations is possible to a certain extent. The expressions for the expected change of gene frequency given in (10) and in (12) are general for any single locus in an epistatic system, as long as $P_1 = P + d_1\phi(Y_0)$ and $P_2 = P + d_2\phi_2(Y_0)$ hold approximately. Such is the case when $\phi(Y)$, $\phi_1(Y)$ and $\phi_2(Y)$ are similar in shape, and d_1 and d_2 are small fractions. This generalization holds true for the expressions of the expected variance of gene frequency changes.

The formulae for the expected gain in (18) and (18)′ are specific for one-locus situations. When no epistasis and no linkage disequilibrium are assumed, they can be combined over more than one locus as

$$E(\Delta\bar{Y}) = k\sigma_A^2 \pm \tfrac{1}{2}k^2 \sum_i (\sigma_\alpha^2\sigma_\beta)_i \pm \sum_i \{V_{\Delta q}\sigma_\beta/q(1-q)\}$$

where σ_A^2 is the total additive genetic variance as the fraction of σ_t^2 or the heritability. The summation is taken over different loci. The sign is + when the heterozygote of the locus is inferior to the mean of the two corresponding homozygotes, and − when the heterozygote is superior. Main use of this combined formula is to point out what kinds of biases are expected in the well-known prediction equation, $i\sigma_A^2$. The first and obvious bias arises from the difference between k and i. As seen in Table 1, however, this difference is not appreciably

Table 1. *Values of selection differential with various intensities of selection and different sizes of genetic samples drawn from a normal population*

	n/N									
	0·1		0·2		0·3		0·4		0·5	
N	k	$\bar{Y}_s$	k	$\bar{Y}_s$	k	$\bar{Y}_s$	k	$\bar{Y}_s$	k	$\bar{Y}_s$
∞	1·74	1·74	1·40	1·40	1·16	1·16	0·96	0·96	0·79	0·79
50	1·71	1·70	1·37	1·37	1·14	1·14	0·95	0·95	0·79	0·79
40	1·70	1·69	1·36	1·36	1·13	1·13	0·95	0·95	0·78	0·78
30	1·68	1·67	1·35	1·35	1·12	1·13	0·94	0·94	0·78	0·78
20	1·68	1·64	1·33	1·33	1·10	1·11	0·93	0·93	0·77	0·77
10	(1·78)*	1·54	1·27	1·27	1·05	1·07	0·89	0·90	0·73	0·74

k: Computed by the use of the first two terms in the formula derived in the text.

$\bar{Y}_s$: Computed by the average of scores for the ordered data given by Fisher and Yates (1953).

* The value of k for $n = 1$ and $N = 10$ has a large deviation from the true value due to the approximations made in computations.

https://doi.org/10.1017/S0016672300000689 Published online by Cambridge University Press

large. The finiteness of sample size alone, therefore, does not become a serious source of bias. Even for $N = 10$ and $n = 2$, the difference is less than 10%. In the following discussion, for this reason, k is assumed to be not different from i.

With respect to the effect of dominance two extreme cases are considered. When the directions of dominance at different loci balance out in the positives and negatives, the difference between $i\sigma_A^2$ and $E(\Delta \bar{Y})$ should be zero or very small. Under these circumstances the expected gain becomes estimable by $k\sigma_A^2$, although the variance of such estimates can easily be much larger than that expected when all loci do not exhibit dominance.

The other extreme case is given when the directions of dominance are the same for all loci. Now the effects of dominance on $E(\Delta \bar{Y})$ become cumulative only in one direction, positive or negative. In other words, the deviation of $i\sigma_A^2$ from $E(\Delta \bar{Y})$ becomes maximum. This deviation can be partitioned into two parts. One is the deviation due to dominance effects *per se* and the other is the joint effects of dominance and variance of gene frequency change at each locus. The latter is reciprocally proportional to the number of individuals selected and used as parents for the next generation.

Another way to interpret the two parts of the deviation is quite instructive. The first part is the deviation due to the non-linearity of gene action with respect to the allelic substitution at the locus. The linear contribution of gene action to $E(\Delta \bar{Y})$ is $k\sigma_A^2$, and the term $\sum_i \{\sigma_\alpha^2 \sigma_\beta\}_i$ represents a sum of adjustments made on the linear effects at each locus by the corresponding effects of dominance. It should be repeated that this adjustment *is not restricted by the sizes of genetic sample* used. Even in an infinite size selection, this term is non-zero, provided that the directions of dominance at different loci do not cancel in $\pm \sum_i \{\sigma_\alpha^2 \sigma_\beta\}_i$. The second part of the deviation represents an overall effect of inbreeding when genes exhibit dominance. When this term is negative, the effect is often called *inbreeding depression*. Such effect is then liable to the actual number of parents selected for producing the following generation.

In this study only one cycle of selection was treated. With the present-day knowledge of most geneticists it is extremely difficult to extend the theories of one-cycle truncated selection to an arbitrary number of cycles. Experimental studies such as one by Clayton, Morris & Robertson (1957), and empirical studies on high-speed computers by Fraser (1957, 1960), Martin & Cockerham (1960) and some others, may give light on this difficult task.

SUMMARY

A theory of mass selection in a small population was developed, and the mean change in gene frequencies, the variance of gene frequency changes and the expected gain in the mean phenotypic value of an offspring population were formulated in terms of a generalized selection differential and the additive and dominance effects of genes.

The magnitude of the variance of changes in gene frequency was compared with

the magnitude of the variance expected from the genetic random drift in a population with the same gene frequency and of the same size in absence of selection. The former was found to be usually smaller than the latter when the gene frequency ranged from intermediate to high and when selection was directed for a high performance.

The usual prediction equation for gain from selection in an infinite population was compared with the expected gain formula derived for a small population. The size of the population did not cause a serious difference between the two expected gains when there was no dominance effect of genes. Dominance alone could cause the usual prediction to be slightly more biased. The joint effects of the finite size of population and dominance gene action could amount to a considerable bias in the usual prediction equation. Such a bias can be, in the main, accounted for by the inbreeding depression.

REFERENCES

Clayton, G. A., Morris, J. A. & Robertson, A. (1957). An experimental check on quantitative genetical theory. I: Short-term responses to selection. *J. Genet.* **55**, 131–151.

Fisher, R. A. & Yates, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd, London.

Fraser, A. S. (1957). Simulation of genetic systems by automatic digital computers. II: Effects of linkage on rates of advance under selection. *Aust. J. biol. Sci.* **10**, 492–499.

Fraser, A. S. (1960). Simulation of genetic systems by automatic digital computers. V: Linkage, dominance and epistasis. *Biometrical Genetics*, 70–83. Pergamon Press, New York.

Kojima, K. (1959). Role of epistasis and overdominance in stability of equilibrium with selection. *Proc. nat. Acad. Sci., Wash.*, **45**, 984–989.

Martin, F. G. & Cockerham, C. C. (1960). High speed selection studies. *Biometrical Genetics*, 37–45. Pergamon Press, New York.

Ruben, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41**, 200–226.

APPENDIX

Approximate values of $\bar{n}_1$ *and* $\bar{n}_2$

The integral equations in (6) can be evaluated in the following manner:

(1) Expand $\phi(Y_0)$ around $\overline{Y}_0$ which is the mean of the nth observation from the largest in a sample of size $(N-1)$ randomly taken from the parental population.

$$\phi(Y_0) = \phi(\overline{Y}_0)+(Y_0-\overline{Y}_0)\phi'(\overline{Y}_0)+\frac{1}{2!}(Y_0-\overline{Y}_0)^2\phi''(\overline{Y}_0)$$

$$+\frac{1}{3!}(Y_0-\overline{Y}_0)^3\phi'''(\overline{Y}_0)+\frac{1}{4!}(Y_0-\overline{Y}_0)^4\phi''''(\overline{Y}_0)+\ldots$$

(2) The ratio $f(Y_0)/P$ in (6) can be written as

$$\frac{N}{n}\cdot\frac{(N-1)!}{(n-1)!\,(N-n-1)!}P^{n-1}(1-P)^{N-n-1}\phi(Y_0) = \frac{N}{n}.g(Y_0).$$

$g(Y_0)$ is the density function of the nth observation from the largest in a sample of size $(N-1)$.

https://doi.org/10.1017/S0016672300000689 Published online by Cambridge University Press

(3) Hence $\bar{n}_1$ is a sum of the following quantities:

$$nU_1 \int_{-\infty}^{\infty} f(Y_0)\,\mathrm{d}Y_0 = nU_1$$

$$NU_1 d_1 \phi(\bar{Y}_0) \int_{-\infty}^{\infty} g(Y_0)\,\mathrm{d}Y_0 = NU_1 d_1 \phi(\bar{Y}_0)$$

$$NU_1 d_1 \phi'(\bar{Y}_0) \int_{-\infty}^{\infty} (Y_0 - \bar{Y}_0) g(Y_0)\,\mathrm{d}Y_0 = 0$$

$$\frac{1}{2!} NU_1 d_1 \phi''(\bar{Y}_0) \int_{-\infty}^{\infty} (Y_0 - \bar{Y}_0)^2 g(Y_0)\,\mathrm{d}Y_0 = \frac{1}{2!} NU_1 d_1 \phi''(\bar{Y}_0) u_2(Y_0)$$

$$\frac{1}{3!} NU_1 d_1 \phi'''(\bar{Y}_0) \int_{-\infty}^{\infty} (Y_0 - \bar{Y}_0)^3 g(Y_0)\,\mathrm{d}Y_0 = \frac{1}{3!} NU_1 d_1 \phi'''(\bar{Y}_0) u_3(Y_0)$$

. . . . . .
. . . . . .
. . . . . .

where $u_2(Y_0)$, $u_3(Y_0)$, and so on, represent the second, third and higher central moments of the $g(Y_0)$ distribution.

(4) Then

$$\bar{n}_1 = nU_1 + NU_1 d_1 \left\{ \phi(\bar{Y}_0) + \frac{1}{2!} \phi''(\bar{Y}_0) u_2(Y_0) + \frac{1}{3!} \phi'''(\bar{Y}_0) u_3(Y_0) + \ldots \right\}$$

(5) Similarly

$$\bar{n}_2 = nU_2 + NU_2 d_2 \left\{ \phi(\bar{Y}_0) + \frac{1}{2!} \phi''(\bar{Y}_0) u_2(Y_0) + \frac{1}{3!} \phi'''(\bar{Y}_0) u_3(Y_0) + \ldots \right\}$$

https://doi.org/10.1017/S0016672300000689 Published online by Cambridge University Press