University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

2007

Combining genetic test information and correlated phenotypic records for breeding value estimation

Matthew L. Spangler
University of Nebraska-Lincoln, mspangler2@unl.edu

J. Keith Bertrand
University of Georgia, adshead@uga.edu

R. Rekaya
University of Georgia, rrekaya@uga.edu

Spangler, Matthew L.; Bertrand, J. Keith; and Rekaya, R., "Combining genetic test information and correlated phenotypic records for breeding value estimation" (2007). *Faculty Papers and Publications in Animal Science*. 798.
https://digitalcommons.unl.edu/animalscifacpub/798

Combining genetic test information and correlated phenotypic records for breeding value estimation

M. L. Spangler,[1]* J. K. Bertrand,* and R. Rekaya*†‡

*Animal and Dairy Science Department, †Department of Statistics, and ‡Institute of Bioinformatics, University of Georgia, Athens 30602-2771

ABSTRACT: The use of marker assisted selection in the beef cattle industry to date has involved using traditional EPD in tandem with molecular test information. In the current study, a multiple-trait simulation was carried out to create a beef cattle data set using genetic parameter estimates from the literature to identify the best procedure for combining both sources of information and to assess the added benefit of the procedure. To reach these objectives, the following simulation/analysis steps were implemented: (1) varying percentages (100, 5, or 0) of available records for the trait of interest, (2) varying percentages (100, 50, 25, or 0) of animals with molecular information, (3) scenarios where the favorable (F) or the unfavorable (U) allele was more frequent, and (4) analysis of the response due to selection over 5 generations. The data sets included 3 correlated traits in which 2 of them, birth weight and postweaning gain, had complete recording and the availability of records for the third trait (marbling score) varied. It was further assumed that molecular information was available for the third trait for a causative gene that explained 10% of the genetic variation.

Estimates of Pearson correlations between true and predicted breeding values for marbling score declined as the amount of information declined, and instances in which the molecular information was recorded were always closer to the true values than in the case in which the molecular information was absent. When the U allele was more frequent, rank correlation estimates were increased among top sires, low accuracy sires, and high accuracy sires by approximately 24.9, 12.1, and 4.7% with limited marbling score records and complete genotyping compared with limited marbling score records and no genotyping. Similar results were seen when the F allele was more frequent. When there was a complete absence of recording for the trait of interest, the same trends in correlations were observed and were lower than when the trait of interest was recorded. Jointly considering molecular and phenotypic information showed a greater long-term response compared with tandem selection, showing that discrimination of candidates for selection based solely on molecular information is not optimal.

Key words: beef cattle, best linear unbiased prediction, gene-assisted selection

INTRODUCTION

In beef cattle breeding, producers are supplied with traditional EPD and can opt to test animals for the presence or absence of genes that are believed to have large effects on some secondary traits. Given the complexity of these traits, the identification of causative genes alone will not explain the majority of the observed variation and thus will not be an efficient selection tool. Unfortunately, selection decisions using these 2 sources of information are often made by those without technical expertise (Shafer, 2005).

Soller (1978) was among the first to discuss the uses of molecular information and suggested the preselection of animals before progeny tests based on molecular information. Fernando and Grossman (1989) proposed BLUP methodology for obtaining breeding values from a mixed inheritance model. Although several attempts have been made (e.g., Meuwissen and Goddard, 1996) to combine phenotypic and molecular information, the practical impact was limited because of the theoretical and computational complexities.

Currently in beef production, traits like marbling and tenderness are recorded on a limited basis due to the difficulty and expense of collecting these carcass measurements. In the case of sparsely recorded traits, additional information can be garnered not only through molecular information but also through genetic relationships with already available production

traits. Although the collection of carcass traits can be expensive and cumbersome, the collection of correlated traits like birth weight or weaning weight is easily accomplished and routinely done. With this in mind, additional information about difficult-to-measure traits may be available, through genetic correlations, at no additional cost.

Consequently, the objective of the current study was to devise a method of combining molecular and phenotypic sources of information to compute a single breeding value that can be used in a straightforward manner.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the data were obtained from a simulated database.

Let $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', ..., \mathbf{y}_t')'$ be a vector of phenotypic records for t quantitative traits. Without loss of generality, assume that a major gene is segregating only for the last trait. The major gene is assumed to have l different codominant alleles, $A_1, A_2, ..., A_l$, with relative frequencies $p_1, p_2, ..., p_l$, where $\sum_{i=1}^{l} p_i = 1$. For a population in Hardy-Weinberg equilibrium, the genotypic frequencies of the $l(l + 1)/2$ distinguishable genotypes are the product of the corresponding allelic frequencies.

Following the structure presented by Wu et al. (2002), the statistical model that describes the relationship between phenotypes, genotypes, and polygenic effects could be presented in the following hierarchical structure:

$$\mathbf{y}|\beta, \mathbf{g}, \mathbf{a}, \mathbf{u}, \mathbf{R}_0 \sim N(\mathbf{X}\beta + \mathbf{Wa} + \mathbf{Zu}, \mathbf{I} \otimes \mathbf{R}_0),$$

$$g_i|g_{s_i}, g_{d_i} \sim \prod f(g_i|g_{si}, g_{d_i}), \text{ and}$$

$$(g_{si}, g_{di}) \sim \prod h(g_{si}) \prod h(g_{di}),$$

where β, $\mathbf{a}$, and $\mathbf{u}$ are the vectors of systematic effects, major gene effects, and polygenic effects, respectively; $\mathbf{X}$, $\mathbf{W}$, and $\mathbf{Z}$ are the corresponding known incidence matrices; and g_i, g_{si}, and g_{di} are the genotypes for the major gene for animal i, sire of animal i, and dam of animal i, respectively. Function f indicates the Mendelian segregation patterns between progeny and parents, and h is a multinomial distribution with known probability vector determined by the frequencies of the l alleles of the major gene.

Given that a major gene was assumed segregating only for the last trait, the additive effects could be written as

$$\mathbf{u}_t^* = \mathbf{a} + \mathbf{u}_t,$$

with $\mathbf{a}$ and $\mathbf{u_t}$ representing the major gene and polygenic gene effects for trait t, respectively.

Furthermore, assuming independence between the major gene and polygenic effects, the following equality holds true:

$$\text{var}(\mathbf{u}_t^*) = \text{var}(\mathbf{a} + \mathbf{u}_t) = \text{var}(\mathbf{a}) + \text{var}(\mathbf{u}_t).$$

Additionally, it was assumed that the major gene effect was independent of the polygenic effects in the first $(t\text{-}1)$ traits. Thus, if G_0 is the matrix of genetic (co)variances among all traits, and the major gene explains c percentage of the total additive variance for the last trait, then the resulting genetic (co)variance matrix between the 3 polygenic effects, G_0^*, will be

$$G_0 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & . & . & \sigma_{1t} \\ & \sigma_2^2 & . & . & \sigma_{2t} \\ & & . & . & . \\ & & & . & . \\ & & & & \sigma_t^2 \end{bmatrix} G_0^* = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & . & . & \sqrt{(1-c)}\sigma_{1t} \\ & \sigma_2^2 & . & . & \sqrt{(1-c)}\sigma_{2t} \\ & & . & . & . \\ & & & . & . \\ & & & & (1-c)\sigma_t^2 \end{bmatrix}.$$

In other words, the variance and associated (co)variances of the trait for which there is a causative gene are reduced according to the amount of variation explained by the gene.

Unknown Genotypes

Given the hierarchical structure of the genotypes presented above, and that genotypes in this study were assigned at random in the population, it is possible to extract additional genotypic information from the pedigree. Animals with missing genotypic information can be assigned 1 or both alleles given parental, progeny, or mate information. Given this trio of information sources and following an algorithm similar to Qian and Beckmann (2002) and Tapadar et al. (2000), imputation on missing genotypes were made and additional genotypic information was garnered.

Simulation

A multiple-trait simulation was carried out using 3 traits. The simulation was designed to mimic a beef cattle data set. The traits included were birth weight, postweaning gain, and the trait of interest, marbling score (**MS**). The genetic parameters for these 3 traits were within the bounds of published values (Woodward et al., 1992; Marshall, 1994; Splan et al., 1998). The genetic and residual correlations between traits are given in Table 1.

The phenotypic means (variances) used in the simulation were 40 (25), 1 (0.1), and 3.7 (0.20) for birth weight, average postweaning gain, and MS, respectively. Additionally, the simulation model included the fixed effect of herd (300 levels). Herd effects were gen-

Table 1. Parameters for the simulated traits[1]

	BrW	PWG	MS
BrW	0.3	0.25	−0.20
PWG	0.6	0.3	0.40
MS	0.35	0.35	0.3

[1]Heritability is on the diagonal; genetic correlations are on the upper off-diagonal; and residual correlations are on the lower off-diagonal. BrW = birth weight; PWG = average postweaning gain; and MS = marbling score.

Table 2. Description of the data sets

Data set	Favorable allele frequency	Marbling score record[1]	Molecular information[2]
U-100-100	0.4	100	100
U-5-100	0.4	5	100
U-5-50	0.4	5	50
U-5-25	0.4	5	25
U-5-0 (Tandem)	0.4	5	0
F-100-100	0.6	100	100
F-5-100	0.6	5	100
F-5-50	0.6	5	50
F-5-25	0.6	5	25
F-5-0	0.6	5	0
U-0-100	0.4	0	100
U-0-50	0.4	0	50
U-0-25	0.4	0	25
U-0-0	0.4	0	0

[1]Percentage of animals in the data file with a marbling score record.
[2]Percentage of animals in the data file and pedigree with molecular information recorded.

erated from U[40, 65], U[1, 1.1], and U[3.7, 3.9] for birth weight, average postweaning gain, and MS, respectively. Boundaries for these uniform distributions were set heuristically to generate phenotypes that are similar in magnitude to the observed responses for these traits. Marbling score was assumed to be recorded such that traces = 3.00, slight = 4.00, small = 5.00, modest = 6.00, and moderate = 7.00. The causative gene was assumed to be biallelic and to account for 10% of the genetic variation in MS. The frequencies were assumed to be 0.6 and 0.4 for the favorable (**F**) or unfavorable (**U**) alleles, respectively. A latter scenario explored the case when the allele frequencies were reversed. The values for the fixed gene effect were heuristically set to −0.15, 0, and 0.15, for the U homozygote, heterozygote, and F homozygote, respectively, meaning that these values, on average, provided 10% of the genetic variance in MS such that the total variance of MS (gene effect plus polygenic effects) was achieved.

For continuous traits, as is the case in this study, a 3-trait mixed linear model (Henderson and Quaas, 1976) was implemented. For the first 2 traits, no molecular information was assumed available and only the polygenic effects were included in the model. For the third trait, a major gene and the remaining polygenic effects were considered.

In matrix notation, the following mixed inheritance model was used:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{ZWa} + \mathbf{e},$$

where **y** is the vector of phenotypic observations for the 3 traits; β and **u** are the vectors of fixed and random effects, respectively; **a** is the vector of genotype means (included in the model only for the third trait); and **e** is the vector of residual terms. The **X** and **Z** are known incidence matrices, and **W** is a matrix that includes the genotype of each animal. For the random effects, it was assumed that:

$$\mathbf{u} \sim N(\mathbf{0},\mathbf{A} \otimes \mathbf{G}) \text{ and } \mathbf{e} \sim N(\mathbf{0},\mathbf{I} \otimes \mathbf{R}),$$

where **G** and **R** are 3×3 genetic and residual (co)variance matrices, respectively, and A is the additive relationship matrix.

Data Sets for Analysis

A detailed description of the data sets created for analysis can be found in Table 2. The allelic frequencies, amount of phenotypic data for the trait of interest, and the number of animals for which molecular information was recorded were varied. Animals with missing records or missing genotypes were randomly assigned. Data sets were labeled to represent the given scenario, with the F or U allele was more frequent, varying percentages (100, 5, or 0) of recorded phenotypes for the trait of interest, and varying percentages (100, 50, 25, or 0) of animals with molecular information. For example, data set U-5-100 would represent the case in which the U allele is at a frequency of 0.6, 5% of the MS records are observed, and all animals in the pedigree have molecular information. There are an infinite number of scenarios that could be explored by changing allele frequencies, observed genotypes, and observed phenotypes. However, the purpose of the current study was to explore the differences in extreme scenarios regarding the availability of molecular and phenotypic information, and to explore what appears to be a plausible scenario for the presence of MS records.

The beef cattle industry is currently in a situation in which traits such as MS are recorded sparsely in breed association databases. This could be due to the relative newness of these traits compared with traditional traits such as birth weight or weaning weight. The cost and difficulty of recording, via carcass data or live animal ultrasound, surely contributes to this problem as well. Consequently, a simulation scenario that allows for 5% of the MS records to be observed is certainly germane. It is possible that there may be subsets of populations in which no records for MS would be observed. However, the use of correlated traits to predict missing MS records is not possible

unless a minimal amount of these records are observed to ascertain the genetic relationships of the traits of interest with other production traits.

The pedigree file consisted of 4 generations with 1,000 animals in the base population and 3 subsequent generations consisting of 3,000 animals each, for a total of 10,000 animals of which 9,000 had records and were distributed in 300 herd classes. The correlation estimates were averaged over 10 replications. To estimate the unknown genetic parameters, Gibbs sampling was utilized. A long chain of 20,000 iterations with a burn-in period of 5,000 iterations was implemented. A postGibbs analysis showed that the 15,000 rounds were sufficient for the prediction of breeding values in this study. As indicated earlier, missing genotypes were predicted from known parental genotypes and from known progeny genotypes. Animals in the base population for which the genotype could not be predicted by progeny information were assigned alleles based on the assumed allelic frequencies.

Genetic Progress Analysis

Genetic progress over 5 generations of selection was assessed using the simulated data sets created from the scenario in which the U allele was more frequent. Although 5 generations may appear to be short-term selection, it seems misleading to study the genetic progress over extremely long periods of time because breeding goals and objectives change rapidly. Further, after 5 generations of selection pressure for increased MS, the F allele is close to fixation. Each generation, 3,000 more animal records were created so that by the end of generation 5 the pedigree included 25,000 animals and the data file included 24,000 animals with records. Additionally, each generation 100 herd classes were added. In the case of the data sets U-100-100, U-5-100, U-5-50, and U-5-25, dams for the next generation were selected from the 3 previous generations (9,000 animals). Dams were further discriminated against based on their predicted MS breeding value.

It is logical to assume that producers would set a threshold breeding value for the trait of interest and select animals from within that group perhaps based on other traits (e.g., health, structural correctness, disposition, etc.) rather than selecting the top 3,000 females. Although these traits were not simulated, to mimic this scenario dams were chosen at random from those who were in the top 67% of their predicted MS breeding value. This random draw from the pool of 6,000 dams could be thought of as selection among females of acceptable genetic merit for other phenotypic traits (e.g., weight, structural correctness, disposition, etc.). Consequently, 6,000 females were candidates for selection, of which 3,000 contributed to the next generation. Approximately 30% of these females were from the last generation and could be considered as replacement females. The replacement rate of 30% used here is slightly greater, but reasonably close, to the equilibrium replacement rate of 21.19% described by Greer et al. (1980). These females were then mated to the top 100 sires, chosen from all possible sires outside of the base population (i.e., first 1,000 animals) as determined by their predicted breeding value for MS. Selection of males and females was done without consideration of the accuracy associated with the predicted breeding value.

For the case of tandem selection, females and sires were chosen based on their genotype for the causative gene alone, such that animals having the undesirable genotype were not considered regardless of their polygenetic merit. In the early stages of selection there were not sufficient numbers of animals having the desirable genotype, so candidates for selection included those homozygous for the F allele and those who were heterozygotes but in the top 67% for their predicted breeding value for MS.

The analysis was carried out as previously stated with the exception of U-5-50 and U-5-25. In these 2 cases, the allele frequency used to assign genotypes to missing animals in the base population were updated each generation such that the original frequency of 0.6 for the U allele was changed to reflect the new allelic frequency in the entire population each generation. This change was made because the underlying assumption was that information concerning allelic frequencies in previous generations is unknown and the only known information is the molecular information currently available in the updated data file.

The results, MS breeding values and allelic frequencies, were averaged over 10 replicates. The average breeding values and allelic frequencies reported by generation are the average of the true values from all animals in the last generation (the last 3,000 animals) with the exception of generation 0, which is the average over the 10,000 animals in the pedigree before selection.

RESULTS AND DISCUSSION

Minimal Records for Marbling Score

Pearson Correlations. Correlation estimates between true and predicted MS breeding values from the analysis of the differing data sets when the U allele is more frequent can be found in Table 3. As expected, U-100-100 yielded the highest Pearson correlation among all animals with the simulated true breeding values. This is due to the fact that all available information, complete phenotypic records for all 3 traits and complete molecular information, is used for the prediction of breeding values. Of more interest are the differences between the remaining data sets. The only difference between these data sets is the amount of molecular information available, and thus the correlations increase with the number of animals genotyped. There is a 20.8% increase in the accuracy of prediction

Table 3. Correlation estimates between the true values and those predicted among data sets when limited marbling score (MS) records were observed[1]

Item[2]	Data set[3]				
	U-100-100	U-5-100	U-5-50	U-5-25	U-5-0
All	0.694	0.493	0.479	0.459	0.408
Sires with >100 progeny	0.859	0.755	0.753	0.733	0.721
Sires with <20 progeny	0.816	0.591	0.580	0.555	0.527
Top sires	0.598	0.421	0.398	0.375	0.337
Bottom sires	0.511	0.294	0.274	0.256	0.244

[1]The estimates are the average over 10 replicates.

[2]All = Pearson correlations among all animals; top sires = Spearman rank correlations among sires within the top 20% for their true MS breeding value; bottom sires = Spearman rank correlations among sires within the bottom 20% for their true MS breeding value.

[3]The initial letter corresponds to which allele is more frequent (U = unfavorable), the first number corresponds to the percentage of available records for MS, and the second number corresponds to the percentage of animals with molecular information.

when all animals are genotyped compared with the case when no animals are genotyped. A decided advantage of the inclusion of molecular information still remains when only 25% of the animals have molecular information (approximately 12.5% increase vs. no molecular information). This is certainly encouraging, given the fact that limited molecular information, assigned to animals at random, can lead to more accurate estimates of breeding values when compared with traditional methods. These results are only for 1 generation and are in agreement with short-term response results from Verrier (2001) who used marker assisted selection in a 2-trait analysis. The author found a slight increase in the accuracy (0.8 to 3%) of breeding value prediction using marker assisted selection for the trait with missing values, depending on the recombination rate, over traditional BLUP values averaged over 4 generations of selection. Verrier (2001) found that the advantage of marker assisted selection, or breeding value prediction in general, was aided by the use of correlated records when 1 trait had missing values. Similar results in early generations were found by Villanueva et al. (1999) who reported an advantage in the total genetic gain using truncation selection when major gene information was used compared with when it was ignored in early generations.

Rank Correlations. The possibility of the reranking of certain pertinent subgroups of any population is of perhaps more interest when considering the selection of parents for the next generation. Sires with more than 100 offspring in the pedigree file were considered to be influential, thus the rank correlation between their true values and those predicted from the analysis of the various scenarios is recorded in Table 3. These twenty-two animals sired an average of 136 offspring. The differences are minor, but as in the case with the Pearson correlations, the largest estimate comes from the case when all information is recorded and declines with the loss of information. These small differences can be explained by the fact that these sires have numerous offspring and consequently are high accuracy

sires. The fact that the estimates for the rank correlations between U-5-100 and U-5-50 are similar (0.755 and 0.753) is not surprising. Although only 50% of the genotypes were observed in U-5-50, approximately 80% of the animals had their genotype correctly predicted during the analysis due to the use of pedigree information.

Similarly, Spearman rank correlations were estimated among sires that had less than 20 offspring. These 60 sires would generally be younger animals and associated with less confidence in their breeding value. On average these younger sires had 15 progeny. Again, the estimates declined with the decline in information available. There is an approximate 4.5% increase in the rank correlation in favor U-5-50 over U-5-25, and a 5.3% increase in favor of U-5-25 over U-5-0. This shows that the inclusion of limited molecular information could aid in the accurate selection of young sires and perhaps allow for their use at earlier ages, thus reducing the generation interval.

The top sires, sires in the top 20% for their true breeding for MS, and the undesirable sires, those in the bottom 20% for their breeding value for MS, are of the most interest to those looking to select future parent stock. Again, the results show an increase in the correlation along with the increase in information. Of the most significance is the 11.3% increase in the rank correlation among top sires between the estimated and the true value when only 25% of the genotypes are known when compared with the case when genotypic information is unknown. This shows that the inclusion of molecular information in the prediction of breeding values can significantly aid in the selection of the best candidate sires.

Avoiding the sires that offer the least potential for genetic improvement for the trait of interest, in this case MS, is also important. In this comparison, the most significant differences in rank correlations were 7.3 and 7.0% between U-5-100 and U-5-50, and between U-5-50 and U-5-25, respectively.

Table 4. Correlation estimates between the true values and those predicted among data sets when no marbling score (MS) records were observed[1]

Item[2]	Data set[3]			
	U-0-100	U-0-50	U-0-25	U-0-0
All	0.453	0.435	0.412	0.342
Sires with >100 progeny	0.442	0.426	0.419	0.367
Sires with <20 progeny	0.542	0.538	0.509	0.445
Top sires	0.301	0.288	0.283	0.251
Bottom sires	0.246	0.230	0.218	0.146

[1]The estimates are the average over 10 replicates.

[2]All = Pearson correlations among all animals; top sires = Spearman rank correlations among sires within the top 20% for their true MS breeding value; and bottom sires = Spearman rank correlations among sires within the bottom 20% for their true MS breeding value.

[3]The initial letter corresponds to which allele is more frequent (U = unfavorable), the first number corresponds to the percentage of available records for MS, and the second number corresponds to the percentage of animals with molecular information.

Complete Absence of Marbling Score Records

The scenario in which only 5% of the records for the trait of interest are observed is not ideal; however, a much less desirable, but possible, scenario is the case in which there is a complete absence of records for the trait of interest in a particular set of data. If genetic correlations exist between the missing traits and other available traits, it can still be possible to obtain breeding values for the missing trait. The following results (Table 4) are obtained under this scenario.

Pearson Correlations. The correlations are lower than the case in which records for MS are observed and decline as the number of animals with known genotypes decreases. Most notable is the fact that even in the absence of MS records, there was stronger association between the true and predicted values obtained from the analysis of U-0-100 and U-0-50 compared with U-5-0. Although these gains are slight, approximately 11.0 and 6.6%, respectively, it still shows that the joint consideration of available correlated traits and molecular information is a favorable method compared with recording minimal phenotypic records for the trait of interest and not simultaneously considering available genotypic information.

Rank Correlations. In general, the rank correlations among all subgroups follow the same patterns as when there are records available for the trait of interest. Overall, the rank correlation estimates are lower than when there are limited records available for MS. The lower estimates of the top and bottom sire rank correlations compared with U-100-100 can be explained by the fact that when all phenotypes are observed the breeding values for MS can be easily differentiated. However, in the presence of missing records, animals who are superior in their true breeding value but have no record themselves and very little progeny information are regressed toward the mean.

Minimal Records for Marbling Score and Increased Incidence of the Favorable Allele

In the previous scenarios the F allele was assumed to have a frequency of 0.4. In the following scenario the frequencies are switched such that the F allele is at a frequency of 0.6. Logically it can be justified that either allele could be more frequent. If there has been phenotypic selection for increased marbling then it is reasonable to assume that the F allele would be in higher frequency. However, if this gene also negatively affects external fat (increases external fat) then it is possible that the gene has been selected against. The results from scenarios where the F allele has a frequency of 0.6 are in Table 5. The changes in the results for Pearson correlations are minimal due to the fact that the frequencies are very close to being equal. The Pearson correlation estimates are slightly lower than those in Table 3, but the differences are negligible.

Rank Correlations. The rank correlations among sires with large amounts of progeny information and among those with small numbers of offspring are very similar to the results found in Table 3 and in general tend to be slightly smaller. The same trend can be identified among sires who are the least desirable for their true MS breeding values. The largest differences occur in the rank correlations among the top sires. Although the estimates among these subgroups follow the same trends (i.e., decreasing with decreasing information), the estimates are much lower when the F allele is more frequent. These results are still favorable to the case presented in Table 4 when there are no records for the trait of interest observed.

The differences between the cases when the F allele is more frequent or less frequent are small. However, in this case the frequencies are 0.6 and 0.4. As the frequencies diverge from each other it is possible that these differences may become more apparent. As selection over time occurs, the initial frequencies may change due to the rapid or slow fixation of the F allele or the possible loss of the allele depending on selection strategies.

Genetic Progress

Average MS breeding values by generation can be found in Table 6. The changes in MS breeding values for the 10 replications are depicted in Figure 1. Across all generations, U-100-100 had higher true breeding values for MS. In the early stages of selection, generations 1 and 2, the tandem procedure led to higher breeding values with its greatest advantage occurring in generation 1. However, this advantage dissipated by generation 3 and by generation 5 the tandem procedure yielded the least desirable results. In fact at generation 5, the average of the true MS breeding values in the youngest generation of U-5-25, U-5-50, and U-5-100 were 24.7, 37.0, and 43.6% larger, respectively, than that of the tandem procedure.

Table 5. Correlation estimates between the true values and those predicted among data sets when limited marbling score (MS) records were observed[1]

Item[2]	Data set[3]				
	F-100-100	F-5-100	F-5-50	F-5-25	F-5-0
All	0.695	0.496	0.486	0.470	0.414
Sires with >100 progeny	0.873	0.689	0.660	0.655	0.639
Sires with <20 progeny	0.780	0.537	0.533	0.518	0.490
Top sires	0.618	0.344	0.328	0.316	0.295
Bottom sires	0.567	0.272	0.257	0.255	0.198

[1]The estimates are the average over 10 replicates.

[2]All = Pearson correlations among all animals; top sires = Spearman rank correlations among sires within the top 20% for their true MS breeding value; and bottom sires = Spearman rank correlations among sires within the bottom 20% for their true MS breeding value.

[3]The initial letter corresponds to which allele is more frequent (F = favorable); the first number corresponds to the percentage of available records for MS; and the second number corresponds to the percentage of animals with molecular information.

The magnitude of change in the breeding values of birth weight and average postweaning gain varied depending on the selection strategy. For the tandem procedure, the starting value of the average birth weight breeding value was 0.06, and at generation 5 the value was –0.46. The starting value of postweaning gain was –0.005, and at generation 5 the value was 0.08. The majority of the change took place between generations 0 and 2. This was due to the fact that in the first 2 generations of selection some pressure was applied to the MS breeding value due to the selection of some heterozygotes in the absence of enough homozygotes to be kept as replacements. This caused a correlated response in birth weight and average postweaning gain. The starting values for the average breeding values were all equal, and at generation 5 the average birth weight breeding values of the animals in the last generation were equal to –1.59, –1.59, –1.53, and –1.49 for the scenarios of U-5-25, U-5-50, U-5-100, and U-100-100, respectively. The breeding value averages for average postweaning gain at generation 5 were equal

to 0.29, 0.29, 0.30, and 0.22 for the scenarios of U-5-25, U-5-50, U-5-100, and U-100-100, respectively. The change in the breeding values for these 2 traits is simply a correlated response that can be explained by the genetic correlation between them and MS.

The average allelic frequency of the F allele can be found in Table 7. The changes in the frequency of the F allele averaged over the 10 replicates are depicted in Figure 2. The change in the frequency of the F allele was most rapid in the tandem procedure and generally declined as the percentage of animals with known genotypes declined. The U-100-100 procedure was intermediate between U-5-100 and U-5-50. At the fifth generation the frequency of the F allele in the youngest

Table 6. Average marbling score breeding values by generation[1]

Generation	Data set				
	U-100-100	U-5-100	U-5-50	U-5-25	Tandem
0	–0.062	–0.062	–0.062	–0.062	–0.062
1	0.250	0.167	0.151	0.133	0.209
2	0.469	0.314	0.306	0.270	0.339
3	0.671	0.468	0.431	0.408	0.431
4	0.943	0.651	0.627	0.565	0.515
5	1.192	0.837	0.799	0.727	0.583

[1]The initial letter corresponds to which allele is more frequent (U = unfavorable); the first number corresponds to the percentage of available records for MS; and the second number corresponds to the percentage of animals with molecular information. The estimates are the average over 10 replicates. The average breeding values per generation were calculated based on the true breeding values of the 3,000 animals in that given generation, with the exception of generation 0, in which the 10,000 animals in the pedigree before selection were used.
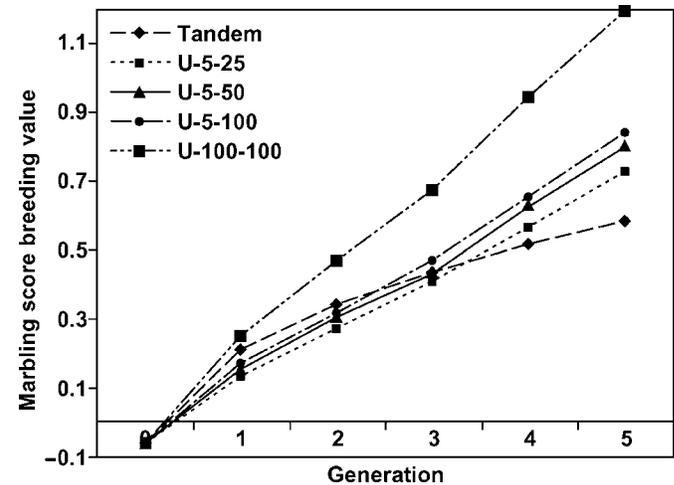

Figure 1. Genetic progress by generation for marbling score for differing methods. Breeding values are the average of replications 1 to 10. For data sets, the initial letter corresponds to which allele is more frequent (U = unfavorable); the first number corresponds to the percentage of available records for MS; and the second number corresponds to the percentage of animals with molecular information.

Table 7. Average favorable allele frequency by generation[1]

	Data set				
Generation	U-100-100	U-5-100	U-5-50	U-5-25	Tandem
0	0.44	0.44	0.44	0.44	0.44
1	0.56	0.59	0.56	0.52	0.69
2	0.63	0.68	0.63	0.58	0.78
3	0.70	0.74	0.68	0.64	0.85
4	0.77	0.81	0.76	0.69	0.89
5	0.84	0.87	0.82	0.74	0.92

[1]The initial letter corresponds to which allele is more frequent (U = unfavorable); the first number corresponds to the percentage of available records for MS; and the second number corresponds to the percentage of animals with molecular information. The estimates are the average over 10 replicates. The average breeding values per generation were calculated based on the true breeding values of the 3,000 animals in that given generation, with the exception of generation 0 in which the 10,000 animals in the pedigree before selection were used.

generation was 0.92, 0.74, 0.82, 0.87, and 0.84 for tandem, U-5-25, U-5-50, U-5-100, and U-100-100, respectively. The high frequency of the F allele created by the tandem procedure is not surprising given that molecular information was the main criteria for selection. The fact that the tandem procedure had the highest frequency of the F allele at generation 5, and the lowest MS breeding value shows that the rapid fixation of an allele in a population is not conducive to genetic improvement. Dekkers (2004) explained that the loss of genetic response due to tandem selection decreases
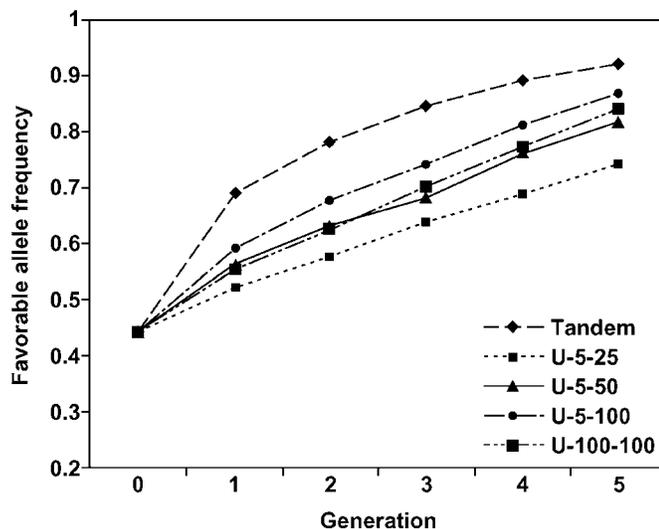
to zero as the effect of the gene becomes larger. However, in the case of polygenic traits of economic value in livestock, this is not likely because a major gene may, at most, account for 10% of the genetic variation for the targeted trait. Furthermore, it may be possible that single genes with large effects, such as the one simulated here, may have deleterious pleiotropic effects or may be closely linked to other genes with detrimental effects (Lande and Thompson, 1990). Consequently, ignoring phenotypes in selection can cause unexpected and undesirable changes.

IMPLICATIONS

This study has shown that the inclusion of molecular information in the prediction of breeding values can lead to more accurate evaluation of candidates for selection. This can be achieved with far less than complete genotyping; however, the authors concede that genotyping even just 25% of a pedigree may not be a practical solution. More work will need to be done to minimize genotyping costs and maximize genotypic information by strategically testing key animals. The use of available, correlated traits is critical in the procurement of accurate breeding values in the absence of records for the targeted trait. For selection purposes, a tandem procedure is not optimal in improving the trait of interest in the long-term, and animals should not be discriminated against based solely on molecular information. More will need to be accomplished to further define the practical usefulness and economic impact of such an approach.

Figure 2. Changes in allele frequency by generation for differing methods. Frequencies are the average of replications 1 to 10. For data sets, the initial letter corresponds to which allele is more frequent (U = unfavorable); the first number corresponds to the percentage of available records for marbling score; and the second number corresponds to the percentage of animals with molecular information.

LITERATURE CITED

Dekkers, J. C. M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. J. Anim. Sci. 82(Suppl. E):E313–E328.

Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467–477.

Greer, R. C., R. W. Whitman, and R. R. Woodward. 1980. Estimation of probability of beef cows being culled and calculation of expected herd life. J. Anim. Sci. 51:10–19.

Henderson, C. R., and D. Quaas. 1976. Multiple trait evaluation using relatives records. J. Anim. Sci. 43:1188–1197.

Lande, R., and R. Thompson. 1990. Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics 124:743–756.

Marshall, D. M. 1994. Breed differences and genetic parameters for body composition traits in beef cattle. J. Anim. Sci. 72:2745–2755.

Meuwissen, T. H. E., and M. E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. Genet. Select. Evol. 28:161–176.

Qian, D., and L. Beckmann. 2002. Minimum-recombinant haplotyping in pedigrees. Am. J. Hum. Genet. 70:1434–1445.

Shafer, W. 2005. Educating beef cattle breeders on the use of genomic technology for quantitative traits. J. Anim. Sci. 83(Suppl. 1):326–327. (Abstr.)

Soller, M. 1978. The use of loci associated with quantitative effects in dairy cattle improvement. Anim. Prod. 27:133–139.

Splan, R. K., L. V. Cundiff, and L. D. Van Vleck. 1998. Genetic parameters for sec-specific traits in beef cattle. J. Anim. Sci. 76:2272–2278.

Tapadar, P., S. Ghosh, and P. P. Majumder. 2000. Haplotyping in pedigrees via a genetic algorithm. Hum. Hered. 50:43–56.

Verrier, E. 2001. Marker assisted selection for the improvement of two antagonistic traits under mixed inheritance. Genet. Sel. Evol. 33:17–38.

Villanueva, B., R. Pong-Wong, B. Grundy, and J. A. Woolliams. 1999. Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genet. Sel. Evol. 31:115–133.

Woodward, B. W., E. J. Pollak, and R. L. Quaas. 1992. Parameter estimation for carcass traits including growth information of Simmental beef cattle using restricted maximum likelihood with a multiple-trait model. J. Anim. Sci. 70:1098–1109.

Wu, S. S., C. Ma, R. Wu, and G. Casella. 2002. A hierarchical statistical model for estimating population properties of quantitative genes. BMC Genet. 3:10.