tions and quantify deviations from the assumptions. Data from the 1000 Bulls Project Run9 (n = 6191 genomes) will be presented examining multiple alleles, private alleles, genotype concordance between runs and variant filtering thresholds.

Key Words: cattle, sequence, imputation

1047 Mixed-model GWAS on milk production traits of 1.16M genotyped Holstein cattle. J. Jiang*[1], J. Cheng[1], C. Maltecca[1], L. Ma[2], P. M. VanRaden[3], and J. R. O'Connell[4], [1]*Department of Animal Science, North Carolina State University, Raleigh, NC,* [2]*Department of Animal and Avian Sciences, University of Maryland, College Park, MD,* [3]*Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD,* [4]*Department of Medicine, University of Maryland School of Medicine, Baltimore, MD.*

Genome-wide association studies (GWAS) have been widely used for elucidating the genetic basis of complex traits. The mixed-model method is usually needed to account for sample relatedness and polygenic effects in GWAS, but it is computationally challenging to apply it to large-scale samples. We here present a new solution to mixed-model GWAS, which we refer to as SLEMM (https://github.com/jiang18/slemm), and apply it to the largest-to-date GWAS on milk production traits by using data from the US Council on Dairy Cattle Breeding. SLEMM enables million-scale genomic restricted maximum likelihood estimation and accurate approximation of mixed-model association statistics. We used deregressed estimated breeding values and ~76K autosomal SNP genotypes of ~1.16M Holstein cattle in mixed-model association analysis. The mixed model's polygenic term was accounted for by ~48K LD-pruned SNPs. Single-marker association statistics were computed for the 76K SNPs. This GWAS identified few new associations on milk production traits compared with our previous analysis with only 27K Holstein bulls. GWAS with subsamples of 50K, 100K, 150K, and 200K individuals showed that the increase in sample size has a bigger effect on *P*-values of significant SNPs than nonsignificant ones; that is, nonsignificant SNPs rarely become significant as the sample size increases. In summary, this study suggests that dairy GWAS in Holsteins reach saturation at relatively small sample sizes.

Key Words: GWAS, mixed model, milk production

1048 SLEMM: Million-scale genomic best linear unbiased predictions with window-based SNP weighting. J. Cheng*[1], C. Maltecca[1], P. Vanraden[2], J. O'Connell[3], L. Ma[4], and J. Jiang[1], [1]*North Carolina State University, Raleigh, NC,* [2]*Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD,* [3]*University of Maryland School of Medicine, Baltimore, MD,* [4]*University of Maryland, College Park, MD.*

The amount of animal genomic data is increasing exponentially. Using a large number of genotyped and phenotyped animals for genomic predictions is appealing yet challenging. The genomic best linear unbiased prediction (GBLUP) model and various SNP-based Bayesian alphabet models such as Bayes R remain widely popular for genomic prediction. The Bayesian models are typically advantageous for traits that have genes of large effect. However, the Markov chain Monte Carlo (MCMC) sampling method often used by Bayesian models is time-consuming. Here we present an alternative approach in a framework of multi-step evaluation for million-scale genomic predictions, which we refer to as SLEMM. Unlike MCMC, SLEMM relies on an efficient implementation of the stochastic Lanczos algorithm for REML and BLUP. We further develop a window-based SNP weighting method to improve prediction accuracy. SLEMM was compared with GBLUP and Bayes R in terms of prediction accuracy. Extensive data analyses, covering a spectrum of polygenic traits in multiple plant and animal species, show that SLEMM had comparable accuracies with Bayes R (0.3% lower than Bayes R and 3% greater than GBLUP for animals; 2% greater than Bayes R and 0.2% greater than GBLUP for plants, where most traits are highly polygenic). SLEMM was further applied on a large-scale Holstein cow data set (5 milk production traits from about 300K animals with 60K SNPs) from the Council on Dairy Cattle Breeding. Prediction accuracies using SLEMM were consistently greater than using Bayes R (0.1 to 2% greater) and GBLUP (0.3 to 1% greater). Simulation analyses show that SLEMM can complete genomic predictions for 0.5M genotyped animals and 50K SNPs in ~0.4 h with 9 GB of memory while Bayes R used ~6.6 h with 24.5 GB of memory. SLEMM used ~5.5 h and 63 GB of memory for prediction of 3M animals whereas Bayes R had a limitation of 0.5M animals in this case. In short, SLEMM paves the way for million-scale genomic predictions. Further comparison with single-step GBLUP will be evaluated. SLEMM is freely available at https://github.com/jiang18/ssgp.

Key Words: genomic prediction, BLUP, SLEMM