GENETIC CHANGES WITH GENERATIONS OF ARTIFICIAL SELECTION

LUIS SILVELA

Departamento de Genética, Instituto Nacional de Investigaciones Agrarias, Carretera de La Coruña, Km. 7, Madrid-35, Spain

Manuscript received April 11, 1979
Revised copy received April 8, 1980

ABSTRACT

Using conditional probabilities and moment-generating matrices, I derived approximate algebraic equations that give expectations of gene frequency, population mean, gene frequency variance within lines, or heterozygosity, and gene frequency variance between lines, or drift, for repeated cycles of recurrent selection in populations of finite size. For genes of large effect, the responses to selection differ substantially from the classical expectations, and equations are derived that give quantitative estimates of asymmetry of response when selection is done in opposite directions. Particular cases of the derived formulae yield equations given by other authors. The error involved in the approximations is discussed in the APPENDIX.

THE moment-generating matrix method has proven to be quite useful for studying genetic changes in populations of finite size. The method was developed by ROBERTSON (1952) to study the change of the additive genetic variance of a recessive gene in a random-mating population with no selection. In a later paper, ROBERTSON (1960) applied the method to follow changes in gene frequency for a gene of small effect in a population under selection. He dealt with additive and recessive genes. The theory was developed assuming that the selective advantage of the gene was known. The results were then applied to artificial selection by using a formula, originally derived by HALDANE (1931), relating the metric scale, where artificial selection is applied, to the selective advantage that is conferred to the gene by this selection.

In this paper, the matrix method is extended to slightly more complex situations. It always deals with one-locus models in populations of finite size under selection. We consider any degree of dominance including overdominance, genes of small effect and, whenever possible, also genes of large effect. The theory is developed on the metric scale so that it directly concerns artificial selection.

We follow changes in the mean including asymmetry of response when it occurs, changes in gene frequency variance within lines, or heterozygosity, and changes in gene frequency variance between lines, or drift. The results obtained provide a better understanding of the changes of genetic statistics in populations simultaneously subjected to drift and selection.

THE MODEL

Consider a one-locus diploid model with two alleles in a random-mating population of an ideal species that is monoecious with the possibility of self-fertilization.

In each generation, a random-mating population of N parents produces M progeny. Each single progeny is produced by taking two parents at random with replacement to allow for the possibility of self-fertilization. The best N phenotypes in the progeny are selected to form the parental population in the next generation. Random mating to obtain M progeny is statistically equivalent to taking a random sample of size M from the conceptually infinite population of all possible progeny, which will of course be in Hardy-Weinberg proportions.

Let this infinite population be that shown in Table 1, where q is the gene frequency among the parents, u is the gene effect and a the degree of dominance. We use g_k for the deviation of a genotypic value from the mean. The population mean is $\mu = 2qu + 2q(1-q)au$, the additive genetic variance $\sigma_a^2 = 2q(1-q)[1 + (1-2q)a]^2u^2$ and the total genotypic variance $\sigma_g^2 = \sum_k f_k g_k^2$, as given by COMSTOCK and ROBINSON (1948).

The phenotype of an individual is considered to be the sum of the overall population mean, the genotypic deviation in the locus considered, g_k, and a residual or environmental deviation, r. The residual deviation is assumed to be a standard normal variate and the genotypic and the residual deviations are considered independent, $\mathrm{Cov}(g_k,r) = 0$. We use p for the phenotypic deviation, or the difference of a phenotype from the population mean, so that $p = g_k + r$.

The three distributions involved in the selection process are thus

$$\mathrm{Prob}(g_k) = f_k\,,\quad \Psi_R(r) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right)\quad \text{and}\quad \Psi_P(p) = \sum_k f_k\,\Psi_R(p-g_k)$$

PROBABILITY OF SELECTING A GIVEN GENOTYPE

Let us consider a fixed phenotypic value p. The genotypic content of p could be any of the three possibilities. In statistical terms, it is a random variable, the probability of which is distributed as follows: Let $S(g_k|p)$ be the conditional probability of g_k given p, $\mathrm{Prob}(g_k \times p)$ the joint probability of g_k and p, and for

TABLE 1

Genotypic frequencies, values and deviations

Genotype	Frequency	Genotypic values	Genotypic deviations
LL	$f_1 = q^2$	$2u$	$g_1 = 2u-\mu$
Ll	$f_2 = 2q(1-q)$	$u+au$	$g_2 = u+au-\mu$
ll	$f_3 = (1-q)^2$	0	$g_3 = -\mu$

the sake of simplicity let Prob( ) be either probability or density depending on the distributions involved.

$$S(g_k|p) = \frac{\text{Prob}(g_k \times p)}{\text{Prob}(p)} = \frac{f_k\, \Psi_R(p-g_k)}{\Psi_P(p)} = \frac{f_k \dfrac{1}{\sqrt{2\pi}} \exp\left[-\dfrac{(p-g_k)^2}{2}\right]}{\sum\limits_k f_k \dfrac{1}{\sqrt{2\pi}} \exp\left[-\dfrac{(p-g_k)^2}{2}\right]} =$$

$$f_k \frac{1 + p\, g_k + \dfrac{p^2-1}{2} g_k^2 + \ldots}{1 + \dfrac{p^2-1}{2}\sigma_g^2 + \ldots} = f_k \left[1 + p\, g_k + \frac{p^2-1}{2}(g_k^2 - \sigma_g^2) + \ldots\right] .$$

If the phenotype selected is not fixed, but rather a random variable with distribution $\Theta(p)$, the unconditional probability of a given genotype being selected $S(g_k)$ is

$$S(g_k) = \int_{-\infty}^{\infty} S(g_k|p)\, \Theta(p)\, dp = f_k \left[1 + E(p) g_k + \frac{E(p^2) - 1}{2}(g_k^2 - \sigma_g^2) \ldots\right], \quad (1)$$

where the expectations $E(p)$ and $E(p^2)$ are taken over the distribution $\Theta(p)$.

For example, if a phenotype is taken at random from the original population, $\Theta(p) = \Psi_P(p)$ and

$$S(g_k) = \int_{-\infty}^{\infty} \frac{f_k\, \Psi_R(p-g_k)}{\Psi_P(p)}\, \Psi_P(p)\, dp = f_k \int_{-\infty}^{\infty} \Psi_R(p-g_k)\, dp = f_k ,$$

as it should.

If the phenotype selected is the best one of a random sample of size M, then $\Theta(p)$ is the distribution of the first-order statistic of a random sample of size M taken from $\Psi_P(p)$, and $E(p)$ and $E(p^2)$ are the first two moments of the distribution $\Theta(p)$.

If the upper fraction of the sample or the top N phenotypes are selected, as will be the case in this paper, then $E(p)$ and $E(p^2)$ become the averages of the moments of the distributions of the top N-order statistics.

If u is small, then $\Psi_P(p)$ is approximately normal, and $E(p)$ and $E(p^2)$ can be approximated by the averages of normal-order statistics. KOJIMA (1961) calls $E(p)$ the generalized selection differential, but usually it is referred to as just the intensity of selection. In the following, we shall use i for $E(p)$ and i_2 for $E(p^2)$ without necessarily implying that $\Psi_P(p)$ is normal.

The reasoning above implies that the frequencies of the three possible genotypes in the selected fraction have a multinominal distribution with probabilities $S(g_1)$, $S(g_2)$ and $S(g_3)$ (KOJIMA 1961; GALLEY and CURNOW 1972). Extension of the theory to multiple or even infinite genotypes is straightforward.

Changes in gene frequency: If we assume that the gene effect u is small enough that terms in u^2 and higher powers can be ignored, we get from (1) the well-known formula

$$S(g_k) = f_k(1 + ig_k) \ , \qquad (2)$$

which shows the linear relationship between the metric scale and the fitness or selective scale (HALDANE 1931; ROBERTSON 1963).

KIMURA and CROW (1978) have shown that, for infinite population size and truncation selection, if i is the ratio of the ordinate at the truncation point to the proportion selected, this last formula is valid even if the phenotypic distribution is far from normal, as long as it is differentiable.

The probability that an individual with a genetic value $(\mu + g_k)$ is selected is given by $S(g_k)$, and then the expected genotypic mean and variance of the top N individuals selected are

$$\mu_N = \sum_k (\mu + g_k) S(g_k) = \mu + i\sigma_g^2 \ ,$$

always larger than the population mean μ, and

$$\sigma_N^2 = \frac{N-1}{N} \sum_k (\mu + g_k - \mu_N)^2 \, S(g_k) = \frac{N-1}{N} (\sigma_g^2 + i\mu_3) \ ,$$

which may be larger or smaller than the population genotypic variance σ_g^2, depending on the size and sign of the population third-order genotypic central moment μ_3. For random selection $i = 0$ and $\mu_N = \mu$, $\sigma_N^2 = \frac{N-1}{N} \sigma_g^2$ as they should.

Expectations in the progeny of the N selected individuals can be obtained from multinomial distribution theory. Unless otherwise specified the expectations in this paper refer to the conceptually infinite progeny populations.

Ignoring terms of order u^2 and letting $E[\,]_1$ be the expectation in the progeny, the following set of equations, given in matrix form, can be derived.

$$\begin{bmatrix} E[q]_1 \\ E[q(1-q)iu]_1 \\ E[q(1-q)(1-2q)aiu]_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & \left(1-\frac{1}{2N}\right) & 0 \\ 0 & 0 & \left(1-\frac{1}{2N}\right)\left(1-\frac{2}{2N}\right) \end{bmatrix} \begin{bmatrix} q \\ q(1-q)iu \\ q(1-q)(1-2q)aiu \end{bmatrix}$$

This is a recurrence relation, and the transition matrix is called a moment generating matrix. The vector of expectations at generation t will be obtained by multiplying the initial vector t times by the transition matrix. This method was obtained by ROBERTSON (1952, 1960), and many conclusions of this section have already been reported by him.

For the first few generations, the approximation requirements are just those in (2), *i.e.*, the gene effect u (or, to be more specific, iu and aiu) should be small enough that their squares can be ignored. As the number of generations increases, the approximation deteriorates; and, for the formulae to be valid, Niu and $Naiu$ should be small and their squares negligible (ROBERTSON 1960; KIMURA 1962).

A check on how good these approximations are as the number of generations increases is given in the APPENDIX.

Letting $A = (1 - \frac{1}{2N})$ and $B = (1 - \frac{2}{2N})$, the expected gene frequency at generation t results

$$E[q]_t = q + \frac{1-A^t}{1-A} q(1-q)iu + \frac{1-A^tB^t}{1-AB} q(1-q)(1-2q)aiu . \qquad (3)$$

Formula (3) was given by ROBERTSON (1960) for the special cases of recessive and additive gene action ($a = -1$, $a = 0$), his notation corresponding to the selective or fitness scale where $s = 2iu$.

Making $t = \infty$ and entering the values of A and B in (3), we get the selection limit

$$E[q]_\infty = q + 2Niu\, q(1-q) \left[1 + \frac{N}{3N-1} (1-2q)a \right] . \qquad (4)$$

For additive ($a = 0$) and recessive ($a = -1$) gene action, (4) yields the formulae reported by ROBERTSON (1960). For overdominant loci ($a > 1$) and initial equilibrium frequency $\bar{q} = \frac{1+a}{2a}$, (4) becomes the expansion of KIMURA's (1957) equation given by ROBERTSON (1962).

In equation (3), a appears in the product $(1 - 2q)a$, so that for $q = 0.5$, a common situation in breeding species with the possibility of self-fertilization, the whole selection process as defined by $E[q]_t$ is the same for additive, dominant and overdominant loci.

The rate of advance is

$$R[q]_t = E[q]_{t+1} - E[q]_t = q(1-q)iu\, [A^t + A^t + A^tB^t(1-2q)a] .$$

For any gene frequency, as selection proceeds t increases and A^tB^t will become much smaller than A^t. That is, a point will be reached at which the effects of the degree of dominance a upon selection will be exhausted. Thereafter, changes will occur at a relative rate of $A^t = \left(1 - \frac{1}{2N}\right)^t$. Therefore, the effects of the degree of dominance will be readily exhausted if N is small. As a matter of fact, if $N = 1$, as can be the case for an autogamous species, $AB = 0$ and the rate of advance becomes independent of a after the very first generation.

Rates of advance, ratios of initial to final responses and the half-life of the selection process have been extensively treated by ROBERTSON (1960).

Changes in second-order moments: Let O_1, O_2, O_3 be the observed frequencies of the three genotypes in the selected fraction. One can easily derive

$$E[q(1-q)]_1 = E[(O_1 + \tfrac{1}{2}O_2)(1 - O_1 - \tfrac{1}{2}O_2)] = \left(1 - \frac{1}{2N}\right)(1+iua)q(1-q) +$$

$$\left(1 - \frac{1}{2N}\right)q(1-q)(1-2q)iu - \left(1 - \frac{1}{2N}\right)\frac{8N-6}{2N-1} q^2(1-q)^2iua .$$

Proceeding as before, we get

$$\begin{bmatrix} E[q(1-q)]_1 \\ E[q(1-q)(1-2q)iu]_1 \\ E[q^2(1-q)^2iua]_1 \end{bmatrix} = \begin{bmatrix} & 1+iua & 1 & -\frac{8N-6}{2N-1} \\ A & 0 & B & 0 \\ & \frac{A}{2N}iua & 0 & BC \end{bmatrix} \begin{bmatrix} q(1-q) \\ q(1-q)(1-2q)iu \\ q^2(1-q)^2iua \end{bmatrix}$$

where $C = \left(1 - \frac{3}{2N}\right)$. After t generations of selection, we obtain

$$E[q(1-q)]_t = q(1-q)A^t \left[1 + iuat + \frac{4N-3}{5N-3}\left(\frac{1-B^tC^t}{1-BC} - t\right)iua + \frac{1-B^t}{1-B}(1-2q)iu - \frac{8N-6}{2N-1}\frac{1-B^tC^t}{1-BC}q(1-q)iua\right] . \quad (5)$$

The expression $E[q(1-q)]_t$ gives the within-line variance or heterozygosity at generation t. For additive genes, $2u^2 E[q(1-q)]_t$ is the additive genetic variance or the heritability since, within the approximation limits, the phenotypic variance equals one.

For the early generations, if N is relatively large, $\left(1 - \frac{1}{N}\right)^t \simeq 1 - \frac{t}{N}$ and,

$$E[q(1-q)]_t = q(1-q)\left(1 - \frac{t}{2N}\right)\left[1 + iut4a\left(q - \frac{1}{2}\right)\left(q - \frac{1+a}{2a}\right)\right] .$$

If there is no selection and only drift occurs, the initial heterozygosity decreases as $q(1-q)\left(1 - \frac{t}{2N}\right)$: This reduction can be accelerated or retarded by selection depending on the sign of $iut\ 4a\left(q - \frac{1}{2}\right)\left(q - \frac{1+a}{2a}\right)$. For $a > 0$, the reduction will be accelerated if q is within the interval $\frac{1}{2} < q < \frac{1+a}{2a}$ and retarded if q is outside the interval. As a increases, $\frac{1+a}{2a}$ approaches $\frac{1}{2}$ and the acceleration interval tends to disappear. For $-1 \leq a \leq 0$, there will also be acceleration for $q > \frac{1}{2}$. For $q = \frac{1}{2}$ or $q = \frac{1+a}{2a}$, selection will have no effect on the reduction of heterozygosity.

The retardation may be such as to reverse the direction of change, for $E[q(1-q)]_t$ will increase as long as $iu\ 4a\left(q - \frac{1}{2}\right)\left(q - \frac{1+a}{2a}\right)\left(2 - \frac{2t+1}{N}\right) > \frac{1}{N}$, *i.e.*, for genes with small initial frequencies and with large values of u and i and also for highly heterotic genes at high initial frequencies, provided N is large.

For N relatively large and during the early generations,

$$Vq_t = E[q]_t^2 - [E[q]_t]^2 = q(1-q)\frac{t}{2N}\left[1 + iut\,4a\left(q - \frac{1}{2}\right)\left(q - \frac{1+a}{2a}\right)\right] .$$

With no selection, the variance due to drift alone always increases as $V_{q_t} = q(1-q)\frac{t}{2N}$. This effect of drift may be attenuated or enhanced by selection. It will be attenuated if $a > 0$ and $\frac{1}{2} < q < \frac{1+a}{2a}$ or if $-1 \leq a \leq 0$ and $q > \frac{1}{2}$.

The variance at the beginning always increases, but if $E[q]_\infty = 1$, then $V_{q\infty} = 0$; therefore, V_{q_t} necessarily passes through a maximum and then decreases. If $E[q]_\infty = \frac{1}{2}$, the variance at the limit has a maximum value of $V_{q\infty} = \frac{1}{4}$.

Using equations (3) and (5) one can also follow changes in the mean $\mu = 2u[q + q(1-q)a]$,

$$\frac{E[\mu]_t}{2u} = E[q]_t + E[q(1-q)]_t a = q + \frac{1-A^t}{1-A}q(1-q)iu + \frac{1-A^tB^t}{1-AB}q(1-q)$$
$$(1-2q)iua + A^t q(1-q)a[1+iuat + \frac{4N-3}{5N-3}\left(\frac{1-B^tC^t}{1-BC} - t\right)iua + \qquad (6)$$
$$\frac{1-B^t}{1-B}(1-2q)iu - \frac{8N-6}{2N-1}\frac{1-B^tC^t}{1-BC}q(1-q)iua) .$$

For $t = 1$, (6) equals the formula given by KOJIMA (1961).

Again, for early generations and N relativity large, the rate of response

$$\frac{R[\mu]_t}{2u} = \frac{E[\mu]_{t+1}}{2u} - \frac{E[\mu]_t}{2u} = q(1-q)\left[[1 + (1-2q)a]^2 iu - \frac{a}{2N} - \right.$$
$$\left. \frac{2t+1}{N}2a^2\left(q - \frac{1}{2}\right)\left(q - \frac{1+a}{2a}\right)iu\right] .$$

For infinite population size, $R[\mu]_t = i2q(1-q)[1 + (1-2q)a]^2u^2 = i\sigma_a^2$, as it should.

The rate of response, and therefore the response itself, is the result of several and sometimes opposing forces as was shown algebraically for one generation by KOJIMA (1961) and through computer matrix iteration for repeated cycles of selection by HILL (1969b). The first term, $[1 + (1-2q)a]^2 iu$, is always positive and due only to selection. The second term, $-\frac{a}{2N}$, of opposite sign to that of a, is the inbreeding depression. It is independent of selection and is due to the finiteness of the population. The third term can be positive or negative depending on the values of q and a. It is due to the joint action of selection and finite population size and, as selection proceeds, increases in absolute value with t, while the other two terms remain constant in size and sign. If this third term is positive,

an initially negative rate of response might later become positive and the early loss could be recovered and exceeded. If a is positive, N not too large and either iu or $[1 + (1-2q)a]$ small, then the early response can be negative.

The rate of advance changes as selection proceeds, the acceleration being

$$\frac{RR[\mu]_t}{2u} = \frac{R[\mu]_{t+1}}{2u} - \frac{R[\mu]_t}{2u} = -\frac{4a^2}{N} q(1-q)\left(q - \frac{1}{2}\right)\left(q - \frac{1+a}{2a}\right)iu \ .$$

For example, for recessive genes ($a = -1$) at low initial frequencies, selection response will be accelerated during the early generations, especially for small values of N and large values of iu. This change will be reflected in increases in the additive genetic variance and is due to the finiteness of the population (ROBERTSON 1952) as well as to selection. The acceleration will be maximum for $q = 0.25$.

The advance in the first generation and the final advance can be obtained from (6), and their ratio is given by

$$\frac{\Delta\mu_\infty}{\Delta\mu_1} = \frac{E[\mu]_\infty - \mu}{E[\mu]_1 - \mu} =$$

$$\frac{-aq(1-q) + 2Niuq(1-q)\left[1 + \dfrac{N}{3N-1}(1-2q)a\right]}{\dfrac{-aq(1-q)}{2N} + iuq(1-q)\left[[1 + (1-2q)a]^2 - \dfrac{2a^2}{N}\left(q - \dfrac{1}{2}\right)\left(q - \dfrac{1+a}{2a}\right)\right]} \ .$$

For the final advance expression to be valid, Niu must be small.

For additivity ($a = 0$), the ratio is (ROBERTSON 1960)

$$\frac{\Delta\mu_\infty}{\Delta\mu_1} = \frac{2Niuq(1-q)}{iuq(1-q)} = 2N \ .$$

For nonadditive gene action and a larger in absolute value than Niu, we have again

$$\frac{\Delta\mu_\infty}{\Delta\mu_1} \simeq \frac{-aq(1-q)}{-aq(1-q)/2N} = 2N \ .$$

For recessive genes ($a = -1$), this is an exception to the results obtained by ROBERTSON (1960). In practice, however, small Niu values imply very small population sizes, which could exceptionally be the case when fast (but not very fast) inbreeding is wanted ($N = 2,3$) or in autogamous species ($N = 1$).

GENES OF LARGER EFFECT

Changes in one generation: For genes of larger effect, we have to keep, in expansion (1), terms in u^2 and ignore only terms in u^3. The probability of a given genotype being selected becomes then

$$S(g_k) = f_k\left[1 + ig_k + \frac{i_2 - 1}{2}(g_k^2 - \sigma_g^2)\right] .$$

For infinite population size, LATTER (1965), using the same model, derived a formula for the frequency of a genotype following selection to which $S(g_k)$ is linked as follows: If a standard normal distribution is truncated at x_o, then it can be shown that $i_2 = x_o i + 1$, and since the phenotypic distribution is approximately normal, then

$$S(g_k) \simeq f_k\left[1 + ig_k + \frac{x_o i}{2}(g_k^2 - \sigma_g^2)\right] \simeq f_k \frac{1 + ig_k\left[1 + \frac{g_k x_o}{2}\right]}{1 + \frac{1}{2} i x_o \sigma_g^2},$$

which is LATTER's (1955) expression in our notation.

Proceeding as before, with the multinomial probabilities $S(g_k)$, the derivations of expectations after one cycle of selection are straightforward. The results, except for the cases of nonadditive gene action or small population size, do not add much to LATTER's.

For recessive genes, the changes in gene frequency and in the mean after one generation of selection are

$$\Delta q = 2q^2(1-q)iu + (i_2 - 1)\left(\frac{1}{2} - q^2\right)\frac{\sigma_a^2}{2q}$$

$$\Delta_\mu = \sigma_a^2 i\left[1 + \frac{i_2 - 1}{i}\left(\frac{1}{2} - q^2\right)2u + iq(1-q)u\right] + \frac{q(1-q)u}{N}\left[1 + 6iq\left(\frac{2}{3} - q\right)u - 12(i_2 - 1)q\left(\frac{2}{3} - q\right)\left(q^2 - \frac{1}{2}\right)u^2 - i\sigma_a^2\right].$$

In Δq, the second term will give rise to asymmetry of response when selection with the same intensity is done in the opposite direction $(-i)$ and it is proportional to σ_a^2; as pointed out by ROBERTSON (1977), the responses will be symmetrical when both genotypic values are equally frequent $q^2 = \frac{1}{2}$.

For infinite population size, $q = 0.20$, $u = 0.25$, $i = 1.76$ and $i_2 = 3.25$, the response in the mean is $\Delta\mu = 1.36\, i\sigma_a^2$. The classical formula underestimates the true response by 36%, a surprisingly large error.

For small population sizes, the response in the mean may be insignificant or even opposite to the direction of selection. For $N = 10$, $q = 0.20$, $u = 0.20$ and upward selection with $i = 1.73$ and $i_2 = 3.18$, the change in the mean is $\Delta\mu\{+\} = 4.15\ \sigma_a^2$, whereas for downward selection with the same intensity $i = -1.73$, $i_2 = 3.18$, then $\Delta\mu\{-\} = 0.09\ \sigma_a^2$. That is, when small population size and nonadditive gene action occur at the same time, the asymmetry of response, one of the most conspicuous features of the responses due to genes of large effect (LATTER 1965), may be greatly enhanced.

Repeated cycles of selection: For additivity, transition matrices can be obtained

that are small enough to be handled. The second-order formula for the expected gene frequency at generation t is:

$$E[q]_t = q + \frac{1-A^t}{1-A} q(1-q) iu + \left[\frac{1-A^tB^t}{1-AB}\frac{i_2-1}{2} + \left[\frac{A^t-A^tB^t}{A-AB} - \frac{1-A^tB^t}{1-AB}\right]\frac{i^2A}{A-1}\right] q(1-q)(1-2q)u^2 ,$$

where the terms in second-order moments (i_2, i^2) will give the asymmetry of response to selection in the opposite direction ($-i$).

DISCUSSION

The standard theory of selection and that of inbreeding due to finite population size were developed separately. The joint action of selection and finite size has been treated by several authors in several different ways (KIMURA 1957; ROBERTSON 1960; KOJIMA 1961; HILL 1969a).

Although many of the results were implicit in previous works, the simultaneous mathematical treatment of both theories developed in this paper allows the setting up of explicit critical regions within or without which the rate of change and the acceleration of genetic moments are positive or negative. They permit us to know when selection enhances or opposes the effects of inbreeding or *vice versa* and to determine which share of the response can be accounted for by selection, which by inbreeding and which by the joint action of both, and how these components will change with time.

Most times the critical regions are defined by parameter values such as $q = \frac{1}{2}$ or $q = \frac{1+a}{2a}$, gene frequencies well known in standard selection theory, or expressions like $\left(1 - \frac{1}{2N}\right)$ which appear throughout the theory of inbreeding.

Some of the main features of the theories of inbreeding due to finite size and of selection in infinite populations are as follows:

With inbreeding, the mean gene frequency does not change; whereas, with selection, it always increases (ignoring overdominance).

The gene frequency variance within lines always decreases with inbreeding, but, with selection, it increases if $q < \frac{1}{2}$ and decreases otherwise. The between-lines variance increases steadily with inbreeding and does not change with selection.

With inbreeding, the mean at generation t is

$$E[\mu]_t = 2qu + 2q(1-q)au\left(1 - \frac{1}{2N}\right)^t ,$$

which increases or decreases with time if $a < 0$ or $a > 0$, respectively. The ratio of total to initial change is

$$\frac{\Delta\mu_\infty}{\Delta\mu_1} = \frac{-2q(1-q)au}{-2q(1-q)ua/2N} = 2N ,$$

and the half-life of the process is

$$t_{1/2} = \frac{\ln 1/2}{\ln\left(1 - \frac{1}{2N}\right)} ,$$

which for relatively large N is $t_{1/2} \simeq 1.4N$. Both the ratio and half-life are independent of u, q and a.

However, with selection in infinite populations, the mean always increases in the direction of selection, the ratio of the total to the initial change being

$$\frac{\Delta_{\mu_\infty}}{\Delta_{\mu_1}} = \frac{2[1 - q - q(1-q)a]u}{2iq(1-q)\,[1 + (1-2q)a]^2u^2} ,$$

which increases as i, u, and q decrease, especially so for recessive genes ($a = -1$). Likewise, the half-life of the selection process increases as i, u and q decrease, again, more so for recessive genes.

When selection and finite population size occur at the same time, the resulting properties are intermediate to that of the theories of inbreeding and selection. New features appear, such as the response being opposite to the direction of selection. For small Niu values, the response is mainly controlled by inbreeding and for large Niu values, by selection.

Standard selection theory has been developed at the first level of approximation, ignoring terms in u^2. Changes in the mean are explained in terms of the variance. As the gene effect increases, genetic moments of increasingly higher order start contributing significantly to selection response. In the formula, along with u^2, second-order moments of the intensity of selection (i^2, i_2) appear, which, always being positive, give rise to asymmetry of response when selection is done in opposite directions (LATTER 1965). Specific values of the parameters N, a, i, i_2, q and u may result in responses to selection in opposite directions being of the same sign.

More than the quantitative accuracy of the formulae, which breaks down as selection proceeds, it is their diagnostic value that may contribute to a better understanding of the properties of this model so widely used in genetics. The application of these methods to two-loci models will presumably yield some interesting results.

I thank BARTOLOME JODAR for the derivation of the relation of our formula and LATTER's, and ANTONIO BARRERA and JAVIER TEJIDO for the computer work. Two anonymous reviewers also made constructive comments.

LITERATURE CITED

COMSTOCK, R. E. and H. F. ROBINSON, 1948 The components of genetic variance in populations. Biometrics **4**: 254–266.

GALLEY, S. J. and R. N. CURNOW, 1972 The effects of finite population size and selection on the correlation between gene frequency changes at two different loci and on the amount of linkage disequilibrium. Theoret. Appl. Genet. **42**: 335–345.

HALDANE, J. B. S., 1931 A mathematical theory of natural and artificial selection. Part VII. Selection intensity as a function of mortality rate. Proc. Cambridge Phil. Soc. **27**: 131–136.

HILL, W. G., 1969a On the theory of artificial selection in finite populations. Genet. Res. **13**: 143–163. ——, 1969b The rate of selection advance for non-additive loci. Genet. Res. **13**: 165–173.

KIMURA, M., 1957 Some problems of stochastic processes in genetics. Ann. Math. Stat. **28**: 882–901. ——, 1962 On the probability of fixation of mutant genes in a population. Genetics **47**: 714–719.

KIMURA, M. and J. F. CROW, 1978 Effect of overall phenotypic selection on genetic change at individual loci. Proc. Natl. Acad. Sci. U.S. **75**: 6168–6171.

KOJIMA, K., 1961 Effects of dominance and size of population on response to mass selection. Genet. Res. **2**: 177–188.

LATTER, B. D. H., 1965 The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. Aust. J. Biol. Sci. **18**: 585–598.

NARAIN, P. and A. ROBERTSON, 1969 Limits and duration of response to selection in finite populations: The use of transition probability matrices. Ind. J. Heredity **1**: 1–19.

ROBERTSON, A., 1952 The effect of inbreeding on the variation due to recessive genes. Genetics **37**: 189–207. ——, 1960 A theory of limits in artificial selection. Proc. Roy. Soc. B. **153**: 234–249. ——, 1962 Selection for heterozygotes in small populations. Genetics **47**: 1291–1300. ——, 1963 Some comments on quantitative genetic theories. In: *Statistical Genetics and Plant Breeding.* Edited by W. D. HANSON and H. F. ROBINSON. Natl. Acad. Sci., Natl. Res. Council. ——, 1977 The non-linearity of offspring-parent regression. pp. 297–304. In: Proc. Intern. Conf. on Quantitative Genetics. Edited by E. POLLAK, O. KEMPTHORNE and T. B. BAILEY. The Iowa State University Press, Ames.

Corresponding editor: B. S. WEIR

APPENDIX

The degree of approximation of first- and second-order formulae is checked against expectations obtained with a transition probability matrix developed by HILL (1969a, b) and shown to be a suitable approximation for the transition matrix in which the selection process is exactly described. The matrix is obtained assuming that the gene frequencies after one generation of selection are binomially distributed around their mean. We use second-order approximations for the mean, and limiting values of the expectations are obtained using the matrix of transient states following NARAIN and ROBERTSON (1969).

Let us use subindices 1, 2 and h to differentiate expectations coming from first- and second-order formulae and HILL's transition matrix.

In Table 2 we give the number of generations t for which the differences $E[q]_{t,1} - E[q]_{t,h}$ and $E[q]_{t,2} - E[q]_{t,h}$ are smaller than $\varepsilon = 0.01$, assuming additivity and a 25% selection pressure.

First-order approximate formulae are valid for much longer when $q = 0.50$. So are the second-order approximate formulae for small q. For large Niu values our formulae are valid only for the first few generations. If $q = 1/2$, first- and second-order formulae are identical.

Expectations for second-order moments are obtained under the same assumption that u^2 is small enough to be ignored, through a similar moment generating matrix and their degree of approximation is of the same order of magnitude. A few particular cases of changes in the within-line variance are given in Table 3.

TABLE 2

Number of generations t *for which* $E[q]_{t,1}$ *and* $E[q]_{t,2}$ *(in brackets) lie within* $\varepsilon = 0.01$ *of* $E[q]_{t,h}$

Gene effect	Gene frequency								
	$q = 0.10$ Population size			$q = 0.20$ Population size			$q = 0.50$ Population size		
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$u = 0.05$	$\infty(\infty)$	11(∞)	9(>20)	$\infty(\infty)$	10(>20)	8(>20)	$\infty(\infty)$	>20(>20)	15(15)
$u = 0.10$	6(∞)	4(>20)	4(>20)	5(>20)	4(13)	5(11)	5(11)	7(7)	7(7)
$u = 0.15$	3(17)	3(15)	3(16)	3(9)	2(8)	2(8)	5(5)	5(5)	4(4)
$u = 0.25$	2(9)	1(10)	1(5)	2(4)	2(5)	2(4)	3(3)	2(2)	2(2)

TABLE 3

$E[q(1-q)]_{t,I}$ *and* $E[q(1-q)]_{t,h} \times 1000$ *for different parameter values (the expectations are indicated as* $E_{t,I}$ *and* $E_{t,h}$*)*

	Parameter values									
N	15		15		15		5		10	
u	0.10		0.10		0.10		0.10		0.05	
q	0.20		0.20		0.50		0.50		0.30	
a	0.00		0.00		0.00		0.00		1.00	
i	1.00		1.54		1.54		1.21		1.12	
Generations										
t	$E_{t,h}$	$E_{t,1}$	$E_{t,h}$	$E_{t,1}$	$E_{t,h}$	$E_{t,1}$	$E_{t,h}$	$E_{t,1}$	$E_{t,h}$	$E_{t,1}$
0	160	160	160	160	250	250	250	250	210	210
1	164	164	169	169	240	242	224	225	205	206
2	167	167	176	176	228	234	200	203	200	201
3	168	169	181	182	215	226	178	182	194	197
4	169	170	183	186	201	218	158	164	188	191
5	169	171	184	190	187	211	140	148	182	186
6	167	170	183	192	173	204	124	133	176	180
10	146	165	166	193	122	178	77	87	151	158
15	133	152	128	182	76	150	42	51	122	131

For the first set of parameter values, selection overrides the effects of inbreeding for the first few generations, and the heterozygosity, and thus the heritability, increases. Our formula would predict a maximum heterozygosity of 171 at generation 5, whereas actually a maximum 169 is attained at generation 4. The second set of parameters shows that a higher intensity of selection enhances this effect, and heterozygosity increases more and for a longer period of time. A maximum of 194 at generation 8 (not shown in the table) is predicted, while it turns out to be 184 at generation 5. The coincidence is not good, but the diagnostic value of our formula is clear.

In set 3, for initial gene frequency $q = 0.50$, the heterozygosity is maximum at the beginning and decreases more rapidly. The absolute errors of approximation are higher.

For small population sizes or smaller gene effects the agreement between expectations is much better as shown by the last two sets. As before, the degree of approximation deteriorates with t in all cases, more rapidly for high Niu values.