Abstracts of the 2017 American Dairy Science Association® Annual Meeting

June 25–28, 2017
Pittsburgh, PA

Journal of Dairy Science®
Volume 100, Supplement 2

was in average 0.63 across generations and chromosomes, Genomic relationship matrices were computed from a) unweighted regular SNP, b) unweighted regular SNP and + QTN, c) regular SNP with variances from GWA, d) unweighted regular SNP and QTN with known variances, e) as before but only using 10% of the largest QTNs, and f) using only QTNs with known variances. Accuracies for the 11th generation were computed by BLUP and ssGBLUP. To ensure full rank, raw genomic relationship matrices (GRM) were blended with 1% or 5% of numerator relationship matrix, or 1% of the identity matrix. Rank of GRM with 100 QTN as determined by the number of eigenvalues explaining 90% variation in GRM was 8,497 for unweighted GRM, increased to 9,553 after blending,decreased to 4,054 with weighted GRM and 10% QTN included, and was 76 when only causative QTNs were used. The accuracy for the last genotyped generation with BLUP was 0.32. For ssGBLUP, that accuracy increased to 0.49 with a regular GRM, to 0.53 after adding unweighted QTN, to 0.63 when QTN variances were estimated, and to 0.89 when QTN variances were assumed known. When GRM was constructed from QTNs only, the accuracy was 0.95 with 5% blending rising to 0.99 with 1% blending. Accuracies assuming 1000 QTN were generally lower, with a similar trend. Accuracies using the APY inverse were equal or higher than those with a regular inverse. The rank of weighted GRM is between the rank of unweighted GRM and that computed with causative SNP only. Single-step GBLUP can account for causative SNP when variances of causative QTN are known.

Key Words: genomic relationship matrix, genomic prediction, causative variant

468 Impact of pedigree truncation on accuracy and convergence of ssGBLUP in a population with long pedigree when only a fraction of animals are phenotyped. I. Pocrnic*[1], D. A. L. Lourenco[1], H. L. Bradford[1], C. Y. Chen[2], and I. Misztal[1], [1]*Department of Animal and Dairy Science, University of Georgia, Athens, GA,* [2]*Genus PIC, Hendersonville, TN.*

In a genomic evaluation, it is desirable to have low computing cost while retaining high accuracy of evaluation for young animals. When the population is large but only few animals have phenotypes, especially for low heritability traits, the convergence rate of BLUP or single-step genomic BLUP (ssGBLUP) can be very slow. While eliminating old pedigrees can seriously affect (G)EBV for old animals, usually only younger animals are candidates for selection. This study investigates the effect of pedigree truncation on convergence rate and accuracy of prediction for young animals. The data consisted of 216k, 221k, 722k, and 579k phenotypes on 4 traits (T1, T2, T3, T4) from a purebred pig line. Heritabilities were <0.1 for T1 and T2, and >0.2 for T3 to T4. A total of 2.4 million animals born from 1971 to 2016 were included in the complete pedigree. Genotypes were available for 33,502 animals and consisted of 60,003 SNP. A bivariate animal model was fit for T1–2, and T3–4, separately. Computations were done by BLUP or ssGBLUP, and were conducted with complete pedigree or different levels of pedigree depth (Pn), where n = 1, 2, 3, 4, 5. Pedigree depth n was defined as n ancestral generations from the animals with phenotypes. The number of pedigree animals for T1–2 (T3–4) varied from 226k (760k) for P1 to 228k (767k) for P5. Genomic relationship matrix was inverted either by a regular or the algorithm for proven and young (APY). GEBV between runs with the complete and pruned pedigrees for genotyped animals were correlated at >0.99 for P2 to P5. For T1–2 (T3–4), convergence required up to 7,381 (1,421) rounds with the complete pedigree; this number decreased for different levels of pedigree depth up to less than 1,730 (854) rounds for P2. Use of the APY inverse in ssGBLUP improved convergence up to 25% on average, without affecting accuracy. Pedigree pruning and the APY algorithm are important tools to reduce the computing cost of ssGBLUP without negatively impacting accuracy of predictions.

Key Words: algorithm for proven and young, pedigree depth, single-step genomic BLUP

469 Bayesian whole-genome prediction and genome-wide association analysis with missing genotypes using variable selection. C. Chen*[1], K. A. Weigel[2], E. E. Connor[3], D. M. Spurlock[4], M. J. VandeHaar[1], C. R. Staples[5], and R. J. Tempelman[1], [1]*Michigan State University, East Lansing, MI,* [2]*University of Wisconsin-Madison, Madison, WI,* [3]*USDA-ARS, Beltsville, MD,* [4]*Iowa State University, Ames, IA,* [5]*University of Florida, Gainesville, FL.*

Single-step genomic best linear unbiased Predictor (ssGBLUP) has become increasingly popular for whole-genome prediction (WGP) modeling as it utilizes any available pedigree and phenotypes on both genotyped and non-genotyped individuals. The WGP accuracy of ssGBLUP has been demonstrated to be greater than or equivalent to popular Bayesian regression models. However, these assessments have not typically included phenotypes of non-genotyped individuals in the Bayesian regression analyses, making the interpretation of these comparisons difficult. Increasingly, ssGBLUP has been used for genome-wide association (GWA) studies, although there is no clear guidance on how to determine statistical significance in these analyses. We address this issue and additionally propose a GWA based on a Bayesian single-step stochastic search and variable selection (ssSSVS) model that allows for phenotypes on non-genotyped animals. Our study was based on a dairy consortium data set including 3,186 Holstein cows from 6 US research stations based on the 60671 USDA-ARS bovine SNP panel. In a replicated simulation study using these same genotypes, a different number of causal variants (n_c = 30, 300, or 3,000) were randomly assigned to the markers, masking 20% of cows as non-genotyped, for a trait having a heritability of 0.25. We determined that ssSSVS had greater ($P < 0.05$) WGP accuracy than ssGBLUP with n_c = 30 or n_c = 300. Moreover, ssSSVS always performed better ($P < 0.05$) than ssGBLUP for GWA measured as partial area under a receiver-operating characteristic (ROC) curve (pAUC) up to a false positive rate of 5%. In a 10-fold within-station cross-validation study using phenotypes from the dairy consortium, we determined that ssSSVS had greater ($P < 0.05$) WGP accuracies in milk fat compared with ssGBLUP for genotyped individuals, although no such differences were detected for body weight. No differences between ssSSVS and ssGBLUP for prediction accuracies for non-genotyped individuals were determined for either trait. Overall, ssSSVS is a promising method for both WGP and GWA, particularly for genetic architectures characterized by a few genes with large effects.

Key Words: Bayesian variable selection, genome wide association, whole genome prediction

470 SSGP: SNP-set based genomic prediction to incorporate biological information. J. Jiang*[1], J. O'Connell[2], P. VanRaden[3], and L. Ma[1], [1]*Department of Animal and Avian Sciences, University of Maryland, College Park, MD,* [2]*University of Maryland School of Medicine, Baltimore, MD,* [3]*Animal Genomics and Improvement Laboratory, ARS-USDA, Beltsville, MD.*

Genomic prediction has emerged as an effective approach in plant and animal breeding and in precision medicine. Including biological information into the genomic model can be of great advantage. Due to the statistical and computational challenges in large genomics studies,

however, a fast and flexible method to incorporate such external information is still lacking. Here, we proposed a linear mixed model that can incorporate biological information in a flexible way and developed a fast variational Bayes-based software package named SSGP. In our model, whole genome markers can be split into groups in a user-defined manner, and each group of markers is given a common effect variance. Since previous functional genomics studies have accumulated much evidence on which genes, genomic regions or pathways are more/less important for a trait of interest, we can divide genome-wide SNPs into several groups based on their levels of importance and then use the predefined SNP sets in SSGP. Additionally, each marker has a prespecified weight for which the rule can be flexibly assigned, e.g., based on minor allele frequency or LD pattern. The model was implemented with the parameter expanded variational Bayesian method. For testing purpose, we analyzed a large cattle data set consisting of ~24k bulls (20k in training set and 4k in validation set) and ~760k whole-genome SNP markers. By simply grouping markers based on proximity (markers were divided into continuous, non-overlapping chunks, each containing 1k SNPs) and considering only additive effects, SSGP already performed better than Bayes A in all 5 milk traits analyzed, with an increase of up to 8 percent points in prediction accuracy. Meantime, it took only ~5h for each trait with 20 threads. We also analyzed many simulation data sets and the WTCCC heterogeneous stock mice data set for which the results of many existing methods had been reported. Generally, SSGP could achieve similar prediction performance compared with the best approaches reported, though only proximity was used for grouping SNPs. Collectively, the method and software show great potential to increase accuracy in genomic prediction, particularly in the future when more useful biological information is becoming available.

Key Words: genomic prediction, SNP set, biological information