adjustments for a genomic base. Backsolving genomic predictions to SNP effects may require only a group of genotyped animals representing the dimensionality of the genomic information. The results obtained in this study are applicable to large genotyped populations.

Key Words: algorithm for proven and young, direct genomic value, interim evaluations

168 Potential benefits from using a new reference map in genomic prediction. D. J. Null*[1], P. M. VanRaden[1], D. M. Bickhart[2], J. B. Cole[1], J. R. O'Connell[3], and B. D. Rosen[1], [1]*USDA Animal Genomics and Improvement Laboratory, Beltsville, MD,* [2]*USDA Dairy Forage Research Center, Madison, WI,* [3]*University of Maryland-Baltimore, Baltimore, MD.*

Many genomic studies in cattle have used the 2009 reference assembly from the University of Maryland (UMD3.1). A new USDA Agricultural Research Service-University of California, Davis (ARS-UCD) assembly based on longer DNA reads from the same cow (Dominette) should improve sequence alignment, imputation, and genomic prediction. To test imputation, markers were converted from their previous map locations to new map locations in a process called liftover. Flanking sequences from array manifests or from the UMD3.1 map were remapped to the new assembly. For the 60,671 markers currently used in US genomic evaluations, > 99% aligned in forward direction to the same chromosome, but a few previously used markers were no longer usable. The new locations were then used to impute the 60,671 markers from genotype subsets available on most arrays for 1,748,453 Holsteins (HO), 215,800 Jerseys (JE), 32,724 Brown Swiss (BS), 4,834 Ayrshires (AY), and 3,517 Guernseys (GU). Average numbers of distinct haplotypes per segment decreased 5% for HO and from 1 to 30% for other breeds. Many previous problem areas no longer have excess numbers of haplotypes, particularly on the X chromosome and the pseudoautosomal region of X. Truly lethal haplotypes were more cleanly separated from false candidate haplotypes. Percentage of haplotypes with parent-progeny noninheritance dropped from 3.7 to 3.1 for HO, 4.4 to 3.9 for JE, 1.4 to 1.2 for BS, 2.5 to 1.5 for AY, and 1.6 to 1.4 for GU. Percentage of inherited haplotypes with 1 mistake dropped from 4.9 to 3.7 for HO, 4.3 to 3.6 for JE, 2.9 to 2.8 for BS, 3.5 to 2.8 for AY, and 3.0 to 2.7 for GU. Only a few segments such as on the left end of chromosome 8 had poorer properties for all breeds. Several regions of UMD3.1 were previously known to be on incorrect chromosomes and were excluded from use but can now be used with ARS-UCD. To test sequence alignment, paired-end reads from a HO bull were aligned to both maps, and 2.3% more paired reads aligned in the correct orientation within 5,000 base pairs. The new map improves genotype imputation, sequence alignment, and marker locations.

Key Words: reference assembly, genomic prediction, liftover

169 Implications of limited dimensionality of genomic information on persistency of genomic predictions and GWAS. I. Misztal*, I. Pocrnic, and D. Lourenco, *University of Georgia, Athens GA.*

The purpose of this study was finding possible explanation on peculiarities of dimensionality (M) of genomic information. The gene content matrix derived from 35 to 60k SNP chips has a limited M as determined by singular value decomposition; identical results are obtained with eigenvalues of genomic relationship matrix. Even with a very large number of animals, M ranges from about 4,000 for commercial pigs and broiler chicken, to about 15,000 in Holsteins. This number is normally attributed to the expected number of chromosome junctions as derived by Stam: M = 4NeL, where Ne is effective population size and L is genome size. However, approximation of realized accuracies assuming M for animals with same information is not accurate. Accuracies of genomic prediction assuming M/4 animals in genomic recursions and the APY algorithm are >90% of those assuming full dimensionality. These recursions also suggest that predictions based on M animals with very high reliability should be both very accurate and persistent, and predictions from large national evaluations in Holsteins could converge. However, the real accuracies seem lower than expected. The genome in a population can be visualized in 2 ways. First, as Ne haplotypes within each 1/4 Morgan segment. Second, as 4NeL sequential segments. Eigenvalues analyses of the genomic information shows that popular segments cluster along the genome. Subsequently the number of segments can be higher than determined by singular values. In particular, M/4 clusters could account for 90% of segments. SNP selection decreases the dimensionality; the minimum is the number of causative SNP. SNP selection can eliminate clusters without substantial variation but point to clusters with high variation, potentially creating high GWAS signals not related to QTL. Some ideas in this study were derived from simulated populations assuming complete genome coverage and an additive model. It remains to be seen whether accuracy predictions in real populations are affected by additional factors such as incomplete genome coverage and non-additive effects. Singular value analysis of gene content (or eigenvalue analysis of genomic relationship matrix) helps understand the complexity of genomic selection.

Key Words: genomic selection, APY algorithm, dimensionality

170 Modelling uncertain paternity to address differential pedigree accuracy. H. L. Bradford*[2,1], Y. Masuda[1], J. B. Cole[2], I. Misztal[1], and P. M. VanRaden[2], [1]*University of Georgia, Athens, GA,* [2]*Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD.*

The objective was to implement uncertain parentage models to account for differences in daughter pedigree accuracy. Elite sires have nearly all daughters genotyped resulting in correct paternity assignment. Bulls of lesser genetic merit have fewer daughters genotyped creating the possibility for more paternity errors in their daughters. Data were simulated with QMSim for a moderately (0.3) heritable, sex-limited trait. We created missing pedigrees by removing 8% of sires and 18% of dams. In total, 15 bulls were selected each generation, and the daughters of the best 5 bulls had accurate pedigrees. Daughters of the remaining 10 bulls had 9% sire and 3% dam pedigree errors. Data included 164,500 pedigree animals, 90,000 phenotypes, and 18,000 genotypes and were modeled with an overall mean, additive genetic, and residual effects using single-step genomic BLUP (ssGBLUP) with unknown parent groups. The uncertain parentage model partitioned contributions in A^{-1} to the parent on record (90 to 100%) and to the appropriate unknown parent group (0 to 10%) depending on the type of animal. We validated predictions based on the youngest animals (n = 14,950) without phenotypes. Accuracy was the correlation between true and estimated breeding values. Accuracies (SE) were nearly identical with 0.65 (0.01) for ssGBLUP and 0.64 (0.01) for uncertain parentage. Dispersion was the regression of true on estimated breeding values, and no differences existed between the models with dispersion (SE) of 0.83 (0.01) for ssGBLUP and 0.84 (0.01) for uncertain parentage. Bias was the difference between true and estimated breeding value and was scaled by the genetic standard deviation. Both models had bias (SE) of 0.24 (0.01). Similarly, bias differences were small when evaluating subsets