Comparison Of Genetic Models For Predicting Response To Selection On Long-term By Computer Simulation

M. Satoh[*], O. Sasaki[*], K. Ishii[*] and A. Nishiura[*]

Introduction

Predicting the response to selection and inbreeding is important to compare different strategies of selection with respect to genetic gain. Stochastic simulation is a useful predictive tool. The factors that affect response to selection can be roughly classified into two groups. One includes those associated with real-life selection, such as population size, selection rate, selection method, and mating system. The other includes gene and genetic effects such as genetic parameters, numbers of loci and alleles, genotypic values, gene frequencies, linkages, and mutations. Use of a genetic model that assumes finite population and finite loci effectively covers all of the considerable number of effects of these real-life factors and genetic assumptions. On the other hand, with the genetic model that assumes finite population and infinite loci it is easy to calculate responses to selection without using genetic assumptions. The objectives of this study were to compare genetic models used in simulations based of rates of genetic change under various conditions and to assess the characteristics of these models.

Material and methods

Four genetic models for predicting long-term response to selection were compared on the basis of population mean, genetic variance, and skewness with different population sizes and numbers of loci. Heritability of the selected trait in the base population was 0.5 and the selection rate was 30% (10% for males and 50% for females). Selection was based on the phenotypic value of the individual, and animals were mated only at random within a generation. For the sake of simplicity, our presentation was restricted to the selected trait under completely additive genetic control with two alleles and an initial gene frequency of 0.5 at all loci. Unlinked loci contributing equally to the variance of the selected trait were considered.

Model assuming finite population and finite loci. This model can treat various real-life factors such as population size, selection rate, selection method, mating system, and genetic assumptions such as genetic parameters, numbers of loci and alleles, genotypic values, and gene frequencies, linkages, and mutations.

Model assuming finite population and infinite loci. Consider now a large number of unlinked loci contributing equally to the variance of the selected trait. If u_i is the breeding value of animal i, then

$$u_i = \frac{u_s + u_d}{2} + \varepsilon_i$$

---

[*] National Institute of Livestock and Grassland Science, Tsukuba 305-0901, Japan

where u_s and u_d are the breeding values of the sire and dam, respectively, and ε_i represents Mendelian sampling. If both parents of animal i are known, then

$$a_{ii}\sigma_a^2 = \frac{1}{4}[a_{ss}\sigma_a^2 + a_{dd}\sigma_a^2 + 2a_{sd}\sigma_a^2] - \mathrm{var}(\varepsilon_i)$$

where a_{ii}, a_{ss}, a_{dd}, and a_{sd} are elements of the additive relationship matrix and σ_a^2 is the additive genetic variance. Since $F_i = 0.5a_{sd}$, then

$$\varepsilon_i \sim \left(0, \left(\frac{1}{2} - \frac{F_s + F_d}{4}\right)\sigma_a^2\right)$$

where F_s and F_d are the inbreeding coefficients of the sire and dam, respectively. The phenotypic values were generated as

$$y_i = u_i + e_i,\ e_i \sim N(0, \sigma_e^2)$$

where y_i is the phenotypic value of the i^th animal, e_i is the environmental deviation, and σ_e^2 is the environmental variance.

Model assuming infinite population and finite loci. Consider N unlinked loci with two alleles contributing equally to the variance of the selected trait and m favorable genes with the same frequency (q). When the number of homozygous loci for favorable genes is x, the joint probability of m and x with N loci is given as:

$$p(m, x) = p(m)p(x|m) = {}_{2N}C_m q^m (1 - q)^{2N-m}\ {}_{N}C_x \cdot {}_{N-x}C_{m-2x} \cdot 2^{m-2x}/{}_{2N}C_m \qquad (1)$$

where p(m) is the probability of m favorable genes and p(x|m) is the conditional probability of x, given m favorable genes (Takeda et al., 1988). The base population is constructed from equation (1). The total genetic value of the i^th animal is:

$$g_i = (m - N)a + (m - 2x)d$$

where +a and –a are the genotypic values of one and the other homozygotes, and d is the genotypic value of the heterozygote. When we think of the genotype conferring a certain value on the i^th animal and the environment causing a deviation from this, then symbolically,

$$y_i = g_i + e_i,\ e_i \sim N(0, \sigma_e^2). \qquad (2)$$

Because the probability of truncation selection based on phenotypic value is $p(S \le y_i)$, using the above equation we can obtain the following equation:

$$p(S \le y_i) = p((S - g_i)/\sigma_e \le \varepsilon) \qquad (3)$$

where ε is a random variable having the standard normal distribution. We can obtain the value of S by solving the following equation:

$$\sum_i p(g_i) \int_{S-g_i}^{\infty} f(\varepsilon)d\varepsilon = \int_S^{\infty} f(\varepsilon)d\varepsilon$$

where $f(\varepsilon)$ is the density function of the standard normal distribution. The probability of there being n favorable genes in a gamete is given by:

$$p_g(n) = \sum\sum {}_{m-2x}C_{n-1}\,(1/2)^{m-2x} p_s(m, x) \qquad (4)$$

where $p_s(m, x)$ is the joint probability of m and x in the selected animals. After the coupling of gametes,

$$p(m) = \sum_{i=0}^{m} p_{g_m}(i) \cdot p_{g_f}(m - i) \qquad (5)$$

where $p_{g_m}(n)$ and $p_{g_f}(n)$ are the probabilities of there being n favorable genes in the male and female gametes. Repeat of the procedures from (2) to (5) can predict the response to long-term selection. This makes it possible to expand the model to include loci for different genotypic values.

Model assuming infinite population and infinite loci. The expected response to selection, $E(\Delta g)$, is given by the classical formula:

$$E(\Delta g) = i r_{a\hat{a}} \sigma_a$$

where i is the intensity of selection, $r_{a\hat{a}}$ is the correlation between the selection criterion and the true breeding value, and σ_a is the additive genetic standard deviation. Unfortunately, for many reasons this prediction is not always correct. Especially if the population size is limited, modifications are necessary (Verrier et al., 1991).

Results and discussion

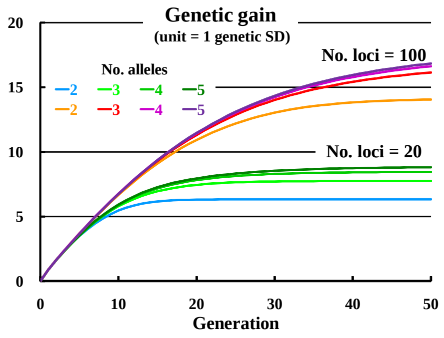
Figure 1 shows changes in the generation mean and genetic variance with various population sizes when the number of loci is 100. When the population size was 10 males and 50 females or more, the population reached its selection limit between about generations 40 and 50. When the population size was 5 males and 25 females or fewer, some alleles reached fixation due to random drift before reaching their maximum selection limit. The observed evolution clearly showed an initial decrease due to selection (the so-called Bulmer effect).

Figure 2 shows changes in the generation mean and genetic variance with various numbers of loci when the population size was 10 males and 50 females per generation. When the numbers of loci were assumed to be 50 and 100, the population reached its selection limit at about generations 35 and 50, respectively. When the number of loci was more than 500, the population did not reach the selection limit before the selection of 100 generations.

Table 1: Characteristics of genetic models assuming finite or infinite population and loci

Population size	Number of loci	
	Finite	Infinite
Finite	· Possible to treat various conditions in real-life situations · No criteria for genetic assumptions such as numbers of loci and alleles, genotypic value, etc.	· Suitable for medium-term evaluation · Overestimates genetic change and genetic variance under bias from true genetic parameters
Infinite	· Suitable for long-term evaluation · Unsuitable under more complex situations such as BLUP selection	· Easy to predict response to selection · Unsuitable for repeated selection without some modifications

Table 1 summarizes the characteristics of the genetic models assessed here. The model that assumes finite population and infinite loci is often used for predictions in long-term selection in animal breeding. However, the model that assumes infinite population and finite loci was suitable for most medium-sized and large breeding livestock herds. This model is suitable for deterministic simulations and is easily used to calculate for long-term selection (e.g. over 10,000 generations), but it is not possible to use it to evaluate responses to selection by complex methods such as BLUP. On the other hand, the model that assumes finite population and infinite loci overestimated genetic variance and responses to selection if the number of loci was less than the real number of loci.


Genetic gain
(unit = 1 genetic SD)
No. loci = 100
No. alleles
2
3
4
5
2
3
4
5
No. loci = 20
20
15
10
5
0
0
10
20
30
40
50
Generation