USE OF BOVINE SNP50 DATA FOR FEED EFFICIENCY SELECTION DECISIONS IN ANGUS CATTLE

Megan M. Rolf[*], Jeremy F. Taylor [*], Robert D. Schnabel[*], Stephanie D. McKay[*], Matthew C. McClure[*], Sally L. Northcutt[§], Monty S. Kerley[*] and Robert L. Weaber[*]

[*]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA

[§]American Angus Association, St. Joseph, MO 64506, USA

Abstract

The past few years have led to a rapid increase in the use of molecular genetic technologies in the beef industry. With any new technological advance, the methods for implementation must be characterized and tested in populations of beef cattle. Recently, large panels of single nucleotide polymorphism (SNP) markers have become available and a multitude of animals have been genotyped. The best use of these data will likely be in the form of genomic selection, where the marker information is incorporated into the current system of genetic prediction and EPDs (expected progeny differences) will continue to be reported by the breed associations. Genomic selection methods will be exceptionally valuable for traits that are difficult and expensive to measure (such as residual feed intake, or RFI) or that are measured late in life (such as longevity/stayability). One method to utilize information from reduced marker panels is to utilize genomic relationship matrices (GRMs) in place of traditional pedigree-derived relationship matrices in genetic evaluation. Traditional pedigree derived matrices (numerator relationship matrices, NRM) contain a number for each pair of animals describing the average proportion of DNA two animals share identical by descent, or their "relatedness". The data in a GRM may more accurately reflect the kinship between two animals because it is calculated directly from genomic data. This method is particularly useful for animals that have missing pedigree data, such as in populations of commercial cattle. We used a GRM to test genomic selection for feed efficiency traits, quantified the number of markers needed to calculate a GRM, performed a genome scan for regions influencing feed efficiency and tested model predicted feed intakes against individual feed intake data in a commercial Angus cattle population.

Introduction

The beef industry has made enormous strides in improving genetic merit for economically relevant traits (ERTs) such as calving ease, growth and carcass quality over the last several decades. Much of this improvement has been made possible by the availability of EPDs published by almost every purebred beef breed association (Crews, 2005), which are based on best linear unbiased prediction (BLUP) methods outlined by Henderson (1975). Most of these ERTs focus on outputs from the production system. However, production inputs, such as feed inputs, can have a significant influence on profitability and these traits have remained essentially unselected.

Feed efficiency is a trait with enormous economic importance, but selection for efficiency has remained elusive due to the difficulty and expense of gathering phenotypic data (Archer et al., 1997). In the past, increased growth rate has been selected for in the beef industry because growth and efficiency are correlated (Koch et al., 1963). Because of the correlated

selection response, selection for increased growth rates can result in unintended consequences like increases in mature size and increased maintenance requirements in the cowherd (Archer et al., 1999; Okine et al., 2004).

Model predicted feed intakes from programs such as the Cornell Value Discovery System (CVDS; Guiroy et al., 2001) and the Decision Evaluator for the Cattle Industry (DECI; Williams et al., 2003a,b) have the potential to increase the number of feed efficiency observations produced on progeny of registered animals. These models are designed to predict the differences in intake for cattle that are fed in pens by allocating the total feed fed to the entire pen to individual animals based on their performance for traits related to growth and carcass composition. A study by Williams et al. (2006) showed high phenotypic (0.947-0.933) and genetic (0.97-0.99) correlations between CVDS and DECI predicted dry matter required (pDMR) values, so only the CVDS model was used in our study. These model predicted intakes have the potential to serve as an indicator trait for feed efficiency, much the same way that ultrasound data are indicator traits for carcass quality and yield. Indicator traits such as these do not necessarily impact revenue or risk themselves, but are easier and more cost effective to record and are genetically correlated with the ERT of interest (Crews, 2005). As more records are obtained on an indicator or causal trait, it will become more effective to incorporate these data into genetic evaluation systems in the beef industry.

Since the first national genetic evaluation in 1974 (Willham, 1993), the beef industry has been collecting phenotypes and incorporating them into genetic evaluation systems. The incorporation of feed intake data into genetic evaluation has the potential to dramatically influence selection on maintenance efficiency and genomic selection has the potential to make the most of limited data for genetic prediction on a large number of animals using either large marker panels (such as the 50K or 800K chips) or smaller panels of markers associated with ERTs. Genomic selection is the ability to (theoretically) select for desirable alleles at all genes in the genome that influence a trait by using markers spread throughout the entire genome. This approach has several significant advantages over marker assisted selection. It explains a larger portion of the genetic variance than a single marker, provides an easy, familiar interface (EPDs) and the danger suggested by Spangler et al. (2007) whereby producers select only for a few markers and disregard EPDs is entirely avoided.

Feed Intake Data

Individual feed intake records were collected for average daily feed intake (AFI), residual feed intake (RFI) and average daily gain (ADG) on 862 commercial Angus steers born between 1998 and 2005 at either the Circle A Ranch (Iberia, Stockton and Huntsville, MO) and research farms participating in the MFA Inc. feeding trials (Thompson and Greenley, MO). Intake data were collected using Calan gates (Circle A Ranch steers) or GrowSafe feeding systems (MFA steers, fed at the University of Missouri) and live weights were taken three times (beginning, mid-test and final) during the course of the feeding trial. DNA was available for genotyping and analysis on 698 of the steers as no blood was collected during the first year of the trial. Cryopreserved semen units were obtained on 1,721 Angus AI sires born between 1956 and 2003 that were used in artificial insemination (AI) within the United States. These animals included the sires of the steer calves and their male ancestors. Complete 62-generation pedigrees were provided by the American Angus Association. Half-sib family sizes derived from sire

information ranged from 1 to 81 progeny. Due to the fact that this is a population of commercial animals, dams were unregistered and available pedigree information was determined to be unreliable based on attempts to phase chromosomes and infer missing genotypes. As not all maternal grandsires had been genotyped, correct parentage could not be assigned. This population structure is suitable for testing methods of genomic selection on commercial populations.

Data Acquisition

Residual feed intake was calculated as the difference between observed and expected feed intake ($\widehat{FI}$), which was predicted from the regression of average daily feed intake (AFI) on ADG and metabolic midweight (MMW: mid-weight$^{0.75}$) as follows:

$$RFI = AFI - \widehat{FI}$$

$$\widehat{FI} = b_0 + b_1 ADG + b_2 MW^{0.75}$$

Weights were taken at three different times during the feeding trial (first day of the test, mid-test and end of test). These cattle were commercially owned and the specific ration composition is unknown, however all of the animals within a feeding group were fed the same ration. RFI was calculated individually for each feeding group and the mean R^2 value for the regression models was 0.49.

Genotypes were acquired using the Illumina BovineSNP50 assay and were screened for Mendelian inheritance to verify the accuracy of the sire pedigrees. Genotypes for nine sires were found to be inconsistent with their paternal pedigree and an additional two animals were split embryos (identical twins), so these animals were removed from the dataset. Quality control was performed for genotypes so that the minor allele frequency (MAF) was $\geq$0.05 and call rate was >95%. Quality control constraints resulted in 41,028 SNPs being retained for analysis on 698 steers and 1,707 AI sires. Missing genotypes (0.58%) were imputed using fastPHASE (Scheet and Stephens, 2006) with Btau4.0 positions.

Additive Effects Analysis

A numerator relationship matrix (NRM) was generated using pedigree information on 862 Angus steers, their dams (where available) and 34,864 identified parental ancestors. Variance components, breeding values and residuals were estimated using an animal model in the multiple trait derivative free restricted maximum likelihood (MTDFREML; Boldman et al., 1995) program. Convergence for models using a NRM was assumed when the variance of the -2*log-likelihood was $<1x10^{-12}$. The model fit feeding pen as a fixed effect (year and season of birth were nested within pen, so only one effect was included) and breeding values and residuals were assumed to be uncorrelated.

Genomic Relationship Matrix

The method of calculating a genomic relationship matrix (GRM) used in this analysis was proposed by VanRaden (2008). It is a regression method that uses a fraction of the population with complete and accurate pedigree data (in the form of a NRM on those individuals) to calibrate the allele sharing to the expected value of the relationship matrix, in this case, E[G]=A. In this dataset, the NRM was generated on 1,707 Angus AI sires with complete and accurate pedigree data. Complete genotypes for 698 Angus steers and 1,707 AI sires were assembled into a 2,405 x 41,028 genotype matrix (M) with animals in rows and SNPs in columns. The elements in M are -1, 0 and 1 for AA, AB and BB genotypes, respectively. The GRM was calibrated by finding the regression of the upper triangular elements of MM' on the corresponding elements of A for the 1,707 AI sires only. The estimated slope and intercept were used to calibrate the GRM for all 2,405 animals as:

$$G = \left[\frac{MM' - g_0(11')}{g_1}\right]$$

Estimates of these parameters were 9,731.9±0.65 and 15,198±7.26 for g_0 and g_1, respectively. The mean molecular inbreeding coefficient over all animals was 0.079. We estimated variance components, fixed effects, breeding values and residuals using restricted maximum likelihood under an animal model where the NRM was replaced by the GRM. Convergence was assumed on models including the GRM when heritability estimates had converged from above and below to three significant figures.

Variance component estimates and heritabilities are shown in Table 1. Estimated breeding values (EBVs) and residuals were retained for further analysis. The estimated heritabilities (AFI 0.14; RFI 0.14; ADG 0.09) reported here were much lower than literature estimates (AFI 0.45; RFI 0.39; ADG 0.28; MacNeil et al., 1991; Arthur et al., 2001) and standard errors were fairly high, possibly due to sampling effects resulting from the small number of animals used in this study. The mean accuracy for all 2,405 animals was 0.32 for AFI and RFI, while the mean accuracy for ADG was 0.23. Mean accuracies for steers (NRM 0-0.46 vs GRM 0.36-0.43) and sires of steers (NRM 0-0.45 vs 0.37-0.44) indicate that similar accuracies were achieved when using either the NRM or GRM in this dataset, however the GRM accuracy was achieved given an approximately 20% loss in phenotypic data. This is most likely the result of the ability of the GRM to extract information from the genotypes related to the identity by descent information among the steers due to the relationships among their dams, which was missing in the NRM analysis. Accuracies for the GRM analyses were lower than those previously reported for genomic selection (Hayes et al., 2009; VanRaden et al., 2009; VanRaden, 2008; Schaeffer, 2006; Meuwissen et al., 2001), presumably due to the small number of animals with phenotypic observations and the lower heritabilities estimated in this study. Even though there are limitations within this dataset, it would be possible to combine datasets from other Angus research populations in an effort to increase the heritability estimates and obtain more accurate EBVs.

Marker Panel Subsets

MATLAB (The Mathworks, Natick, MA) was used to test the number of markers necessary to precisely estimate the GRM using the approach outlined above. Subsets of n markers were randomly sampled with replacement (see Figure 1) from the full set of 41,028 markers. This approach ensures full representation of the entire genome within the marker subsets. For each of the 50 replicates (i=1,…,50) for each subset of n markers, a GRM (G_{ni}) was estimated using the regression approach proposed by VanRaden (2008). Correlations were estimated between the upper triangular elements of G_{ni} and G (the full GRM estimated from all available SNPs) for all 2,405 animals and between G_{ni} and A for all 1,707 AI sires and averages were produced across replicates. Mean correlation of the NRM and the full GRM was approximately 0.86 when considering the 1,707 Angus AI sires. The mean correlation of the full GRM and the G_{ni} exceeded 0.86 when between 1,000 and 2,500 markers were utilized for the calculation of the GRM. Minimal increases in the correlation coefficients were seen when greater than 10,000 SNPS were included in the calculation of the GRM, which is illustrated in Figure 1. Tables containing the complete correlation results for this analysis can be found in Rolf et al. (2010).

It is likely that smaller panels of 384 or 1,536 markers will be utilized in the beef industry until genotyping costs decrease to more affordable levels for most producers. Consequently, we performed 200 bootstrap replicates of 384 or 1,536 randomly sampled SNPs from the full set of 41,028 markers. Minimum, mean and maximum correlations between the bootstrap samples and the full GRM were 0.60, 0.65 and 0.68 and 0.85, 0.87 and 0.88 for 384 SNP and 1,536 SNP panels, respectively. The mean correlation using 1,536 randomly sampled markers met the mean correlation between the NRM and full GRM, indicating that in the absence of pedigree data in commercial herds, a GRM constructed from a panel of 1,536 SNPs may be a viable alternative to calculate EBVs for genetic selection. One potential caveat of this approach is that the panels used in the beef industry will not be randomly sampled SNPs, but rather panels of SNPs that are associated with various economically relevant traits. The efficacy of this approach will depend on the distribution of linkage disequilibrium among the markers and the minor allele frequencies.

Genome-Wide Association Analysis

Estimated breeding values (EBVs) and residuals obtained previously from mixed model estimation with a GRM were utilized for genome-wide association analysis (GWAS). Traits (RFI, AFI and ADG) were analyzed on 698 animals as either EBVs or phenotypes (EBV + residual). The same SNP set was utilized for the GWAS analysis and included 41,028 SNP with an average MAF of 0.28 and an average spacing of 65.73±68.45 kb for the 39,484 autosomal and 487 X chromosome loci. SNP that mapped to unassigned contigs (ChrUn; n=1,057) were also included in the analysis.

Multiple hypothesis testing is always an issue when performing GWAS studies, so a permutation analysis (Churchill and Doerge, 1994) was performed to obtain a genome-wide significance threshold of 0.05 to control the rate of type I error in this study. The permutation analysis consisted of 10,000 dependent variable permutations per trait accompanied by a GWAS on the permuted dependent variable to determine the largest F-statistics obtained by chance with the data provided.

The GWAS was performed using custom code developed and implemented in MATLAB (MathWorks, Natik, MA) and was comprised of three steps. The first step consisted of individual one-way analysis of variance (ANOVA) tests for each SNP, trait and dataset. Analyses using EBVs were weighted by their corresponding accuracies and fitted additive effects only and analyses utilizing phenotypes fit either additive effects alone or additive and dominance effects simultaneously. Genotypes were coded 1, 0 or -1 for additive effects and 0, 1 or 0 for dominance values, corresponding to AA, AB and BB genotypes, respectively. The second step included all SNPs that met or exceeded the pre-determined genome-wide significance threshold previously described. These SNPs were included in a forward-selection analysis which was performed on a chromosome-by chromosome basis. The SNP with the highest F-statistic was sequentially added to the ANOVA model for each chromosome until no more SNPs could be added that met or exceeded the significance threshold. All of the SNPs selected in the chromosome-by-chromosome analysis were then combined into a final model to estimate the amount of variance explained by the selected SNPs in the third step.

GWAS utilizing phenotypes and modeling only additive effects yielded either few or zero SNPs in the final analysis models for all traits. Because of this, modeling both additive and nonadditive effects explained a larger portion of the phenotypic variance (AFI, 49.802%; RFI, 25.494%; ADG, 27.093%). As a result, any further discussion of results will pertain only to EBV analyses or analyses of phenotypes including both additive and dominance effects. Analysis of steer EBVs yielded a larger number of SNPs in the final model for RFI and ADG. The markers included into the final analysis for all three traits explained a fairly large portion of the additive genetic variance; however, because of the reduced power of this dataset, the amount of phenotypic variance explained was less than optimal. To facilitate comparison of results between phenotype and EBV analyses, Figure 2 shows a side-by-side comparison of the regions of the genome detected for AFI from the analysis of both steer EBVs (presented in panel A) and phenotypes (presented in panel B).

The number of SNPs included in the final analyses and the concordance between different traits in the analysis can be found in Tables 2 (EBV) and 3 (phenotypes). SNPs were considered concordant if they fell within the range of ±0.5 Mb of the position of the selected SNP to better account for the linkage disequilibrium in cattle populations (McKay et al., 2007) as well as selection of SNPs that are pleiotropic or closely linked and could show correlated responses with one another when used in selection. Of particular interest is the percentage of SNPs that were included in the final model for one trait, but also above the significance threshold (forward selected) for another trait. The concordance of SNPs between traits was fairly consistent with the magnitude of expected genetic correlations between traits. Interestingly, there was a slight concordance between AFI and RFI, evidenced by the fact that in the EBV analysis approximately 13% of SNPs included in the RFI model were forward selected for ADG and 16% of SNPs in the ADG final model were forward selected in the RFI analysis. Similar results were also observed in the phenotypic analysis, but of a smaller magnitude. This suggests that despite the phenotypic independence between RFI and ADG, genetic independence is not guaranteed between these two traits, as suggested by Kennedy et al. (1993). A high concordance was observed between AFI and ADG in the EBV analysis (30% of SNPs in the AFI model and 20% of SNPs in the ADG model), which was expected based on the moderate genetic correlation between these two traits. It appears to be possible to identify QTL for feed intake which are independent of ADG in this dataset. Selection on only these QTL would theoretically

allow genetic improvement for feed efficiency without a problematic correlated response in the growth rate of the growing animal or in mature size in the cowherd.

Pathway Analysis

Even with appropriate consideration for multiple hypothesis testing, it is possible to select spurious SNPs in GWAS analyses. Many previous studies of feed efficiency in beef cattle have used linkage analysis resulting in large confidence intervals, or the SNP positions were not reported. Due to the lack of ability to accurately compare results from previous studies with those achieved in this study, a pathway analysis was conducted. The purpose of the pathway analysis is to have a form of independent validation of the GWAS results. If regions of the genome are identified that perform functions related to growth and metabolism, then it is likely that a real association has been discovered. Due to the larger number of annotations of human genes, we mapped human annotations to the bovine genome using the UCSC genome browser. Because of the limited range of LD in beef cattle (McKay et al., 2007), we identified regions of interest surrounding the SNPs using a 1 Mb window (SNP position±0.5 Mb). Genes within these regions were identified in the Database for Annotation, Visualization and Integrated Discovery (DAVID; Huang et al., 2009, Dennis et al., 2003) and queried against the KEGG Pathway Database (Kanehisa et al., 2010, 2006; Kanehisa and Goto, 2000). Results from the KEGG database were summarized into global pathways and their corresponding sub-categories using the KEGG Atlas. A summary of the KEGG pathway analysis findings is provided in Table 4. The analysis was most successful utilizing those traits with the highest heritabilities (AFI and RFI), suggesting that in well-powered studies with large numbers of animals and large SNP lists, pathway analysis is very useful. Many of the regions detected in the GWAS appear to harbor genes which are involved in growth or metabolic functions (69% AFI and 85% RFI as a percentage of the total number of identified pathways).

Model-Predicted Feed Intakes

The Cornell Value Discovery System (CVDS) was used to predict the dry matter required (pDMR) for the 862 (698 with DNA) Angus steers used in this study using a growth and maintenance model. No ration information was available, so animals were all assumed to have eaten a diet with an equivalent composition and nutrient density. Pen feed intakes were obtained by pooling the average individual feed intakes of each animal within the pen. Sex, growth promotant implant status, date on test, carcass data (ribeye muscle area, yield grade, hot carcass weight, fat thickness and marbling score) and weight data were input into CVDS and were used to specify growth and maintenance model parameters and account for composition of gain in the calculation of pDMR. The phenotypic correlation between pDMR and AFI was 0.78 ($p<0.0001$) in this dataset, which is consistent with that reported by Williams et al. (0.784; 2006).

Breeding values and residuals were estimated for pDMR on all animals using the previously outlined procedure. Variance components for pDMR were 0.1648 and 4.0933 for additive and phenotypic variance, respectively. Heritability was 0.04, which was lower than literature estimates, but consistent with the rest of the traits.

The largest concordance between SNPs in the final model for pDMR and SNPs forward selected in the other analyzed traits was found between pDMR and ADG (EBV 73%, phenotypes

77%). This result was expected given the dependence of pDMR on growth data. SNPs in the final model for pDMR also showed concordance with SNPs forward selected for AFI (EBV 21%, phenotypes 22%) and RFI (EBV 15%, phenotypes 7%). These results indicate that while there may be significant overlap between pDMR and ADG, the moderate concordance of these predicted measurements with AFI and RFI merit further exploration of this trait as an indicator trait in genetic selection procedures.

Conclusions

These data on commercial Angus steers were useful for implementing a method of genomic selection for feed efficiency, not only in the 698 steers with observations, but also for generation of EBVs with moderate accuracies on 1,707 of the most widely used Angus AI sires in the United States. We suggest that studies utilizing GRM for producing breeding value estimates utilize at least 1,500 SNPs and preferably, 10,000 SNPs per animal. Inclusion of additional SNPs into the calculation of the GRM yielded only marginal improvement over a GRM calculated with 10K SNPs.

A large number of SNPs have been identified in these analyses which could be included in commercial marker panels for use in Angus cattle for selection on feed efficiency traits. These models account for large amounts of genetic (AFI 54%, RFI 62%, ADG 54% and pDMR 56%) or phenotypic (AFI 49%, RFI 25%, ADG 27% and pDMR 30%) variation in these populations. The estimates of the variance explained and the SNP effects are biased due to population sampling, as the SNPs most strongly associated in this population may not be representative of the Angus breed as a whole. A pathway analysis was the first step towards validation of these SNP associations; however these studies should be repeated and compared using an independent population of animals to produce an unbiased estimate of the amount of genetic variation explained by these SNPs in feed efficiency traits.

To the authors' knowledge, this is the first work to examine the use of a predicted feed efficiency phenotype in a genome wide association analysis to compare model predictions to observed phenotypic records in beef cattle. Additional comparisons of pDMR with results using actual feed intake data, gain and RFI in studies with larger numbers of animals and larger heritabilities will be essential to further explore the use of these data for genetic evaluation and selection decisions in commercial beef cattle populations.

Table 1: Descriptive statistics and estimated variance components for NRM and GRM analyses of three feed efficiency traits (adapted from Rolf et al., 2010).

Trait[a]	N	Mean	Min	Max	Var	σ^2_A	σ^2_e	h^2
AFI	862	11.0326	6.0599	15.2116	3.0323	0.1436	0.7786	0.16
RFI	862	0.0026	-3.3386	4.9952	0.7626	0.1147	0.4364	0.21
ADG	862	1.5363	0.0231	2.3443	0.1077	0.000002	0.552	0.00
AFI	698[b]	10.8943	6.0599	15.2116	3.1608	0.1404	0.8680	0.14
RFI	698[b]	-0.0201	-3.3412	4.9952	0.8255	0.0849	0.5286	0.14
ADG	698[b]	1.5175	0.0231	2.2941	0.1105	0.0053	0.0528	0.09

[a]Average daily feed intake, AFI; residual feed intake, RFI; and average daily gain, ADG; all measured in units of kg/d.
[b]DNA samples were available on only 698 of the 862 phenotyped steers. Variance components for these three analyses were estimated using the GRM.

Table 2: Number of SNPs included in the final models or above the significance threshold (Forward Selected) for GWAS analysis of feed efficiency EBVs. Numbers to the right are the number and percentage of SNPs in the final model for the trait in the row that were above the significance threshold for the trait in the column.

			Forward Selected		
	No. in Model	No. Fwd Selected	AFI	RFI	ADG
AFI	53	178	-	37[a] 69.81%	16[a] 30.19%
RFI	66	281	35[a] 53.03%	-	9[a] 13.64%
ADG	68	274	14[a] 20.59%	11[a] 16.18%	-

Table 3: Number of SNPs included in the final models or above the significance threshold (Forward Selected) for GWAS analysis of feed efficiency phenotypes. Numbers to the right are the number and percentage of SNPs in the final model for the trait in the row that were above the significance threshold for the trait in the column.

	No. in Model	No. Fwd Selected	Forward Selected		
			AFI	RFI	ADG
AFI	65	83	-	11[a] 16.92%	3[a] 4.62%
RFI	18	21	10[a] 55.56%	-	1[a] 5.56%
ADG	24	33	3[a] 12.50%	3[a] 12.50%	-

Table 4: Results from DAVID and KEGG Atlas for pathway analyses.

Global Pathway	Sub Category	Number of Pathways/No Genes[a]							
		AFI EBV[b]	AFI Phen[b]	RFI EBV[b]	RFI Phen[b]	ADG EBV[b]	ADG Phen[b]	pDMR EBV[b]	pDMR Phen[b]
Metabolism	Carbohydrate Metabolism*					1/1		2/2	
	Energy Metabolism*							1/1	
	Lipid Metabolism*		3/9	4/4	2/8				
	Nucleotide Metabolism*	2/2							
	Amino Acid Metabolism*	1/1		4/5					
	Metabolism of Other Amino Acids*			2/2					
	Glycan Biosynthesis and Metabolism*			3/3		2/4		4/4	
	Metabolism of Cofactors and Vitamins*		1/1						
	Biosynthesis of Secondary Metabolites*		2/4	2/2	2/4				
	Xenobiotics Biodegradation and Metabolism*		1/4	3/3	1/4				
Genetic Information Processing	Translation	1/1							
	Folding, Sorting and Degradation*	2/3							
Environmental Information Processing	Signal Transduction*	1/2	6/12	2/4	4/5	3/3		4/4	
	Signaling Molecules and Interaction*		1/5			1/3		1/3	
Cellular Processes	Transport and Catabolism*			1/1					
	Cell Motility			1/2		1/1			
	Cell Growth and Death*	1/1	1/4	1/1	1/1				
	Cell Communication		3/5	4/7	1/1			1/8	
	Endocrine System*		4/6	4/5	2/2			2/2	
	Immune System	1/1	5/9	3/3	1/1		1/1		1/1
	Nervous System		1/2	2/3		1/1		1/1	
	Development*		1/1		1/1				
Human Diseases	Cancers*	1/1	5/8	5/5	4/4				
	Neurodegenerative Diseases		3/5		1/1	2/3			
	Metabolic Disorders*		2/6	1/1					
	Infectious Diseases			1/1				2/2	

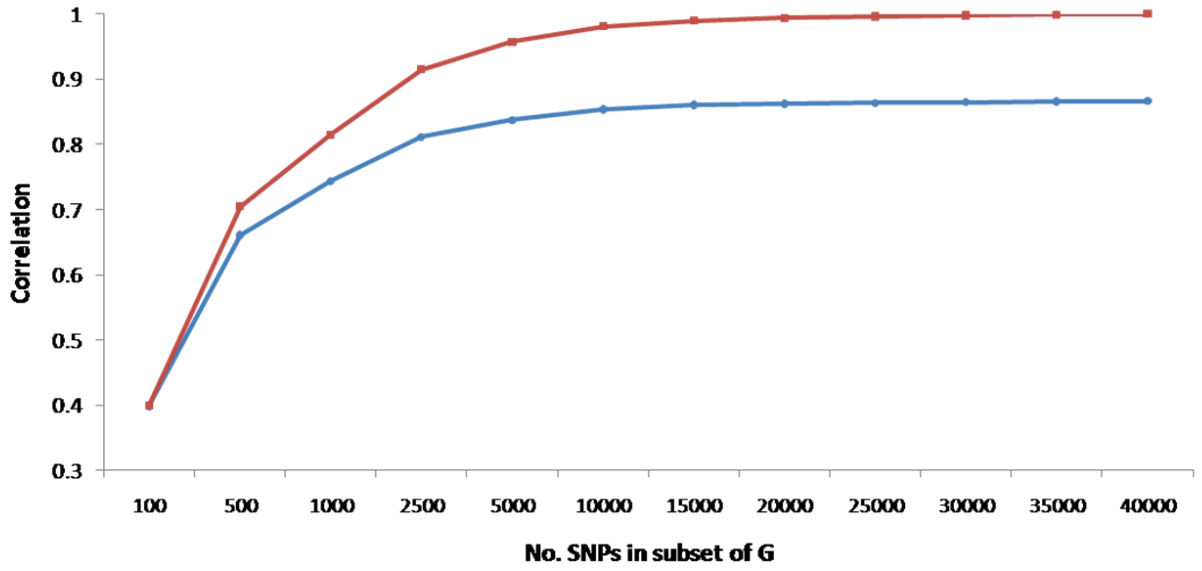

Figure 1: Correlation between the G_{ni} estimated from bootstrap samples of reduced marker panels versus the full GRM calculated with all markers available (n=41,028). The red line indicates the correlation between relationship coefficients between full and reduced marker sets for all 2,405 animals. The blue line indicates the correlation between relationship coefficients between reduced marker sets and the NRM for all 1,707 AI sires (adapted from Rolf et al., 2010).

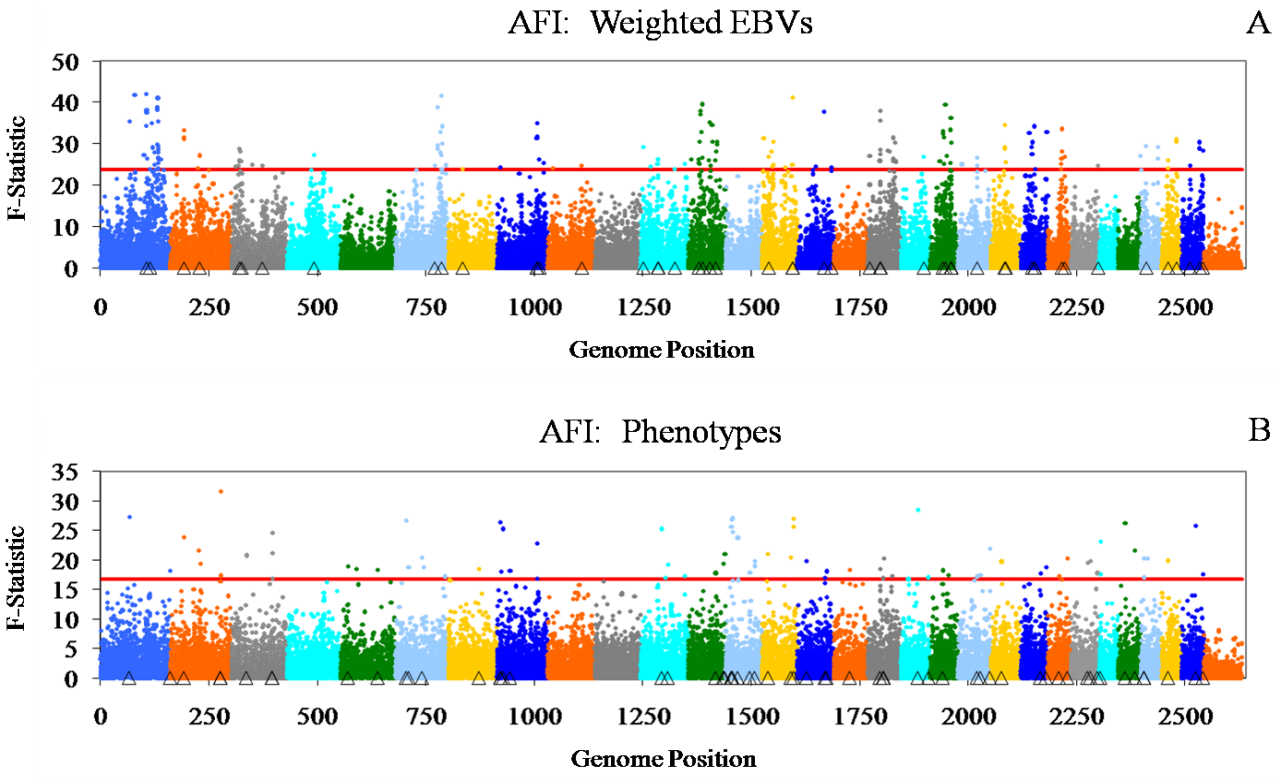

Figure 2: Concordance between EBV and phenotype analyses for AFI.

Literature Cited

Archer, J. A., P. F. Arthur, R. M. Herd, P. F. Parnell and W.S. Pitchford. 1997. Optimum postweaning test for measurement of growth rate, feed intake, and feed efficiency in British breed cattle. J. Anim. Sci. 75:2024-32.

Archer, J.A., E. C. Richardson, R. M. Herd and P. F. Arthur. 1999. Potential for selection to improve efficiency of feed use in beef cattle: A review. Aust. J. Exp. Agr. 50:147-61.

Arthur, P.F., J. A. Archer, D. J. Johnston, R. M. Herd, E. C. Richardson and P. F. Parnell. 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency and other postweaning traits in Angus cattle. J. Anim. Sci. 79:2805-11.

Boldman, K. G., L. A. Kriese, L. D. Van Vleck, C. P. Van Tassell and S. D. Kachman. 1995. A manual for use of MTDFREML. A set of programs to obtain estimates of variance and covariance. USDA, Agriculture Research Service, Clay Center, NE.

Churchill, G.A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-71.

Crews, D.H., Jr. 2005. Genetics of efficient feed utilization and national cattle evaluation: A review. Genet. Mol. Res. 4:152-65.

Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki. 2003. DAVID: Database for Annotation, Visualization and Integrated Discovery. Genome Biol. 4(5):P3.

Guiroy, P.J., D. G. Fox, L. O. Tedeschi, M. J. Baker and M. D. Cravey. 2001. Predicting individual feed requirements of cattle fed in groups. J. Anim. Sci. 79:1983-95.

Hayes, B.J., P. J. Bowman, A. J. Chamberlain and M. E. Goddard. 2009. Invited Review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433-43.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-47.

Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe and M. Hirakawa. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucl. Acids Res. 38:D355-D360.

Kanehisa, M., S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki and M. Hirakawa. 2006. From genomics to chemical genomics: New developments in KEGG. Nucl. Acids Res. 34:D354-D357.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucl. Acids Res. 28:27-30.

Huang, D. W., B. T. Sherman and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat. Protoc. 4(1):44-57.

Kennedy, B. W., J. H. van der Werf and T. H. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci. 71:3239-50.

Koch, R. M., L. A. Swiger, D. Chambers and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486-94.

MacNeil, M. D., D. R. Bailey, J. J. Urick, R. P. Gilbert and W. L. Reynolds. 1991. Heritabilities and genetic correlations for postweaning growth and feed intake of beef bulls and steers. J. Anim. Sci. 69:3183-9.

McKay, S. D., R. D. Schnabel, B. Murdoch, L. K. Matukumalli, J. Aerts, W. Coppieters, D. H. Crews, Jr., E. D. Neto, C. A. Gill, C. Gao, H. Mannen, P. Stothard, Z. Wang, C. P. Van Tassell, J. L. Williams, J. F. Taylor and S. S. Moore. 2007. Whole genome linkage disequilibrium maps in cattle. BMC Genet. 8:74-85.

Meuwissen, T. H., B. J. Hayes and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-29.

Okine, E. K., J. Basarab, L. A. Goonewardene and P. Mir. 2004. Residual feed intake and feed efficiency: Difference and implications. In: Florida Ruminant Nutrition Symposium, pp. 27-38.

Rolf, M. M., J. F. Taylor, R. D. Schnabel, S. D. McKay, M. C. McClure, S. L. Northcutt, M. S. Kerley and R. L. Weaber. 2010. Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle. BMC Genet. 11:24.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Brdg. Genet. 123:218-23.

Scheet P and M. Stephens. 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. Am. J. Hum. Genet. 78:629-644.

Spangler, M.L., J. K. Bertrand and R. Rekaya. 2007. Combining genetic test information and correlated phenotypic records for breeding value estimation. J. Anim. Sci. 85:641-649.

VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414-23.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor and F. S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16-24.

Williams, C.B., G. L. Bennett, T. G. Jenkins, L. V. Cundiff and C. L. Ferrell. 2006. Using simulation models to predict feed intake: Phenotypic and genetic relationships between observed and predicted values in cattle. J. Anim. Sci. 84:1310-6.

Williams, C.B. and T. G. Jenkins. 2003a. A dynamic model of metabolizable energy utilization in growing and mature cattle. I. metabolizable energy utilization for maintenance and support metabolism. J. Anim. Sci. 81:1371-81. 103

Williams, C.B. and T. G. Jenkins. 2003b. A dynamic model of metabolizable energy utilization in growing and mature cattle. II. metabolizable energy utilization for gain. J. Anim. Sci. 81:1382-9.

Willham, R. L. 1993. Ideas into action: A celebration of the first 25 years of the beef improvement federation. University Printing Services, Oklahoma State University, Stillwater, OK.