Theor Appl Genet (1988) 75: 333–339

Inbreeding effective population size under some artificial selection schemes

1. Linear distribution of breeding values

Hyun Kang[1] **and Gene Namkoong**[2]

[1] North Central Forest Experiment Station, U.S. Department of Agriculture-Forest Service and Department of Forestry, University of Wisconsin, Madison, WI 53706, USA
[2] Southeastern Forest Experiment Station, USDA-Forest Service and Genetics Department, North Carolina State University, Raleigh, NC 27695, USA

Received May 15, 1987; Accepted September 3, 1987
Communicated by P. M. A. Tigerstedt

Summary. It is well known that truncation selection is the most efficient form of directional selection in terms of changing gene frequency. In this paper we show circumstances where truncation selection followed by a balanced mating generates inbreeding effective population size smaller than that generated by a selection that assigns mating frequencies to individuals according to their breeding values, where both selection schemes give the same expected performance of selected individuals (selection differential). Breeding values of selected individuals and the weight used to determine mating frequencies are assumed to be linearly distributed on a performance scales, x. To assign mating frequencies to the individuals in the weighting system, the selected individuals are grouped using a constant δ, and i^{th} group in the interval x_i, $x_i + \delta$. With small number of groups, say 2 or 3, the weighting system in general generates inbreeding effective population size that is larger than that generated by a truncation selection. As the number of the groups increases, truncation selection generates larger effective numbers.

Key words: Inbreeding effective population size – Truncation selection – Weighting system

Introduction

Effective population size concept (Wright 1931) evolved as a means of relating observed numbers and population properties such as increase of homozygosis or the change in gene frequency due to small population size (Crow 1954; Kimura and Crow 1963). Instead of observed numbers, expected contribution of full-sib families or individuals to the gene pool of the progeny population can be used to determine the effective population size (Robertson 1961).

In most cases two well known population sizes, inbreeding and variance effective population size (Crow 1954), yield similar results. However, they deal with different reference populations. The inbreeding effective population size deals with the parent (or grandparent) population, while the variance effective population size is directly related to the number in the progeny population (Kimura and Crow 1963). This distinction is useful in deciding on the effective number to be used in different situations.

In selective breeding the determination of the variance effective number usually involves describing a complete cycle of selection with reference points at the time of reproduction. Therefore, the effective number is primarily useful in learning the impact of small breeding population size on multiple-generation selective breeding or studies of selection limits. Inbreeding effective population size does not require describing a complete cycle of selective breeding because reference is made to the parent population. Therefore, the number can be applied in multiplication activities following a selective breeding cycle as well as in a multiple-generation selection. For example, in forestry, seed orchards are widely used to produce seeds to be used in commercial planting programs. Inbreeding effective population size of the trees in the seed orchard can be useful guide to predicting the expected inbreeding level of the seeds collected. In this case, the actual size of the progeny population is not the critical component in determining the effective number. The inbreeding effective population size used in this paper deals with situations similar to the seed orchard example.

The term "selective breeding cycle" as used here represents a combined action of selection and mating the selected individuals. When developing selection techniques breeders frequently assume that random or structured random mating will follow, where structured random mating refers to artificially organized matings with random components (Harris et al. 1984). This assumption is equivalent to saying that the fertility of selected individuals is independent of the selected trait. Recently, Lindgren (1986) suggested assigning a mating frequency weight to an individual according to its performance or breeding value to further increase genetic gain. This approach amounts to a two-stage selection: truncation selection followed by fertility selection, where fertility is a function of the performance of breeding value. However, the genetic gain can be increased by simply increasing the selection differential in truncation selection followed by a balanced structured random mating. The issue, therefore is which selection-breeding combination generates greater gain.

Other authors have addressed an equivalent question by relating selection differential (I), selection coefficient (s), and average excess of a gene (a) (Milkman 1978, Kimura and Crow 1978, Crow and Kimura 1979). While discussing the relation $s = ia$ widely used by Griffing (1960), Milkman (1978) offered a qualitative explanation as to why truncation selection is the most efficient form of rank-order selection for fixed w, where w represents the mean fitness of the population after selection. He stated that "to modify truncation selection while keeping w constant, it is necessary to transfer some reproduction from individuals of a higher rank to those of lower rank". Crow and Kimura (1979) quantitatively showed that truncation selection is the most efficient form of directional selection in changing gene frequency for a given effect of the gene on the character. They showed that for a given 's/a' truncation selection yields smaller 'I' than other selections. This could be rephrased as saying that for a given 'I' and 'a', truncation selection yields larger s than other selection schemes.

Instead of s, it is possible to use inbreeding effective population size to compare different selection schemes. The objective of this paper is to present situations where truncation selection generates smaller inbreeding effective population size than other selective breeding schemes for a given selection differential.

Inbreeding effective population size

We will use Robertson's (1961) definition of effective population size,

$$Ne = \frac{(\Sigma\, u_i)^2}{\Sigma\, u_i^2}\,, \tag{1}$$

where u_i represents the expected contribution of i^{th} full-sib group or i^{th} individual. Although Robertson derived the formula as a variance effective number, Kimura and Crow (1963) showed that the same formula can be used for inbreeding effective number. They noted that the a priori probability that two randomly chosen gametes are produced by the same parent (1/Ne) is $\Sigma_i k_i^2/(\Sigma_i k_i)^2$ where k_i represents the expected size of a progeny group. If observed number h_i was used, the probability would be $\Sigma_i h_i (h_i - 1)/N h (N h - 1)$, where N represents the number of parents and h represents the mean family size. Resulting inbreeding effective numbers of monoecious species are:

$$Ne = \begin{cases} \dfrac{Nk}{k + V_k/k} & \text{using expected numbers,} \quad (2\,a) \\[2ex] \dfrac{Nh - 1}{h - 1 + V_h/h} & \text{using observed numbers,} \quad (2\,b) \end{cases}$$

where V represents the variance of the family size. Equation (2 b) is discussed by Kimura and Crow (1963). The most important property of Equation (2 a) is that Ne is always less than or equal to N. Equality is obtained when the progeny family size is constant ($V_k = 0$). On the other hand, when $V_h = 0$, Ne in Equation (2 b) becomes $2N - 1$ under constant population size over generation ($h = 2$) and approaches infinity as h approaches 1. Therefore, the properties of the inbreeding effective population size as used here are different from those of the traditional definition (2 b).

Expression (1) is obtained by replacing k_i in the definition of the a priori probability $[1/Ne = \Sigma_i k_i^2/(\Sigma_i k_i)^2]$ with $u_i T$, where T represents the expected total progeny number ($= \Sigma_i k_i$), which drops out in the expression.

Noting $\Sigma_i u_i = 1$,

$$Ne = 1/\Sigma_i u_i^2\,. \tag{3}$$

This definition of Ne, along with the property that $Ne < N$, implies that any selection scheme that generates differences in the expected size of progeny groups will reduce the inbreeding effective number, even if all N individuals are involved in producing the progeny population.

Model and assumption

Suppose N monoecious individuals are selected according to truncation selection, and S represents the expected performance of the selected individuals. If the breeder wishes to increase S by increasing the gametic contribution of the individuals with higher performance, then the inbreeding effective number (N_w) will be less than N, while the expected performance is increased, say from S to S*. Alternatively, the breeder could define a new truncation point that corresponds to S*. This will also result in a smaller number of selected individuals (N_t) than N. If the new set of individuals contributes equally to the following generation, then N_t becomes the inbreeding effective number under the new truncation selection. For breeders who are willing to accept a smaller Ne to achieve S^*, the question is whether the same S^* should be obtained by simply truncating at a lower N_t, or if the unbalanced mating of N individuals (resulting in N_w) can achieve equivalent results. Thus, if we can find $N_w > N_t$ for a given S^*, then there are unbalanced mating systems that are in this sense "better" than merely increasing the truncation level.

In this paper, the term "truncation selection" will be used to represent a truncation selection where all selected members have equal probability of contributing to the progeny population. The alternative form of selection, where the degree of contribution of selected parents to the progeny population varies according to their performance (or breeding value), will be referred to as the "weighting system".

Let f(x) represent the frequency distribution of the breeding value of individuals that fall between the initial truncation point m and a value M > m, where M is the largest breeding value of the sampled individuals, and $\int_m^M f(x)\,dx = C$. Without loss of generality, we can let $C = 1$. Also, let g(x) represent a function that determines the mating contribution of a selected individual with breeding value x to the progeny population.

We assume that the breeder may either not wish, or be able to estimate exactly an individual's true breeding value, x_j, but can assign an individual to a group in an interval of length δ. Within each group the individuals would have equal gametic representation, but the group's contribution would be determined by a weight q_i, which is a function of g(x). The interval length is determined by the breeder. We envision this model to be applicable to many actual breeding situations, specifically to designing seed orchards in forest tree breeding. We assume that the selected individuals are grouped using a constant δ, and the i^th group in the interval x_i, $x_i + \delta$, and;

$$p_i = \int_{x_i}^{x_i+\delta} f(t)\,dt, \quad q_i = \int_{x_i}^{x_i+\delta} g(t)\,dt,$$

$$\Sigma_i p_i = \Sigma_i q_i = 1, \quad \text{and} \quad N_i = N p_i,$$

where

p_i = the fraction of individuals falling between x_i and $x_i + \delta$,
q_i = the weight assigned to the individuals from i^th group,
N_i = the expected number of individuals in the i^th group,
x_i = m,
x_{K+1} = M,
δ = $(M - m)/K = d/K$, and
K = number of groups.

The weight q_i may be viewed as the frequency of sampling an individual from i^th group with replacement. Therefore, if a total of N^* individuals are to be sampled from the pool of N selected individuals, the expected number of samples from i^th group is $N^* q_i$. The probability that a particular individual to be sampled given i^th group is $1/N_i$. If the N^* samples contribute equally to the progeny gene pool, then the expected contribution of j^th individual from i^th group to the following generation is

$$u_{ij} = q_i/N_i\,.$$

Because individuals in the same group have the same expected contribution, $u_{ij} = u_{ik}$, we will use $N_i u_i$ in place of $\Sigma_j u_{ij}$.

By replacing u_{ij} in (3), the effective population size is,

$$\begin{aligned} N_w &= 1/\Sigma_{ij} u_{ij}^2 \\ &= 1/\Sigma_i N_i u_i^2 = N/\Sigma_i (q_i^2/p_i)\,. \end{aligned} \tag{4}$$

The expected performance of the new sample is

$$\begin{aligned} S^* &= \Sigma_i \bar{x}_i N_i u_i \\ &= \Sigma_i u_i q_i, \quad \text{where} \end{aligned} \tag{5}$$

$$\bar{x}_i = \int_{x_i}^{x_i+\delta} t f(t)\,dt\,.$$

Given S^*, we need to search for a new truncation point m^*, which satisfies

$$S^* = \int_{m^*}^{M} t f(t)\,dt, \quad \text{and} \quad m^* > m\,.$$

The expected number of individuals included in the selection with m^* as the truncation point is

$$N_t = N \int_{m^*}^{M} f(t)\,dt\,.$$

Because all N_t individuals contribute equally to the progeny population gene pool, N_t is the effective population size. Given N_t and N_w, we can determine conditions which satisfy

$$\theta = N_t/N_w < 1\,.$$

Where both f and g are linear functions of x

As an approximation of a truncated normal and other distributions, breeding values of the individuals in the parental population are assumed to be distributed according to a linear function f(x). f(x) is, however, defined between m and M, and x does not include any value less than the point corresponding to the apex of the distribu-

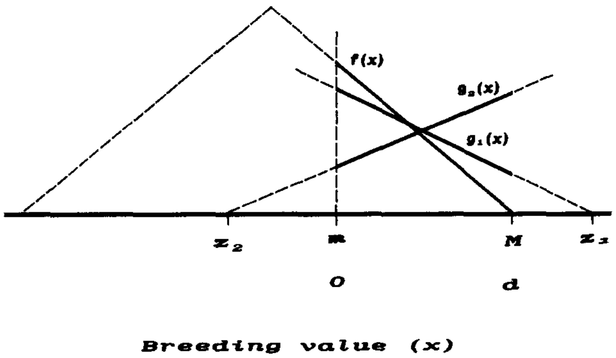

Fig. 1. f and g as linear functions of x, where x represents the performance scale. f(x) and g(x) are defined in [0, d] or [m, M]

tion (Fig. 1). To simplify the expression further without loss of generality, we will replace m and M with 0 and d, respectively. Then,

$$f(x) = 2(d - x)/d^2, \quad \text{and}$$

$$g(x) = 2(x - y\,d)/d^2(1 - 2y),$$

where $y = z/d$, and z is the intersection of g(x) with the x axis.

We also obtain

$$p_i = (2/d^2) \int_{x_i}^{x_i+\delta} (d - t)\,dt$$
$$= [2(K - i) + 1]/K^2$$

$$q_i = [2/d^2(1 - 2y)] \int_{x_i}^{x_i+\delta} (t - y\,d)\,dt$$
$$= (2i - v)/K^2(1 - 2y),$$

where

$v = 1 + 2Ky$, and $y < 1/2K$ or $y > (2K - 1)/2K$.

If $1/2K < y > (2K - 1)/2K$, then at least one $q_i < 0$, meaning that at least one group will not be included in the mating.

From (4), the effective population size under this scheme is (Appendix),

$$N_w = N/\Sigma_i(q_i^2/p_i)$$
$$= N/\Sigma_i[(2i - v)/K^2(1 - 2y)]^2/K^2[2(K - i) + 1]$$
$$= N/\{1 + 4(Q - 1)[(1 - y)/(1 - 2y)]^2\} \qquad (6)$$

where $Q = \Sigma_i 1/(2i - 1)$, which is a constant for given K.

From (5), the expected performance of the weighted samples is (Appendix),

$$S^* = \Sigma_i(2i - v)[(i - 1/2)\delta - \varepsilon_i]/K^2(1 - 2y)$$
$$= [d/(1 - 2y)][2/3 - y - (1 - y)Q/3K^2] \qquad (7)$$

where $\varepsilon_i = \bar{x}_i - (x_i + \delta/2)$.

Given S*, we can search for a new truncation point m*. When f(x) is a linear function of x,

$$S^* = m^* + d^*/3$$
$$= 2m^*/3 + d/3, \qquad (8)$$

where m* is the new truncation point and $d^* = M - m^*$.

Using (7) and (8),

$$m^* = 3[d/(1 + 2y)][2/3 - y - (1 - y)Q/3K^2]/2 - d/2$$
$$= R\,d/2(1 - 2y),$$

where $R = (1 - y)(K^2 - Q)/K^2$.

Given the new truncation point, we can determine the expected number of individuals (N_t) with values greater than or equal to x_m^*.

$$N_t = N[1 - (2\,d\,m^* - m^{*2})/d^2]$$
$$= N\{1 - R/(1 - 2y) + [R/(1 - 2y)]^2/4\}$$
$$= N[1 - R/2(1 - 2y)]^2$$
$$= N[1 - (1 - y)(K^2 - Q)/2K^2(1 - 2y)]^2.$$

In this case all the individuals have the same probability of contributing to the following generation, and N_t is the desired effective population size.

The final parameter of interest is the ratio between N_t and N_w.

$$\theta(K, y) = N_t/N_w$$
$$= [1 - (1 - y)(K^2 - Q)/2K^2(1 - 2y)]^2$$
$$\cdot \{1 + 4(Q - 1)[(1 - y)/(1 - 2y)]^2\} \qquad (9)$$

We will analyze Equation (9) by looking for a domain of y that results in $\theta < 1$ for given K. This can be done by determining the roots of $\theta - 1 = 0$, which is a quartic function of $s = (1 - y)/(1 - 2y)$. It is apparent that $s = 0$ (i.e. when $y = 1$) is a root of the equation. When $y = 1$ then $f(x) = g(x)$. The quartic equation $\theta - 1 = 0$ then reduces to a cubic equation

$$s^3 - (2/a)s^2 + [(a^2 + b)/a^2 b]s - 2/a\,b = 0, \qquad (10)$$

where $a = (k^2 - Q)/2K^2$ and $b = 4(Q - 1)$.

The three roots of Equation (10), in terms of y, are shown for K values between 2 and 20, and at 200 in Table 1. Figure 2 shows the range of y values where $\theta < 1$ for K values between 2 and 10. It is apparent from the results that θ is strongly influenced by the number of groups. When $K = 2$ or 3, $\theta < 1$ exists for both positive and negative values of y. When y is positive the intersection of g with x (i.e. z) is located on the positive side of the x axis (for example, $g_1(x)$ in Fig. 1). Starting $K = 4$, $\theta < 1$ exists only for positive y values. Starting $K = 5$, the domain of y becomes substantially smaller. Figure 3 shows θ between $y = 1$ and $y = 2$. It turns out that the minimum exists in this range for all K values, and the minimum θ of all the groups is 0.907 at $y = 1.873$ for $K = 2$.

Discussion

The above example clearly shows circumstances where truncation selection on a population with linearly distributed breeding values produces a smaller inbreeding effective population size than a linear weighting of mating frequencies combined with groupings. The surprising part, however, is that θ is highly dependent on the number of groups. Under the linear assumption of f(x) and g(x), and given y, the S* tends to be smaller when the number of groups is small. As the number of groups increases, S* eventually converges to $(2/3 - y)\,d/(1 - 2y)$. On the other hand, the effective population size (N_w) decreases as the number of groups increases. Therefore grouping itself can be viewed as a technique to be considered in breeding. However, the minimum value θ can take (0.907) is not overwhelming, and the result does not stand on its own as a new technique to be used. Its practical

Table 1. Roots, in terms of y, of $\theta - 1 = 0$. K = number of groups

K	Root 1	Root 2	Root 3
2	0.4215	−0.2965	0.2500
3	0.3982	−1.8263	0.1707
4	0.3874	12.2248	0.1746
5	0.3814	3.1383	0.1851
6	0.3777	2.2633	0.1946
7	0.3750	1.9317	0.2025
8	0.3731	1.7565	0.2089
9	0.3717	1.6475	0.2143
10	0.3705	1.5730	0.2188
11	0.3696	1.5186	0.2226
12	0.3689	1.4769	0.2260
13	0.3682	1.4440	0.2289
14	0.3677	1.4172	0.2315
15	0.3672	1.3949	0.2338
16	0.3667	1.3761	0.2359
17	0.3664	1.3599	0.2378
18	0.3660	1.3458	0.2395
19	0.3657	1.3335	0.2410
20	0.3655	1.3226	0.2425
200	0.3587	1.1295	0.2784

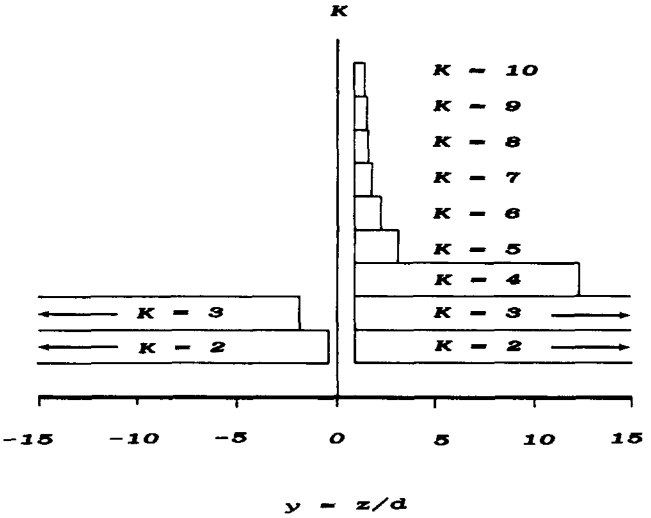

Fig. 2. Region of y where $\theta < 1$ for different number of groups (K)

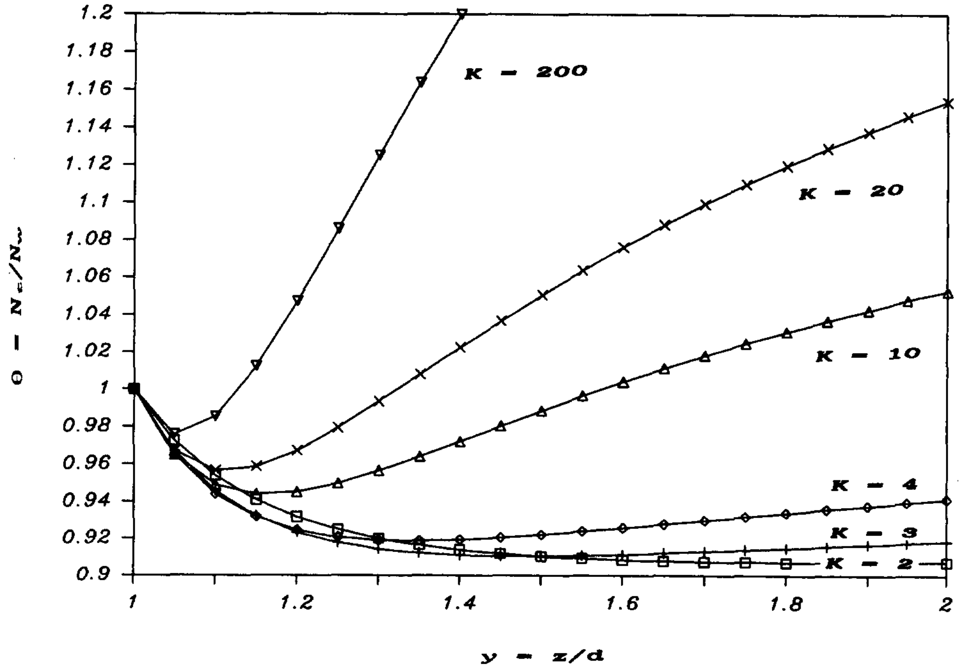

Fig. 3. θ between $y = 1$ and $y = 2$ for different number of groups (K)

value has to be judged in conjunction with other factors influencing the breeder's decisions.

The assumption that f(x) is a linear function of x is used to make this analysis possible. However, this analysis was made on individuals with extreme values, and a linear approximation could be as good as or better than normal approximation at that range.

We have numerically examined (Kang, in preparation) the model under normal distribution and different weighting schemes. The results are different than those found under a linear model, but the main conclusion that selection schemes other than truncation can generate a greater inbreeding effective population size for a given S* remains. We also found some other weighting schemes that could generate the ratio θ smaller than those found in this analysis.

No attempt has been made to relate the inbreeding effective population size found here with the variance effective population size of Robertson (1961).

The number strictly deals with parental population size without any regard to the progeny population size. When expectation instead of observed values is used, the variance and inbreeding effective size are identical (Kimura and Crow 1983). However, to translate the current finding to variance effective population size, the description of a complete cycle of selection is necessary.

Contrary to the strong generalized statement possible on the superiority of truncation selection in changing gene frequencies (Kimura and Crow 1978; Crow and Kimura 1979), the statement on the influence of different selection methods on inbreeding effective population size is much more restrictive. Under the linear distribution of f, we are able to say that with small number of groups, say 2 or 3, the linear weighting system in general generates inbreeding effective population size larger than that generated by a truncation selection that has the same S*. As the number of groups increases, truncation selection generates larger effective numbers. However, the finding that depending on the circumstance, alternatives more desirable than truncation selection exist, is significant because it necessitates the use of the condition under which truncation selection is the most efficient form of directional selection.

Appendix

Derivation of equation (6)

$$N_w = N/\Sigma_i (q_i^2/p_i)$$

$$\Sigma_i (q_i^2/p_i)$$

$$= \Sigma_i [(2i - v)/K^2(1 - 2y)]^2/\{1/[2(K - i) + 1]\}$$

$$= [1/(1 - 2y)^2]\,\Sigma_i (2i - v)^2/\{K^2[2(K - i) + 1]\}$$

By replacing i with K + 1 − i, we get

$$\Sigma_i (q_i^2/p_i)$$

$$= [1/(1 - 2y)^2]\,\Sigma_i [2K(1 - y) - (2i - 1)]^2/[K^2(2i - 1)]$$

$$= [1/(1 - 2y)^2][4(1 - y)^2 Q - 3 + 4y],$$

$$\text{where}\quad Q = \Sigma_i [1/(2i - 1)]$$

$$= [1/(1 - 2y)^2][(1 - 2y)^2 + 4(Q - 1)(1 - y)^2]$$

$$= 1 + 4(Q - 1)[(1 - y)/(1 - 2y)]^2$$

$$N_w = N/\{1 + 4(Q - 1)[(1 - y)/(1 - 2y)]^2\}.$$

Derivation of equation (7)

$$S^* = \Sigma_i q_i \bar{x}_i = \Sigma_i q_i [(i - 1/2)\delta - \varepsilon_i],$$

$$\text{where}\quad \varepsilon_i = \bar{x}_i - (x_i + \delta/2)$$

$$= \Sigma_i q_i (i - 1/2)\delta - \Sigma_i q_i \varepsilon_i$$

$$S^* + \Sigma_i q_i \varepsilon_i$$

$$= [\delta/K^2(1 - 2y)]\,\Sigma_i (2i - v)(i - 1/2)$$

$$= [\delta/K^2(1 - 2y)][2K(K + 1)(2K + 1)/6$$

$$- K(K + 1)(1 + v)/2 + vK/2]$$

$$= [\delta/6K^2(1 - 2y)][4K^3 + 3K^2(1 - v) - K]$$

$$\varepsilon_i = \bar{x}_i - (x_i + \delta/2) = \delta/6[2(K - i) + 1],$$

$$\Sigma_i q_i \varepsilon_i = \Sigma_i [\delta/6K^2(1 - 2y)](2i - v)/[2(K - i) + 1]$$

$$= [\delta/6K^2(1 - 2y)]\,\Sigma_i (2i - v)/[2(K - i) + 1].$$

By replacing i with K + 1 − i, and simplifying,

$$= [\delta/6K^2(1 - 2y)]\,\Sigma_i [2K(1 - y) - (2i - 1)]/(2i - 1)$$

$$= [\delta/6K^2(1 - 2y)][2K(1 - y)Q - K]$$

$$= [(1 - y)Q/3K(1 - 2y) - 1/6K(1 - 2y)]\,\delta$$

$$S^* = \{[4K^3 + 3K^2(1 - v) - K]/6K^2(1 - 2y)$$

$$- (1 - y)Q/3K(1 - 2y) + 1/6K(1 - 2y)\}\,\delta$$

$$= [d/(1 - 2y)][2/3 - y - (1 - y)Q/3K^2].$$

References

Crow JF (1954) Breeding structure of populations. II. Effective population number. In: Kempthorne O, Bancroft TA, Gowan JW, Lush JL (eds) Statistics and mathematics in biology. Iowa State College Press, pp 543–556

Crow JF, Kimura M (1979) Efficiency of truncation selection. Proc Natl Acad Sci USA 76:396–399

Griffing B (1960) Theoretical consequences of truncation selection based on the individual phenotype. Aust J Biol Sci 13:307–343

Harris DL, Rich SS, Wang C (1984) Effective breeding population size for structured random mating with random or directional selection. In: Hinkelmann K (ed) Experimental design, statistial models, and genetic statistics. Marcel Dekker, New York, pp 321–343

Kimura M, Crow JF (1963) The measurement of effective population number. Evolution 17:279–288

Kimura M, Crow JF (1978) Effect of overall phenotypic selection on genetic change at individual loci. Proc Natl Acad Sci USA 75:6168–6171

Lindgren D (1986) How should breeders respond to breeding values? Proc IUFRO Meeting Working Parties, Oct 12–18, 1986, Williamsburg VA

Milkman R (1978) Selection differentials and selection coefficients. Genetics 88:391–403

Robertson A (1961) Inbreeding in artificial selection programmes. Genet Res 2:189–194

Wright S (1931) Evolution in mendelian populations. Genetics 16:97–159