GENOMIC METHODOLOGY

0939 (M053) Signature of selection reveals large difference in selection traits. X. Zhang*[1], I. Misztal[1], M. Heidaritabar[2], J. W. M. Bastiaansen[3], R. Hawken[4], R. Okimoto[4], R. L. Sapp[4], H. H. Cheng[5], D. A. L. Lourenco[1], and W. M. Muir[6], *[1]University of Georgia, Athens, [2]Wageningen University, Netherlands, [3]Animal Breeding and Genomics Centre, Wageningen University, Netherlands, [4]Cobb-Vantress Inc., Siloam Springs, AR, [5]USDA, ARS, ADOL, East Lansing, MI, [6]Purdue University, West Lafayette, IN.*

Selection on animals changes the population-wide frequency spectrum of genes related to the traits under selection. With the aid of single-nucleotide polymorphism (SNP) methods, it is possible to inspect for changes in allelic frequencies directly. To reveal the impact of recent selection on genetic variation, we compared the allele frequencies before and after three generations of selection on an index of three traits in two lines (F and M) sampled from commercial broiler chicken. Line M animals are from a sire line that was selected mainly for growth traits, and line F animals are from a dam line that was selected mainly for reproductive traits. Selection was performed by applying single-step Genomic Best Linear Unbiased Prediction (ssGBLUP). Genotypes were used in this study for allele frequency analysis. The M and F lines consisted of 4922 and 4904 genotyped animals, respectively. After quality control, genotypes included information on 52,742 and 52,639 SNPs in line M and F, respectively. Selection was for an index consisting of body weight at 6 wk, ultrasound measurement of breast meat, and leg score. The average allele frequency change for both lines on autosomes was 0.049. Threshold value for detecting selected regions, where allele frequency changes exceeded expectations under drift were 0.140 and 0.136 for line M and F, respectively. There were 25 and 17 selection regions detected on line M and F, respectively, without any overlap of regions between the lines. Average 4heterozygosity change in line F was greater compared to line M (0.008 vs. 0.003, $P < 0.01$). The putative selected regions between line M and F are different. The results we present indicate that in newly selected populations, the genotype frequencies across chromosomes change differently according to the selection lines even if animals are selected for same traits.

Key Words: SNP, allele frequency change, genomic selection

0940 (M054) Weighted single-step genomic BLUP: an iterative approach for accurate calculation of breeding values and SNP effects. X. Zhang*, D. A. L. Lourenco, and I. Misztal*, University of Georgia, Athens.*

The purpose of this study was to explore options for genome wide association analysis (GWAS) with single-step GBLUP (ssGBLUP). In GWAS by ssGBLUP, GEBV are converted to marker (SNP) effects. Unequal variances for markers are then derived from SNP solutions and subsequently incorporated into a weighted genomic relationship matrix. Improvements on the SNP weights can be obtained iteratively either by recomputing the SNP effects only or by also recomputing the GEBV. Four options were used to calculate the weights: 1) proportional to $2p_i(1-p_i)u_i^2$, where p_i and u_i are frequency and effect of the i-th SNP; 2) proportional to $2p_i(1-p_i)u_i^2$+ constant; 3) weights as in 1, but updating only the top 25 SNP; 4) updating only the top 5 SNP. A simulated data set was used that included 15,600 animals in five generations, of which 1540 were genotyped for 50k SNP. The simulation involved phenotypes for a trait with heritability of 0.5 potentially affected by 5 QTL. Accuracy between TBV and GEBV for genotyped animals in generation 5 was used for evaluation. Comparisons also involved BayesC with deregressed proofs and $\pi = 0.9999$. In single-step, SNP effects were tracked along 10 iterations and weights were equal to 1.0 in the first iteration. Results showed option 3 as the best in identifying simulated QTL without background noise and with precision in most of the regions, as well as BayesC; after two iterations, the accuracy of GEBV reached a plateau and was 0.91 as opposed to 0.88 for BayesC. Testing also included a commercial data set with 200k animals and 15K genotypes for 39k SNP. For one of the traits, Manhattan plots with option 3 and BayesC looked identical showing six large peaks and very small background noise. However, the realized accuracy was 0.16 in the first round and 0.14 in the subsequent rounds, as opposed to 0.19 for BayesC. For the other traits, the accuracy by BayesC was lower and Manhattan plots did not have clear peaks. The option to compute weights for SNP in ssGBLUP with the top 25 SNP gives a good identification of top segments. However, further work is required to compute weights to maximize accuracy for a variety of cases. In addition, a choice for GWAS in single-step approach is based on simplicity and flexibility in case of complex models.

Key Words: weighted SNP, ssGBLUP, BayesC

0941 (M055) Derivation of Bayes and Minimax decision rules for allelic frequencies estimation in biallelic loci. C. A. Martinez*[1,2], K. Khare[2], and M. A. Elzo[1], *[1]Dept of Animal Sciences, University of Florida, Gainesville, [2]Dep. of Statistics, University of Florida, Gainesville.*

In population genetics, allelic frequencies are typically estimated via maximum likelihood (MLE). Under this setting, allele frequencies are treated as unknown fixed parameters. However, population genetics theory indicates that allele frequencies vary at random, thus they should be treated as random variables. The aim of this study was to derive Bayes and Minimax estimators (ME) of allele frequencies for biallelic loci using decision theory. Because an optimal decision rule with uniformly smallest risk rarely exists, an approach is to establish principles that allow ordering of decision rules according to their risk function. Two general methods were used to obtain average risk optimality: the Bayes and the Minimax principles. Briefly, given a loss function and a prior distribution, the Bayes principle looks for an estimator minimizing the posterior risk, while the Minimax principle consists of finding decision rules that minimize the supremum (over the parameter space) of the risk function (the worst scenario). For an arbitrary locus, the sampling model was a trinomial distribution for numbers of individuals for each genotype and the prior was a Beta distribution, chosen because of mathematical convenience, flexibility and genetic interpretation of its parameters. Three types of loss functions were considered: square error (SEL), Kullback-Leibler (KLL), and a quadratic error loss (QEL). The SEL and KLL yielded the same estimator, which was a convex combination of the prior mean and the MLE. Using the Bayes estimator from QEL, a ME was derived by applying a theorem that states that a Bayes estimator with constant risk is also Minimax. The constant risk was obtained by finding appropriate hyperparameter values. This estimator was shown to be equivalent to MLE. The prior associated with this ME was uniform [0, 1]. One consequence of using the previous theorem on the derivation of ME is that the uniform distribution is a least favorable prior, that is, it causes the greatest average loss. Extension to several loci under linkage equilibrium and independent priors was discussed. The estimators derived here have the appealing property of allowing variation in allelic frequencies, which is more congruent with the reality of finite populations exposed to evolutionary forces. In addition, from a Bayesian perspective they permit modeling uncertainty and incorporation of previous genotypic information from the population.

Key Words: allele frequencies, average risk optimality, decision theory

0942 (M056) Adjusting genomic relationship matrices in single-step genomic BLUP for crossbred evaluations. D. Lourenco, and I. Misztal, *University of Georgia, Athens.*

Different breed-specific genomic relationship matrices (GB) were compared to the standard across-breed genomic relationship matrix (G) used in single-step genomic evaluations. Datasets were simulated that resembled a terminal-cross population. Two purebred lines were separated by 50 generations. Three scenarios considered selection based on high EBV, high phenotypes, and no selection. The datasets used for evaluations contained phenotypes and pedigrees for the last 15 generations and genotypes for the last eight generations of purebreds. Data on F1 animals were from a single generation. Number of purebred parents genotyped varied from 3100 to 3300 depending on the scenario, and number of genotyped F1 was 1200. The heritability for the simulated trait was 0.30. Testing involved four genomic matrices: GB1 considered specific allele frequencies (AF) for each pure and crossbred; GB2 used AF for crossbred calculated based on AF from the two purebreds; GB3 and GB4 had AF as in GB2 and GB1, respectively; however, each element was scaled by breed-specific scaling factors. Across-breed and breed-specific correction factors for G and all GB were also used to account for the non-random genotyping caused by selection. The validation was done in F1 animals and parameters of the regression of TBV on GEBV were used to assess the accuracy of evaluations. For G and all GB, coefficients of determination (R^2) and regression were higher when no artificial selection was applied. When no correction factor was used, R^2 for G, GB1, GB2, GB3, and GB4 for EBV selection were 0.33, 0.03, 0.37, 0.37, and 0.03, respectively; for mass selection were 0.23, 0.33, 0.37, 0.38, and 0.33, respectively; for no selection were 0.47 for G and 0.46 for all GB. However, after using breed-specific correction factors, the difference between G and GB was reduced and GB1 and GB4 gave similar results to G (R^2 = 0.40 under EBV and mass selection; R^2 = 0.47 under no selection), while GB2 and GB3 had slightly worse performance. Most unbiased predictions were with G and the correction factor applied, which regressions were close to 1.0 for the purebreds and from 0.65 to 1.0 for the crossbreds; the highest inflation was with the EBV selection and no phenotypes on crossbreds. Breed-specific genomic matrices provide little benefits for genomic evaluations in a terminal cross model. The best performance is with standard G corrected for an average selection across breed types.

Key Words: ssGBLUP, breed-specific, allele frequency