

Asymptotic Genetic Response Due to Two-Stage Genomic Index Selection Derived from Correlated or Independent Indices during Both Stages

Kenji Togashi^{1,*}, Kazunori Adachi², Kazuhito Kurogi², Toshio Watanabe¹, Syohei Toda¹, Takanori Yasumori³, Satoshi Matsubara⁴, Kiyohide Hirohama², Tsutomu Takahashi², Shoichi Matsuo², and Toshikazu Ijichi²

¹Maebashi Institute of Animal Science, Livestock Improvement Association of Japan, 316 Kanamaru-machi, Maebashi, Gunma, 371-0121, Japan

²Livestock Improvement Association of Japan, 11-17 Fuyuki, Koto-ku, Tokyo 135-0041, Japan

³Maebashi Bull Center, Livestock Improvement Association of Japan, Maebashi, Gunma, 371-0121, Japan

⁴Tokachi Bull Center, Livestock Improvement Association of Japan, 173-8, Sarubetsu, Makubetsucho, Nakagawagun, Hokkaido, 089-0625, Japan

*Corresponding author: Kenji Togashi, Maebashi Institute of Animal Science, Livestock Improvement Association of Japan, 316 Kanamaru-machi, Maebashi, Gunma, 371-0121, Japan, Tel: +81-0272692440; Fax: +81-0272692440; E-mail: k-togashi@liaj.or.jp

Received: 06 Aug, 2021 | Accepted: 30 Oct, 2021 | Published: 15 Nov, 2021

Citation: Togashi K, Adachi K, Kurogi K, Watanabe T, Toda S, et al. (2021) Asymptotic Genetic Response Due to Two-Stage Genomic Index Selection Derived from Correlated or Independent Indices During Both Stages. *J Anim Sci Res* 5(2): dx.doi.org/10.16966/2576-6457.154

Copyright: © 2021 Togashi K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

We developed equations to predict the asymptotic response due to two-stage selection, where first-stage selection was performed by using GEBVs based solely on genotypes and second-stage selection was performed with GEBVs that combined genotypes and phenotypes. The situation that we considered involved four-path selection executed as sires to breed sons, sires to breed daughters, dams to breed sons, and dams to breed daughters. We established two procedures to predict the response. The first incorporated correlated indices during the first and second-stage selections of two-stage selection. The other procedure used independent indices during two-stage selection. The response per generation in the initial generation was greater for the correlated indices than for the independent indices. However, the asymptotic response per generation was slightly greater for the independent indices than for the correlated indices. The asymptotic response per generation was lower during two-stage selection than during single-stage selection. However, the asymptotic response per year was greater for two-stage selection than for single-stage selection. In addition, that trend was more conspicuous when the economic weight was 1:3 for the first ($h^2=0.3$) to second ($h^2=0.05$) index trait compared with economic weights of 1:1 and 3:1. However, the magnitude of the response to the aggregate genotype relative to single-stage selection but to absolute magnitude was greater at an economic weight of 3:1 than at those of 1:1 and 1:3. The reduction in genetic variance from the initial to an asymptotic generation was greater for a scenario where young parents selected at the first-stage accounted for 30% of all parents in two-stage selection than where they accounted for 70%. The reduction in genetic variance of the aggregate genotype over generations was smaller for independent indices than correlated indices during two-stage selection. Our new formula for predicting genetic response applies to any combination of accuracies of GEBVs and intensities of selection. Therefore, the formula presented is a general equation for predicting genetic response over generations due to two-stage genomic index selection.

Keywords: Two-stage selection; Asymptotic selection response; Genomically enhanced breeding value; Economic weight; Aggregate genotype

Introduction

In livestock breeding, the selection strategy chosen enables the performance of the offspring to be predicted. In turn, future selection strategies can be optimized in light of current performance predictions. In other words, predicting genetic gain is not only about predicting future performance but also about evaluating breeding decisions before they are put into practice. In an open dairy cattle breeding program, a proportion of the cattle are owned by just a few breeders or a breeding company. These animals can be treated as a closed population from which are selected the sires and the dams of sires for the next generation. Thus, this situation provides a reference point for understanding the results of more complex open breeding programs that account for genotype \times environment interaction, the migration of animals into and from the target population, and genetic mutation.

Animal breeders often assume that any selection-induced changes of genetic and phenotypic parameters are negligible. However, in the long run, selection ultimately minimizes variability in genetic parameters. In particular, index selection will not only reduce the genetic (co)variance of index traits but also change the phenotypic (co)variances. The selection-associated changes in the parameter estimates may justify re-prediction of the response. Traditional analysis of covariance between relatives or pedigrees is unlikely to be sufficiently sensitive to detect small changes in genetic variability. However, assessing covariance according to a marker-based genomic relationship matrix [1] is more sensitive than the traditional analysis. Therefore, an equation that accurately predicts genetic response over time has enormous practical importance for producers and should be dynamically reconstructed from generation to generation.

Maximum genetic gain for a particular breeding goal can be accomplished by using a single-stage selection index in which all candidates were measured for all component traits before selection. However, the main advantage of multistage over single-stage selection is cost savings. The profitability of breeding operations can be improved by reducing the cost of measuring traits expressed later in life, which is the cost of maintaining the animal for a longer period of time in single-stage selection. Cotterill PP and James JW [2] presented a general approximation of genetic gain after two stages of selection under the assumption that the distribution of traits after stage-one selection remains roughly normally distributed, and Cerón-Rojas JJ, et al.[3,4] have presented an optimum multistage linear phenotypic selection index that reflects this assumption. In addition, Cerón-Rojas JJ, et al. [3,4] have created a decorrelated multistage linear phenotypic selection index-that is, one that uses independent indices for selection during a generation to maximize the breeding goal. Xu SZ and Muir WM [5,6] developed a fairly general approach that is based on transforming the indices during a generation to a set of orthogonal values that is, independent selection indices. Togashi K, et al. [7,8] developed formulas for calculating the asymptotic response from single-stage index selection for differential selection among male and female parents and of parents to produce male *versus* female replacements (namely, four-path selection programs). The index traits were based on genomically enhanced breeding values (GEBVs) computed from multiple-trait best linear unbiased prediction (MT-BLUP). The application of two-stage index selection to an index composed of GEBVs or to an asymptotic response has not previously been reported.

The main purpose of our current study was to develop formulas for predicting deterministically the genetic response due to two-stage selection composed of GEBVs from generation to generation until an asymptotic genetic response was reached. One formula incorporates the assumption that the traits in stages subsequent to the first round of selection remain roughly normally distributed [2], whereas the other formula uses independent indices during first- and second-stage selection [3,4]. Our secondary aims were to apply the formulas to several scenarios and to compare the asymptotic responses due to two-stage selection with those of single-stage selection, focusing on differences in economic weights, the selection percentage of young parents to all parents to produce the next generation, and the heritability of the component traits of the index.

Methods

Two-stage selection under assumption of normal distribution

Here, we attempt to provide a simple and practical method for determining gain from two-stage selection in the same way as Cotterill PP and James JW [2], who assumed that traits after stage one selection remain roughly normally distributed.

Suppose that stage one selects on an index of several traits composed of GEBV (x) for all traits with no phenotypes (that is, a bull's GEBV with no progeny [with no records] or a heifer's GEBV with no records); that stage two selects on an index of several traits composed of GEBVs (y), which incorporates the milk production records of a bull's daughters and a dam's own lactation records, in addition to the genomic information available at the first stage; and that the goal is the optimal improvement of the breeding value for merit-that is, the "aggregate genotype (H)"[9]. To demonstrate the application of our formulas, the milk production records at the second stage are assumed to be generated only from daughters of a test bull; obtaining the milk

production records from all daughters of the sons of a test bull is prohibitively time-intensive. However, all available data with regard to genomic and phenotypic information at the first- and second-stage selections would be included during actual selection using single-step genomic BLUP or ssGBLUP [10-12].

In the same way as did Cotterill PP and James JW [2], we assumed that (x; y; H) is initially trivariate normal and that the distribution of (y; H) remains roughly normal after stage one of selection on x. Let the selection index at the first-stage, comprising GEBVs without records, and the index at the second stage, composed of GEBVs incorporating records from progeny (for sires) or an individual (for dams), be I_y and I_{pr} , respectively. Let the reliability of GEBV for an index trait i of I_y and that of the estimated genetic value (\hat{G}) based on progeny (for sires) or individual (for dams) records without genomic information be r_{GEBV_i, I_y}^2 and $r_{i, progeny or individual records}^2$, respectively. The approximated reliability of GEBV for an index trait i ($r_{GEBV_i, I_{pr}}^2$) of I_{pr} at the second stage in which the reliabilities of GEBVs are combined for the first and second stages can be calculated according to the methods of Buch LH, et al.[13] and Togashi K, et al. [14] as:

$$\begin{bmatrix} r_{GEBV_i, I_y}^2 & r_{GEBV_i, I_y, r_{i, progeny or individual records}^2} \\ \text{symmetric} & r_{i, progeny or individual records}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} r_{GEBV_i, I_y}^2 \\ r_{i, progeny or individual records}^2 \end{bmatrix}$$

$$r_{GEBV_i, I_{pr}}^2 = b_1 r_{GEBV_i, I_y}^2 + b_2 r_{i, progeny or individual records}^2.$$

Therefore, the reliability of GEBV at the second stage ($r_{GEBV_i, I_{pr}}^2$) is greater than that at the first stage (r_{GEBV_i, I_y}^2) due to the combination of genomic and phenotypic information. Here, we used the approximated reliability of GEBV; however, the actual model-based reliability would be obtained by using all data for phenotypic records, and pedigree and genomic relationships in ssGBLUP.

The reliability of progeny testing without genomic information for trait i ($r_{i, progeny records}^2$) is computed from

$$r_{i, progeny records}^2 = \frac{0.25nh^2}{1 + 0.25(n-1)h^2}$$

where n is the number of progeny per sire, which we set at 25. The reliability for a dam's second selection on the basis of three of her own records without genomic information ($r_{i, individual records}^2$) is computed from

$$r_{i, individual records}^2 = \frac{nh^2}{1 + (n-1)\text{repeatability}}$$

where n is the number of the dam's own records with the assumption of repeatability= $h^2+0.1$. Covariance components of GEBV for two traits $\sigma_{\hat{g}_{12}}$ are derived by assuming that sufficient data are used to estimate marker effects [7,8,15]; that is, covariance components of GEBV ($\sigma_{\hat{g}_{12}}$) in I_{pr} ,

$$\sigma_{\hat{g}_{12}} = r_{GEBV_1, I_{pr}}^2 r_{GEBV_2, I_{pr}}^2 \sigma_{g_{12}}$$

where $\sigma_{g_{12}}$ is the genetic covariance between traits 1 and 2.

Therefore, when information for m traits is available for selection:

$$Cov(GEBV_{I_{pr}})_{m \times m} = \begin{bmatrix} r_{GEBV1,I_{pr}}^2 \sigma_{G_1}^2 & r_{GEBV1,I_{pr}}^2 r_{GEBV2,I_{pr}}^2 \sigma_{G_{12}} & \dots & r_{GEBV1,I_{pr}}^2 r_{GEBVm,I_{pr}}^2 \sigma_{G_{1m}} \\ \cdot & r_{GEBV2,I_{pr}}^2 \sigma_{G_2}^2 & \dots & r_{GEBV2,I_{pr}}^2 r_{GEBVm,I_{pr}}^2 \sigma_{G_{2m}} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ Symmetry & \cdot & \dots & r_{GEBVm,I_{pr}}^2 \sigma_{G_m}^2 \end{bmatrix}, (1)$$

where $GEBV_{I_{pr}}$ is a vector of m known GEBVs in I_{pr} .

Similarly to $cov(GEBV_{I_{pr}})_{m \times m}$, a covariance matrix of GEBV in I_y is:

$$Cov(GEBV_{I_y})_{m \times m} = \begin{bmatrix} r_{GEBV1,I_y}^2 \sigma_{G_1}^2 & r_{GEBV1,I_y}^2 r_{GEBV2,I_y}^2 \sigma_{G_{12}} & \dots & r_{GEBV1,I_y}^2 r_{GEBVm,I_y}^2 \sigma_{G_{1m}} \\ \cdot & r_{GEBV2,I_y}^2 \sigma_{G_2}^2 & \dots & r_{GEBV2,I_y}^2 r_{GEBVm,I_y}^2 \sigma_{G_{2m}} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ Symmetry & \cdot & \dots & r_{GEBVm,I_y}^2 \sigma_{G_m}^2 \end{bmatrix}$$

where $GEBV_{I_y}$ is a vector of m known GEBVs in I_y . We used the averaged reliability of GEBVs in a population as a characteristic or parameter of the population. However, $cov(GEBV_{I_{pr}})$ must be adjusted to account for the effect of first selection on I_y . $Cov(GEBV_{I_{pr}})$, which was adjusted to account for the effect of first selection on I_y is given in the supplementary methods section [S1 Text](#).

As a result,

$$cov(GEBV_{I_{pr}})_{m \times m} = cov(GEBV_{I_y})_{m \times m} - \frac{1}{var(I_y)} cov(GEBV_{I_{pr}}, GEBV_{I_y}) a a' cov(GEBV_{I_y}, GEBV_{I_{pr}})$$

where $GEBV_{I_{pr}}$ is an $m \times 1$ vector of GEBVs in second-stage selection (I_{pr}); $Cov(GEBV_{I_{pr}})_{m \times m}$ is as shown in equation (1); a is an $m \times 1$ vector of economic value for m traits in the aggregate genotype; and $cov(GEBV_{I_{pr}}, GEBV_{I_y})$ is the transpose of $cov(GEBV_{I_y}, GEBV_{I_{pr}})$.

In the same way, a covariance matrix between the GEBVs in I_{pr} and I_y , that is, $cov(GEBV_{I_{pr}}, GEBV_{I_y})$ is given as:

$$Cov(GEBV_{I_{pr}}, GEBV_{I_y})_{m \times m} = Cov(GEBV_{I_{pr}}, GEBV_{I_y})_{m \times m} - \frac{k}{var(I_y)} Cov(GEBV_{I_{pr}}, GEBV_{I_y}) a a' Cov(GEBV_{I_y}, GEBV_{I_{pr}})$$

where $Cov(GEBV_{I_y}, GEBV_{I_{pr}})$ can be computed as $var(GEBV_{I_y})$, as shown previously.

Therefore,

$$Cov(I_{pr}, I_y) = \sigma_{I_{pr}I_y} = a' Cov(GEBV_{I_{pr}}, GEBV_{I_y}) a,$$

$$var(I_{pr}) = \sigma_{I_{pr}}^2 = a' Cov(GEBV_{I_{pr}}) a$$

Hence,

$$r_{I_y, I_{pr}} = \frac{\sigma_{I_y, I_{pr}}}{\sqrt{\sigma_{I_y} \sigma_{I_{pr}}}}$$

That is, these two selection indices (I_y and I_{pr}) are not independent, because the information for the second index includes the information for the first index.

The selection indices for the first- and second-stage selection, I_y and I_{pr} , are not independent but correlated. The selection intensity for I_{pr} is explained in the appendix, and that for I_y is from the work

of Falconer and Mackay [16]. Genetic gain for the aggregate genotype (H) due to first selection on I_y is given as $\hat{I}_{I_y} \sigma_{I_y}$, where \hat{I}_{I_y} is the selection intensity for I_y . Genetic (co)variances for the component traits in the aggregate genotype are affected by first selection on I_y .

The genetic (co)variance matrix for m-component traits after first selection (G^*) is expressed as:

$$G^* = G - \frac{k_{I_y}}{var(I_y)} cov(g, GEBV_{I_y}) a a' cov(GEBV_{I_y}, g')$$

where G is an $m \times m$ genetic (co)variance matrix for m-component traits before first selection; g is a vector of true genetic value for m-component traits in the aggregate genotype before first selection;

$k_{I_y} = i_{I_y} (i_{I_y} - v_1)$, \hat{I}_{I_y} is the selection intensity for I_y ; v_1 is a standardized truncation point for I_y ; and

$$Cov(g, GEBV_{I_y}) = \begin{bmatrix} r_{GEBV1,I_y}^2 \sigma_{G_1}^2 & r_{GEBV1,I_y}^2 r_{GEBV2,I_y}^2 \sigma_{G_{12}} & \dots & r_{GEBV1,I_y}^2 r_{GEBVm,I_y}^2 \sigma_{G_{1m}} \\ \cdot & r_{GEBV2,I_y}^2 \sigma_{G_2}^2 & \dots & r_{GEBV2,I_y}^2 r_{GEBVm,I_y}^2 \sigma_{G_{2m}} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ symmetry & \cdot & \dots & r_{GEBVm,I_y}^2 \sigma_{G_m}^2 \end{bmatrix}$$

according to MT-BLUP properties [17] and deriving covariance GEBV between traits i and j [15]. Genetic gain for the aggregate genotype (H) due to second selection on I_{pr} is given as $\hat{I}_{I_{pr}} \sigma_{I_{pr}}$ where $\hat{I}_{I_{pr}}$ is the selection intensity for I_{pr} according to the Appendix.

Because GEBVs in I_{pr} are given according to single-step genomic BLUP, the selection index coefficients composed of GEBVs in I_{pr} are the same as the economic weights of the aggregate genotype [7,8,18].

We assumed that the indices and aggregate genotype had multivariate normal distribution at each stage during two-stage selection. Under this assumption, the regression of the aggregate genotype on any linear function of the genetic values is linear [19], and the total selection response for two stages is the sum of the response obtained at each stage [20,21]. Consequently, genetic gain for the aggregate genotype during two-stage selection is the sum of $\hat{I}_{I_y} \sigma_{I_y}$ and $\hat{I}_{I_{pr}} \sigma_{I_{pr}}$.

The response to the i^{th} component trait due to index selection (I_y) (that is, ΔG_{I_y}) and index selection (I_{pr}) (that is, $\Delta G_{I_{pr}}$) is shown as:

$$\Delta G_{I_y} = \frac{\hat{I}_{I_y}}{\sigma_{I_y}} a' \begin{bmatrix} r_{GEBV1,I_y}^2 r_{GEBV1,I_y}^2 \sigma_{G_{11}} \\ r_{GEBV2,I_y}^2 r_{GEBV1,I_y}^2 \sigma_{G_{12}} \\ \cdot \\ r_{GEBVm,I_y}^2 \sigma_{G_i}^2 \\ \cdot \\ r_{GEBVm,I_y}^2 r_{GEBV1,I_y}^2 \sigma_{G_{1m}} \end{bmatrix} \text{ and}$$

$$\Delta G_{I_{pr}} = \frac{\hat{I}_{I_{pr}}}{\sigma_{I_{pr}}} a' \times i^{th} \text{ column of } cov(GEBV_{I_{pr}})_{m \times m}$$

according to MT-BLUP properties [17]. The total response to trait i during two-stage selection is the sum of ΔG_{I_y} and $\Delta G_{I_{pr}}$.

The genetic (co)variance matrix for m-component traits after two-stage selection (G^{**}) is expressed as:

$$G^{**} = G^* - \frac{k_{I_{prs}}}{\text{var}(I_{prs})} \text{cov}(g^*, GEBV'_{I_{prs}}) a a' \text{cov}(GEBV_{I_{prs}}, g^*)$$

where G^* is an $m \times m$ genetic (co)variance matrix for m-component traits after the first selection; g^* is a $m \times 1$ vector of true genetic value for m-component traits in the aggregate genotype after first selection; $k_{I_{prs}} = i_{I_{prs}}(i_{I_{prs}} - u_2)$, $i_{I_{prs}}$ is the selection intensity for I_{prs} ; u_2 is a standardized truncation point for I_{prs} and

$\text{cov}(g^*, GEBV'_{I_{prs}}) = \text{cov}(GEBV_{I_{prs}}, g^*) = \text{cov}(GEBV_{I_{prs}})$ according to MT BLUP properties [17].

Two-stage selection according to decorrelated constrained selection indices

Cerón-Rojas JJ, et al. [3,4] developed decorrelated constrained selection indices for multi-stage selection. These indices minimize the mean-squared difference between the index and the aggregate genotype at each stage under the restriction that the covariance between the indices at different stages is zero, thus preventing correlation between index values at different stages. Under this restriction, the selected individual index values after the first selection stage can be normally distributed [6].

According to Cerón-Rojas JJ, et al.[3,4], the index coefficients for i^{th} stage selection (that is, β_i) are: $\beta_i = K_i b_i$,

$$\text{where } K_i = (I - \varphi_i), \varphi_i = Q_{ii}^{-1} S_{i(i-1)} [S_{(i-1)} Q_{ii}^{-1} S_{i(i-1)}]^{-1} S_{(i-1)}, b_i = Q_{ii}^{-1} A_i a;$$

a is a vector of economic value for component traits of the aggregate genotype; and the other definitions in the preceding abbreviations are explained by Cerón-Rojas JJ, et al. [3,4].

When we applied the formulas from the work of Cerón-Rojas JJ, et al. [3,4] to two-stage genomic selection composed of GEBVs in the current study,

$$\beta_1 = IQ_{11}^{-1} A_1 w = \text{Cov}(GEBV)_{I_1}^{-1} \text{Cov}(GEBV_{I_1}, g^*) a = \text{Cov}(GEBV)_{I_1}^{-1} \text{Cov}(GEBV)_{I_1} a = a,$$

according to MT BLUP properties [17]. Because,

$$K_{2,2m \times 2m} = I_{2m \times 2m} - Q_{22}^{-1} S_{2,1} [S_{1,2} Q_{22}^{-1} S_{2,1}]^{-1} S_{1,2},$$

according to Cerón-Rojas JJ, et al. [3,4], the components of $K_{2,2m \times 2m}$ herein are expressed as:

$$Q_{22,2m \times 2m} = \begin{bmatrix} \text{Cov}(GEBV_{I_y})_{m \times m} & \text{Cov}(GEBV_{I_y}, GEBV'_{I_{prs}})_{m \times m} \\ \text{Cov}(GEBV_{I_{prs}}, GEBV'_{I_y})_{m \times m} & \text{Cov}(GEBV_{I_{prs}})_{m \times m} \end{bmatrix}$$

$$S_{2,1,2m \times 1} = \begin{bmatrix} \text{Cov}(GEBV_{I_y})_{m \times m} \\ \text{Cov}(GEBV_{I_{prs}}, GEBV'_{I_y})_{m \times m} \\ \dots \\ a_m \end{bmatrix}, \text{ and}$$

$$S_{1,2,1 \times 2m} = [a_1 \ a_2 \ \dots \ a_m] [\text{Cov}(GEBV_{I_y})_{m \times m} \ \text{Cov}(GEBV_{I_y}, GEBV'_{I_{prs}})_{m \times m}],$$

where $\text{Cov}(GEBV_{I_y}, GEBV'_{I_{prs}})$ is the transpose of $\text{Cov}(GEBV_{I_{prs}}, GEBV'_{I_y})$, as shown earlier.

In addition, $b_{2,2m \times 1}$ in our study here can be written as:

$$b_{2,2m \times 1} = Q_{22}^{-1} A_2 a = Q_{22,2m \times 2m}^{-1} \begin{bmatrix} \text{Cov}(GEBV_{I_y})_{m \times m} \\ \text{Cov}(GEBV_{I_{prs}})_{m \times m} \end{bmatrix} a,$$

according to MT BLUP properties [17]. So

$$\beta_{2,2m \times 1} = K_{2,2m \times 2m} b_{2,2m \times 1} \dots \dots \dots (2)$$

Therefore, the selection indices for the first (I_y) and second (I_{prs}) selections are:

$$I_y = a' GEBV_{I_y} \text{ and}$$

$$I_{prs} = \begin{bmatrix} \beta'_{2,1,2,\dots,m} & \beta'_{2,m+1,m+2,\dots,m+m} \end{bmatrix} \begin{bmatrix} GEBV_{I_y} \\ GEBV_{I_{prs}} \end{bmatrix}, \text{ respectively.}$$

The responses to the aggregate genotype due to first- and second-stage selection are $i_{I_y} \sigma_{I_y}$ and $i_{I_{prs}} \sigma_{I_{prs}}$, respectively.

The responses to trait i due to index selection (I_y) (that is, ΔG_{I_y}) and index selection (I_{prs}) (that is, $\Delta G_{I_{prs}}$) are shown as:

$$\Delta G_{I_y} = \frac{i_{I_y}}{\sigma_{I_y}} a' \begin{bmatrix} r_{GEBV1,I_y}^2 r_{GEBV1,I_y}^2 \sigma_{G_{i1}} \\ r_{GEBV2,I_y}^2 r_{GEBV2,I_y}^2 \sigma_{G_{i2}} \\ \dots \\ r_{GEBVi,I_y}^2 \sigma_{G_i} \\ \dots \\ r_{GEBVm,I_y}^2 r_{GEBVm,I_y}^2 \sigma_{G_{im}} \end{bmatrix} \text{ and}$$

$$\Delta G_{I_{prs}} = \frac{i_{I_{prs}}}{\sigma_{I_{prs}}} \beta_2' \times i^{th} \text{ column of } \begin{bmatrix} \text{Cov}(GEBV_{I_y})_{m \times m} \\ \text{Cov}(GEBV_{I_{prs}})_{m \times m} \end{bmatrix}_{2m \times m},$$

according to MT BLUP properties [17]; $\delta_{I_{prs}}^2$ is as shown earlier, and

$$\delta_{I_{prs}}^2 = \begin{bmatrix} \beta'_{2,1,2,\dots,m} & \beta'_{2,m+1,m+2,\dots,m+m} \end{bmatrix} Q_{22,2m \times 2m} \begin{bmatrix} \beta_{2,1,2,\dots,m} \\ \beta_{2,m+1,m+2,\dots,m+m} \end{bmatrix}.$$

The total response to trait i during two-stage selection is the sum of ΔG_{I_y} and $\Delta G_{I_{prs}}$. The first-and second-selection indices are independent; thus the first selection does not affect the second selection. Therefore, the genetic (co)variance matrix for m-component traits after two-stage selection (G^{**}) is expressed as:

$$G^{**} = G - \frac{k_y}{\text{var}(I_y)} \text{Cov}(g, \text{GEBV}'_y) aa' \text{Cov}(\text{GEBV}'_y, g) - \frac{k_{prs}}{\text{var}(I_{prs})} \text{Cov}(g, \text{GEBV}'_{prs})_{\text{for } 1, 2, \dots, m^{\text{th}} \text{ column}, \text{GEBV}'_{prs}}_{\text{for } m+1, m+2, \dots, m+m^{\text{th}} \text{ column}} \beta_2 \beta_2' \text{Cov} \left[\begin{array}{l} \text{GEBV}'_{prs} \text{ for } 1, 2, \dots, m^{\text{th}} \text{ row} \\ \text{GEBV}'_{prs} \text{ for } m+1, m+2, \dots, m+m^{\text{th}} \text{ row} \end{array} \right]_{2m \times m}$$

where

$$\text{Cov}(g, \text{GEBV}'_{y_j} \text{ for } 1, 2, \dots, m^{\text{th}} \text{ column}, \text{GEBV}'_{I_{prs}} \text{ for } m+1, m+2, \dots, m+m^{\text{th}} \text{ column})_{m \times 2m} = \left[\text{Cov}(\text{GEBV}'_{I_y})_{m \times m}, \text{Cov}(\text{GEBV}'_{I_{prs}})_{m \times m} \right],$$

according to MT-BLUP properties [17], and

$$\text{Cov} \left[\begin{array}{l} \text{GEBV}'_{y_j} \text{ for } 1, 2, \dots, m^{\text{th}} \text{ row} \\ \text{GEBV}'_{I_{prs}} \text{ for } m+1, m+2, \dots, m+m^{\text{th}} \text{ row} \end{array} \right]_{2m \times m} g'$$

is the transpose of

$$\text{Cov}(g, \text{GEBV}'_{I_y} \text{ for } 1, 2, \dots, m^{\text{th}} \text{ column}, \text{GEBV}'_{I_{prs}} \text{ for } m+1, m+2, \dots, m+m^{\text{th}} \text{ column})_{m \times 2m} = \left[\text{Cov}(\text{GEBV}'_{I_y})_{m \times m}, \text{Cov}(\text{GEBV}'_{I_{prs}})_{m \times m} \right]$$

Asymptotic response to four-path selection due to two-stage index selection

Four-path selection: Four-path selection due to SS (selection path: sires to breed sires), SD (selection path: sires to breed dams), DS (selection path: dams to breed sires), and DD (selection path: dams to breed dams) is considered as in the work of Togashi K, et al. [7,8]. In the previous section, we considered two-stage selection and defined the first- and second-selection indices as I_y and I_{prs} , respectively. However, we did not consider two-stage selection for the selection path DD, because selection is usually weak in that path [22]. Therefore, we considered the first- and second-selection indices as $I_{y_{SS}}$ and $I_{prs_{SS}}$ respectively, for the SS selection path. In the same way, $I_{y_{SD}}$ and $I_{prs_{SD}}$ are defined as the first- and second-selection indices in SD, and $I_{y_{DS}}$ and $I_{prs_{DS}}$ are respectively those for DS. The first-selection indices for SS and SD are identical, because the component traits in the indices are the same between SS and SD. Therefore, $I_{y_{SS}}$ and $I_{y_{SD}}$ are represented as I_{y_S} . The economic value for the component traits of the aggregate genotype is the same in all four selection paths. In this situation, two genetic variances need to be distinguished—one for the male population and one for the female population [23]. This step must be taken because, in two-path selection, selected bulls and dams produce the offspring in the next generation without distinguishing the male or female offspring. However, in four-path selection, selected bulls or dams are distinguished according to whether they produce the male or female offspring in the next generation. In contrast, the same genetic (co)variance matrix is taken for the male and female populations in the initial generation before selection, that is, generation 0.

The genetic (co)variance matrices of G , G^* , and G^{**} are defined before two-stage selection, after first selection, and after second selection, respectively. These matrices are derived from correlated indices during two stages under the assumption of trivariate normal distribution for the true genetic value before first selection and the first- and second-selection indices. The adjusted genetic (co)variance matrices of G , G^* , and G^{**} for SS and SD are shown as:

$$G^*_{SS \text{ or } SD,t} = G_{S,t} - \frac{k_{y_{S,t}}}{\text{var}(I_{y_{S,t}})} \text{Cov}(g_S, \text{GEBV}'_{y_{S,t}}) aa' \text{Cov}(\text{GEBV}'_{y_{S,t}}, g'_S)_t,$$

$$G^{**}_{SS,t} = G^*_{SS \text{ or } SD,t} - \frac{k_{I_{prs_{SS,t}}}}{\text{var}(I_{prs_{SS,t}})} \text{Cov}(g^*_S, \text{GEBV}'_{I_{prs_{SS,t}}})_t aa' \text{Cov}(\text{GEBV}'_{I_{prs_{SS,t}}}, g^*_S)_t, \text{ and}$$

$$G^{**}_{SD,t} = G^*_{SS \text{ or } SD,t} - \frac{k_{I_{prs_{SD,t}}}}{\text{var}(I_{prs_{SD,t}})} \text{Cov}(g^*_S, \text{GEBV}'_{I_{prs_{SD,t}}})_t aa' \text{Cov}(\text{GEBV}'_{I_{prs_{SD,t}}}, g^*_S)_t,$$

where subscript t indicates generation t, and subscript s means a sire or male population.

Note that the second-stage selection indices for SS ($I_{prs_{SS,t}}$) and SD ($I_{prs_{SD,t}}$) are the same, because the first-selection indices for SS and SD are identical. Variance of the first selection index for SS and SD ($I_{y_{S,t}}$) and that of the second-stage selection index for SS ($I_{prs_{SS,t}}$) and SD ($I_{prs_{SD,t}}$) are given in the Supplementary Methods section [S2 Text](#).

The reduction factor (k) in the procedure is constant during generations according to decorrelated constrained selection indices, whereas it varies during generations in the second-stage selection based on correlated indices, because $r_{I_y I_{prs}}$ changes with successive generations (Appendix).

In the same way as for a male population, G , G^* , and G^{**} are defined for a female population. Variance of the first selection index for DS ($I_{y_{DS,t}}$) and DD ($I_{y_{DD,t}}$) and that of the second-stage selection index for DS ($I_{prs_{DS,t}}$) are given in the Supplementary Methods section [S3 Text](#).

The prediction error variance is not affected by selection and is determined only by the amount of information (that is, number of records; [24,25]). Accounting for prior selection, the reliability of the EBV (r^2) is given according to Dekkers JCM [23] and Bijma P [26] as:

$$r^{*2} = 1 - \left(1 - r^2\right) \sigma_0^2 / \sigma^{*2}$$

where r^2 is the unadjusted reliability, σ^{*2} is the genetic variance adjusted for prior selection, and σ_0^2 is the genetic variance in the base unselected and random population. Therefore, the reliability of the GEBV for trait i in the selection index for the first selection ($I_{y,S}$) in the sire population of generation t ($r^2_{\text{GEBV}_i, I_{y_{S,t}}}$) is shown as:

$$r^2_{\text{GEBV}_i, I_{y_{S,t}}} = 1 - \frac{\left(1 - r^2_{\text{GEBV}_i, I_{y_{S,0}}}\right) \sigma_{G_0}^2}{\sigma_{G_{S,t}}^2}.$$

In the same way, the reliability of the GEBV for trait i in the selection indices for the first selection ($I_{y,DS}$) and ($I_{y,DD}$) in the dam population of generation t ($r^2_{\text{GEBV}_i, I_{y_{DS,t}}}$ and $r^2_{\text{GEBV}_i, I_{y_{DD,t}}}$) is shown as:

$$r^2_{\text{GEBV}_i, I_{y_{DS,t}}} = 1 - \frac{\left(1 - r^2_{\text{GEBV}_i, I_{y_{DS,0}}}\right) \sigma_{G_0}^2}{\sigma_{G_{D,t}}^2} \text{ and}$$

$$r^2_{\text{GEBV}_i, I_{y_{DD,t}}} = 1 - \frac{\left(1 - r^2_{\text{GEBV}_i, I_{y_{DD,0}}}\right) \sigma_{G_0}^2}{\sigma_{G_{D,t}}^2}, \text{ respectively.}$$

Let the proportion of individuals coming from the first selection to produce the next generation in the SS pathway be piy_{SS} and that of those coming from the second selection be $1 - piy_{SS}$. In the same way as for the SS pathway, piy_{SD} and $1 - piy_{SD}$ are defined as the proportions of individuals in the SD pathway coming from the first and second selections, respectively. In addition, piy_{DS} and $1 - piy_{DS}$ are defined in the DS pathway in the same ways as for the SS and SD pathways.

Selection introduces so-called gametic phase disequilibrium [16], that is, negative covariance in gene effects at the two loci, and this effect reduces the genetic variance in the group of selected individuals. However, the variance due to Mendelian sampling is unaffected by selection and is equal to $\frac{1}{2}\sigma_{g0}^2$, where σ_{g0}^2 is the genetic variance in the unselected and non-inbred base population. Mendelian sampling variance represents the variation created by sampling one of a pair of parental alleles at each locus. This sampling process is unaffected by selection. Therefore, genetic (co)variances reach an asymptote in single-trait phenotypic selection [27]. We made the standard assumption of the infinitesimal model so that changes in allele frequency can be ignored and so that all changes of genetic variances are due solely to gametic phase disequilibrium (that is, the “Bulmer effect”). Therefore, asymptotic response refers to the response when genetic (co)variances have reached an asymptote.

Thus, the additive genetic (co)variance matrices in the sire ($G_{S,t+1}$) and dam ($G_{D,t+1}$) populations of generation $t + 1$ are:

$$G_{S,t+1} = \frac{1}{4}\{piy_{SS}[G_{SS,t}^*] + (1 - piy_{SS})[G_{SS,t}^{**}]\} + \frac{1}{4}\{piy_{DS}[G_{DS,t}^*] + (1 - piy_{DS})[G_{DS,t}^{**}]\} + \frac{G_0}{2} \text{ and}$$

$$G_{D,t+1} = \frac{1}{4}\{piy_{SD}[G_{SD,t}^*] + (1 - piy_{SD})[G_{SD,t}^{**}]\} + \frac{G_{DD,t}}{4} + \frac{G_0}{2}, \text{ respectively.}$$

The genetic (co)variance in the unselected base population during generation 0 is the same between the sire and dam populations and is shown as G_0 . The last term ($\frac{G_0}{2}$) is the within-family (or Mendelian) sampling variance, which is not affected by selection in the absence of inbreeding [27]. The genetic correlation of the aggregate genotype between the first (H_1) and asymptotic (H_a) generations is computed as $\frac{\sigma_{H_a}^2}{\sqrt{\sigma_{H_1}\sigma_{H_a}}}$,

where $\sigma_{H_1}^2$ and $\sigma_{H_a}^2$ are the genetic variance of the aggregate genotype at the first generation and an asymptote, respectively.

Responses to the aggregate genotype: The responses to the aggregate genotype due to first- and second-stage selection are $i_{I_y}\sigma_{I_y}$ and $i_{I_{prs}}\sigma_{I_{prs}}$, respectively. Consequently, the genetic gain of the aggregate genotype (ΔH /generation) for four-pathway selection is expressed as:

$$\frac{1}{4} \times \left[\frac{piy_{SS}i_{I_y}(\sigma_{I_{m,ss}} + i_{I_y}\sigma_{I_y}) + (1 - piy_{SS})(\sigma_{I_{m,ss}} + i_{I_y}\sigma_{I_y}) + piy_{SD}i_{I_y}(\sigma_{I_{m,ss}} + i_{I_y}\sigma_{I_y}) + (1 - piy_{SD})(\sigma_{I_{m,ss}} + i_{I_y}\sigma_{I_y}) + piy_{DS}i_{I_y}(\sigma_{I_{m,ss}} + i_{I_y}\sigma_{I_y}) + (1 - piy_{DS})(\sigma_{I_{m,ss}} + i_{I_y}\sigma_{I_y}) + \sigma_{I_{m,ss}}i_{I_y} }{4} \right] = \frac{\Delta H_{SS} + \Delta H_{SD} + \Delta H_{DS} + \Delta H_{DD}}{4}$$

where

$$\Delta H_{SS} = piy_{SS}\Delta H_{I_y,SS} + (1 - piy_{SS})(\Delta H_{I_{prs,SS}} + \Delta H_{I_y,SS}),$$

$$\Delta H_{SD} = piy_{SD}\Delta H_{I_y,SD} + (1 - piy_{SD})(\Delta H_{I_{prs,SD}} + \Delta H_{I_y,SD}),$$

$$\Delta H_{DS} = piy_{DS}\Delta H_{I_y,DS} + (1 - piy_{DS})(\Delta H_{I_{prs,DS}} + \Delta H_{I_y,DS}), \text{ and}$$

$$\Delta H_{DD} = \sigma_{I_y,DD}i_{I_y,DD} = \Delta H_{I_y,DD}$$

Note that two-stage selection was not considered for the selection path DD.

Suppose that L_{SS} , L_{SD} , L_{DS} , and L_{DD} are the respective generation intervals for the SS, SD, DS, and DD pathways. Consequently, ΔH /year can be expressed as:

$$\frac{\Delta H}{Year} = \frac{\Delta H_{SS} + \Delta H_{SD} + \Delta H_{DS} + \Delta H_{DD}}{L_{SS} + L_{SD} + L_{DS} + L_{DD}},$$

where

$$L_{SS} = piy_{SS}L_{SS,I_y} + (1 - piy_{SS})L_{SS,I_{prs}},$$

$$L_{SD} = piy_{SD}L_{SD,I_y} + (1 - piy_{SD})L_{SD,I_{prs}}, \text{ and}$$

$$L_{DS} = piy_{DS}L_{DS,I_y} + (1 - piy_{DS})L_{DS,I_{prs}},$$

where L_{SS,I_y} is the generation interval for the first selection in SS, and $L_{SS,I_{prs}}$ is the generation interval for the second selection in SS, L_{SD,I_y} , $L_{SD,I_{prs}}$, L_{DS,I_y} , and $L_{DS,I_{prs}}$ are defined in the same way in SD and DS selection paths.

The selection accuracies for $I_{y,S,t}$, $I_{y,DS,t}$, $I_{y,DD,t}$ are $\sigma_{I_{y,S,t}}/\sigma_{H_{S,t}}$, $\sigma_{I_{y,DS,t}}/\sigma_{H_{D,t}}$, $\sigma_{I_{y,DD,t}}/\sigma_{H_{D,t}}$ respectively,

$$\text{where } \sigma_{H_{S,t}}^2 = a'G_{S,t}a \text{ and } \sigma_{H_{D,t}}^2 = a'G_{D,t}a.$$

The selection accuracies for $I_{prs,SS,t}$, $I_{prs,SD,t}$, and $I_{prs,DS,t}$ are,

$$\frac{\sigma_{I_{prs,SS,t}}}{\sigma_{H_{S,t}}^*}, \frac{\sigma_{I_{prs,SD,t}}}{\sigma_{H_{S,t}}^*}, \text{ and } \frac{\sigma_{I_{prs,DS,t}}}{\sigma_{H_{DS,t}}^*}, \text{ respectively.}$$

$$\text{where } \sigma_{H_{S,t}}^2 = a'G_{SS \text{ or } SD,t}a \text{ and } \sigma_{H_{DS,t}}^2 = a'G_{DS,t}a$$

Genetic gain for the i^{th} component trait of the index: The genetic gains for the i^{th} component trait of generation t from the correlated selection indices during two stages are shown as:

$$\Delta G_{i_{y,S,t}} = \frac{i_{y,S,t}}{\sigma_{I_{y,S,t}}} a' \begin{bmatrix} r_{GEBV_{y,S,1,t}}^2 I_{GEBV_{y,S,1,t}}^2 \sigma_{G_{1,S,t}} \\ r_{GEBV_{y,S,2,t}}^2 I_{GEBV_{y,S,2,t}}^2 \sigma_{G_{2,S,t}} \\ \vdots \\ r_{GEBV_{y,S,i,t}}^2 \sigma_{G_{i,S,t}} \\ \vdots \\ r_{GEBV_{y,S,m,t}}^2 I_{GEBV_{y,S,m,t}}^2 \sigma_{G_{m,S,t}} \end{bmatrix}$$

$$\Delta G_{i_{prs,SS,t}} = \frac{i_{prs,SS,t}}{\sigma_{I_{prs,SS,t}}} a' \times i^{th} \text{ column of } Cov(GEBV_{I_{prs,SS}})_{l,m \times m}, \text{ and}$$

$$\Delta G_{i_{prs,SD,t}} = \frac{i_{i_{prs,SD,t}}}{\delta_{i_{prs,SD,t}}} a_i \times i^{th} \text{ column of } \text{cov}(\text{GEBV}_{i_{prs,SD,t}})_{1,m \times m}$$

where $\Delta G_{i_{y,S,t}}$ is the genetic gain based on $I_{y,S,t}$, $\Delta G_{i_{prs,SS,t}}$ is the genetic gain based on $I_{prs,SS,t}$, and $\Delta G_{i_{prs,SD,t}}$ is the genetic gain based on $I_{prs,SD,t}$.

$\Delta G_{i_{y,DS,t}}$, $\Delta G_{i_{y,DD,t}}$, and $\Delta G_{i_{prs,DS,t}}$ are given by replacing subscript S (sires) with D (dams).

In addition, $\text{cov}(\text{GEBV}_{i_{prs,SS,t}})_{1,m \times m}$ is the same as $\text{cov}(\text{GEBV}_{i_{prs,SD,t}})_{1,m \times m}$.

The genetic gains for the i^{th} component trait of generation t from independent selection indices during two stages are shown as:

$$\Delta G_{i_{y,S,t}}, G_{i_{y,DD,t}}, \text{ and } \Delta G_{i_{y,DS,t}}$$

which are identical to those of the correlated indices mentioned earlier.

$$\Delta G_{i_{prs,SS,t}} = \frac{i_{i_{prs,SS,t}}}{\sigma_{i_{prs,SS,t}}} \beta_{2,S} \times i^{th} \text{ column of } \begin{bmatrix} \text{cov}(\text{GEBV}_{i_{y,SS,t}})_t \\ \text{cov}(\text{GEBV}_{i_{prs,SS,t}})_t \end{bmatrix}_{2,m \times m}$$

$$\Delta G_{i_{prs,SD,t}} = \frac{i_{i_{prs,SD,t}}}{\sigma_{i_{prs,SD,t}}} \beta_{2,S} \times i^{th} \text{ column of } \begin{bmatrix} \text{cov}(\text{GEBV}_{i_{y,SD,t}})_t \\ \text{cov}(\text{GEBV}_{i_{prs,SD,t}})_t \end{bmatrix}_{2,m \times m}, \text{ and}$$

$$\Delta G_{i_{prs,DS,t}} = \frac{i_{i_{prs,DS,t}}}{\sigma_{i_{prs,DS,t}}} \beta_{2,D} \times i^{th} \text{ column of } \begin{bmatrix} \text{cov}(\text{GEBV}_{i_{y,DS,t}})_t \\ \text{cov}(\text{GEBV}_{i_{prs,DS,t}})_t \end{bmatrix}_{2,m \times m}$$

where $\beta_{2,S}$ and $\beta_{2,D}$ correspond to equation (2) for the sire and dam populations, respectively.

Therefore, the response to trait i due to the selection of generation t ($R_{index,i,t}^*$) derived from four selection paths (SS, SD, DS, and DD), regardless of whether the indices are correlated or independent, can be calculated as:

$$R_{index,i,t}^* = \frac{1}{4} \left[\begin{aligned} & p i y_{SS} \Delta G_{i_{y,S,t}} + (1 - p i y_{SS}) (\Delta G_{i_{y,S,t}} + \Delta G_{i_{prs,SS,t}}) + p i y_{SD} \Delta G_{i_{y,S,t}} + (1 - p i y_{SD}) \\ & (\Delta G_{i_{y,S,t}} + \Delta G_{i_{prs,SD,t}}) + p i y_{DS} \Delta G_{i_{y,DS,t}} + (1 - p i y_{DS}) (\Delta G_{i_{y,DS,t}} + \Delta G_{i_{prs,DS,t}}) + \Delta G_{i_{DD,D,t}} \end{aligned} \right]$$

Consequently, $\Delta G_i/\text{year}$ can be expressed as:

$$\frac{\Delta G_i}{\text{year}} = R_{index,i,t}^* \left/ \frac{L_{SS} + L_{SD} + L_{DS} + L_{DD}}{4} \right.$$

Example applications of the formula

We used two quantitative traits to compare the efficiency of index selection based on GEBVs and assumed that the number of traits was equal for both the index and the breeding goal (H). Trait 1 was assumed to be moderately heritable, with $h^2=0.3$, whereas trait 2 was assumed to have low heritability, with $h^2=0.05$. The breeding goal (H)

comprising traits 1 and 2 was defined as $H = a_1 g_1 + a_2 g_2$, where g_i is the true genetic value and a_i is the relative economic weight for trait i, and the genetic correlation ($r_{g_1 g_2}$) between the two traits was assumed as $r_{g_1 g_2} = 0.3$. In addition, three sets of relative economic weights between traits 1 and 2 were considered, that is, $a_1 : a_2 = 1 : 1, 3 : 1$, and $1 : 3$. The reliability of the GEBV in the population during generation 0 was calculated according to Daetwyler HD, et al. [28] and Van Grevenhof EM, et al. [29], under the assumption that the historical effective size of the unselected population (N_e) and the size of the genome (L, in Morgans) are 100 and 30, respectively. The reference population was assumed to comprise 5,000 progeny-tested sires, with 50 offspring per test bull. Therefore, given these assumptions, the reliabilities of the GEBVs for traits 1 and 2 in generation 0 were 0.4006 and 0.2441, respectively.

The number of progeny per sire was set at 25. Therefore, the reliabilities of progeny testing for proven bulls without genomic information for traits 1 and 2 in generation 0 were 0.6696 and 0.2404, respectively. As a result, the reliabilities of GEBVs of the proven bull when combining his genomic information and his daughters' records for traits 1 and 2 in generation 0 were 0.7294 and 0.3900, respectively. The repeatability for a dam's second selection on the basis of three of her own records was set at the value of heritability+0.1, so that the reliabilities without genomic information for traits 1 and 2 in generation 0 were 0.5 and 0.1154, respectively. As a result, the reliabilities of GEBVs of the dam after combining genomic information and three of her own lactation records for traits 1 and 2 in generation 0 were 0.6252 and 0.3119, respectively.

The first selection of two-stage selection used only genotypes with no phenotypes for GEBVs in the SS, SD, and DS selection paths. In the second selection of two-stage selection, genotypes and the progeny's milk-production records were used in the GEBVs in the SS and SD selection paths, and genotypes and three of the dam's own lactation records were used in the GEBV in the DS selection path. Genotypes and three of the dam's own lactation records were used in GEBV in the DD selection path, but selection was performed only once per generation. Note that dams did not use their daughters' records, because doing so would be too time-intensive. In comparison with the two-stage selection described above, single-stage selection in SS and SD was performed only once per generation and was based only on the first-lactation records of 25 of the test bull's daughters without genomic information, as done in traditional progeny testing. In the same way as SS and SD, single-stage selection in DS and DD was performed only once per generation on the basis of three of the dam's own lactation records.

For SS and SD, the selection percentage for the first selection, which was based on only genotypes with no phenotypes, was assumed as 20%, and those for the second selection, which was based on both genotypes and phenotypes, were assumed as 5% and 15%, respectively, of the bulls that passed first-stage selection. The selection percentages for DS were assumed as 40% for the first selection and as 1% of the dams that passed first-stage selection for the second selection, which was based on genotypes and the dam's own records. Selection for DD was performed only once per generation at an assumed selection percentage of 70%.

We compared the responses to the aggregate genotype and the genetic responses to each component trait of the index over generations with those to single-stage selection performed only once per generation without genomic information. To compare the genetic gains between single- and two-stage selections, they must all have the

same final intensity of selection. Let the selected proportions at the first- and second-stage selections be p_1 and p_2 , respectively, such that the final selected proportion (p) is $p=p_1 \times p_2$. For example, the selected proportions at the first- and second-stage selections in SS were 0.2 and 0.05, respectively; the equivalent selected proportion during single-stage selection in SS, that is, traditional progeny testing, was $0.2 \times 0.05=0.01$. The genetic responses from traditional single-stage selection over generations were computed according to the method of Togashi K, et al. [7]. The generation intervals for the first and second selections in SS and SD were assumed as 2.0 and 6.5 years, respectively, and in DS were assumed as 2.0 and 6.0 years, respectively. The generation interval in DD was assumed as 6.0 years. The proportion of parents from the first selection to produce the next generation in SS, SD, and DS (that is, p_{iySS} , p_{iySD} , and p_{iyDS}) was set at 0.3 and 0.7 for the two example scenarios. Note that we refer to the parents that were selected at the first stage and that had only genotypes “young parents.”

Results and Discussion

Changes in genetic (co)variances over generations

The changes in the genetic (co)variances over generations due to correlated indices during two-stage and single-stage selection in the sire population in which the economic weight was 1:1 are shown in table 1. An asymptote in genetic (co)variances was reached quickly, because we considered a closed dairy cattle selection scheme.

The initial genetic variances during generation 0 for the first and second traits were 0.3 and 0.05, respectively. The initial genetic covariance between the two index traits for the initial genetic correlation of 0.3 was 0.0367. Genetic (co)variances reached an asymptote after approximately six generations in both the single-stage and two-stage index selections. Furthermore, genetic (co)variances in other scenarios-where we modified the sire or dam population or the relative economic weight between two index traits-reached asymptotes at almost the same generation, as shown in table 1. It would not take much time to reach an asymptote in genetic (co)variances, since we considered a closed dairy cattle selection scheme. Genetic (co)variances at an asymptote in the correlated two-stage selection in which the proportion of young parents to all parents in SS, SD, and DS was 0.3 were smaller than in single-stage selection, and these (co)variances under single-stage selection were lower than those for two-stage selection in which the proportion of young parents to all parents in SS, SD, and DS was 0.7. Therefore, the (co)variances at an asymptote were smallest in two-stage selection where the proportion of young parents was 0.3. This result occurred because selection accuracy at the second stage in two-stage selection based on combining genotype and phenotypic records was higher than that at the first-stage in two-stage selection based solely on genotypes; that is, the higher accuracy, the greater the reduction in genetic (co)variances. Conversely, the (co)variances at an asymptote were largest in two-stage selection where the proportion of young parents was 0.7. This effect emerged because the selection accuracy for the first selection (that is, selection in young parents) was lower when derived from GEBVs without records than from GEBVs with records.

Asymptotic genetic responses to selection

The genetic variances of the aggregate genotype (H) in the first and sixth generations are shown in table 2 for the correlated and independent indices for both stages. The sixth generation is expressed as an asymptotic generation, as shown in table 1. The genetic correlation of H between the first and asymptotic (sixth) generations ($\frac{\sigma_{H_0}^2}{\sqrt{\sigma_{H_0} \sigma_{H_6}}}$) is also shown in table 2. The variances at the sixth generation

or an asymptote in both the correlated and independent two-stage selections were smaller in which the proportion of young parents to all parents in SS,SD, and DS was 0.3 than in which the proportion of young parents to all parents in SS, SD, and DS was 0.7. This trend agreed with that of the genetic variances in male populations, as shown in table 1.

The reduction in the genetic variance of the aggregate genotype over generations (that is, 6:1 in table 2) was smaller for independent indices than for correlated indices. This might be because independence during two-stage selection decreases genetic variance during subsequent generations without each variance influencing the other. Likewise the tendency for the reduction over generations mentioned earlier was recognized in terms of the genetic correlation in the first and sixth generations (r_g). Therefore, the genetic correlations in the first and asymptotic generations can be treated as easy and simple criteria for understanding the reduction in genetic variance over generations.

Genetic gains per generation for the aggregate genotype in the first six generations, in units of genetic standard deviation during generation 0, are shown in table 3. For two-stage selection, genetic gains in the initial generation, that is, generation 1, for the aggregate genotype were greater for correlated indices than for independent indices. This tendency is in agreement with the findings of Cerón-Rojas JJ, et al. [3,4]. In addition, the restriction procedure on independence for two indices is similar to a restricted selection index. Therefore, the resulting response is expected to be less than the maximized response without such restrictions. However, at an asymptote (that is, sixth generation), the genetic gains per generation for the aggregate genotype were slightly greater for independent indices than for correlated indices. This result occurred because the asymptotic genetic variances were greater for independent indices than with correlated indices (Table 2).

The genetic gains for the aggregate genotype per year or generation from two-stage selection are shown in table 4 and are compared with the genetic gains from single-stage selection in units of genetic standard deviation during generation 0 (that is, σ_{H_0}). The genetic gain per generation ($\Delta H/\text{generation}$) with economic weights (1:1 and 3:1) was lower in two-stage selection than in single-stage selection, in agreement with Young SSY [21], Saxton AM [30], and Cerón-Rojas JJ, et al.[3,4]. However, the genetic gain per generation ($\Delta H/\text{generation}$) in two-stage selection with economic weights (1:3) in which young parents accounted for 30% of all parents was almost the same as that of single-stage selection. This result occurred because the selection accuracy of single-stage selection for the low-heritability trait (trait 2; $h^2=0.05$), that is, due to progeny-tested sires with 25 progeny records, was 0.49, which was much smaller than that of second-stage selection in SS or SD when derived from combining their genotypes and progeny records (that is, 0.62). This outcome was in agreement with Togashi K, et al. [7], who found that the low-heritability trait selected on GEBV demonstrated a greater asymptotic genetic response when its relative economic weight was greater than that of the moderately heritable index trait.

Conversely, annual genetic gain ($\Delta H/\text{year}$) at an asymptote or sixth generation was greater for two-stage selection with the proportion of young parents comprising 0.3 than for two-stage selection with young parents comprising 0.7; $\Delta H/\text{year}$ when the proportion of young parents 0.7 was still higher than in single-stage selection. This result is first due to the reduction in the generation interval: the generation intervals for single-stage selection, two-stage selection in which young parents accounted for 30%, and two-stage selection with a young-parent proportion of 70% to all parents were 6.25, 5.28, and 3.98

years, respectively. Second, it occurred because the selection accuracy of the second stage in two-stage selection was much higher than that of the first stage, as mentioned earlier; that is, the weighted selection accuracy for the first and second stages in two-stage selection with a young-parent proportion of 30% was higher than that with young parents comprising 70% of all parents. For example, weighted selection accuracies for the first and second stages in two-stage selection with equal economic weights (that is, 1:1) in which young parents accounted for 30% and 70% of all parents were 0.727 and 0.640, respectively, in generation 1.

Compared with that for single-stage selection, annual genetic gain ($\Delta H/\text{year}$) was highest for an economic weight of 1:3 between the first ($h^2=0.3$) and second ($h^2=0.05$) traits; tested economic weights were 1:1, 3:1, and 1:3. This result occurred because the reliabilities of the low heritable trait (that is, trait 2, $h^2=0.05$) during generation 0 for the GEBV without phenotypic records, progeny testing in single-stage selection with 25 progeny records per test bull, and the dam's own three records were 0.2441, 0.2404 and 0.1154, respectively. In contrast, the reliabilities of the moderately high-heritability trait (that is, trait 1, $h^2=0.3$) during generation 0 for the GEBV without phenotypic records, for progeny testing in single-stage selection with 25 progeny records per test bull, and for three of the dam's own records were 0.4006, 0.6696, and 0.5, respectively; that is, the relative magnitude of reliability of GEBV without phenotypic records compared with that of progeny records or the dam's own records was higher for the low-heritability trait (trait 2; $h^2=0.05$) than for the moderately high-heritability trait (trait 1, $h^2=0.3$). Thus, the low-heritability trait demonstrated relative superiority in genetic gain for genomic selection, compared with single-stage selection without GEBVs, in agreement with Lande R and Thompson Rand Muir WM [31,32]. In contrast, the absolute (not relative) magnitude of genetic gain ($\Delta H/\text{year}$) in units of genetic standard deviation during generation 0 was highest for an economic weight of 3:1 among the assessed combinations of economic weights between the two traits (1:1, 3:1, and 1:3), in agreement with Togashi K and Lin CY [33]. This is because the asymptotic genetic variance of the aggregate genotype was largest for an economic weight of 3:1 (Table 2). Consequently, improving the aggregate genotype is easier and faster when high-heritability traits have greater economic weight than low-heritability traits.

Annual genetic gains (in units of genetic standard deviation during generation 0) for the component traits in correlated indices during two-stage selection were compared with those from single-stage selection (Table 5). The annual genetic gain for the first trait ($h^2=0.3$) during two-stage selection was nearly equal to that of single-stage selection. The reliability of GEBV for the first trait during generation 0 (that is, 0.4006) was set lower in the current study than in that of Togashi K, et al. [7,8]; consequently genetic gain for even moderately heritable traits ($h^2=0.3$) will be increased when the reliability of GEBV is set higher than in the current study. The annual genetic gains for the second trait ($h^2=0.05$) after two-stage selection were 53% to 97% greater than those from single-stage selection. This finding agrees with the trend shown in table 4, in which the low-heritability trait demonstrated relatively superior genetic gain for genomic selection compared with single-stage selection without GEBVs. In addition, this increase was greater when the proportion of young parents was 0.3 than when that proportion was 0.7, owing to the higher weighted accuracy for the first- and second-stage selections with young parents as 30% of all parents than that with young parents as 70%, as explained earlier (Table 4). This tendency was seen not only in the correlated indices but also in the independent indices during two stage selection.

The selection accuracies for the SS and SD pathways in the initial and asymptotic generations (Table 6) were lower in the first selection of genomic two-stage selection than in single-stage selection without GEBVs, which in turn were lower than in the second selection of genomic two-stage selection. This is because we set the reliability of GEBV during generation 0 lower in this study than in the work of Togashi K, et al. [7,8]. In addition, individuals at the first-stage of genomic selection do not have records, whereas progeny records or their own records are available for individuals at second-stage genomic selection and single-stage selection. Furthermore, selection accuracy was highest during the second genomic selection among all evaluated combinations of selection (first-stage genomic selection, second-stage genomic selection, and single-stage selection without GEBVs). This result occurred because the reliability of the GEBV at second-stage genomic selection was combined with that of the GEBV at first selection without records and that for progeny or own records at second stage.

Reliability in an asymptotic generation (that is, sixth generation) was greater at an economic weight of 3:1 than 1:1 and was greater still than that for an economic weight of 1:3. That is, the heavier the economic weight on the trait with moderately high heritability (that is, trait 1, $h^2=0.3$), the more accurate was selection in the asymptotic generation. The decrease in reliability from the first to the sixth generation was more marked for first-stage genomic selection without records than for single-stage selection based on records or for second-stage genomic selection based on GEBVs and progeny records or the parents' own records. This is because the reduction in the reliability of GEBVs results from the reduction in genetic variance over generations, whereas the reliability for single-stage selection is based on progeny records or parents' own records and that of second-stage genomic selection reflects the combination of GEBVs and records. In the current study, we assumed no renewal of the reference population during generations. Consequently, updating the reference population over generations is important to prevent a decrease in the reliability of GEBVs due to the reduction in genetic variance over generations. Conversely, selection accuracy in generation 0 at the second stage in the two-stage selection remained nearly the same even at an asymptote when young parents accounted for 70% of all parents; it was 89% to 96% of the accuracy in the first generation (Table 6). During practical efforts to increase annual genetic gain, the generation interval has a tendency to decrease [34]; that is, more young parents tend to be used to produce the next generation. Our current findings indicate that selection at an asymptote that allows for a young-parent population of 70% of all parents in a generation maintains almost the same selection accuracy as that of the initial generation, as long as progeny records or parents' own records are incorporated with GEBVs at the second-stage selection. Therefore, the importance of progeny records or parents' records in increasing or maintaining selection accuracy throughout generations is even greater in the genomic era than previously.

However, annual genetic gain ($\Delta H/\text{year}$) for two-stage selection was greater with a young-parent proportion of 0.3 than with young parents comprising 0.7 of all parents, owing to the higher weighted selection accuracy for the first- and second-stage selections with young parents as 30% of all parents than with the young parents as 70% (Table 4). Therefore, annual genetic gain in two-stage selection was influenced by the weighted accuracy between the first and second stages, the proportion of young parents among all parents, the generation interval, the selection intensity, and the genetic variance of the trait that is to be improved. It merits further study to examine these effects with a view toward maximizing annual genetic gain in two-stage selection.

Table 1: Genetic (co)variances for the first seven generations in the sire population with economic weights of 1:1 for two index traits.

	Generation							
		1	2	3	4	5	6	7
Single-stage selection ¹	σ_{G1}^2	0.3	0.2263	0.2222	0.2217	0.2216	0.2216	0.2216
	σ_{G12}	0.0367	0.0315	0.0299	0.0293	0.029	0.0289	0.0289
	σ_{G2}^2	0.05	0.0496	0.0493	0.0492	0.0492	0.0492	0.0492
Two-stage genomic selection ² , young parents=0.3	σ_{G1}^2	0.3	0.2139	0.2164	0.2162	0.2161	0.2161	0.2161
	σ_{G12}	0.0367	0.0257	0.0233	0.0226	0.0223	0.0222	0.0222
	σ_{G2}^2	0.05	0.0486	0.0476	0.0474	0.0473	0.0473	0.0473
Two-stage genomic selection, young parents=0.7	σ_{G1}^2	0.3	0.24	0.2363	0.2357	0.2356	0.2355	0.2355
	σ_{G12}	0.0367	0.0291	0.0269	0.0261	0.0259	0.0258	0.0257
	σ_{G2}^2	0.05	0.049	0.0484	0.0482	0.0481	0.0481	0.0481

¹Single-stage selection is based on records without GEBVs.

²Two-stage selection involves correlated indices for both stages.

Table 2: Genetic variance of the aggregate genotype (H) in the first and sixth generations.

Generation	Correlated indices for both stages				Independent indices for both stages			
	1	6	6:1 ¹	r_g^2	1	6	6:1	r_g
Economic weight	Single-stage selection							
1:1	0.423	0.329	0.776	0.881	0.423	0.329	0.776	0.881
3:1	2.970	2.211	0.744	0.863	2.970	2.211	0.744	0.863
1:3	0.970	0.815	0.840	0.917	0.970	0.815	0.840	0.917
	Two-stage selection (young parents, 0.3)							
1:1	0.423	0.308	0.727	0.853	0.423	0.331	0.782	0.885
3:1	2.970	2.117	0.713	0.844	2.970	2.287	0.770	0.877
1:3	0.970	0.739	0.762	0.873	0.970	0.785	0.808	0.899
	Two-stage selection (young parents, 0.7)							
1:1	0.423	0.335	0.791	0.890	0.423	0.347	0.819	0.905
3:1	2.970	2.314	0.779	0.883	2.970	2.398	0.807	0.899
1:3	0.970	0.794	0.818	0.904	0.970	0.816	0.841	0.917

¹The ratio of genetic variance in the sixth generation to that of the first generation

² r_g , The genetic correlation of H between the first and sixth generations

Table 3: Genetic gain of the aggregate genotype (ΔH) per generation in units of genetic standard deviation during generation 0 during the first six generations¹.

Generation	Correlated indices	Independent indices
Economic weight, 1:1		
1	1.386	1.279
2	1.053	1.088
3	1.040	1.053
4	1.035	1.045
5	1.033	1.044
6	1.033	1.043
Economic weight, 3:1		
1	1.479	1.395
2	1.054	1.142
3	1.056	1.101
4	1.054	1.095
5	1.054	1.094
6	1.054	1.094
Economic weight, 1:3		
1	1.231	1.154
2	1.021	1.019
3	0.979	0.981
4	0.970	0.971
5	0.968	0.968
6	0.967	0.967

¹The proportion of young parents to all parents in selection pathways SS (sires to breed sons), SD (sires to breed daughters), and DS (dams to breed daughters) is 0.3.

Correlated indices for two-stage selection assume that traits remain roughly normally distributed after first-stage selection. In this context, Cerón-Rojas JJ, et al. [3,4] showed that the null hypothesis (that is, non-normal distribution) was rejected when the first selection percentage was 0.22 or 0.55. In the current study, the selection percentages of the first selection in the SS (SD) and DS pathways were 20% and 40%, respectively-nearly the same as, or lower than, those of Cerón-Rojas JJ, et al. [3,4]. Weak first selection, such as in the current study, would not affect the normality of the GEBVs at second-stage selection.

Although further validation by stochastic simulation is required, covering a greater variety of parameters (e.g., reliability of GEBV, heritability, genetic correlation, economic weight, selection percentage, and generation interval), our methodology provides an easy means to deterministically evaluate the potential benefits of four-path two-stage selection based on GEBVs and to optimize selection strategies by using available marker data.

For most traits, the objective is continuous improvement, but for some traits, the goal is to reach an intermediate value; that is, maximum genetic improvement is not economically desirable for all traits. For example, cows with extremely high peak yield tend to experience increased stress and are less resistant to disease. Therefore, it is economically desirable to improve milk production but limit peak yield at the same time. The restricted index procedure has been

Table 4: Genetic gain of the aggregate genotype (ΔH) per year or generation in units of genetic standard deviation during generation 0 due to two-stage genomic selection compared with single-stage selection based on records without GEBV.

	Young parents, 0.3				Young parents, 0.7			
	ΔH /year		ΔH /generation		ΔH /year		ΔH /generation	
	ΔH^1	Ratio ²	ΔH^3	Ratio ⁴	ΔH	Ratio	ΔH	Ratio
Economic weight=1:1								
Independent index								
G⁵								
1	0.242	1.060	1.279	0.894	0.228	0.998	0.908	0.635
6	0.198	1.128	1.043	0.952	0.185	1.058	0.737	0.673
Correlated index								
1	0.263	1.149	1.386	0.969	0.240	1.049	0.954	0.667
6	0.196	1.117	1.033	0.943	0.186	1.062	0.739	0.675
Economic weight=3:1								
Independent index								
G								
1	0.264	1.058	1.395	0.893	0.247	0.989	0.982	0.629
6	0.207	1.110	1.094	0.937	0.194	1.037	0.770	0.659
Correlated index								
G								
1	0.280	1.122	1.479	0.947	0.256	1.025	0.954	0.667
6	0.200	1.069	1.054	0.903	0.191	1.023	0.739	0.675
Economic weight=1:3								
Independent index								
G								
1	0.219	1.198	1.154	1.011	0.204	1.119	0.812	0.712
6	0.183	1.238	0.967	1.045	0.172	1.162	0.685	0.739
Correlated index								
G								
1	0.233	1.278	1.231	1.079	0.213	1.165	0.845	0.741
6	0.183	1.238	0.967	1.045	0.173	1.170	0.689	0.744

¹ ΔH /year.

²The ratio of ΔH /year due to two-stage selection to that of single-stage selection.

³ ΔH /generation.

⁴The ratio of ΔH /generation due to two-stage selection to that of single-stage selection.

⁵G: generation.

developed to achieve a selection goal subject to various restrictions or constraints [35].

In the current study, we assessed both single- and two-stage selections. Single-stage selection was assumed as traditional selection such as progeny testing without genomic information. However, single-stage selection without second selection-that is, selection based on GEBVs derived from genotypes alone, without phenotypes-can be considered as an alternative method. Furthermore, the overall selection procedure currently used in dairy industry selection mixes the two-stage selection we followed in this study with single-stage selection based on GEBVs derived from genotypes alone, without phenotypes. Further study to extend the procedure we developed here is warranted. However, the procedure we developed serves as a reference point for better understanding the outcomes of the more complex selection currently practiced in the dairy industry.

Table 5: Genetic gain/year for each component trait in correlated indices for two stages and its comparison with that of single stage selection in units of genetic standard deviation in the first generation.

G ¹	young parents=0.3				young parents=0.7			
	$\Delta G1(h^2=0.3)$		$\Delta G2(h^2=0.05)$		$\Delta G1(h^2=0.3)$		$\Delta G2(h^2=0.05)$	
	$\Delta G1^2$	$\Delta G1^{*3}$	$\Delta G2$	$\Delta G2^*$	$\Delta G1$	$\Delta G1^*$	$\Delta G2$	$\Delta G2^*$
economic weight=1:1								
1	0.277	1.093	0.087	1.906	0.253	0.999	0.079	1.731
2	0.198	1.011	0.096	1.935	0.196	0.999	0.084	1.694
3	0.198	1.043	0.089	1.838	0.190	0.999	0.082	1.681
4	0.197	1.044	0.088	1.830	0.189	0.999	0.080	1.677
5	0.197	1.044	0.087	1.831	0.189	0.999	0.080	1.676
6	0.197	1.044	0.087	1.832	0.189	0.998	0.080	1.676
economic weight=3:1								
1	0.288	1.113	0.047	1.835	0.263	1.017	0.042	1.643
2	0.202	1.015	0.054	1.965	0.202	1.012	0.046	1.648
3	0.203	1.051	0.052	1.894	0.195	1.011	0.045	1.650
4	0.203	1.053	0.051	1.883	0.195	1.011	0.045	1.648
5	0.203	1.054	0.051	1.879	0.194	1.011	0.045	1.647
6	0.203	1.054	0.051	1.878	0.194	1.011	0.045	1.646
economic weight=1:3								
1	0.223	1.015	0.161	1.806	0.202	0.923	0.147	1.652
2	0.185	1.029	0.133	1.600	0.174	0.968	0.128	1.532
3	0.177	1.020	0.128	1.615	0.167	0.961	0.123	1.544
4	0.175	1.018	0.127	1.627	0.165	0.957	0.121	1.553
5	0.175	1.018	0.127	1.633	0.164	0.956	0.121	1.557
6	0.175	1.018	0.127	1.635	0.164	0.956	0.121	1.558
¹ G:Generation								
² $\Delta Gi = \Delta G$ per year for trait i								
³ $\Delta Gi^* = \Delta G$ for trait i in two stage selection/ ΔG for trait i in single stage selection.								

For creating a long-term goal for genetic improvement, the formula we developed here may be useful only for a large-scale producer or a region. However, when the traits are the same between the selection index and the aggregate genotype, selection index coefficients composed of GEBVs that are computed from MT-BLUP are the same as the economic weights of the aggregate genotype in any generation [7,8,18]. Hence, although the accuracy of GEBVs may vary from time to time, it is unnecessary to reconstruct the index to accommodate these changes. Nevertheless, it remains necessary to predict the genetic response due to selection according to changes in the accuracy of GEBVs. In addition, when the number of individuals saved (N) is small; the preceding formulae may overestimate the actual selection intensity because of sampling effects in finite population size. The expected selection intensity for a finite population size can be computed from the work of Burrows PM [36]. With truncation selection, the variance due to Mendelian sampling is unaffected by selection, but the expected reduction in the between-family component is reduced by the proportion 1/N owing to sampling of parents with replacement.

Goddard ME [37] developed a formula to calculate the accuracy of GEBVs under a conceptual framework where QTLs are in mutual linkage equilibrium (LE) and where a single marker is associated with each QTL. In this stylized setting, the entire genome is represented as independent QTL-marker pairs. When a large proportion of the markers used in the analysis are in LE with QTL and when linkage disequilibrium spans short regions, the accuracy of GEBVs will be overestimated [38]. In the example scenarios in which we applied the formula we developed in the current study, the accuracy of GEBVs during the initial generation was calculated under the same conceptual framework as that of Goddard ME [37] by using the formulas of Daetwyler HD, et al. [28] and Van Grevenhof EM, et al. [29]. On the other hand, a closer relationship between test and reference animals led to higher reliability [39-41]. Because Daetwyler HD, et al. [28] assumed that individuals in the reference population were not closely related to selection candidates, the accuracy of the GEBVs that we used here may be conservative. In addition, the use of model-based accuracy for animals with phenotypes, that is, the animals at the second-stage selection in this study, may have mitigated the potential overestimation of the accuracy of GEBVs because of the complement of polygenes in the phenotypic variance at second-stage selection. Furthermore, the formula for predicting genetic response that we developed in this study applies to any combination of accuracies of GEBVs and intensities of selection. Therefore, the formula presented here is a general equation for predicting genetic response over generations due to two-stage genomic index selection. However, caution needs to be exercised when setting accuracy of GEBVs in the initial generation considering the accuracy of GEBVs in real population. However, the purpose of this study is to develop new formula to predict genetic responses over generations selected on index composed of GEBVs rather than to examine the accuracy of GEBVs. The example in this study is to demonstrate our new formula which can deal with any combination of accuracies of GEBVs and intensities of selection.

Conclusion

Here, we developed formulas to calculate the responses from two-stage index selection based on GEBVs. The first formula involves correlated indices during two-stage selection, and the other uses independent indices. We then extended these formulas such that they could be used to predict genetic responses not only to a single generation but also to successive generations along four-path selection programs. We applied these formulas to several scenarios involving two index traits with differing heritabilities and relative economic weights. Then, we compared the asymptotic two-stage genomic selection responses (that is, responses when genetic (co)variances reached an asymptote) with those of single-stage selection based on records without GEBVs. The genetic response per generation to the aggregate genotype in the initial generation was greater for the correlated indices than for independent indices. However, the genetic response at an asymptote was slightly greater for the independent indices than for the correlated indices. The response per generation was smaller in two-stage selection, but the response per year was greater in two-stage selection than in single-stage selection. In addition, the relative superiority of the genetic response based on two-stage genomic index selection to single-stage selection without GEBVs was greater when the economic weight between the first ($h^2=0.3$) and second ($h^2=0.05$) traits was 1:3 than when it was 1:1 or 3:1. The reduction in the genetic variance of the aggregate genotype over generations was smaller for independent indices than for correlated indices. The

Table 6: Selection accuracy for SS, SD, DS, and DD pathways.

Economic weight		1:1			3:1			1:3		
Generation		1	6	6:1 ¹	1	6	6:1	1	6	6:1
Single-stage selection without GEBVs										
	SS	0.728	0.655	0.899	0.790	0.709	0.898	0.596	0.545	0.914
	SD	0.728	0.655	0.899	0.790	0.709	0.898	0.596	0.545	0.914
	DS	0.615	0.509	0.828	0.679	0.558	0.822	0.470	0.397	0.844
	DD	0.615	0.509	0.828	0.679	0.558	0.822	0.470	0.397	0.844
Two-stage genomic selection ² (young parents, 0.3)										
First selection	SS	0.574	0.392	0.683	0.613	0.375	0.612	0.509	0.398	0.782
	SD	0.574	0.392	0.683	0.613	0.375	0.612	0.509	0.398	0.782
	DS	0.574	0.424	0.739	0.613	0.421	0.687	0.509	0.417	0.819
Second selection	SS	0.793	0.742	0.936	0.864	0.783	0.905	0.692	0.652	0.943
	SD	0.793	0.742	0.936	0.864	0.783	0.905	0.692	0.652	0.943
	DS	0.738	0.662	0.897	0.807	0.701	0.868	0.631	0.577	0.914
	DD ³	0.717	0.646	0.901	0.767	0.681	0.888	0.618	0.562	0.908
Two-stage genomic selection (young parents, 0.7)										
First selection	SS	0.574	0.452	0.788	0.613	0.458	0.747	0.509	0.431	0.847
	SD	0.574	0.452	0.788	0.613	0.458	0.747	0.509	0.431	0.847
	DS	0.574	0.456	0.795	0.613	0.463	0.756	0.509	0.437	0.858
Second selection	SS	0.793	0.759	0.956	0.864	0.807	0.934	0.692	0.663	0.959
	SD	0.793	0.759	0.956	0.864	0.807	0.934	0.692	0.663	0.959
	DS	0.738	0.677	0.918	0.807	0.722	0.894	0.631	0.588	0.932
	DD*	0.717	0.659	0.920	0.767	0.698	0.910	0.618	0.573	0.927

¹The ratio of selection accuracy in the first generation to that of the sixth generation

²Two-stage selection involves correlated indices during both stages.

³DD* indicates single selection based on SNP genotypes+dam's own three records.

formulae that we developed here to predict genetic response applies to any combination of accuracies of GEBVs and intensities of selection. Therefore, the formula is a general prediction equation for computing genetic response over generations due to two-stage genomic index selection.

Conflict of Interest

The authors declare no conflict of interest.

References

1. VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91: 4414-4423.
2. Cotterill PP, James JW (1981) Optimising two-stage independent culling selection in tree and animal breeding. *Theor Appl Genet* 59: 67-72.
3. Cerón-Rojas JJ, Toledo FH, Crossa J (2019a) The Relative Efficiency of Two Multistage Linear Phenotypic Selection Indices to Predict the Net Genetic Merit. *Crop Sci* 59: 1037-1051.
4. Cerón-Rojas JJ, Toledo FH, Crossa J (2019b) Optimum and Decorrelated Constrained Multistage Linear Phenotypic Selection Indices Theory. *Crop Sci* 59: 2585-2600.
5. Xu SZ, Muir WM (1991) Multistage selection for genetic gain by orthogonal transformation. *Genet* 129: 963-974.
6. Xu S, Muir WM (1992) Selection index updating. *Theor Appl Genet* 83: 451-458.
7. Togashi K, Kurogi K, Adachi K, Tokunaka K, Yasumori T, et al. (2020a) Asymptotic response to four path selection due to index and single trait selection according to genomically enhanced breeding values. *Livestock Sci* 231: 103846.
8. Togashi K, Adachi K, Kurogi K, Watanabe T, Nurimoto, M, et al. (2020b) Asymptotic four-path genomic index selection response with or without accounting for the uncertainty of the predictions. *Livestock Sci* 240: 104139.
9. Hazel LN (1943) The Genetic Basis for Constructing Selection Indexes. *Genet* 28: 476-490.

10. Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. *J. Dairy Sci* 92: 4656-4663.
11. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, et al. (2010) Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* 93: 743-752.
12. Christensen OF (2012) Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet Sel Evol* 44: 37.
13. Buch LH, Kargo M, Berg P, Lassen J, Sørensen AC (2012) The value of cows in reference populations for genomic selection of new functional traits. *Anim* 6: 880-886.
14. Togashi K, Adachi K, Kurogi K, Yasumori T, Tokunaka K, et al. (2019) Effects of preselection of genotyped animals on reliability and bias of genomic prediction in dairy cattle. *Asian-Australas J Anim Sci* 32: 159-169.
15. Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. *J Anim Breed Genet* 124: 331-341.
16. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Pearson Education, Harlow, UK.
17. Henderson CR (1984) *Applications of linear models in animal breeding*. University of Guelph, Guelph, ON, Canada.
18. Schneeberger M, Barwick SA, Crow GH, Hammond K (1992) Economic indices using breeding values predicted by BLUP. *J Anim Breed Genet* 109: 180-187.
19. Kempthorne O, Nordskog AW (1959) Restricted selection indices. *Biometrics* 15: 10-19.
20. Cochran WG (1951) Improvement by means of selection. In: J. Neyman, (ed) *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA.
21. Young SSY (1964) Multi-stage selection for genetic gain. *Heredity* 19: 131-145.
22. Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* 123: 218-223.
23. Dekkers JCM (1992) Asymptotic response to selection on best linear unbiased predictors of breeding values. *Anim Prod* 54: 351-360.
24. Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31: 423-447.
25. Henderson CR (1982) Best linear unbiased prediction in populations that have undergone selection. *Proceedings of the world congress of sheep and beef cattle breeding*, New Zealand 1: 191-201.
26. Bijma P (2012) Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *J Anim Breed Genet* 129: 345-358.
27. Bulmer MG (1971) The effect of selection on genetic variability. *Am Nat* 105: 201-211.
28. Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3: e3395.
29. Van Grevenhof EM, Van Arendonk JAM, Bijma P (2012) Response to genomic selection: The Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. *Genet Sel Evol* 44: 26.
30. Saxton AM (1983) A comparison of exact and sequential methods in multi-stage index selection. *Theor Appl Genet* 66: 23-28.
31. Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genet* 124: 743-756.
32. Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet* 124: 342-355.
33. Togashi K, Lin CY (2010) Theoretical efficiency of multiple-trait quantitative trait loci-assisted selection. *J Anim Breed Genet* 127: 53-63.
34. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, et al. (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc Natl Acad Sci U S A* 113: E3995-E4004.
35. Togashi K, Lin CY (2004) Development of an optimal index to improve lactation yield and persistency with the least selection intensity. *J Dairy Sci* 87: 3047-3052.
36. Burrows PM (1972) Expected selection differentials for directional selection. *Biometrics* 28: 1091-1100.
37. Goddard ME (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136: 245-257.
38. de los Campos G, Sorensen D, Gianola D (2015) Genomic heritability: what is it? *PLoS Genet* 11: e1005048.
39. Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol* 42: 5-17.
40. Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci* 95: 389-400.
41. Wu X, Lund MS, Sun D, Zhang Q, Su G (2015) Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. *J Anim Breed Genet* 132: 366-375.

Appendix

I_y and I_{prs} are standardized, dividing I_y and I_{prs} by σ_{I_y} and $\sigma_{I_{prs}}$, respectively.

Standardized I_y and I_{prs} are termed as $st(I_y)$ and $st(I_{prs})$, respectively. Standardized I_y and I_{prs} are shown in figure 1.

First, we obtain the values of v_1 and v_2 from two independent standard normal distributions as truncation points for intended selection percentage of first and second selection, respectively.

The conditional distribution of $st(I_{prs})$ given $st(I_y) = v_1$ distributes with average of $r_{st(I_y),st(I_{prs})}$ and with standard deviation of $\sqrt{1-r_{st(I_y),st(I_{prs})}^2}$, where $r_{st(I_y),st(I_{prs})}$ is correlation between $st(I_y)$ and $st(I_{prs})$. On the other hand, v_2 is defined as the truncation point of a standard normal random variable with average of 0 and with standard deviation of 1 for the intended selection percentage of the second selection. So the truncation point of the conditional distribution of $st(I_{prs})$ given $st(I_y) = v_1$ is $r_{st(I_y),st(I_{prs})}v_1 + v_2\sqrt{1-r_{st(I_y),st(I_{prs})}^2}$. Consequently, that truncation point corresponds to that of a standard normal random variable with average of 0 and with standard deviation of 1, that is, $st(I_{prs})$. This truncation point for second selection is in agreement with Cerón-Rojas JJ, et al [3,4].

Consequently, selection intensity for $st(I_{prs})$ becomes $\frac{1}{\sqrt{2\pi}} \frac{\exp(-0.5u_2^2)}{p_2}$, where $u_2 = r_{st(I_y),st(I_{prs})}v_1 + v_2\sqrt{1-r_{st(I_y),st(I_{prs})}^2}$, $r_{st(I_y),st(I_{prs})} = r_{I_y I_{prs}}$ and p_2 is the selected proportion at the second-stage. Note that selection intensity for the second-stage selection varies during generations

because $r_{I_y I_{prs}}$ varies during generations, although the intended selection percentage of the second-stage selection is constant during generations.

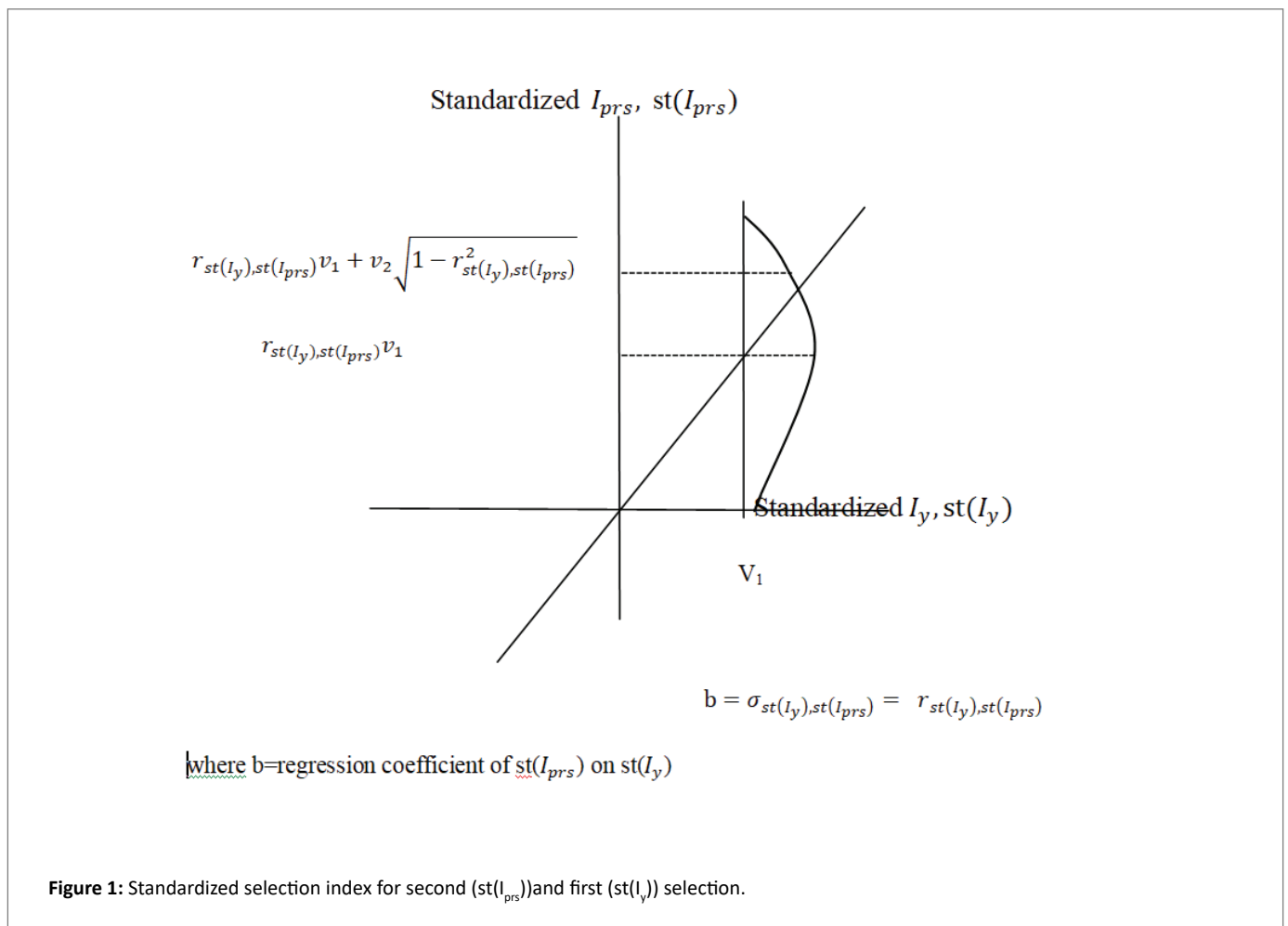


Figure 1: Standardized selection index for second ($st(I_{prs})$) and first ($st(I_y)$) selection.

$$\text{var}(I_{y_{DS,t}}) = [a_1, a_2, \dots, a_m] \text{Cov}(GEBV_{y_{DS,t}}) \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} =$$

$$[a_1, a_2, \dots, a_m] \begin{bmatrix} r_{GEBV1,y_{DS,t}}^2 \sigma_{G_{1,D,t}}^2 & r_{GEBV1,y_{DS,t}}^2 r_{GEBV2,y_{DS,t}}^2 \sigma_{G_{2,D,t}} & \dots & r_{GEBV1,y_{DS,t}}^2 r_{GEBV m,y_{DS,t}}^2 \sigma_{G_{m,D,t}} \\ \cdot & r_{GEBV2,y_{DS,t}}^2 \sigma_{G_{2,D,t}}^2 & \dots & r_{GEBV2,y_{DS,t}}^2 r_{GEBV m,y_{DS,t}}^2 \sigma_{G_{m,D,t}} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ symmetry & \cdot & \dots & r_{GEBV m,y_{DS,t}}^2 \sigma_{G_{m,D,t}}^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \dots\dots(a)$$

where $r_{GEBV i,y_{DS,t}}^2$ is the reliability of GEBV in $I_{y_{DS,t}}$ for the i^{th} trait in the dam population of generation t, $\sigma_{G_{i,D,t}}^2$ is the genetic variance

for the i^{th} trait in the dam population of generation t, and $\sigma_{G_{i,j,D,t}}$ is the genetic covariance for traits i and j in the dam population of generation t. In addition,

$$\text{var}(I_{prs_{DS,t}}) = [a_1, a_2, \dots, a_m] \text{Cov}(GEBV_{I_{prs_{DS,t}}}) \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix},$$

where $\text{cov}(GEBV_{I_{prs_{DS,t}}})_t$ is obtained by replacing I_y with I_{prs} in equation (a) and

$$\text{var}(I_{y_{DD,t}}) = [a_1, a_2, \dots, a_m] \begin{bmatrix} r_{GEBV1,y_{DD,t}}^2 \sigma_{G_{1,D,t}}^2 & r_{GEBV1,y_{DD,t}}^2 r_{GEBV2,y_{DD,t}}^2 \sigma_{G_{2,D,t}} & \dots & r_{GEBV1,y_{DD,t}}^2 r_{GEBV m,y_{DD,t}}^2 \sigma_{G_{m,D,t}} \\ \cdot & r_{GEBV2,y_{DD,t}}^2 \sigma_{G_{2,D,t}}^2 & \dots & r_{GEBV2,y_{DD,t}}^2 r_{GEBV m,y_{DD,t}}^2 \sigma_{G_{m,D,t}} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ symmetry & \cdot & \dots & r_{GEBV m,y_{DD,t}}^2 \sigma_{G_{m,D,t}}^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

where $r_{GEBV i,y_{DD,t}}^2$ is the reliability of GEBV in $I_{y_{DD,t}}$ for the i^{th} trait in the dam population of generation t.