Breeding and Genetics

293 Methods to implement ancestor discovery in the US dairy cattle database. J. Nani*1, 2, J. Cole2, and P. VanRaden2. 1Instituto Nacional de Tecnología Agropecuaria EEA Rafaela, Rafaela, SantaFe, Argentina, 2Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD.

Accurate and complete pedigrees are, even in the post genomics era, fundamental to plant and animal breeding because accurate genetic and genomic evaluations often rely on making genomic and pedigree relationships consistent. Program Fixped uses haplotypes to accurately confirm or discover distant relatives, such as maternal grandsires (MGS) and maternal great-grandsires (MGGS), with improved efficiency compared with independent SNP methods. In the US dairy cattle database, around 300,000 animals with no dam ID can be linked to their discovered MGS and MGGS by creating a constructed dam or maternal granddam (MGD) ID to fill in the missing pedigree information. Program Finddam creates the constructed dam and/or MGD ID to link calves to MGS and MGGS in the pedigree. This ID consists of 3 parts: 1) as currently, a 3-letter country code so that each country can construct their own IDs (i.e., USA), 2) DAM or MGD following the country code for the calf dam and maternal granddam respectively, and 3) the numeric portion of the calf ID (animal key) to ensure stability of data processing that can be expanded to 9 digits in the next few years. Program Finddam also allows linking calves to ~60,000 discovered MGS and MGGS not previously added because their dam and MGD ID were already reported. Recently expanded features of Fixped are discovery and confirmation of close relatives such as sires, dams, full and half sibs, clones, and paternal grandsires. Implementation of Fixped will increase the speed of genotype loading and avoid processing delays near deadlines because the current uploading program can then only confirm if the reported ancestors are correct and avoid searching the whole database for genotypes of relatives. Finally, when a real dam ID is found outside the database, those IDs will be preferred over constructed IDs unless the reported dam or MGD do not match the genotypes of the calf and grandsire. Pedigree providers will have an option to remove discovered relationships that they believe to be incorrect.

Key Words: ancestry discovery, pedigree, genomics

294 Bias of dairy sheep evaluations using BLUP and single-step genomic BLUP with metafounders and unknown parent groups. F. L. Macedo1, 2, O. F. Christensen1, J. M. Astru2, I. Aguilar3, Y. Masuda4, and A. Legarra5. 1INRA, Toulouse, France, 2UdelaR, Montevideo, Uruguay, 3Aarhus University, Aarhus, Denmark, 4DELE, Toulouse, France, 5INIA, Montevideo, Uruguay, 6University of Georgia, Athens, GA.

Bias is a problem in pedigree-based and genomic-based predictions, and it hampers correct selection procedures. Assessing bias for small dairy cattle breeds, sheep, and goat is difficult. Also, there is a plethora of options to integrate Unknown Parent Groups in Single Step GBLUP. In this work we quantify possible biases in predictions for a dairy sheep breed (Manchego), which is expected to enable a continued high level of genotype validation and relative discovery for many years as the genotype collection expands.

Key Words: dairy cattle, parentage discovery, genotype validation

295 Parent and grandsire discovery in a rapidly expanding collection of genotypes. G. Wiggans*, Council on Dairy Cattle Breeding, Bowie, MD.

For genomic selection based on SNP, the Council on Dairy Cattle Breeding (Bowie, MD) has collected over 3.9 million genotypes. In 2019, over 67,000 genotypes were added monthly on average. To assure that a genotype is assigned to the correct animal and that the pedigree is correct, parents are verified, and each genotype is compared with other genotypes to detect unreported parents or progeny or a duplicate genotype. To speed this discovery, a set of 4,668 SNP was defined based on their presence on nearly all genotyping chips, parent-pedigree consistency, and minor allele frequency. Assessment is done after 96 and 1,000 SNP so that comparison can stop if a relationship is unlikely. If both parents are confirmed, only genotypes from potential relatives born < 500 d earlier are checked to detect duplicates. Each SNP genotype is represented by 2 bits rather than 1 byte to save storage space. Because discovered relationships are recorded based on genotype-specific identification, they are unaffected by the assignment of the genotype to a different animal. This process improves efficiency by comparing 2 genotypes only once, using sequential memory access when doing comparisons, and limiting discovery to just once a day to reduce setup time. The design allows use of the presence of progeny and date loaded to exclude comparisons with genotypes unlikely to be related. When a parent is not confirmed, the grandsire may be designated as unlikely using the same SNP set. For unlikely and unknown grandsires, discovery is done weekly based on haplotype matching, which relies on imputation done for weekly evaluations. This haplotype analysis also discovers other relationships, which can provide a check on SNP-based discovery. These changes in discovery method were developed to address the ever-increasing computing time needed as the number of genotypes in the US genetic evaluation system rapidly grows. The new discovery design is expected to enable a continued high level of genotype validation and relative discovery for many years as the genotype collection expands.

Key Words: dairy cattle, parentage discovery, genotype validation

296 Profiles of causative SNP in a genome-wide association study. I. Misztal*, I. Pocrnic2, M. Perez-Enciso3, and D. A. L. Lourenco1. 1University of Georgia, Athens, GA, 2The Roslin Institute, Midlothian, United Kingdom, 3CRAG, Barcelona, Spain.

The purpose of this study was to see the impact of causative SNP on GWAS with different populations with different effective population size. Three populations were simulated assuming 100 equidistant causative SNP with identical substitutions effects. Causative SNP were included in 50 k SNP genotypes. Ten generations were simulated, with the last 3 genotype. Population NE60 was composed of 2000 animals per generation with effective population size 60. Population NE600 was composed of the same number of animals but with effective population size 600. NE60_3x was as NE60 but with 6000 animals per generation. Analyses from 2007 to 2017. Then we compared (G)EBVs from “partial” and (G) EBVs of young rams from “whole” data across several pairs of cutoff dates, resulting in 65 comparisons. All models resulted in some overestimation of the genetic trend of 0.20 – 0.40 genetic standard deviations. As for the slope (over/underdispersion of (G)EBVs) BLUP_MF, BLUP_UPG, SSGBLUP_MF and SSGBLUP_UPG were unbiased (slopes near 1 with s.e. ~0.02 across comparisons) whereas SSGBLUP_EUPG was biased (slope 0.87 with s.e. 0.02). This is probably due to double counting. One particular truncation year (2008) showed bias for all methods (~0.70 for SSGBLUP_MF and ~0.90 for the other methods) and the likely reason was suboptimal collect of young males that particular year.

Key Words: genomic, bias, sheep
were performed using single step GBLUP, with solutions converted to SNP values and subsequently to p-values for each SNP, in a GBLUP context, p-values are equivalent to those in standard GWAS methodology, where each SNP is treated as fixed effect, and a genomic relationship matrix accounts for the population structure. Manhattan plots for standardized SNP solutions showed large values for few of the 100 causative SNP and were very noisy. Manhattan plots for p-values were similar to those for SNP solutions. The number of SNP effects with p-values over the statistical threshold was smallest for NE60, larger for NE60_3X, and the largest for NE600. SNP profiles were created by averaging SNP solutions ± 100 SNP around causative SNP. The profiles showed distinct peak for the causative SNP, with smaller signals for adjacent SNP. The peak was smallest for NE60 and largest for NE600. Each causative SNP influenced about 50 adjacent SNP for NE60 and NE60_3X, and about 10 SNP for NR600. The profiles help understand that the effective use of causative SNPs requires knowing their exact positions and either boosting their variance in analyses or elimination of SNPs adjacent to causative SNPs.

**Key Words:** genomic selection, causative SNP, sequence data

**297 Predicted producing value: Formula to account for actual inbreeding in a mating program framework.** S. Westberry*, C. Heuer, N. Deeb, and D. Kendall, STgenetics, Navasota, TX.

Increasing herd profitability is essential to dairy producers desiring to withstand the ever-changing dairy industry climate. When optimizing future herd performance based on parent average, a crucial piece of the mating value is missing. The actual relationship of each mating pair needs to be determined and how the resulting inbreeding will affect the producing value of the progeny. In the US, PTA evaluations produced by the CDCB are penalized for expected inbreeding depression based on the Expected Future Inbreeding (EFI) of the animals. The Predicted Producing Value (PPV) formula first removes the inbreeding penalty from sire and dam PTA values. The next step of the formula is to determine the expected progeny inbreeding or actual relationship of each mating pair which utilizes the genomic relationship between all pairs in the proposed matings. The expected inbreeding of the calf can then be multiplied by the inbreeding depression of the selected trait for optimization by utilizing the inbreeding depression factors published by the USDA. In a recent study, STgenetics compared utilizing a parent average PTA optimization mating program and a PPV optimization mating program in Chromosomal Mating. The mating scenario included optimizing Lifetime Net Merit (NMS) across 26,500 females and 100 bulls. The PPV\textsubscript{NMS} of the projected progeny was calculated for both optimization strategies so that the 2 results could be compared. The mating that optimized PTA\textsubscript{NMS} had an average progeny PPV\textsubscript{NMS} of 1604, while the mating that optimized PPV\textsubscript{NMS} had an average progeny PPV\textsubscript{NMS} of 1631. This means that optimizing for PPV\textsubscript{NMS} increased the value of the progeny by $27 over the lifetime of the progeny on average. The mating that optimized PTA\textsubscript{NMS} also yielded projected progeny with 1.3% lower inbreeding on average compared with the projected progeny from optimizing PTA\textsubscript{NMS}. Utilizing PPV to optimize progeny value in a mating program allows producers to make the most accurate mating decisions based on profitability.

**Key Words:** inbreeding, mating program, genomic selection

**298 Inbreeding depression due to different age classes of inbreeding on production and fertility traits in Canadian Holsteins.** B. O. Makanjuola*1, C. Maltecca2, F. Miglior3,1, F. S. Schenkel1, and C. F. Baes1,4, 1Centre for Genomic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, 2Department of Animal Science and Genetics Program, North Carolina State University, Raleigh, 3Ontario Genomics, Toronto, ON, Canada, 4Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland.

The reduction in the mean phenotypic performances of livestock animals could be attributed to rising inbreeding levels. However, this observed decline may not be caused by the total number of inbreeding. Therefore, partitioning inbreeding into different age classes could help in assigning detrimental effects to different classes. The aim of this study seeks to investigate the effect of recent and ancient inbreeding on production and fertility traits in Canadian Holstein cattle. Phenotypic records of 46,430 cows with birth year ranging from 2007 to 2017 were available for production and fertility traits. These animals had 50K genotype data and pedigree records, which comprised of 259,871 individuals. Inbreeding coefficients were estimated using traditional pedigree measures (FPED) and genomic pedigree measures using segment-based (FROH) and marker-by-marker (FGRM) based approaches. Additionally, both pedigree and genomic inbreeding were partitioned into different classes by tracing the pedigree back to a specific generation and using the specific length of homozygous segments to represent different classes, respectively. Inbreeding depression was found for all production and most fertility traits, for example, every 1% increase in FPED, FROH and FGRM was observed to cause a −44.71, −40.48 and −48.72 kg reduction, respectively, in 305-d milk yield. Similarly, an extension in the first service to conception (FSTC) of 0.29, 0.24 and 0.31 d in heifers was found for every 1% increase in FPED, FROH and FGRM, respectively. Partitioning both pedigree and genomic inbreeding into age classes resulted in recent age classes showing more unfavorable inbreeding effects, while more distant age classes caused a more favorable effect. For example, we observed a −1.56 kg loss in 305-d protein yield for every 1% increase in the most recent pedi-

**Key Words:** inbreeding depression, recent and ancient inbreeding, pedigree and genomic inbreeding