293 Methods to implement ancestor discovery in the US dairy cattle database. J. Nani*[1,2], J. Cole[2], and P. VanRaden[2], [1]*Instituto Nacional de Tecnologia Agropecuaria EEA Rafaela*, *Rafaela, SantaFe, Argentina*, [2]*Animal 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 individual 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. Macedo[1,2], O. F. Christensen[3], J. M. Astruc[4], I. Aguilar[5], Y. Masuda[6], and A. Legarra*[1], [1]*INRA*, *Toulouse, France*, [2]*UdelaR*, *Montevideo, Uruguay*, [3]*Aarhus University*, *Aarhus, Denmark*, [4]*IDELE*, *Toulouse, France*, [5]*INIA*, *Montevideo, Uruguay*, [6]*University 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 (Manech Tete Rousse). This breed has a selection scheme for milk yield based on performance recording, progeny testing and Artificial Insemination (AI). The data comprises ~35 years, 1,842,295 performance records and 540,999 individuals in pedigree. In pedigree, there are 70% animals with sire and dam known, 15% with missing sire and 15% missing both. We defined 13 Unknown Parent Groups (or Metafounders). 3007 AI males were genotyped with 50k SNP chip. We tested models with and without genomic information (BLUP and SSGBLUP) and using 3 strategies to handle missing pedigree (Unknown Parent Groups (UPG), "Exact" UPG (EUPG), and Metafounders (MF). The Gamma relationship matrix across MF was estimated by GLS from genotypes of rams and showed mild correlation across MF. To quantify bias, we used method LR. We generated "partial" data deleting most recent records at every year from 2005 to 2014. Then we created "whole" data deleting records with cut-off years 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

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-progeny 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*[1], I. Pocrnic[1,2], M. Perez-Enciso[3], and D. A. L. Lourenco[1], [1]*University of Georgia*, *Athens, GA*, [2]*The Roslin Institute*, *Midlothian, United Kingdom*, [3]*CRAG*, *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 genotyped. 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