Integrating molecular data into NCE: expectations, benefits, and needs

M. Spangler

Department of Animal Science, University of Nebraska, Lincoln, NE 68583-0908

Introduction

Genomic information, in the form of Single Nucleotide Polymorphisms, holds the promise to not only increase the accuracy of Expected Progeny Differences (EPD) but also add new and "novel" traits to our suite of traits included in National Cattle Evaluations (NCE). For the most part, genomic information for complex traits (those controlled by many genes) is available to producers in a disjoined context in that it is not seamlessly integrated into EPD estimations and is published separately from EPD. Understanding the benefits of the inclusion of genomic information into EPD first requires knowledge of the differences between an EPD and the results of genomic test (called Molecular Breeding Values or MBV). An EPD is half of the summation of all the independent additive gene effects that cause variation in a given trait (half because each animal only passes on half of their alleles at random). However, with an EPD the specific sources of variation are unknown and for some traits the collection of phenotypes is either cost prohibitive (i.e. tenderness) or it takes a long time to observe a record (i.e. stayability). A MBV, on the other hand, is the summation of the additive SNP effects (multiplied by the number of copies of a given SNP allele) that have been shown through association studies to explain variation in a given trait. SNP are not genes, but serve as markers. The benefit is that DNA, and thus MBV, can be garnered early in life regardless of the trait.

The Value of Improving Accuracy

Several advancements in this technology have occurred with regard to complex traits (i.e. production, carcass, and reproduction traits) including the number of markers included in a given panel, reporting styles of the results, the number of traits for which a diagnostic test exists, and recently, the inclusion of this information for the first time in National Cattle Evaluation (NCE) in the Angus breed.

The promise of the inclusion of marker information into EPD calculations holds three primary benefits:

1. Increased accuracy for young animals (i.e. yearling bulls), which is particularly beneficial when selecting on traits that are measured late in life (e.g., stayability)
2. Shortened generation intervals
3. EPD values for novel traits (i.e. efficiency, end-product healthfulness, disease susceptibility) that may have, at best, sparse collection of phenotypes

The uncertainty surrounding early predictions of genetic merit arise as a result of Mendelian sampling. Every animal is passed a random sample of alleles from each parent, half coming from the dam and half from the sire. We have an estimate of the average effect of what was passed from parent(s) to offspring in the form of pedigree estimates, but the certainty with which we know this estimate is correct (i.e., the accuracy) is low. As more information is collected, such as an individual's own record and data from progeny, accuracy increases. For lowly heritable traits like measures of reproduction, it can take a considerable number of offspring to reach high BIF accuracy levels, given that the BIF scale is more conservative than true accuracy (r) as illustrated in Table 1. To calculate r in the context of progeny test sires the following equation can be used where *n* is the number of progeny:

$$r = \sqrt{\frac{nh^2}{4 + (n-1)h^2}}$$

To convert BIF accuracy to true accuracy (r) the following equation can be used:

$$r = \sqrt{1-(1-BIF)^2}$$

Table 1. Approximate number of progeny needed to reach accuracy levels (true (r) and the BIF standard) for three heritabilities (h^2).

Accuracy		Heritability Levels		
r	BIF	h^2 (0.1)	h^2 (0.3)	h^2 (0.5)
0.1	0.01	1	1	1
0.2	0.02	2	1	1
0.3	0.05	4	2	1
0.4	0.08	8	3	2
0.5	0.13	13	5	3
0.6	0.2	22	7	4
0.7	0.29	38	12	7
0.8	0.4	70	22	13
0.9	0.56	167	53	30
0.999	0.99	3800	1225	700

One primary benefit of molecular information is that it can be garnered much earlier in life (before a phenotypic record can be collected). This knowledge can, in part, reveal a portion of the black box that is Mendelian sampling in young animals. This results in higher accuracy values for young animals, which potentially increases the use of these younger animals in seedstock systems, thus decreasing the generation interval. The equation below predicts the rate of genetic change per year and is dependent on selection intensity, the accuracy of selection, genetic variation, and the length of the generation interval. From this it is apparent that if the generation interval is decreased and /or accuracy is increased this will lead to faster genetic change.

$$\frac{[(\text{Accuracy of Selection})*(\text{Selection Intensity})*(\text{Genetic Standard Deviation})]}{\text{Generation Interval}}$$

However, the magnitude of these benefits will depend on the proportion of variation explained by a given marker panel. Without the seamless integration of this technology into EPD calculations, we find ourselves in the current context of being faced with two disjoined pieces of information: traditional EPD and marker panel results. In this scenario, it is impossible to directly compare EPD to marker panel

results. This is because the molecular scores only explain a portion of the additive genetic variation. Further, some of the marker panel results have a metric of accuracy associated with them. At the current time, this metric is not directly comparable to the Beef Improvement Federation (BIF) accuracy value associated with EPD simply due to differences in the way they are computed. Table 2 shows the relationship between the genetic correlation (true accuracy), %GV and BIF accuracy.

Table 2. The relationship between true accuracy (r), proportion of genetic variation explained (%GV), and Beef Improvement Federation (BIF) accuracy.

r	%GV	BIF
0.1	1	0.005
0.2	4	0.020
0.3	9	0.046
0.4	16	0.083
0.5	25	0.132
0.6	36	0.200
0.7	49	0.286

In contrast to the thought process of DNA marker panel results being a separate and disjoined piece of information, these test results should be thought of as a potentially useful indicator that is correlated to the trait of interest. As such, the MBV can be included in NCE as a correlated trait following methods of Kachman (2008). Other methods have been proposed including using large (50,000+) SNP panels to form a genomic relationship matrix that could allow for known relationships between animals based on genotypes across SNP loci. Combining these sources of information, molecular tools and traditional EPD, has the potential to allow for the benefits of increased accuracy and increased rate of genetic change as discussed earlier.

MacNeil et al. (2010) utilized Angus field data to look at the potential benefits of including both ultrasound records and MBV for marbling as correlated traits in the evaluation of carcass marbling score. MacNeil and colleagues used a 114 SNP marker panel that was developed using 445 Angus animals and calculated to have a genetic correlation (r) of 0.37 with marbling (i.e. the test explained $(0.37)^2 = 0.137$ or 13.7% the additive genetic variation). For animals with no ultrasound record or progeny data, the marker information improved the BIF accuracy of the Angus marbling EPD from 0.07 to 0.13. Assuming a heritability of 0.3 for marbling, a BIF accuracy of 0.13 is equivalent to having approximately 5 progeny carcass records on a young animal or an ultrasound record on the individual itself. In this particular study, both ultrasound records and MBV were found to be beneficial indicators of carcass marbling. The genetic correlation between MBV and ultrasound was found to be 0.80. Since the initiation of MA-EPD by AAA, the SNP panel has evolved and now accounts for 42% of the GV for marbling. The amount of information provided by genomics to NCE will continually change as new products enter the market place and SNP panels are retrained overtime.

Figures 1 and 2 illustrate the benefits of including a MBV into EPD (or EBV which is twice the value of an EPD) accuracy (on the BIF scale) when the MBV explains 10 or 40% of the genetic variation (GV), which is synonymous with R^2 values of 0.1, and 0.4. The darker portion of the bars shows the EPD accuracy before the inclusion of genomic information and the lighter colored portion shows the increase in accuracy after the inclusion of the MBV into the EPD calculation. As the %GV increases, the increase in EPD accuracy becomes larger. Additionally, lower accuracy animals benefit more from the inclusion of genomic information and the benefits decline as the EPD accuracy increases. Regardless of the %GV assumed here, the benefits of including genomic information into EPD dissipate when EPD accuracy is between 0.6 and 0.7. On the other hand, when %GV is 40 an animal with 0 accuracy could go to over 0.2 accuracy with genomic information alone. From table 1, this would be the same as having approximately

4 progeny for a highly heritable trait or 7 progeny for a moderately heritable trait.

Figure 1. Increase in accuracy from integrating genomic information that explains 10% of the genetic variation into Estimated Breeding Values (EBV).

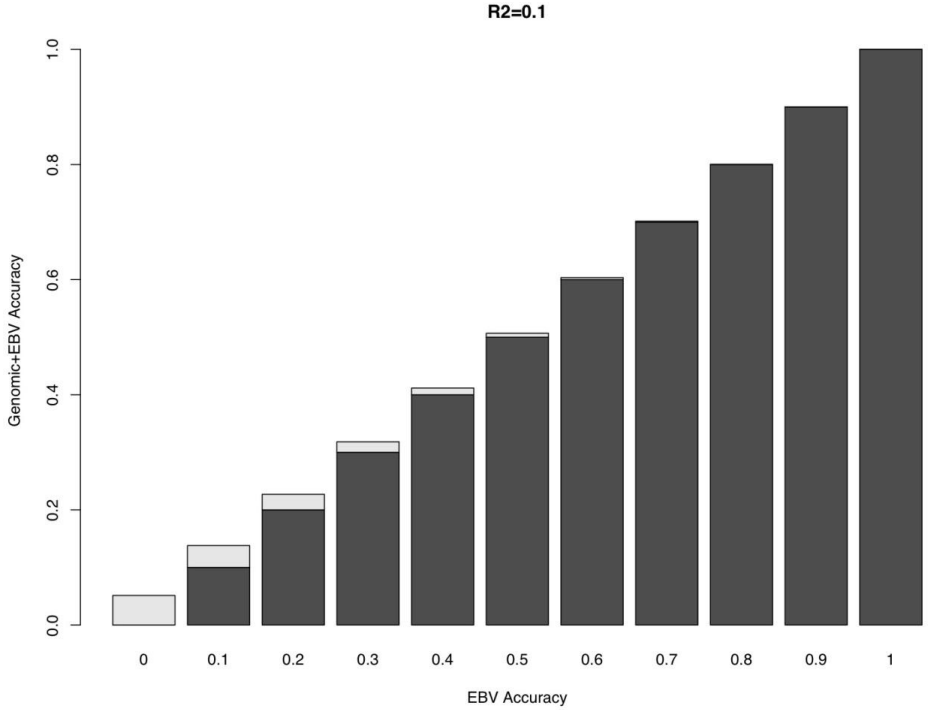
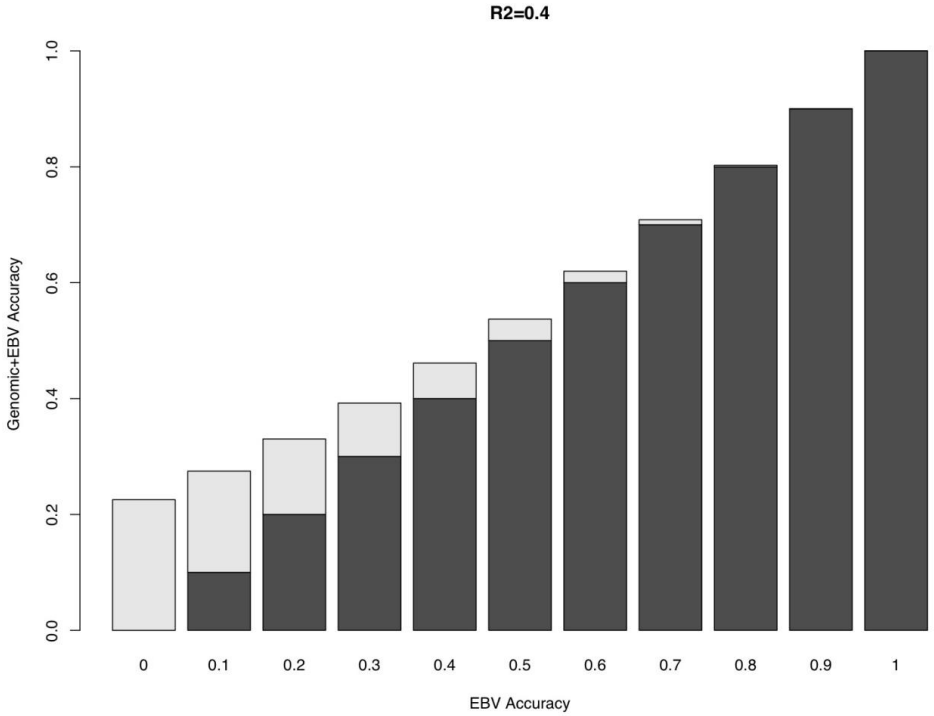


Figure 2. Increase in accuracy from integrating genomic information that explains 40% of the genetic variation into Estimated Breeding Values (EBV).



It is important to understand some limitations in the current application of Marker Assisted Selection. For instance, current marker panels are likely to work best in the populations where discovery occurred, but will potentially decrease in predictive power as the target population becomes more genetically distant from the discovery population (de Roos et al., 2008). The same erosion in accuracy is likely to occur overtime as well (i.e. over generations if panels are not retrained).

Discovery	Target	
Angus	Angus	Closest relationship
Angus	Charolais	↓
Angus	*Bos indicus*	Most distant relationship

In order to investigate the robustness of SNP predictions across breeds, a unified research and outreach project (called the Weight Trait Project; Spangler et al., 2011) was initiated utilizing both industry and academic/ARS resources. Weaning weight records (n=3,328) of calves from the US Meat Animal Research Center (USMARC) were used in the selection of SNP from the Bovine SNP50 associated with adjusted weaning weight. The total pedigree included 5,222 animals. Of the 3,328 calves in the training population, the average breed contributions were 26% Angus, 19% Hereford, and 6.5% each of Red Angus, Simmental, Charolais, Limousin, and Gelbvieh.

Breed associations representing the seven breeds (Table 2) in the USMARC Cycle VII population identified seedstock producers in the region surrounding USMARC to provide DNA samples (tail hair) from calves born in 2009 and their dams. A reduced panel of 192 SNP was constructed based on the most significant SNP from the USMARC association analysis with the addition of 192 SNP from IGENITY® (96 trained on yearling weight in an Angus population and the other 96 from the IGENITY parentage panel). In total, the reduced panel consisted of 384 SNP. IGENITY® served as the genetic service provider partner in this project and genotyped animals with the reduced panel. After editing SNP based on deviation from Hardy-Weinberg Equilibrium and call rates, a total of 159 of the diagnostic SNP (non parentage) were used in the analysis. The genotype data had an average call rate of 85.2% (11.3-100%). Bull calves (n=3,500) from the twenty collaborating herds were genotyped with the reduced panel and MBV were calculated based on prediction equations derived at USMARC for weaning weight (WW) and post weaning gain (PWG). Data including a four-generation pedigree, adjusted weaning weight phenotypes, and pedigree index EPD were obtained from the respective breed associations for each herd in the project. MBV were fit as a correlated trait in both two- and three-trait animal models. Contemporary group effects included herd and sex of calf. Weaning weight was fit with both a direct and maternal component while MBV were assumed to have only a direct genetic component.

Given the partial nature of the genotypes produced by the WTP due to the newness of the genotyping platform used at that time, methodology was developed to account for partial genotypes in the analysis (Kachman et al., 2011). For animal a the proportion, Pa, of the complete genotype (CG) MBV variance accounted for by partial genotypes (PG) is the ratio of the variances calculated by summing over the partial and the complete set of markers. Similarly, the genetic covariance between a trait and PG MBV is also proportional to Pa. The proportion of CG covariance between animals a and b with PG was assumed to be proportional to PaPb. The PG model for the MBV of animal a, scales the CG genetic effect by Pa and adds a missing genotype effect with variance Pa(1-Pa) times the CG genetic variance.

Genetic parameters for weaning weight (direct) and MBV by breed are summarized in Table 3 both before and after accounting for partial genotypes in the analysis. In general, the heritability estimates for WW direct were within expected ranges except for Simmental, which is likely due to the data structure of the Simmental herds in this study. In general, the genetic correlations are low to moderate with relatively

large standard errors. The number of markers used in the current panel might explain the less than desirable performance. Given these correlations, the proportion of genetic variation for weaning weight explained by the panel (r_g^2) ranged from 0 to 7.8% before accounting for PG and 0.09 to 14.44% after. One possible reason for the range in genetic correlations among breeds is that the associations between markers and growth traits are more breed-specific than had been hoped.

Table 3. Heritabilities (SE) by breed for weaning weight (direct) and molecular breeding values (MBV) for weaning weight (WW) direct both Before and After accounting for partial genotypes.

Breed	Heritability Weaning Weight	Heritability Molecular Breeding Value		Genetic Correlation	
		Before	After	Before	After
Angus	0.23±0.02	0.87±0.16	0.75±0.12	0.00±0.10	0.15±0.11
Red Angus	0.24±0.03	0.67±0.16	0.89±0.14	0.10±0.10	0.14±0.11
Charolais	0.12±0.03	0.33±0.16	0.47±0.18	0.28±0.15	0.38±0.16
Gelbvieh	0.22±0.02	0.64±0.18	0.62±0.16	0.25±0.13	0.26±0.14
Hereford	0.14±0.04	0.83±0.15	0.96±0.14	0.20±0.20	0.25±0.21
Limousin	0.27±0.02	0.60±0.19	---	0.24±0.12	---
Simmental	0.75±0.03	0.61±0.16	0.73±0.16	0.05±0.08	0.03±0.09

Summary

It is likely that the list of genetic selection tools will continue to expand in the short-term as this arena is far from stagnant. Although the goal is the consolidation of information into one of two basic forms, EPD and economic index values, the industry has witnessed several intermediate steps in an effort to quickly commercialize technology that has created confusion. Integrated projects such as the WTP that engage researchers, extension personnel, producers, and breed associations are critical to the further development and employment of genomic selection tools. The WTP has created a vast resource that continues to grow in order to investigate the plethora of questions that still exist related to the use of this technology.

Literature Cited

Kachman, Stephen. 2008. Incorporation of marker scores into national cattle evaluations. Proc. 9th Genetic Prediction Workshop, Kansas City, MO, pp. 88-91.

Kachman, S.D., G.L. Bennett, K.J. Hanford, L.A. Kuehn, E.J. Pollak, W.M. Snelling, M.L. Spangler, and R.M. Thallman. 2011. Incorporating molecular breeding values with variable call rates into genetic evaluations. Joint meeting of the American Society of Animal Science (Accepted).

MacNeil, M. D., J. D. Nkrumah, and S. L. Northcutt. 2010. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. J. Anim. Sci. 88:517-522.

de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179:1503-1512.

Spangler, M.L., S. Kachman, K. Hanford, M. Thallman, G. Bennett, W. Snelling, L. Kuehn, and J. Pollak. 2011. Integration of DNA marker information into breeding value prediction. Nebraska Beef Research Report.