Change in Genetic Variance Under Selection in a Self-Fertilizing Population

T. Hayashi and Y. Ukai

Department of Agrobiology, Faculty of Agriculture, The University of Tokyo, Bunkyo, Tokyo 113, Japan

Manuscript received May 4, 1993
Accepted for publication October 25, 1993

ABSTRACT

In this study we show how the genetic variance of a quantitative trait changes in a self-fertilizing population under repeated cycles of truncation selection, with the analysis based on the infinitesimal model in which it is assumed that the trait is determined by an infinite number of unlinked loci without epistasis. The genetic variance is reduced not as a consequence of the genotypic frequency change but due to the build-up of linkage disequilibrium under truncation selection in this model. We assume that the order of the genotypic contribution from each locus is $n^{-1/2}$, where $n$ is the number of loci involved, and investigate the change in linkage disequilibrium resulting from selection and self-fertilization using genotypic frequency dynamics in order to analyze the change in the genetic variance. Our analysis gives recurrence relations of genetic variance among the succeeding generations for the three cases of gene action, i.e., purely additive action, pure dominance without additive effect and the presence of both additive effect and dominance, respectively. Numerical examples are also given as a check on the recurrence formulas.

Based on the infinitesimal model (Fisher 1918), where it is assumed that a quantitative trait is determined by an infinite number of unlinked loci with infinitesimal effects and without epistasis, Bulmer (1971, 1980) investigated the change in the genetic variance of a quantitative trait under selection in a random mating population with statistical regression arguments based on normal distribution theory. He showed that the genetic variance does not change as a consequence of the change in allelic and genotypic frequencies, but instead it is due to linkage disequilibrium induced by selection. Studies of the effects of selection on genetic parameters have been given by several other researchers using the same model (Sorensen and Hill 1983; Tallis 1987; Villanueva and Kennedy 1990; Gomez-Raya and Burns 1990). However, most of the reports were focused on a random mating population and only a few were on a self-fertilizing population. Pederson (1969a,b) reported the change in mean genotype value under repeated cycles of truncation selection in a self-fertilizing population assuming that phenotypic variance remains unchanged by selection, which is not true in practice. The change in genetic variance caused by repeated cycles of selection in a self-fertilizing population has not been much discussed using the infinitesimal model. Although a recurrence relationship for genetic variance between two successive generations under continued truncation selection was established for a random mating population by Bulmer (1971, 1980, p. 154), no similar recursion has been found for a self-fertilizing population. Since most of the major crops, such as rice and soybean, are self-fertilizing, elucidating the change in genetic variance in a self-fertilizing population under selection is important for breeding. In this report, we discuss the change in genetic variance under repeated cycles of truncation selection in a self-fertilizing population with an approach based on the infinitesimal model and using the diallelic case. Dominance effects can be dealt with as well as additive ones in our analysis.

In the conventional argument used in the analysis of a quantitative trait based on the infinitesimal model, the genotypic contribution from a locus, $g$, is assumed to be so small relative to the phenotypic variance, $V_p$, that quantities of order $(g/\sqrt{V_p})^2$ or smaller can be ignored. Instead, our analysis is based on an assumption that the order of genotypic value at each locus is $n^{-1/2}$, where $n$ is the number of loci involved. This assumption is equivalent to the conventional one and leads to the same results when we let $n \to \infty$ (see also Bulmer 1980, p. 150). Since genetic variance changes due to linkage disequilibrium built up by selection, the change in linkage disequilibrium resulting from selection and self-fertilization should be investigated. Our analysis reveals formulas representing the change in genetic variance under selection in a self-fertilizing population. For the succeeding generations derived from an $F_2$ generated by the cross of two inbred lines, recurrence relations for genetic variances are given under some conditions of gene action. Our main result is a recursion for the total genetic variance including both additive and dominance components, which is represented as

$$C(s, t) = C_0(s, t) - c(c - z) \sum_{k=0}^{r-1} C(k, s) C(k, t)/V_p(k),$$

where $C(s, t)$ and $C_0(s, t)$ are the genetic covariances.
between an individual at generation \( t \) and its ancestor at generation \( s \) \((s < t)\) under repeated cycles of selection and in the absence of selection, respectively, and \( V_p(k) \) is the phenotypic variance at generation \( k \). \( C^*(s, t) \) can be given as

\[
C^*(s, t) = \left( 1 - \frac{1}{2^{s+t+1}} \right) (A + D/2^{s+t+1}),
\]

where \( A \) and \( D \) are sums of square of the additive and dominance effects over loci, respectively. \( C(t, t) \) gives the genetic variance at generation \( t \) under selection, \( V_c(t) \). The validity of the recurrence formulae is checked by numerical examples.

**GENETIC VARIANCE**

We deal with a diploid self-fertilizing population with discrete nonoverlapping generations. We assume that the population is in linkage disequilibrium in the initial generation \((\text{generation } 0)\) and the succeeding generations are obtained by repeated cycles of truncation selection and self-fertilization. Selection starts at the initial generation and is operated each succeeding generation, in which individuals with the highest performance are selected as the parents producing the next generation by self-fertilization. Assume that the population size is so large that the effects of drift can be ignored. Consider a quantitative trait controlled by \( n \) unlinked loci with no epistasis, where \( n \) is assumed sufficiently \((\text{infinitely})\) large. It is assumed that both the genotypic and phenotypic values are normally distributed and the environmental effect is independent of the genotypic value. Furthermore we assume that departures of the genotypic value from normality under repeated cycles of self-fertilization and selection can be neglected. Normality would never be distorted by self-fertilization in the absence of selection, thus we should evaluate to what extent normality is maintained under selection to justify this assumption. Departures of the distribution of the genotypic value from normality can be measured by the standardized cumulants of the third or higher order such as skewness or kurtosis which are zero in normal distribution. Bulmer showed \((1980, \text{p. } 154)\) that when all the variance is additive variance, skewness and kurtosis induced by one cycle of truncation selection at the upper 5% truncation point are so small that departures from normality can be safely ignored. Approximate normality would be maintained in several early generations in the case of weak selection and in the presence of dominance and environmental variances to mask additive effects. It can be shown in *Appendix 1* that the third or higher order cumulants of the distribution of genotypic values are related to disequilibria of the same order, \( i.e., \) disequilibria among three or more loci. The assumption of normality implies that only disequilibria between two loci are concerned and those of the third or higher order can be neglected under selection. However, the validity of the assumption of normality of genotypic values should be checked by numerical simulations.

Let there be two alleles at each of \( n \) loci and alleles at the \( l \)th locus be denoted by \( B_l \) and \( b_l \), respectively \((l = 1, \ldots, n)\). Hereafter the three genotypes at the \( l \)th locus, \( B_lB_l, B_lb_l, \) and \( b_l b_l \), are referred to as genotype 1, 2 and 3, respectively. In the \( l \)th generation before selection, the frequency of genotype \( i \) \((i = 1, 2, 3)\) at locus \( l \) is denoted by \( p_i(t) \) and the joint frequency of genotype \( i \) and \( j \) at two loci \( l \) and \( m \), respectively, by \( p_i jlm(t) \). The corresponding genotypic frequencies measured after selection are denoted, adding an asterisk, by \( p_i^*(t) \) and \( p_i j^* jlm(t) \), respectively. In order to evaluate the effects of selection we need also to consider as a contrast another population derived from the same initial generation by \( t \) cycles of self-fertilization without selection. The genotypic frequencies are written for such a population as \( p_i^0(t) \) and \( p_i j^0 jlm(t) \), respectively, with superscript "0." It should be noted that \( p_i j^0 jlm(t) = p_i^0(t) p_j^0(t) \) holds since linkage equilibrium is maintained in a population which has never experienced selection. We assume that the genotypic value of genotype \( i \) at the \( l \)th locus, \( g_i l(t) \), is of order \( n^{-1} \). For any one individual, the total genotypic value, \( G \), is expressed as the sum of the genotypic contributions from each locus,

\[
G = \sum_{l=1}^{n} g_i l(t), \quad (1)
\]

where \( g_i l(t) \) is a random variable indicating the genotypic value at the \( l \)th locus and taking values of \( g_i l(t) \) with probability \( p_i l(t) \), \( p_i^* jlm(t) \), or \( p_i^0 jlm(t) \) corresponding to the situation concerned. For the population derived by \( t \) cycles of selection and self-fertilization from the initial generation, the genetic variance before selection, \( V_c(t) \), is represented as

\[
V_c(t) = \sum_{l=1}^{n} \text{Var}(g_i l(t)) + \sum_{l=1}^{n} \sum_{m=1}^{n} \text{Cov}(g_i l(t), g_j m(t)), \quad (2)
\]

\hspace{1cm} \text{(Bulmer 1971)}

where \( \text{Var}(g_i l(t)) \) and \( \text{Cov}(g_i l(t), g_j m(t)) \) are the variance of \( g_i l(t) \) and the covariance of \( g_i l(t) \) and \( g_j m(t) \neq m \) at generation \( t \), respectively. \( \text{Var}(g_i l(t)) \) is written as

\[
\text{Var}(g_i l(t)) = \sum_{i=1}^{s} p_i l(t)[g_i l(t) - \mu(t)]^2, \quad (3)
\]

where

\[
\mu(t) = \sum_{i=1}^{s} p_i l(t) g_i l(t), \quad (4)
\]

is the mean of \( g_i l(t) \) at generation \( t \). On the other hand, in the population undergoing no selection, the mean of \( g_i l(t) \) at generation \( t, \mu^0(t) \), is given as

\[
\mu^0(t) = \sum_{i=1}^{s} p_i^0 l(t) g_i l(t), \quad (5)
\]
As is shown later, the difference between genotypic frequencies under selection and those in the absence of selection is of order \( n^{-1/2} \). Thus,

\[
\rho_{il}(t) = \rho_{il}^0(t) + O(n^{-1/2}),
\]

where \( O(n^k) \) is the quantity of order \( n^k \) or smaller. By substituting (6) into (4)

\[
\mu_i(t) = \mu_i^0(t) + O(n^{-1}).
\]

By inserting (6) and (7) into (3), we obtain

\[
\text{Var}(g_t; t) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} [p_{ij}(t)p_{jm}(t) - \mu_i(t)\mu_j(t)]^2 + O(n^{-3/2}),
\]

thus

\[
\sum_{i=1}^{n} \text{Var}(g_t; t) = V_C^0(t) + O(n^{-1/2}),
\]

where \( V_C^0(t) = \sum_{i=1}^{n} \sum_{j=1}^{n-1} [p_{ij}(t)p_{jm}(t) - \mu_1(t)]^2 \) is the total genetic variance at generation \( t \) for the population undergoing no selection, since \( g_i \) and \( g_m \) are uncorrelated in the state of linkage equilibrium. In a population under selection, the covariance of \( g_i \) and \( g_m \) is not zero and \( \text{Cov}(g_i, g_m; t) \) is written as

\[
\text{Cov}(g_i, g_m; t) = \sum_{i=1}^{n} \sum_{j=1}^{n} [p_{ij}(t)p_{jm}(t) - \mu_1(t)]p_{ij}g_{ij}g_{jm} \tag{9}
\]

Thus, the covariance term in (11) consists of the sum of \( n(n-1) \) terms of \( g'gD_{lm}(t)g_m \), each of which is of order \( n^{-3} \), it would be sufficient to give an approximation for \( D_{lm}(t) \) up to order \( n^{-3/2} \) or smaller in the analysis. The error of order \( n^{-3/2} \) or smaller in \( D_{lm}(t) \) can be ignored in the sum, \( \sum_{i=1}^{n} \sum_{m=1}^{n} g'_iD_{lm}(t)g_m \).

**CHANGE IN LINKAGE DISEQUILIBRIUM**

We denote the zygotic linkage disequilibrium at generation \( t \) measured after selection by \( D_{ij/m}(t) \). Consider the relation of \( D_{ij/m}(t) \) and \( D_{ij/m}(t) \). At generation \( t \) prior to selection, the mean of individuals with genotype \( i \) at the \( l \)th locus is given as the sum of \( g_{il} \) and the contributions from the other loci, \( \Delta_{il}(t) \), which is induced by linkage disequilibrium. \( \Delta_{il}(t) \) is the conditional mean of the sum of the genotypic values of other loci given \( i \) as genotype at the \( l \)th locus and is represented with linkage disequilibrium as

\[
\Delta_{il}(t) = \sum_{k=1}^{n} \sum_{j=1}^{n} p_{ij}(t)p_{jk}g_{kj}/p_{il} + \sum_{i=1}^{n} \mu_i(t) \tag{12}
\]

Let

\[
\alpha_{il}(t) = g_{il} + \Delta_{il}(t) - \mu(t),
\]

where \( \mu(t) = \sum_{i=1}^{n} \mu_i(t) \) is the population mean of the genotypic value in generation \( t \). \( \alpha_{il}(t) \) is the deviation of the mean value of individuals with genotype \( i \) at the \( l \)th locus from the population mean and may be regarded as the average excess of the individuals. The fitness of the individuals can be related to \( \alpha_{il}(t) \). Assuming that the phenotypic value is normally distributed, the conditional distribution of the phenotypic value of individuals with genotype \( i \) at the \( l \)th locus is also normal, the mean of which is increased by an amount \( \alpha_{il}(t) \). Then a well known formula for the change in allelic frequencies can be rewritten in terms of genotypic frequencies [LATTER (1965), BULMER (1980, Equation 10.4) and see also HASTINGS (1990)]. Thus \( p_{il}(t) \) is related to \( p_{il}(t) \) as follows:

\[
p_{il}(t) = \left[ 1 + \frac{c\alpha_{il}(t)}{V_P(t)} \right] + \frac{cz[\alpha_{il}(t)^2 - E(\alpha_{il}(t)^2)]}{2V_P(t)} + O(n^{-3/2}) \tag{14}
\]

where \( V_P(t) \) is the phenotypic variance in generation \( t \), \( c \) is the selection intensity, i.e., the standardized deviation of the mean of the selected group from the population mean, \( \mu(t) \), \( z \) is the standardized deviation of truncation point from \( \mu(t) \) and

\[
E(\alpha_{il}(t)^2) = \sum_{i=1}^{n} p_{il}(t)\alpha_{il}(t)^2.
\]
Equation 14 shows that one cycle of selection causes genotypic frequencies to change by the order of $n^{-1/2}$. The difference between genotypic frequencies of the population under $t$ cycles of selection and those of the population in the absence of selection is also of order $n^{-1/2}$ if $t$ is much smaller than $\sqrt{n}$, which is assumed in our analysis. Thus (6) is justified.

Let the deviation of the mean value of individuals with genotype $i$ and $j$ at the $l$th and $m$th loci, respectively, from the overall population mean be denoted by $\alpha_{ij/lm}(t)$. Under the assumption of normality, it can be shown (APPENDIX 2) that

$$\alpha_{ij/lm}(t) = \alpha_{ij}(t) + \alpha_{im}(t) + O(n^{-3/2})$$

and replacing $p_{ij}(t)$ by $P_{ij}(t)$ and $\alpha_{ij}(t)$ by $\alpha_{ij/lm}(t)$ in (14),

$$P_{ij/lm}(t) = P_{ij}(t) \left\{ 1 + \frac{c(\alpha_{ij}(t) + \alpha_{im}(t))}{\sqrt{V_{ij}(t)}} + \frac{cz[(\alpha_{ij}(t) + \alpha_{im}(t))^2 - E((\alpha_{ij}(t) + \alpha_{im}(t))^2)]}{2V_{ij}(t)} + O(n^{-3/2}) \right\}$$

where

$$E((\alpha_{ij}(t) + \alpha_{im}(t))^2) = \sum_{i=1}^{3} \sum_{j=1}^{3} P_{ij/lm}(t)[\alpha_{ij}(t) + \alpha_{im}(t)]^2$$

Using (14) and (16) and from the definition of the zygotine linkage disequilibrium, it follows that

$$D_{ij/lm}(t) = P_{ij/lm}(t) - p_{ij}(t)p_{lm}(t)$$

$$= D_{ij/lm}(t) \left\{ 1 + \frac{c(\alpha_{ij}(t) + \alpha_{im}(t))}{\sqrt{V_{ij}(t)}} + \frac{cz[(\alpha_{ij}(t) + \alpha_{im}(t))^2 - E((\alpha_{ij}(t) + \alpha_{im}(t))^2)]}{2V_{ij}(t)} - c(c - z)p_{ij}(t)p_{ij}(t)\alpha_{ij}(t)\alpha_{im}(t)/V_{ij}(t) - czp_{ij}(t)p_{lm}(t)E(\alpha_{ij}(t)\alpha_{im}(t))/V_{ij}(t) + O(n^{-3/2}) \right\}$$

Since $D_{ij/lm}(0) = 0$, the order of $D_{ij/lm}(0)$ is shown to be $n^{-1}$ from (17). It can similarly be shown that $D_{ij/lm}(t)$ is of order $n^{-1}$ for any $t$ and, accordingly, the order of

$$E(\alpha_{ij}(t)\alpha_{im}(t)) = \sum_{i=1}^{3} \sum_{j=1}^{3} P_{ij/lm}(t)\alpha_{ij}(t)\alpha_{im}(t)$$

$$= \sum_{i=1}^{3} \sum_{j=1}^{3} D_{ij/lm}(t)\alpha_{ij}(t)\alpha_{im}(t)$$

is $n^{-2}$. Then (17) can be written as

$$D_{ij/lm}(t) = D_{ij/lm}(t) - c(c - z)p_{ij}(t)p_{ij}(t)\alpha_{ij}(t)\alpha_{im}(t)/V_{ij}(t) + O(n^{-3/2}).$$

From (18), the disequilibrium matrix after selection, $D_{ij/lm}(t) = (D_{ij/lm}(t))(i, j = 1, 2, 3)$ can be represented as

$$D_{ij/lm}(t) = D_{lm}(t) + c(c - z)d_{i}(t)d_{im}(t) + O(n^{-3/2})$$

where

$$d_{i}(t) = (p_{ij}(t)\alpha_{ij}(t), p_{ij}(t)\alpha_{im}(t), p_{ij}(t)\alpha_{im}(t))/\sqrt{V_{ij}(t)}.$$
Since (26) gives the relation of $\text{Cov}(g_i, g_m; t)$ and $\text{Cov}(g_i, g_m; t + 1)$ with (10) and (11) it is possible to obtain a recursion formula for the genetic variance, $V_G(t)$, under repeated cycles of selection.

CHANGE IN GENETIC VARIANCE

Using (11) and (26) and considering the order of each term, we express genetic variance at generation $t + 1$ as

$$V_G(t + 1) = V_G^0(t + 1) + \sum_{i = 1}^{n} \sum_{m = 1}^{n} g'_i D_{im}(t) g_m + O(n^{-1/2})$$

$$= V_G^0(t + 1) + \sum_{i = 1}^{n} \sum_{m = 1}^{n} g'_i T D_{im}(t) T' g_m$$

By substituting (28) and (29) into (27) and manipulating the resulting equation, we find that

$$V_G(t + 1) = V_G^0(t + 1) + \sum_{i = 1}^{n} \sum_{m = 1}^{n} g'_i D_{im}(t) g_m - c(\epsilon - z)[V_G(t)]^2/V_P(t) + O(n^{-1/2})$$

$$= V_G^0(t + 1) - V_G^0(t)$$

$$+[1 - c(\epsilon - z)V_G(t)/V_P(t)]V_G(t) + O(n^{-1/2}).$$

To emphasize additive effects without dominance, we use the notation, $V_A(t)$, for the genetic variance in place of $V_G(t)$. Neglecting the last term of the above equation, we find that

$$V_A(t + 1) = V_A^0(t + 1) - V_A^0(t)$$

$$+[1 - c(\epsilon - z)V_A(t)/V_P(t)]V_A(t).$$

Let $A = \sum_{i = 1}^{n} a_i$, then $V_A(t) = [1 - (\epsilon/2)^{t+1}]A$, which is the additive genetic variance of the population at the $t$th generation derived from the $F_2$ without selection. If $A$ and the environmental variance, $V_E$, are estimated, we can predict the genetic variance at any generation from (30). $V_A(t)$ can be also regarded as the additive component of the total genetic variance at generation $t$.

Case II. Purely overdominant gene action: When additive genetic effects are absent and there are only dominance effects for each locus, we can set $g'_i = (0, d_i, 0)$. Since $a_i = 0$, $p_{1/2}(t) = p_{3/4}(0)$, and $p_{1/2}(t) - p_{3/4}(t) = O(n^{-1/2})$ for $t \geq 1$, the additive genetic variance is sufficiently small to be ignored and the genetic variance is mainly the dominance variance. Gene action is overdominance and it is shown that

$$g'_i T = g'_i/2.$$  \hspace{1cm} (31)

We obtain a similar relation to (29),

$$\sum_{i = 1}^{n} g'_i T d_i(t) = V_G(t)/[2\sqrt{V_P(t)}].$$  \hspace{1cm} (32)

From (31) and (32), (27) is written as

$$V_G(t + 1) = V_G^0(t + 1) + \sum_{i = 1}^{n} \sum_{m = 1}^{n} g'_i D_{im}(t) g_m/4$$

$$- c(\epsilon - z)[V_G(t)]^2/[4V_P(t)]$$

$$+ O(n^{-1/2})$$

$$= V_G^0(t + 1) - V_G^0(t)$$

$$+[1 - c(\epsilon - z)V_G(t)/V_P(t)]V_G(t)/4 + O(n^{-1/2}).$$

Rewriting $V_G(t)$ as $V_B(t)$ in order to indicate that gene action is entirely dominant in this case and neglecting the last term in the above equation,

$$V_B(t + 1) = V_B^0(t + 1) - V_B^0(t)/4$$

$$+[1 - c(\epsilon - z)V_B(t)/V_P(t)]V_B(t)/4.$$  \hspace{1cm} (33)
In the population at the \( t \)th generation derived by \( t \) cycles of self-fertilization without selection from an \( F_z \),

\[
V_0^g(t) = (\frac{1}{2})^{t+1} [1 - (\frac{1}{2})^{t+1}] D,
\]

where \( D = \sum_{n=1}^{\infty} d^n \). Therefore, if \( D \) and \( V_c \) are given, we can also predict the genetic variance at any generation from (33) in this case.

**Case III. Gene action including both additive and dominance effects:** If genotypic values at some loci contain both additive and dominance effects, the recursion formula for the change in genetic variance is expressed as a more complicated form than in cases I and II. We denote the covariance of an individual in generation \( t \) and its ancestor in generation \( s (s \leq t) \) by \( C(s,t) \), where each generation is derived with self-fertilization and selection from an \( F_z \). In other words, \( C(s,t) \) is the covariance between an individual at generation \( t \) and its ancestor in generation \( s \) in the population at the \( t \)th generation derived by \( t \) cycles of self-fertilization without selection from an \( F_z \).

By the same argument which is used to derive (8) from (5), (6), and (7),

\[
\sum_{i=1}^{n} g_i^r T^{-i} (p\_1, s)[a_i - \mu(s)], p\_2(s)[d_i - \mu(s)],
\]

\[
p\_3(s)[-a_i - \mu(s)],
\]

\[
=p\_4(s)[d_i - \mu(s)],
\]

\[
=p\_5(s)[-a_i - \mu(s)],
\]

\[
= O(n^{-1/2}) + O(n^{-1/2})
\]

where \( C(s,t) \) is the covariance between an individual at generation \( t \) and its ancestor at generation \( s \) in the case of no selection. Therefore, from (34),

\[
C(s,t) = C^0(s,t) + \sum_{i=1}^{n} \sum_{m=1}^{n} g_i^r T^{-i} D_m(s) g_m + O(n^{-1/2}).
\]

By applying (19) and (25) repeatedly to \( D_m(s) \) in (35) and assuming selection intensity, \( c \), constant over all generations,

\[
C(s,t) = C^0(s,t) - c(e - z) \sum_{i=0}^{r-1} \left[ \sum_{k=0}^{n} g_i^r T^{-k} d(k) \right] V_r(k) + O(n^{-1/2}).
\]

From (34), we obtain

\[
\sum_{i=1}^{n} g_i^r T^{-i} d(k) = C(k, s)/\sqrt{V_r(k)}.
\]

By using (37) and neglecting the term of order \( n^{-1/2} \) or smaller, it follows from (36) that

\[
C(s,t) = C^0(s,t) - c(e - z) \sum_{k=0}^{r-1} C(k, s) C(k, t)/V_r(k).
\]

In the absence of selection it is well known that

\[
C^0(s,t) = (1 - \frac{1}{4} e^{-1})(A + D/2)^{t+1}
\]

(see BULMER 1980, p. 65). (39)

For any \( t \), if \( A, D \), and \( V_e \) are given \( V_e (t) = C(t, t) \) can be predicted from (38) and (39) in the following way. As the first step we can calculate \( C(0, s) = C^0(0, s) \) for each \( s (s = 0, 1, \ldots, t) \) by (39). At the second step for each \( s (s = 1, \ldots, t) \), by using (38), we obtain

\[
C(1, s) = C^0(1, s) - c(e - z) C(0, 0) C(0, s)/V_r(0).
\]

Accordingly \( C(1, s) \) can be given by \( C^0(1, s) \), \( C(0, 1) \), \( C(0, s) \) and \( V_r(0) = C(0, 0) + V_e \) determined in the first step. Furthermore, repeating the same procedure as the first and second steps we can get \( C(s, t) \) for each
contribute three cases of gene action corresponding to the last section, = Two alleles were denoted by determined by selection. We compared the predicted and observed intensity, , are given, the genetic variance at each generation can be predicted by (38) under continued phenotypic and genetic variance in the initial generation, . Therefore this recursive calculation gives setting of the assumptions in the infinitesimal model. The validity of the recurrence relations should be checked. Note that (30) and (33) are derived from (38) by . . ., . . ., which is given by (38). Therefore this recursive calculation gives . Note that (30) and (33) are derived from (38) by setting , and respectively. The size of the population is 200, which 100 individuals are selected each generation. The initial heritability is . The standard errors of the observed values over replicates ranged from 9.3% to 28.3% of the values.

NUMERICAL EXAMPLES

The recurrence relations of genetic variance, (30), (33) and (38), are given for a self-fertilizing population derived from the cross of two inbred lines on the basis of the assumptions in the infinitesimal model. The validity of the recurrence relations should be checked in practical situations. Since formula (38) contains (30) and (33) as special cases, we may use (38) alone to predict genetic variance at each generation. If the phenotypic and genetic variance in the initial generation, and , and the standardized selection intensity, , are given, the genetic variance at each generation can be predicted by (38) under continued selection. We compared the predicted and observed genetic variances in a simulated population at each generation. In simulations a quantitative trait was determined by 40 unlinked loci. An derived from the cross of two inbred lines was assumed as the initial generation and succeeding generations were produced by truncation selection and self-fertilization. Two alleles were denoted by and for each locus. Three genotypes, and , were assumed to contribute and to the trait, respectively. We treated three cases of gene action corresponding to the last section, , case I, ; case II, ; and III, . Cases I, II and III mean purely additive gene action, pure over-dominance, and complete dominance, respectively. Population size was kept at 200 before the operation of selection at each generation and the 100 individuals with highest performance were selected as the parents, each of which contributed two progeny to the next generation by self-fertilization. Thus, corresponding to this selection scheme, we obtain and .

### TABLE 1

Simulated values from 200 Monte Carlo replicates and predicted values (equation 38) for the genetic variance (, and for case I, II and III, respectively) with truncation selection

<table>
<thead>
<tr>
<th>Case</th>
<th>Generation</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>80.8</td>
<td>80.0</td>
<td>80.3</td>
</tr>
<tr>
<td>1</td>
<td>93.6</td>
<td>84.6</td>
<td>81.9</td>
<td>121.0</td>
</tr>
<tr>
<td>2</td>
<td>81.4</td>
<td>81.9</td>
<td>17.8</td>
<td>15.0</td>
</tr>
<tr>
<td>3</td>
<td>67.4</td>
<td>65.5</td>
<td>11.0</td>
<td>8.1</td>
</tr>
<tr>
<td>4</td>
<td>55.0</td>
<td>51.7</td>
<td>6.9</td>
<td>4.3</td>
</tr>
<tr>
<td>5</td>
<td>46.0</td>
<td>41.3</td>
<td>4.2</td>
<td>2.3</td>
</tr>
<tr>
<td>6</td>
<td>38.3</td>
<td>33.6</td>
<td>2.6</td>
<td>1.2</td>
</tr>
</tbody>
</table>

In each simulated model there are 40 unlinked loci each with two alleles and . The genotypic effects at each locus are , and corresponding to the three genotypes, , and , respectively. We set , and , respectively. The size of the population is 200. The initial heritability is . The initial heritability is . The standard errors of the observed values over replicates ranged from 9.3% to 28.3% of the values.

### TABLE 2

Simulated values from 200 Monte Carlo replicates and predicted values (equation 38) for the genetic variance with truncation selection

<table>
<thead>
<tr>
<th>Case</th>
<th>Generation</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>79.3</td>
<td>80.0</td>
<td>39.8</td>
<td>40.0</td>
</tr>
<tr>
<td>1</td>
<td>113.5</td>
<td>114.9</td>
<td>30.0</td>
<td>29.4</td>
</tr>
<tr>
<td>2</td>
<td>123.4</td>
<td>124.8</td>
<td>18.4</td>
<td>17.0</td>
</tr>
<tr>
<td>3</td>
<td>126.5</td>
<td>123.1</td>
<td>10.6</td>
<td>9.1</td>
</tr>
<tr>
<td>4</td>
<td>115.9</td>
<td>116.6</td>
<td>6.0</td>
<td>4.7</td>
</tr>
<tr>
<td>5</td>
<td>105.9</td>
<td>108.8</td>
<td>3.2</td>
<td>2.4</td>
</tr>
<tr>
<td>6</td>
<td>95.2</td>
<td>100.9</td>
<td>1.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>

The initial heritability is 0.1. The notation is the same as in Table 1. The standard errors of the observed values over replicates ranged from 9.3% to 28.3% of the values.

a Observed values.

b Predicted values.

and III mean purely additive gene action, pure over-dominance, and complete dominance, respectively.
between the predicted and observed genic variances through generations 0 to 6, although the differences between them are getting larger at later generations. In case II, however, the differences are much larger even in early generations (Table 3). This shows that the change in genotypic frequencies caused by selection in simulated populations is so large that it cannot be neglected and, accordingly, the observed genic variances are much changed by selection. The assumption of the infinitesimal model is invalid in case II and hence unsatisfactory accuracy of prediction of the genetic variance is obtained. Covariances between loci due to linkage disequilibria built up by selection are given as the difference of genetic and genic variances, i.e., \( V_2^0(t) \) and \( V_3^0(t) \) in Tables 1–3. It is noted that in case II covariances are much smaller than cases I and III, which means that almost no linkage disequilibrium is generated in the case of no additive effects. In all cases, when \( h^2(0) = 0.1 \), covariances between loci are smaller and the agreement between the predicted and observed genic variances is better than when \( h^2(0) = 0.5 \), as shown in Tables 1–3, because larger environmental variance masks genotypic values and make the effect of selection on the non-normality of genotypic values smaller when heritability is lower. Prediction formula (38) was obtained under the ideal condition that there is an infinite number of loci and normality is maintained under repeated selection. Results of simulations show that formula (38) is valid and can predict the behavior of genetic variances under selection even in actual situations at least for early generations.

Similar arguments used to investigate the change in genetic variances can also be applied to the change in mean values under selection, i.e., total cumulated selection response over several generations. Under the same assumptions as for the study of genetic variances, it can be shown from (4), (5), (14), and (23) that, after \( t \) cycles of self-fertilization and selection starting at \( F_2 \), the mean genotypic value, \( \mu(t) \), is expressed with the mean value after \( t \) cycles of self-fertilization without selection, \( \mu^0(t) \), as

\[
\mu(t) = \mu^0(t) + c \sum_{k=0}^{t-1} C(k, t) \sqrt{V_\mu(k)}.
\]

\( \mu^0(t) \) can be given as \( 80d/2^{2t+1} \) in simulations, therefore, \( \mu(t) \) can be predicted by (40). Mean genotypic values are also observed in simulations. Predicted and observed mean values under selection at generation 0–6 are shown with \( \mu^0(t) \) in Table 4 with \( h^2(0) = 0.5 \) and in Table 5 with \( h^2(0) = 0.1 \) for all cases. There is remarkably good agreement between predicted and observed means. When genetic effects are purely additive as in case I, mean values increase monotonically from generation 0–6 by selection. In case II where there are only dominance effects without additive effects, mean values are decreasing since heterozygous loci are reduced by self-fertilization, but the rate of decrease is slightly slower under selection which maintains heterozygosity. In case III where there are both additive and dominance effects, mean values decrease from generation 0 to 3 and then increase after heterozygosity is sufficiently reduced by self-fertilization. Selection response depends on additive effects and heritability. The presence of dominance has little effect on the cumulated response at generation 6, as shown in cases I and III of Tables 4 and 5.

### CONCLUDING REMARKS

Since selection response depends on genetic variabilities in a population, understanding how genetic variance changes under selection is important for an effective breeding program. When the genetic variance for succeeding generations is predicted, it is possible to find the optimal selection design where maximum genetic gain is achieved. We investigated
Genetic Variance of Selfing

### TABLE 4

<table>
<thead>
<tr>
<th>Generation</th>
<th>Case</th>
<th>Parameter</th>
<th>Parameter</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10.8</td>
<td>10.8</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15.9</td>
<td>15.9</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20.2</td>
<td>20.4</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>23.8</td>
<td>24.2</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>26.8</td>
<td>27.5</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

The initial heritability is 0.5. The standard errors of the observed values over replicates ranged from 0.6 to 2.2, 0.2 to 0.5, and 0.9 to 2.2 in cases I, II and III, respectively.

* Observed values.

* Predicted values.

### TABLE 5

<table>
<thead>
<tr>
<th>Generation</th>
<th>Case</th>
<th>Parameter</th>
<th>Parameter</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10.8</td>
<td>10.8</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15.9</td>
<td>15.9</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20.2</td>
<td>20.4</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>23.8</td>
<td>24.2</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>26.8</td>
<td>27.5</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

The initial heritability is 0.1. The standard errors of the observed values over replicates ranged from 0.6 to 2.4, 0.1 to 0.5, and 0.8 to 2.4 in cases I, II, and III, respectively.

* Observed values.

* Predicted values.

The change in the genotypic frequency and linkage disequilibrium on the assumption in the infinitesimal model that the order of the genotypic value at each of n loci involved is \( n^{-1} \), in order to evaluate the change in genetic variance under selection for a self-fertilizing population. Our analysis gave recursion formulas for genetic variance in the three cases, i.e., we derived (30) for additive genetic effect without dominance, (33) for dominance effect without additive effect, and (38) for the general case, where both additive and dominance effects are assumed at each locus. In general, however, we cannot deal with the additive and dominance components separately, since the additive effect, \( a_i \), is correlated with the dominance effect, \( d_i \), under selection, thus (38) shows the change in total genetic variance which contains the additive component, the dominance component, and the covariance between additive and dominance effects. We can see this in the above example. Let the predicted variances of case I, II and III be denoted by \( V_A(t) \), \( V_D(t) \) and \( V_C(t) \), respectively. The additive and dominance components of the genetic effect of case III correspond to the effects of cases I and II, respectively, so \( V_C(0) = V_A(0) + V_D(0) \) holds, as shown in Tables 1 and 2, since additive and dominance effects are uncorrelated at the initial generation. However, for \( t \geq 1 \), Tables 1 and 2 show that \( V_C(t) \) is no longer expressed as the sum of \( V_A(t) \) and \( V_D(t) \), because of correlation between additive and dominance effects induced by selection. The difference between \( V_C(t) \) and \( V_A(t) + V_D(t) \) can be regarded as the covariance between additive and dominance effects.

It is shown that in a self-fertilizing population a limiting value of genetic variance is always zero under continued truncation selection from (30), (33), and (38), while there exists a non-zero equilibrium value of the genetic variance in a random mating population. The following explanation can be given to this situation. In a self-fertilizing population the heterozygosity of individuals is halved after each cycle of
self-fertilization. All the loci of individuals will be homozygous at the limit after infinitely many cycles of self-fertilization, so that the population will consist of a set of clones of some homozygous individuals. After repeated cycles of truncation selection a clone of the homozygous individual with the best genotype will survive, and hence the genetic variance of the population will be reduced eventually to zero. In a random mating population, the heterozygosity of individuals remains constant under selection when there are an infinite number of loci, so that genetic variability will not vanish. With repeated cycles of truncation selection an equilibrium situation is eventually achieved in which the genetic variance lost by selection is recovered by free recombination of gametes, so that the non-zero limiting value of the variance is attained.

A formula describing the change in genetic variance under repeated cycles of selection has been given for a random mating population by BULMER (1971, 1980). Bulmer has analyzed the regression model between genotypic values of an individual and its ancestors by using normal distribution theory based on the infinitesimal model. This regression technique can also be applied to a self-fertilizing population to analyze the change in genetic variance. An interpretation can be regarded as the within-family variance, i.e., the variance of progeny means, which is affected by selection. For the case of pure dominance without additive effects, (38) can be interpreted similarly. For the general case where there are both additive and dominance effects, using the statistical regression arguments we find, instead of (38),

\[
V_G(t + 1) - V_G(t) = \frac{1}{V_G(t)} \left( C_G(t + 1) - C_G(t) \right) V_G(t).
\]

This formula, however, can also be derived by complicated manipulation of (38), thus it should be noted that (38) is a more general form. Normal distribution theory is based on the assumption of the infinitesimal model that traits are determined by an infinite number of unlinked loci and this model implies that the non-zero limiting value of the variance is attained. With repeated cycles of truncation selection an equilibrium situation is eventually achieved in which the genetic variance lost by selection is recovered by free recombination of gametes, so that the non-zero limiting value of the variance is attained.

We are grateful to William G. Hill and the anonymous referees for thorough reading and helpful comments. This work was supported by Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists.

LITERATURE CITED


 appendix 1

We consider the moments and cumulants of the distribution of genotypic value in the infinitesimal model when linkage disequilibria are induced by selection. From (1), the genotypic value of an individual, G, at generation t under selection is written as

\[
G = \sum_{i=1}^{n} g_i \mu(t) + \sum_{i=1}^{n} \frac{g_i - \mu(t)}{\gamma_i(t)} = \mu(t) + \sum_{i=1}^{n} \gamma_i(t),
\]

where we set \( \gamma_i(t) = g_i - \mu(t) \). The moment of order m of \( \gamma_i(t) \) is denoted by \( \gamma_m(t) \), which is the moment of \( \sum_{i=1}^{n} \gamma_i \). Noting that the orders of \( \gamma_i(t) \) and
\( D_{ij/lm}(t) \) are \( n^{-1/2} \) and \( n^{-1} \), respectively, we obtain \( \gamma_N(t) \) (\( N = 1, 2, 3 \)) as follows;

\[
\gamma_1(t) = E \left[ \sum_{i=1}^{n} y_i(t) \right] = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij}(t)y_{ij}(t) = 0, \quad (A1)
\]

\[
\gamma_2(t) = E \left[ \left( \sum_{i=1}^{n} y_i(t) \right)^2 \right] = \sum_{i=1}^{n} E[y_i(t)^2] + \sum_{i=1}^{n} \sum_{j=1}^{n} E[y_i(t)y_j(t)] = 0, \quad (A2)
\]

\[
\gamma_3(t) = E \left[ \left( \sum_{i=1}^{n} y_i(t) \right)^3 \right] = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} P_{ijk/m}(t)y_{ij}(t)y_{jk}(t)y_{ik}(t) + O(n^{-1/2}), \quad (A3)
\]

where \( P_{ijk/m}(t) \) is the frequency of individuals with three locus genotype \( i, j \) and \( k \) at loci \( l, m \) and \( h \) and \( D_{ijk/m}(t) \) is the third order disequilibrium, i.e. the disequilibrium among three loci \( i, j \) and \( k \), which is defined as

\[
D_{ijk/m}(t) = P_{ijk/m}(t) - p_{ij}(t)D_{jkl}(t) - p_{jk}(t)D_{ikl}(t) - p_{ki}(t)D_{jml}(t) - \mu(t), \quad (A4)
\]

In the infinitesimal model \( n \) is assumed infinitely large, so that the terms of order \( n^{-1/2} \) or smaller can be neglected. From (A1), (A2) and (A3), the third order cumulant of \( G \) at generation \( t \), \( \kappa_3(t) \), is given as

\[
\kappa_3(t) = \gamma_3(t) - 3\gamma_1(t)\gamma_2(t) + 2\gamma_1(t)^3 = \gamma_3(t), \quad (A5)
\]

In this study we assume that genotypic value remains normal under selection. The implication of this assumption is that \( \kappa_3(t) \) or higher order cumulants can be ignored. Therefore, from (A5), the third order disequilibria should be of order \( n^{-2} \) or less as the orders of disequilibria can be expressed as the form of \( n^{-k/2} \). Then, by the same argument as (A3), the fourth order cumulant of \( G \), \( \kappa_4(t) \), can be given as

\[
\kappa_4(t) = \gamma_4(t) - 3\gamma_2(t)^2 \quad (A6)
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{m=1}^{n} \sum_{h=1}^{n} \sum_{r=1}^{n} \sum_{s=1}^{n} D_{ijkl/mh}(t)y_{ij}(t)y_{jk}(t)y_{ki}(t)y_{rh}(t) - \mu(t),
\]

and

\[
kappa_4(t) = \gamma_4(t) - 3\gamma_2(t)^2 \quad (A6)
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{m=1}^{n} \sum_{h=1}^{n} \sum_{r=1}^{n} \sum_{s=1}^{n} D_{ijkl/mh}(t)y_{ij}(t)y_{jk}(t)y_{ki}(t)y_{rh}(t) - \mu(t),
\]

where \( D_{ijkl/mh}(t) \) is the fourth order disequilibrium, i.e., linkage disequilibrium among four loci, \( i, j, k \) and \( h \), and is defined as

\[
D_{ijkl/mh}(t) = P_{ijkl/mh}(t) - p_{ij}(t)D_{jkl}(t) - p_{jk}(t)D_{ikl}(t) - p_{ki}(t)D_{jml}(t) - \mu(t),
\]

where \( P_{ijkl/mh}(t) \) is the frequency of individuals with four locus genotype \( i, j, k \) and \( h \) at loci \( l, m, h \) and \( s \). Thus, in order that genotypic value is kept normal, the order of \( D_{ijkl/mh}(t) \) should be \( n^{-3/2} \) or less. But it can be shown that the orders of \( D_{ijkl/mh}(t) \) and \( D_{ijkl/mh}(t) \) are \( n^{-3/2} \) and \( n^{-2} \), respectively, assuming that the order of genotypic value is \( n^{-1/2} \). So we can conclude as follows. The normal distribution theory allows the second order disequilibria to be included, but is inconsistent with the third, fourth, or higher order disequilibria, which are necessarily induced by selection. The skewness and kurtosis of the genotypic value would hold approximately under repeated cycles of selection, if a more accurate prediction of genetic variance under selection is desired, another distribution than normal should be assumed for the genotypic value, which takes the third or higher order disequilibria into account.

**APPENDIX 2**

We give the proof of (15) in this appendix. \( \alpha_{ij/m}(t) \) is the deviation of the mean value of individuals with genotype \( i \) and \( j \) at loci \( l \) and \( m \), respectively, and written as

\[
\alpha_{ij/m}(t) = g_{i/l} + g_{j/m} + \sum_{k=1}^{n} \sum_{h=1}^{n} P_{ijkl/mh}(t)g_{k/h}/P_{ijkl/m}(t) - \mu(t), \quad (A7)
\]
Using the third order disequilibria, $D_{jk/m}(t)$, defined by (A4), noting that
\[ \sum_{k=1}^{3} p_{k/h}(t)[g_{k/h} - \mu_k(t)] = 0 \]
and
\[ \frac{1}{P_{ij/m}(t)} = 1/[p_{ij}(t)p_{j/m}(t) + D_{ij/m}(t)] = [1 + O(n^{-1})]/p_{ij}(t)p_{j/m}(t), \]
where we utilize the fact that the order of $D_{ij/m}(t)$ is $n^{-1}$, and from (12), (13) and (A7), it can be obtained that
\[ \alpha_{ij/m}(t) = g_{ij} - \mu_i(t) + g_{j/m} - \mu_m(t) + \sum_{k=1}^{n} \sum_{k=1}^{3} [D_{jk/m}(t)/p_{ij}(t)] (A8) \]
\[ + D_{jk/m}(t)/p_{j/m}(t) + D_{ijk/m}(t)/P_{ij/m}(t)g_{k/h} + O(n^{-3/2}) = g_{ij} - \mu_i(t) + \sum_{h=1}^{n} \sum_{k=1}^{3} D_{kh/m}(t)g_{k/h}/p_{j/k}(t) + g_{j/m} - \mu_m(t) \]
\[ + \sum_{h=1}^{n} \sum_{k=1}^{3} D_{jh/m}(t)g_{k/h}/p_{j/m}(t) + O(n^{-3/2}) \]
\[ + \sum_{h=1}^{n} \sum_{k=1}^{3} D_{jh/m}(t)g_{k/h}/P_{ij/m}(t) \]
\[ = \alpha_{ij}(t) + \alpha_{j/m}(t) + \sum_{h=1}^{n} \sum_{k=1}^{3} D_{jh/m}(t)g_{k/h}/P_{ij/m}(t) + O(n^{-3/2}). \]

As is shown in APPENDIX 1, the assumption that the genotypic value remains normally distributed under selection implies that the third order cumulant, $\kappa_3(t)$, can be neglected and hence, from (A5), the order of $D_{jk/m}(t)$ can be regarded as $n^{-2}$ or smaller. Accordingly, from (A8), we obtain (15), that is
\[ \alpha_{ij/m}(t) = \alpha_{ij}(t) + \alpha_{j/m}(t) + O(n^{-3/2}). \]