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

of animals based on pedigree accuracy. Dairy data sets are complex, especially with regard to differences in daughter parentage accuracies across bulls. This complexity is difficult to simulate, and uncertain parentage models should be tested because of the potential to mitigate bias differences across bulls.

**Key Words:** genomic, simulation, single-step genomic BLUP

**171 Genomic predictability of single-step GBLUP for production traits in US Holstein.** Y. Masuda*[1], I. Misztal[1], P. VanRaden[2], and T. Lawlor[3], [1]*University of Georgia, Athens, GA,* [2]*USDA AGIL, Beltsville, MD,* [3]*Holstein Association USA Inc., Brattleboro, VT.*

The objective of this study was to validate genomic predictability of single-step genomic BLUP for 305-d protein yield for US Holsteins. The genomic relationship matrix was created with the Algorithm of Proven and Young (APY) with 18,359 core animals. The full data set consisted of phenotypes collected from 1989 through 2015 and pedigrees limited to 3 generations back from phenotyped or genotyped animals. The predictor data set was created by cutting off the phenotypes, pedigree animals, and genotypes in the last 4 years from the full data set. Genomic predictions (GPTA2011) were calculated for predicted bulls that had no recorded-daughters in 2011 but had at least 50 such daughters in 2015. We calculated the daughter yield deviations with the full data (DYD2015) for the predicted bulls (n = 3,797). We also used the official GPTA published in 2011 with a multi-step method as a comparison, although the official methods have changed since then. Coefficient of determination ($R^2$) and slope ($b_1$) were calculated from a linear regression of DYD2015 on GPTA2011. We investigated the effect of different unknown parent groups (UPGs) and a weight (ω) on the inverse of the pedigree relationship matrix for genotyped animals ($\mathbf{A}_{22}^{-1}$) to compensate incomplete pedigree. When applying QP-transformation to $\mathbf{A}^{-1}$, the $R^2$ was 0.52 with ω = 1 compared with 0.51 from the official GPTA. The $b_1$ was similar (0.78) to 0.81 from the official GPTA. Using ω = 0.90, the $R^2$ was still similar (0.50) but the $b_1$ was greatly improved (0.96). With QP-transformation in $\mathbf{H}^{-1}$, the $R^2$ was less than 0.4 and the $b_1$ was smaller regardless of ω. Without any UPGs, the predictability and the inflation showed the same level as the official GPTA. The GPTA of a young animal is equivalent to the direct genomic value when many genotypes are included in the evaluation. Fixed UPGs in $\mathbf{H}^{-1}$ added an extra value to GPTA of young animal but this addition is likely redundant in genomic prediction. We should exclude the UPG contributions from GPTA of young genotyped animals when $\mathbf{H}^{-1}$ is QP-transformed.

**Key Words:** genomic evaluation, incomplete pedigree, Holstein

**172 Implementing SNP-level multiple-trait across country genomic evaluation without genotype sharing.** B. Fragomeni*, D. Lourenco, Y. Masuda, and I. Misztal, *The University of Georgia, Athens, GA.*

There is a growing interest of Interbull in releasing a multiple across country genomic evaluation. However, most countries are not able to provide genotypes, and an alternative methodology is required. One strategy called SNP MACE posits a multiple-trait SNP BLUP based on left- and right-hand sides of national SNP BLUP. However, different countries use different sets of SNPs and multiple-trait computations with SNP may be difficult. We propose an alternative model based on reconstructing phenotypes for an independent genotyped population. Each country would submit only SNP effects, the number of reference animals, and average reliabilities of GEBV. This information can be used to create a pseudo-population with pseudo-observations. The combined data can be analyzed by multi-trait GBLUP. Conversion of GEBV would provide SNP effects in scale of every country. Simulations included 30k animals resembling the US Holstein population, with effective population size of 120. Chromosome number and size mimicked the cattle genome. The population was then divided in 3: 2 countries and 1 test population with 10k genotyped animals in each, and a different trait was assigned to each country. For the genotyped animals in the 2 countries, DYD were generated with an average reliability of 0.8. SNP effects were calculated with GBLUP in each one of the 2 countries. With SNP effects from the 2 countries, phenotypes were reconstructed for the test population. A bivariate GBLUP was then fitted, and GEBV/DGV were calculate for the test population for both countries. Accuracies were calculated for the validation population on the scale of 2 countries. When SNP effects of one country were used, the realized accuracy was 0.94 for the same population and 0.69 for the second country. When SNP effects of both countries were used, the accuracy for any country was 0.95. With the use of the APY algorithm, the procedure is computationally viable for any population size and any number of countries. An important issue is creation of pseudo-population that holds the same genomic information as the national population.

**Key Words:** SNP-MACE, genomic MACE, SNP effect, Interbull

**173 Lifetime Net Merit versus annualized net present value as measures of profitability of selection.** M. R. Schmitt*[1], P. M. VanRaden[2], and A. De Vries[1], [1]*Department of Animal Sciences, University of Florida, Gainesville, FL,* [2]*USDA-AGIL, Beltsville, MD.*

Current USDA linear selection indexes such as Lifetime Net Merit (NM) estimate lifetime profit given a combination of 13 traits. In these indexes, every animal gets credit for 2.78 lactations of the traits expressed per lactation, independent of its productive life (PL). Selection among animals with different PL is an example of investment in mutually exclusive projects that have unequal duration. Such projects are best compared with the annualized net present value (ANPV) technique. The objective of this study was to compare the ranking and value differences between NM and ANPV for the top 1,539 Holstein sires for NM available in the December 2017 genetic evaluation from the Council on Dairy Cattle Breeding. To calculate the ANPV, economic weights from USDA estimates were multiplied by the PTA of single event traits. Heifer conception rate was recognized at first calving and livability at the end of life. The economic weight of PL was converted from a marginal value of $21 per lactating month depreciated over the standard length of 2.78 lactations, to a replacement cost (−$1500) at the beginning and a salvage value ($800) at the end of life. All other traits were considered lactation dependent, and the economic weights were multiplied by the number of expected lactations (2.78 + PTA PL/10). The values for all 13 traits were discounted and converted to ANPV to compare animals with different investment horizons on the same common horizon. Correlation and rank correlation between NM and ANPV was 0.993 for the group of 1,539 bulls. However, 32% of bulls with the same ANPV had NM deviations greater than $9.90 from the expected NM. Within the highest 300 NM bulls, correlation and rank correlation between NM and ANPV was 0.964 and 0.943, respectively, and the largest changes in ANPV rank from NM rank were −96 and +117. Bulls with a combination of low lactation traits and high PL resulted in the greatest decrease of ANPV rank compared with NM rank. In conclusion, the re-ranking of bulls based on 2 different measures of profitability suggests that further discussion is warranted about construction of selection indexes for genetic selection.

**Key Words:** investment, profit, genetics