Increasing Long Term Response by Selecting for Favorable Minor Alleles

C. Sun[1] and P.M. VanRaden[2]
[1] National Association of Animal Breeders, Columbia, MO, USA
[2] Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD, USA

Abstract

Long-term response of genomic selection can be improved by considering allele frequencies of selected markers or quantitative trait loci (QTLs). A previous formula to weight allele frequency of favorable minor alleles was tested, and 2 new formulas were developed. The previous formula used nonlinear weights based on square root of frequency of the favorable allele. The new formulas included a parameter δ to balance long- and short-term progress; one used simple linear weights instead of square root. The formulas were tested by simulation of 20 generations (population size of 3,000 for each generation) with direct selection on 3,000 QTLs (100 per chromosome) and a heavy-tailed distribution of allele effects. The optimum δ=0.4 from simulation was applied to actual dairy cattle data to compare differences of adjusted and official genomic evaluations. The previous nonlinear formula with δ=1.0 had slower response than unweighted selection in early generations and did not recover by generation 20. Long-term response was slightly greater with the new formulas than with unweighted selection; the linear formula may be best for routine use because of more progress in early generations compared to nonlinear formula. Official and adjusted U.S. evaluations based on actual genotypes and estimated marker effects were more highly correlated using linear weighting of allele frequency than nonlinear weighting. The difference between adjusted and official evaluations was highly correlated negatively with an animal's average genomic relationship to the population. Thus, strategies to reduce genomic inbreeding could achieve almost as much long-term progress as selection of favorable minor alleles.

Key words: genomic evaluation, long-term response, rare alleles, inbreeding

Introduction

Response to genomic selection can continue for many generations or decline rapidly, depending on the number of QTLs, their frequencies, linkage with markers, and effects on the trait or index selected. As genomic selection proceeds, allele frequencies may shift significantly, making long-term response difficult to predict because future genetic variance depends on future rather than current QTL allele frequencies. Genetic variance increases as frequencies of favorable alleles move from 0 toward 0.5, but decreases as their frequencies move from 0.5 to 1. Based on simulations (Muir, 2007) or deterministic predictions (Goddard, 2009), long-term gains from genomic selection can be less than from phenotypic selection or from selection on pedigree and phenotypes.

Long-term response can be improved by modifying the selection pressure applied to a QTL as its allele frequency changes, as demonstrated for 1 QTL in combination with phenotypic selection (Dekkers and van Arendonk, 1998) and for multiple QTLs using index selection (Jannink, 2010). The weight for each marker or QTL is adjusted according to its current frequency, with more weight given to markers that have a favorable allele with low frequency. Such methods can improve long-term response and will be referred to as favorable minor allele (FMA) selection.

This study proposes simple, improved formulas for weighting favorable minor alleles to increase long-term progress from genomic selection with less reduction of short-term progress. The formulas are applied to both simulated and real data, and correlated responses in genomic inbreeding are documented.

Methods

Favorable alleles with low frequency deserve more attention to increase genetic variance and avoid gene loss. For standard genomic selection, estimated genomic breeding values were calculated as

$$\hat{u}_i = \sum_j \hat{\beta}_j z_{ij}$$

where $\hat{u}_i$ is estimated breeding value for animal *i*, $\hat{\beta}_j$ is estimated allele effect for allele *j* and z_{ij} is a centered genotype. With FMA selection, $\hat{\beta}_j$ was replaced by β'_j (the weighted allele effect for allele *j*)

$$\hat{u}_i = \sum_j \beta'_j z_{ij}.$$

Two new formulas to implement FMA selection were derived as follows. The first used nonlinear weights and square root of frequency of the favorable allele as done by Jannink (2010) but also included a parameter δ that could vary from 0 to 1 to balance long- and short-term progress. The new formula is identical to Jannink's if $\delta = 1$. When $0 < f_j < 1$,

$$\beta'_j = \begin{cases} \hat{\beta}_j[1+(0.5 f_j^{-0.5}-1)\delta] & \text{if } \hat{\beta}_j \geq 0 \\ \hat{\beta}_j\{1+[0.5(1-f_j)^{-0.5}-1]\delta\} & \text{if } \hat{\beta}_j < 0 \end{cases}$$

otherwise, $\beta'_j = \hat{\beta}_j$. The second formula included a parameter δ that could vary from 0 to 2, but simple linear weights were used with more weight for favorable minor and less weight for favorable major alleles proportional to frequency difference from 0.5:

$$\beta'_j = \begin{cases} \hat{\beta}_j[1+(0.5-f_j)\delta] & \text{if } \hat{\beta}_j \geq 0 \\ \hat{\beta}_j[1+(f_j-0.5)\delta] & \text{if } \hat{\beta}_j < 0 \end{cases}.$$

Compared to the linear formula, the nonlinear formula puts less emphasis on alleles with intermediate frequency and more emphasis on extremely rare favorable alleles and is less similar to standard genomic selection. For both nonlinear and linear formulas, $\delta = 0$ corresponded to unweighted genomic selection.

Simulated Selection

Responses to 20 generations of selection were tested using the linear and nonlinear weighting formulas with δ that ranged from 0 to 1. Values of $\delta > 1$ also were tested but provided only losses and no benefits within 20 generations of selection and thus are not shown. A group of 30 bulls and 100 females with pedigrees identical to a group of recently genotyped Holsteins was used to generate 3,000 animals as the first generation for selection. In each subsequent generation, the top 100 males and top 1,000 females were selected and mated randomly to produce 1,500 males and 1,500 females in the next generation. Genotypes were simulated with program genosim.f90 (VanRaden *et al.*, 2011) for 30 chromosome pairs with a length of 1 Morgan each. Initial linkage disequilibrium was generated in the base population, followed by inheritance with recombination in the known, actual pedigree generations and in the next 20 simulated generations. Computation was reduced by using direct selection on 3,000 QTL effects (100 per chromosome) instead of indirect selection on estimated marker effects. Allele effects of QTLs had a heavy-tailed distribution, with the largest effect contributing about 5% of genetic variance. Initial allele frequencies were uniformly distributed from 0 to 1 and were independent of effect size.

Actual Population

Actual genotypes and U.S. marker effect estimates for net merit were used to compare official genomic evaluations from June 2013 with FMA selection. The genotyped animals included 349,572 Holsteins, 41,731 Jerseys, and 8,300 Brown Swiss. Each animal had actual or imputed genotypes for 45,188 SNP markers. The linear and nonlinear formulas were both applied with the parameter value for δ set to 0.4 based on the optimum from simulated data or set to 0 to obtain official rankings. The FMA evaluations were standardized to have the same mean and standard deviation as official evaluations. Evaluation differences (FMA minus official)

were examined for individual animals, and correlations were obtained with expected future inbreeding (EFI; half an animal's mean pedigree relationship to its breed) and genomic future inbreeding (GFI; half an animal's mean genomic relationship to its breed).

Results and Discussion

Simulated Selection

Parameter δ was needed to avoid excessive short-term loss from putting too much emphasis on long-term selection. Simulation results showed that the square root formula of Jannink (2010) and the linear formula with δ = 1 both had large losses in early generations and did not recover these losses within 20 generations. Therefore, the remaining simulations focused on optimizing δ to balance long- and short-term progress.

Maximum response by generation 20 was achieved with δ = 0.4 and δ = 0.6 using nonlinear and linear FMA selection, respectively, but losses were larger in the first few generations with δ = 0.6 than with δ = 0.4 (Figure 1). Long-term response was slightly greater with the nonlinear formula, but at a higher cost in early generations. The linear formula might be best for routine use because few breeders can afford a 20-generation planning horizon.

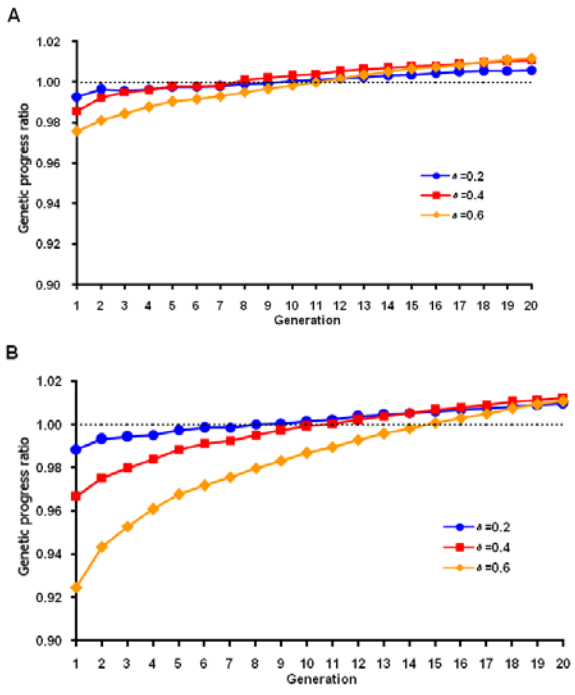

Figure 1. Ratio of adjusted to unadjusted genetic progress by generation for a heavy-tailed QTL distribution. The ratio was calculated as the genetic progress for a simulated population based on adjusted genomic breeding value using various δ in the linear (A) and nonlinear (B) adjustment formula divided by genetic progress based on genomic breeding value from unweighted selection.

Genetic variance decreased across generations as selection proceeded. More genetic variance was maintained across generations by FMA selection (as expected from theory; Figure 2), and higher δ preserved more genetic variance. The linear formula preserved less variance but had higher means than the nonlinear formula in early generations. Jannink (2010) argued that the most immediate cause of the plateau reached by standard genomic selection was the loss of genetic variance, which was more pronounced for small populations.

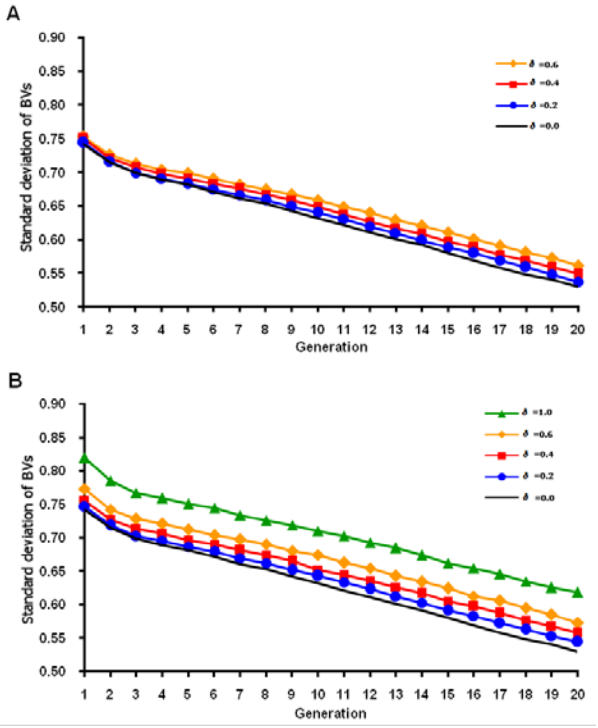

Figure 2. Standard deviation of true breeding value by generation based on a heavy-tailed QTL distribution. True breeding values (BVs) for a simulated population were based on unweighted (δ = 0) or weighted (various δ) genomic selection and calculated using true marker effects. Linear (A) and nonlinear (B) formulas were used to weight allele frequency.

Mean inbreeding coefficients for animals in the last generation were calculated using different allele frequencies (Table 1). Slightly higher genomic inbreeding was found for larger values of δ when true allele frequency was used with both linear and nonlinear FMA selection; inbreeding was slightly lower when using an allele frequency of 0.5 for each locus or using pedigree inbreeding. With FMA selection, larger values of δ preserved more variance and heterozygosity but were not optimal because they slowed fixation of favorable major alleles that deserved to be fixed more quickly.

At the first generation, pedigree and genomic inbreeding using true allele frequency were about 5 and 8.5%, respectively; however, after 20 generations, genomic inbreeding was much higher than pedigree inbreeding, which corresponds to the previous study (Sonesson *et al.*, 2012). Sun *et al.* (2013) developed mating programs by combining the selection and mating steps of optimum contribution theory using linear programming and reported that expected progeny values and progeny inbreeding were improved using genomic breeding values and genomic relationship compared with other strategies that combine breeding values (genomic or traditional BLUP) and relationship matrices (genomic or pedigree).

Table 1. Mean inbreeding coefficients in the final generation calculated using different allele frequencies for simulated populations using 2 QTL distributions

Method	δ	0.500[a]	True[b]	Pedigree[c]
Linear	0.00	0.456	0.256	0.093
	0.20	0.448	0.260	0.092
	0.40	0.441	0.264	0.091
	0.60	0.433	0.269	0.089
Nonlinear	0.20	0.445	0.258	0.092
	0.40	0.433	0.262	0.090
	0.60	0.422	0.266	0.088
	1.00	0.399	0.275	0.085

[a]Mean of diagonal elements of genomic relationship matrix calculated using an allele frequency of 0.5.
[b]Mean of diagonal elements of genomic relationship matrix calculated using true allele frequency in the base population.
[c]Inbreeding based on pedigree information.

Actual Population

Official and FMA evaluations were correlated by 0.994 for Holsteins and Jerseys and by 0.989 for Brown Swiss using linear weighting of allele frequency applied to all animals. Correlations were lower (0.991 for in Holsteins, 0.986 for Jerseys, and 0.978 for Brown Swiss) when nonlinear weighting was applied. If only U.S. animals born in the most recent 5 years were included instead of all animals, Holstein and Jersey correlations did not change, but Brown Swiss correlations were much higher (0.999 with linear and 0.997 with nonlinear weighting). Brown Swiss correlations were higher because most Brown Swiss genotypes are from Europe and include animals with mixed or pure European ancestors that have been separate from the U.S. population for about 25 generations; recent U.S. animals have few European ancestors.

For all 3 breeds, the difference between FMA and official evaluation was highly negatively correlated with GFI but much less correlated with EFI. For recent U.S. animals, the correlations of GFI with evaluation difference were −0.85 for Holsteins, −0.94 for Jerseys, and −0.85 for Brown Swiss with linear weighting and δ = 0.4; correlations of EFI with evaluation difference were only −0.45 for Holsteins, −0.59 for Jerseys, and −0.27 for Brown Swiss. The GFI and EFI correlations changed very little with nonlinear instead of linear weighting. Much of the benefit from FMA selection could be obtained simply by selecting for lower GFI in combination with higher GEBV or by using optimum contribution theory to reduce genomic inbreeding (Sonesson *et al.*, 2012).

The largest differences between FMA and official evaluations were for animals with the lowest or highest GFI (as expected from the highly negative correlations). Animals that gained the most from FMA evaluation were those with ancestors from another breed or from a foreign subpopulation of the same breed.

Breeders have long known that long-term progress can be higher with avoidance of inbreeding, marker-assisted introgression of

favorable alleles from other breeds, or formation of synthetic composites instead of pure breeds. Simulation of FMA selection within a breed indicates only a small (~1%) benefit over 20 generations, but benefits could be larger with across-breed selection or with individual QTLs that explain >5% of genetic variance. Animals with lower genomic relationship to the current population may be more valuable than standard genomic selection assigns, but breeders may need incentives to include those animals in selection programs. The simulation considered only additive effects, and conclusions may differ for QTLs with nonadditive genetic effects. The main benefit of FMA selection is that both the mean and genetic variance in future generations is considered when ranking candidates in the current generation.

Conclusions

Short- and long-term progress were balanced using new formulas for FMA selection. Previous formulas put too much emphasis on rare favorable alleles and resulted in less progress than standard genomic selection over 20 simulated generations. Optimal value was 0.4 for δ when allele effects of QTLs had a heavy-tailed distribution. The linear formula increased long-term response with fewer losses in the first few generations, and could be used for routine evaluation. For actual genotypes and estimated marker effects from U.S. evaluations, individual animal differences between FMA and standard genomic selection were highly correlated to the animal's average genomic relationship to the population. Thus, strategies to reduce genomic inbreeding could achieve almost as much long-term progress as FMA selection.

References

Dekkers, J.C.M. & van Arendonk, J.A.M. 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations. *Genet Res Cambr. 71,* 257–275.

Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica 136,* 245–257.

Jannink J.L. 2010. Dynamics of long-term genomic selection. *Genet Sel Evol 42,* 35-45.

Muir, W.M. 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.

Sonesson, A.K., Woolliams, J.A. & Meuwissen, T.H.E. 2012. Genomic selection requires genomic control of inbreeding. *Genet Sel Evol 44*, 27.

Sun, C., VanRaden, P.M., O'Connell, J.R., Weigel, K.A. & Gianola, D. 2013. Mating programs including genomic relationships and dominance effects. *J Dairy Sci* (Accepted).

VanRaden, P.M., O'Connell, J.R., Wiggans, G.R. & Weigel, K.A. 2011. Genomic evaluations with many more genotypes. *Genet Sel Evol 43,* 10.