Breeding and Genetics: Genomic Selection and Whole-Genome Association

W52 Accuracy and bias of multiple-trait genomic evaluations for linear type traits in US Holsteins. S. Tsuruta*[1], I. Misztal[1], I. Aguilar[2], and T. Lawlor[3], *[1]University of Georgia, Athens, [2]Instituto Nacional de Investigación Agropecuaria, La Piedras, Canelones, Uruguay, [3]Holstein Association USA Inc., Brattleboro, VT.*

The objectives of this study were to conduct multiple-trait (MT) genomic evaluations and to compare bias and accuracy of genomic breeding values (GBV) with those from single-trait (ST) genomic evaluations, using different genomic relationship matrices (G). Genomic evaluations were conducted for 18 linear type traits of US Holsteins using genomic and phenotypic combined data. For each of 6931 bulls, 38,416 SNP markers were used as genomic information. A single-step approach was used to predict GBV with G that assumed various allele frequencies: equal (p = q = 0.5) (GE), base (GB), current (GC), and current with scaled G (GCS). Data sets contained 8,865,120 records for 5,657,787 cows from 2009 data and 7,715,925 records for 4,813,726 cows from 2004 data. Coefficients of determination (R^2) and regressions (δ) of 2004 GBV on 2009 daughter deviations were calculated, as statistics of accuracy and bias, respectively, for 1307 young bulls. Parent averages (PA) were also calculated for 2004 data with no genomic information. With MT (ST) models, average R^2 (%) were 20.5 (18.6) and 37.9 (34.6), 37.3 (34.1), 35.9 (34.2) and 37.7 (34.5), and average δ (× 100) were 76.7 (77.3) and 79.8 (79.4), 76.3 (75.0), 74.7 (73.8) and 76.7 (76.2), for PA and for GBV with GE, GB, GC and GCS, respectively. When the weight for the inverse of the numerator relationship matrix for genotyped animals was reduced to 0.7, R^2 remained almost identical while the average δ increased to 92.7 with GE, 90.7 with GB, 89.2 with GC, and 91.9 with GCS for MT models. The ST models required a stricter convergence criterion of 10^{-14} than that of 10^{-11} for MT models to achieve the same accuracy. The GBV obtained from MT models are more accurate than those from ST models. Modifying the weight for pedigree relationships of genotyped animals reduces bias in GBV. Multiple trait large-scale genomic evaluation is computationally feasible.

Key words: genomic evaluation, linear type trait, US Holstein

W53 Genomic imputation and evaluation using 342 high-density Holstein genotypes. P. M. VanRaden[1], D. J. Null*[1], G. R. Wiggans[1], T. S. Sonstegard[2], and E. E. Connor[2], *[1]Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD, [2]Bovine Functional Genomics Laboratory, ARS, USDA, Beltsville, MD.*

Genomic evaluations for 73,749 Holsteins were computed using 636,967 of the 777,000 markers on the Illumina high density (HD) chip. Observed data included 342 animals with HD genotypes, 54,676 animals with 42,503 marker (50K) genotypes, 17,371 animals with 2,614 marker (3K) genotypes, and 1,360 nongenotyped dams (0K) with >90% of haplotypes imputable from progeny. The HD genotypes were from 180 influential sires, 138 Beltsville research cows, and 24 other females. Percentages of correctly imputed genotypes were estimated using an example simulated chromosome for this same population structure with 1% of genotypes missing and 0.02% incorrect initially from each chip. Over all animals, 94.4% of genotypes were missing initially. After imputation of missing markers, 99.9% of genotypes were correct from HD, 98.0% from 50K, 88.4% from 3K, and 93.8% from 0K genotypes. These imputation rates with version 2 of program findhap.f90 are several percent higher than with version 1. Version 2 begins with long segments to improve haplotype matches for close relatives and ends with short segments to detect matches from more remote ancestors instead of choosing just one optimal segment length. Evaluations were tested using imputation of actual genotypes and August 2007 phenotypes to predict deregressed evaluations of bulls proven after August 2007. For 29 traits tested, estimated genomic reliability averaged 54.3% using 636,967 markers vs. 54.8% using 42,503 regressions vs. 29.9% from traditional parent average. Squared correlations with future data were higher for 10 traits and lower for 19 with HD than with 50K evaluations. The largest marker effects were located at very similar positions, but new markers from the HD chip often had larger effects than the best markers from the 50K chip. Results were less favorable than the 1.6% increase in reliability expected from simulation, but more animals with HD genotypes will improve imputation and reliability. Also, multi-breed evaluation could produce larger gains than the single-breed evaluation investigated here.

Key words: genomic evaluation, imputation, marker density

W54 Genomic evaluation of Angus-Brahman multibreed cattle for feed efficiency and postweaning growth using the Illumina 3k chip. M. A. Elzo*[1], G. C. Lamb[2], D. D. Johnson[1], M. G. Thomas[3], I. Misztal[4], D. O. Rae[1], J. G. Wasdin[1], and J. D. Driver[1], *[1]University of Florida, Gainesville, [2]North Florida Research and Education Center, Marianna, [3]New Mexico State University, Las Cruces, [4]University of Georgia, Athens.*

Lower density chips provide an affordable alternative to evaluate animals and characterize traits. The objective of this research was to evaluate 620 bulls, steers, and heifers ranging from 100% Angus to 100% Brahman (B) using phenotypes from for 4 postweaning feed efficiency and growth traits, pedigree, and SNP genotypes from 2900 loci (Illumina3k chip). Traits were residual feed intake (RFI), daily feed intake (DFI), feed conversion ratio (FCR), and weight gain (WG). Data was collected in a GrowSafe automated feeding facility from 2006 to 2010. Calves remained in pens for the 21-d pre-trial and 70-d trial periods. Animals were assigned to pens by sire group and sex. Concentrate consisted of cottonseed hulls, corn, molasses, and a protein, vitamin, and mineral supplement. Calves were evaluated using a polygenic-genomic mixed model. Fixed effects were contemporary group (year-pen), age of dam, sex of calf, age of calf, B fraction of calf, and heterozygosity of calf. Random effects were animal polygenic (AP; mean zero; variance = A*Vg; A = additive relationship matrix, Vg = additive genetic variance), additive SNP (AS; mean zero; variance = additive SNP variance), and residual effects (mean zero, common residual variance). Variance components and heritabilities were estimated using option VCE (Markov Chain Monte Carlo) of program GS3. Heritabilities were 0.18 for RFI, 0.30 for DFI, 0.16 for FCR, and 0.33 for PWG. The fraction of the additive genetic variance explained by the 2900 markers of the Illumina3k chip was 15% for RFI, 11% for DFI, 37% for FCR, and 20% for PWG. The AS, AP, and TA EBV for all traits tended to decrease as B fraction increased, suggesting that high percent B calves were genetically more efficient (lower RFI), but had lower WG.

Key words: feed efficiency, genomic, multibreed

W55 A neural network approach for association between a low-density whole genome SNP marker panel for 19 traits in beef

cattle. E. Hay*[1], H. Wang[1], X. Liu[1], B. Woodward[2], S. Bauck[2], and R. Rekaya[1], *[1]University of Georgia, Athens, [2]Merial Limited, Duluth, GA.*

The predictive ability of a low-density SNP panel derived from the Illumina Bovine SNP50 and developed for marbling, back fat, carcass weight, rib eye area, yearling weight, and heifer pregnancy rate was evaluated for 19 traits (residual feed, intake, dry matter intake, birth weight, weaning weight, docility, milk yield, average daily gain, yearling weight, carcass weight, marbling scores, rib eye area, back fat, mature weight and height, scrotal circumference, calving ease direct, calving ease maternal, heifer pregnancy). Data consisted of the genotypes of 2,245 Angus animals and their corresponding EPDs with missing rate (on phenotypes) ranging from 0 to 54% (heifer pregnancy). Missing genotypes were replaced with the most likely genotype. Linear regression (LR) and Neural Network (NN) approaches were implemented and compared. For LR, a cross validation procedure was adopted where the data was randomly divided into 5 groups with equal size. In each one of the 5 replicates, 80% of the data was used for training and the remaining 20% for validation. For the NN approach, randomly 2/3 and 1/3 of the data were used for training and validation, respectively. The process was replicated 5 times. A NN is an artificial system of massively interconnected neurons. The network architectures and the learning algorithm define the manner in which the neurons are related and structured. In this study, a feed-forward NN with one hidden layer was used. The parameters of the NN such as the number of neurons in the hidden layer, and learning rate were set heuristically. For both approaches, the lowest correlation between true and estimated breeding values was for docility and (0.52) and scrotal circumference (0.57). The highest accuracy was for yearling weight (0.82 for LR and 0.83 for NN). Across the 19 traits, the differences in performance between LR and NN range from 2 to 5% with superiority for the NN approach, except for weaning weight where the LR was slightly superior (0.27%), the superiority of the NN approach could be due to its ability to intrinsically accommodate the non-linear relationship between variables.

Key words: SNP, whole genome, beef cattle

W56 Whole genome association analyses for ultrasound and carcass merit traits in beef cattle. H. Li*, Z. Wang, P. Stothard, and S. S. Moore, *University of Alberta, Edmonton, Alberta, Canada.*

Carcass merit traits in beef cattle are of high interest in the beef industry and are quantitative in nature. The goal of this study was to identify genetic markers and potential candidate genes associated with meat quality traits in beef cattle. A genome-wide association study (GWAS) for 40,809 single-nucleotide polymorphism (SNP) markers using Bayesian methods was conducted in a total of 922 steers. Five carcass merit traits including carcass weight (CWT), carcass average backfat (CABF), carcass rib eye area (CREA), carcass grade fat (CGF), carcass lean meat yield (CLMY) and 3 ultrasound measurement traits including ultrasound marbling (UMAR), ultrasound backfat (UBF) and ultrasound ribeye area (UREA) were analyzed. The proportion of phenotypic variance explained by markers was 0.50 for CWT, 0.25 for CABF, 0.37 for CREA, 0.27 for CGF, 0.24 for CLMY, 0.39 for UBF, 0.63 for UMAR and 0.37 for UREA. SNPs with large substitution effects indicated that major genes exist for meat quality traits in beef cattle. For instance, the largest SNP on BTA6 explained 6.67% of the phenotypic variance for CWT. Many of the large effect SNPs coincided with previously identified QTL. Positional candidate genes (e.g., 49 for CWT) were identified within 1 Mbp windows flanking the top 10 SNP markers. Potential functional genes (e.g., *NCAPG*, *MED28* on BTA6 and *LYN* on BTA14 for CWT) were selected from these positional genes based on predicted and known functional information associated with the genes or their orthologs. Overall, this WGAS provides a list of markers and potential functional candidate genes associated with beef carcass merit traits that could be used to improve beef production and quality via marker-assisted selection.

Key words: beef cattle, genome-wide association study, carcass merit

W57 Large-scale SNP association analyses for somatic cell score in Canadian Holstein cattle. H. Li*[1], Z. Wang[1], F. S. Schenkel[2], M. Sargolzaei[3], S. S. Moore[1], and P. Stothard[1], *[1]University of Alberta, Edmonton, Alberta, Canada, [2]University of Guelph, Guelph, Ontario, Canada, [3]L'Alliance Boviteq, Saint-Hyacinthe, Québec, Canada.*

Genetic studies in dairy cattle have found that selection for higher milk production brings with it slightly higher rates of mastitis. Mastitis is the most common and most costly disease of dairy cattle and there are no affordable or prevailing methods for directly measuring mastitis. Somatic cell score (SCS) is an excellent indirect trait for selection for mastitis as the genetic correlation between SCS and mastitis is 60% to 80%. The goal of this study was to identify genetic markers and potential candidate genes associated with SCS in dairy cattle. A genome-wide association study for 29,552 single-nucleotide polymorphism (SNP) markers was conducted in a total of 647 Canadian Holstein bulls. The analyses used a combination of Bayesian models to predict individual SNP effects. The proportion of the variance of the bulls' deregressed SCS EBVs explained by markers was 0.76. SNPs with large substitution effects indicated that major genes exist for SCS. The10 SNPs with largest effects were on BTA6, 14, 21, 18, and 19. Many of the large-effect SNPs coincided with previously identified QTL. One hundred and 32 positional candidate genes were identified within 2 Mbp genomic regions flanking the top 10 SNPs. Potential functional genes (e.g., GC and NPFFR2 on BTA6, LY6D on BTA14, MFGE8 and AP3S2 on BTA21, and FOXL1, CYLD and NOD2 on BTA18) were selected from these positional genes based on predicted and known functional information associated with the genes or their orthologs. Functional annotation revealed that these genes are involved in the regulation of immune system activity and inflammation. For instance, the CYLD gene on BTA18 is a crucial negative regulator of inflammation and subsequent immune response in antibacterial defense. This finding reinforces the role of the genetic control on immune response to bacterial infections. The list of SNPs and potential functional candidate genes may prove useful for prevention of mastitis via marker-assisted selection.

Key words: Holstein cattle, genome-wide association study, somatic cell score

W58 Comparison of selective genotyping strategies for prediction of breeding values in a population undergoing selection. A. A. Boligon*[1,2], N. Long[2], L. G. Albuquerque[1], K. A. Weigel[3], D. Gianola[2,3], and G. J. M. Rosa[2], *[1]Department of Animal Sciences, Sao Paulo State University, Jaboticabal, SP, Brazil, [2]Department of Animal Sciences, University of Wisconsin, Madison, [3]Department of Dairy Science, University of Wisconsin, Madison.*

A simulation study was used to evaluate the predictive ability of genomic selection under different strategies of selective genotyping. The genome consisted of 10 chromosomes, each with 202 markers and 100 QTLs evenly spread across 100 cM. After 5,000 generations of random mating with an effective population size of 100 (50 males

and 50 females), a reference population (G0) was generated by a full factorial mating between the 50 males and 50 females. In G0 animals were selected based on their phenotypes (highest phenotypic values) or at random, at different selection intensities, to produce generation G1. Five genotyping strategies were considered to choose 500 animals in G0 as a training set: random sampling (R); highest phenotypic values (H); lowest phenotypic values (L); extreme (low and high) phenotypic values (E); and subset of less related animals (I). The number of individuals in G0 and G1 was fixed at 2,500, and heritability was set to 0.10, 0.25 and 0.50. Additionally, all 5 genotyping strategies were applied also to an indicator trait with a genetic correlation of 0.5 with the target trait. The selective genotyping strategies were compared in terms of their ability of predicting the genetic values of the animals in G1, using a Bayesian Lasso model with all 2020 markers simultaneously. For all simulated traits, the lowest correlation between predicted and true breeding values (GEBV and TBV, respectively) was obtained when using the L genotyping strategy. For E, R, and I strategies, the correlation between GEBV and TBV became slightly higher as selection intensity decreased, and was largest when no selection occurred. These 3 strategies were better than H. In addition, the E, R, and I approaches had lower prediction mean squared errors (PMSE), followed by H and then by L. Overall, genotyping strategy E led to the best predictive ability of breeding values, indicating that animals with extreme phenotypic values in a reference population are the most informative when training genomic selection models.

Key words: Bayesian Lasso, molecular markers, predictive ability

W59 Estimating genomic breeding values in crossbred animals. E. H. Hay*, S. Smith, and R. Rekaya, *University of Georgia, Athens.*

For several livestock species and traits, the accuracies of predicted molecular breeding values are significantly higher than those obtained using parent averages and similar to those obtained for proven animals. These exciting results are unfortunately valid only when animals of the same breed are used in the training and validation sets. Unfortunately, in several applications admixture and crossbred animals are present making the training and validation sets very heterogeneous. Several studies have shown that using SNP estimates from pure breeds on other breeds or crossbred animals have little success. Alternative solutions using different mixture of animals (pure breeds, crossbreds) in training and validation sets were proposed with promising results in some cases. In this study a different approach is adopted where estimates of SNP effects in pure breeds are used to predict molecular breeding values in crossbred animals via the projection of these effects based on the percentage contribution of each pure breed in the crossbred individual. To test the performance of the proposed procedure, a simulation was conducted involving 2 breeds (A and B) with 1,500 animals each and 3 crosses 50/50, 75/25, and 25/75% of A and B breeds, respectively. Genotypes for a low density panel of 384 SNPs were also simulated for all animals without missing. Four analyses were conducted for estimating the SNP effects: 1) using only data from breed A, 2) using only data from breed B, 3) used data from breeds A and B, and 4) using data from A, B and the crossbreds. The results indicated that estimating SNP effects in analysis 4 led to the highest accuracies in crossbreds compared with the other 3 alternatives. However, when our approach was used based on the projection of the SNP estimates from analyses 1 and 2 as a function of the breed composition of the crossbreds, the best results were obtained. In fact, the accuracies in crossbreds increased by 28% compared with the results obtained in analysis 4. Based on this results, it seems that there a possibility of using pure bred estimates of SNP marker effects to predict the molecular breeding values of some crossbred animals.

Key words: genomic, SNP, crossbreds

W60 Accounting for new mutations in the genomic relationship matrix. J. Casellas*, *G2R, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain.*

In this abstract, procedures for calculating the genomic relationship matrix are adapted to accommodate the occurrence of new mutations in the autosomal genome, i.e., the mutational genomic relationship (G_m) matrix. These algorithms derived from the original development by N. R. Wray for the additive numerator relationship matrix and accommodated genomic data from massive genotyping technologies. Assume as starting point g non-overlapping generations with n individuals per generation, all of them being genotyped for m single nucleotide polymorphism (SNP) markers spread across the whole autosomal genome. Moreover, assume $p_{k(l)}$ as the allelic frequency of the kth SNP in the lth generation and $\pi_{k(l)} = 2(p_{k(l)}\text{-}0.5)$ as a correcting factor to set mean values of allele effects to 0; the genotype of the ith individual at the kth SNP (γ_{ik}) was assumed as −1, 0 or 1 (i.e., homozygote, heterozygote and other homozygote, respectively). The general structure of the G_m matrix can be defined as $G_m = G_1 + G_2 + \ldots + G_g$, G_q being a genomic relationship matrix where individuals from generation q were treated as founders and the rows and columns linked to ancestors from previous generations were fixed to 0. To avoid the successive computation of all intermediate G_q matrices, the final mutational genomic relationship between individuals i and j (ρ_{ij}; i being older than j) reduced to the following expression $\rho_{ij} = \sum_{0>r>m}\sum_{0>s>g^*}[(\gamma_{ir} - \pi_{r(s)})(\gamma_{jr} - \pi_{r(s)})]/[2\sum_{0>q>m}p_{q(s)}(1 - p_{q(s)})]$, where g^* was the generation of origin of individual i. This matrix could be straightforwardly implemented within the structure of the genomic BLUP model in order to account for the genetic variability originated from young mutations.

Key words: genomic relationship, matrix, mutation