Breeding and Genetics III: Methods

458 Phenotypic analysis of daily milk, fat, and protein production with geometric morphometrics. A. A. D. Benitez*, J. I. Keller¹, and E. Ezra². ¹Institute of Animal Sciences, Agricultural Research Organization, The Volcani Center, Rishon LeZion, Israel, ²Israel Cattle Breeders Association, Caesarea Industrial Park, Israel.

Concerns have been raised in the past about the Dairy Herd Improvement Association recording frequency, because the interval between samples, about 4 weeks, may not capture the peak production for cows with shorter lactation (less than 10 mo), which have led to the conclusion that these cows have an atypical lactation curve shape. This may be due to sampling frequency rather than biological differences of cows or the influence of biotics and abiotics variables. Geometric morphometrics is a methodology that is used to measure biological shapes and curves, which has already proved its application in biology, medicine, and engineering. This methodology can be used for quantifying, testing, and visualizing shape variation and its co-variation with biotic and abiotic variables. We propose to apply this method to determine the variation and co-variation of lactation curves for milk, fat, and protein production. Daily records of milk production and fat and protein concentration collected by the AFlLab recording system (Afimilk, Kibbutz Afikim, Israel) from January 2014 to January 2017 from 47 large kibbutz (communal) herds distributed throughout Israel, will be analyzed. Lactation data will be plotted into an orbital graph to depict a closed curve. Currently methods to predict future lactation prediction of individual cows are based chiefly on production on the last available test day. Using geometric morphometrics it should be possible to derive more accurate estimates of future production, which can be used both to improve management decisions and genetic evaluations.

Key Words: geometric morphometrics, milk prediction, daily fat recording

459 Genetic parameters of bovine milk color and processing characteristics predicted by mid-infrared spectroscopy. G. Visentin*¹², D. P. Berry², M. De Marchi¹, S. McParland³, A. McDermott¹², S. Scarso¹, M. A. Fenelon², and M. Penasa¹. ¹Department of Agronomy, Food, Natural Resources, Animals, and Environmental (DAFNAE), University of Padova, Legnaro (PD), Italy, ²Animal and Grassland Research and Innovation Center; Teagasc, Moorepark, Fermoy, Co. Cork, Ireland, ³Teagasc Food Research Centre; Teagasc, Moorepark, Fermoy, Co. Cork, Ireland.

Milk color and processing traits are important factors informing the potential and ease to manufacture milk into different dairy products. The objective of the present study was to estimate (co)variance components of milk lightness (L*), redness-greenness (a*), yellowness-blue (b*), rennet coagulation time (RCT), curd-firming time (k20), curd firmness (a30 and a60), heat coagulation time (HCT), casein micelle size (CMS), and pH, measured by mid-infrared spectroscopy on 136,807 test-day records from 9,824 Irish dairy cows between 5 and 305 d-in-milk (DIM) from parities ≤10. Cow breed was defined as the proportion of Holstein, Friesian, Jersey, Norwegian Red, Montbéliarde, and “Other.” Random regression models using Legendre polynomials were performed to describe the change of both additive genetic and within-lactation permanent environmental variances across different DIM. Heritability estimates averaged across all DIM for milk color were 0.31 (L*), 0.11 (a*), and 0.42 (b*); average heritability estimates for processing traits ranged from 0.31 (pH) to 0.49 (k20), except for HCT (0.17). Within-trait genetic correlations approached unity between adjacent DIM, and were <0.40 at the peripheries of lactation. Eigenvalues and associated eigenfunctions of the additive genetic variance of all traits revealed that at least 80% of the total variation was associated with the height of the lactation profile. Average genetic correlations between color traits across all DIM were the weakest between a* and b* (−0.19); fat concentration was strongly genetically associated with b* (0.91), while milk yield was negatively genetically correlated with all color traits. On average, RCT was strongly genetically correlated with both a30 (−0.68) and pH (0.75); genetic correlations between HCT and the other processing traits were, on average, weak ranging from −0.02 (with pH) to 0.28 (with RCT). Milk yield was genetically correlated with both RCT (0.31) and a30 (−0.49). Breeding for milk color and processing traits is possible but with some negative impact on genetic gain for milk yield.

Key Words: milk quality, breeding, infrared spectrometry

460 Genetic parameters of milk fatty acid profile in dairy sheep. J. Serdino, F. Corredu, M. G. Manca, A. Puledda, C. Dimauro, A. Nudda, and N. P. P. Macciotta*, University of Sassari, Sassari, Italy.

Increasing consumer concerns on dairy product nutritional quality have stressed the importance of some features of milk such as the fatty acid (FA) profile. In this study, genetic parameters of milk FA profile of dairy sheep are investigated to evaluate the feasibility of breeding for improving the nutritional quality of sheep milk. Individual milk samples of 989 Sarda ewes farmed in 47 flocks located in the 4 provinces of the Island of Sardinia, Italy, were analyzed for FA composition by gas-chromatography. Genetic parameters of 15 FA (expressed as g/100 g of FA methyl ester) were estimated with an animal model, including fixed effects of lambing type, lambing month, altitude of flock, lactation stage, province and parity, and the random effects of flock-test date (FTD), and animal additive genetic. A generally high contribution of FTD to the phenotypic variance was observed (on average 51%). Heteritability (h²) estimates ranged from 0.03 for C18:3 n-3 to 0.48 for C16:0. Saturated and unsaturated C18 FA showed moderate to low values of h² (from 0.22 to 0.03, for the same FA). On the other hand they exhibited a large contribution of FTD, ranging from 0.46 for C18:0 to 0.82 for C18:3 n-3, respectively. The high heteritability estimate of C16:0 reflects the probable genetic control of its milk content, being partly synthesized de novo in mammary gland. Genetic correlations were negative among C4:0 and short and medium chain SFA. C16:0 showed a negative correlation with most of the investigated FA, and a positive correlation with C4:0, C14:0, C14:0c9 and C16:1c9. The high heritability of C16:0, considered harmful for human health, and its negative genetic correlations with unsaturated C18 FA could be used in genetic strategies to improve the nutritional properties of milk.

Key Words: fatty acids, genetic parameters, sheep milk

461 Genomic predictions for crossbreds from all-breed data. M. E. Tooker*, P. M. VanRaden, and G. C. Fok, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD.

Genomic predictions of transmitting ability (GPTAs) for crossbred animals were computed from marker effects of 5 dairy breeds weighted by each breed’s genomic contribution to the crosses. Estimates of genomic breed composition are labeled breed base representation (BBR)
and are reported since May 2016 for all 1.6 million genotyped dairy animals. Animals with >94% of any breed were rounded to 100%, and contributions of other breeds were set to 0%. All-breed scale GPTAs were first computed for each pure breed for traits milk, fat, protein, productive life, somatic cell score, daughter pregnancy rate, cow conception rate, livability, and net merit. These estimates included foreign information from multi-trait across-country evaluation (MACE) and foreign dams converted from within-breed to the all-breed base. Then, marker effects for each breed were blended by BBR to compute evaluations for crossbreds (<94% purebred) for those same traits. Conformation traits do not have an all-breed scale, so only the Jersey marker effects were applied to the crossbreds, and results seemed reasonable. Calving traits are not predicted for crossbreds, and instead a common mean was used for all crossbreds as is the current practice for breeds other than Holstein and Brown Swiss. All-breed GPTAs were then converted to within-breed GPTAs. Correlations of GPTAs for purebreds computed on the all-breed vs. current within-breed scales were 0.97 to 0.99 for most traits and breeds. Crossbred GPTAs were then computed for 44,023 crossbreds, 20,367 of which had no previous GPTAs because of breed check edits. The new GPTAs were for 1,822 Jersey × Holstein crossbreds with <40% of both breeds (F1 crosses), 75 Brown Swiss × Holstein F1, 7,237 Holstein backcrosses with >67% and <94% Holstein, 7,820 Jersey backcrosses, 313 Brown Swiss backcrosses, 1,763 other crossbreds of various mixtures, and 1,337 purebreds that had previously failed breed checks. Additional automation and redesign of many downstream programs is required for the new all-breed system to be used in weekly, monthly, and full releases. The new system is expected to provide accurate predictions for crosses among the 5 dairy breeds evaluated.

Key Words: crossbreeding, genomic prediction, breed composition

462 Genetic trends from single-step GBLUP and traditional BLUP for production traits in US Holstein. Y. Masuda*1, I. Misztal1, P. M. VanRaden2, and T. J. Lawlor3, 1University of Georgia, Athens, GA, 2USDA, AGIL, Beltsville MD, 3Holstein Association USA Inc., Brattleboro, VT.

The objective of this study was to compare genetic trends from a single-step genomic BLUP (ssGBLUP) and the traditional BLUP (tradBLUP) models for milk production traits in US Holstein. We used 764,029 genotyped animals in this study. Phenotypes were 305-d milk, fat, and protein yield from 21,527,040 cows recorded between January, 1990 and August, 2015. The pedigree file included 29,651,623 animals limited to 3 genotyped animals in this study. Phenotypes were 305-d milk, fat, and protein yield. Monthly and full releases. The new system is expected to provide accurate predictions for crosses among the 5 dairy breeds evaluated.

Key Words: crossbreeding, genomic prediction, breed composition

to or slightly greater than the PTA trend up to 2006. Two trends started to diverge obviously in 2007 and the GPTA trend kept rising while the PTA trend remained at the same level. The single-step method provides very similar genetic trends to the traditional evaluations except for the last few years. The recent lower PTA trend can be due to a downward bias caused with genomic pre-selection of young animals.

Key Words: genomic evaluation, genetic trend, PTA

463 A Genetic Diversity Index method to improve imputation accuracies of rare variants. A. M. Butty1, F. Miglior2, P. Stothard3, F. S. Schenkel1, B. Gredler4, M. Sargolzaei1, and C. F. Baes1, 1Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, 2Canadian Dairy Network, Guelph, ON, Canada, 3Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, 4Qualitas AG, Zug, ZG, Switzerland, 5Semex Alliance, Guelph, ON, Canada.

Different methods to select animals for sequencing have been developed, which rely on pedigree-based relationship matrices, genomic relationships, or on haplotype frequencies. Relationship-based methods select representative key animals of a population whereas haplotype frequency methods aim for better coverage of rare variants. Good average accuracies of imputation from SNP chip to whole-genome sequence (WGS) for common haplotypes were reached with the relationship-based methods. Imputation of rare variants, however, still needs to be improved, which can possibly be accomplished with a newly developed Genetic Diversity Index (GDI). This algorithm optimizes the count of unique haplotypes present in a group of animals composed of already sequenced individuals and a fixed number of sequencing candidates. Optimization is run iteratively, exchanging one candidate at a time and computing the GDI of the new group. Use of the simulated annealing algorithm defines whether the last individual added to the group should be kept. Simulated annealing has the advantage of searching for a global optimum in a situation where multiple local optima are present. The previously mentioned key ancestor and haplotype-based methods for selecting sequencing candidate were assessed and compared with the GDI algorithm using simulated cattle WGS data. Average squared correlation coefficients were used to assess imputation accuracy. A preliminary study showed that the accuracy was 1.5% higher when using GDI to enlarge the reference population than the second-best method. Application of the different methods of selection in North American Holstein data showed that the GDI algorithm selected animals carrying a higher percentage of rare haplotypes than other methods examined. Principal component analysis of the population showed that the animals selected with all tested methods were similarly distributed over the pool of candidates. When representative animals of a population are already sequenced and good overall imputation accuracies are reached, sequencing of genetically diverse animals improved the accuracy of the imputation of rare variants to the WGS density level.

Key Words: sequencing, simulation, imputation

464 Determination of quantitative trait variants by concordance via application of the a posteriori granddaughter design to the US Holstein population. J. I. Weller*1,2, D. M. Bickhart2, G. R. Wiggins2, M. E. Tooker2, J. R. O’Connell4, J. Jiang5, and P. M. VanRaden2, 1Agricultural Research Organization, The Volcani Center, Rishon LeZion, Israel, 2Agricultural Research Service, Beltsville, MD, 3Council on Dairy Cattle Breeding, Bowie, MD, 4University of Maryland, MD.

J. Dairy Sci. Vol. 100, Suppl. 2
Experimental designs that exploit family information can provide substantial predictive power in quantitative trait variant discovery projects. The a posteriori granddaughter design was applied to the US Holstein dairy cattle population. Twenty-nine trait-by-chromosomal segment effects were found with probabilities $< 10^{-20}$ that a segregating quantitative variant was detected by chance. Polymorphism genotypes for 79 grandsires and 16,236 sires were determined by imputation for 3,148,506 polymorphisms across the entire genome; 444 Holstein bulls had complete genome sequence, including 38 of the grandsires. Concordance between quantitative trait locus genotype and polymorphism was determined for all 29 effects. Complete concordance was obtained only for daughter pregnancy rate on chromosome 18 and protein percentage on chromosome 20. For each quantitative trait locus, effects of the 20 polymorphisms with the highest concordance scores for the analyzed trait were computed by stepwise regression. The effects for stature on chromosome 7, daughter pregnancy rate on chromosome 18, and protein percentage on chromosome 20 met the following 3 criteria: complete or nearly complete concordance, significance of the polymorphism effect after correction for all other polymorphisms, and a marker coefficient of determination that was $> 50\%$ of the total multiple-regression coefficients of determination for the 20 polymorphisms with highest concordance. An intronic variant SNP on chromosome 5 at position 93,945,738 explained 7% of the variance for fat percentage and 85% of the total variance explained by the multiple-marker regression. Variants identified in this study are likely to provide improved predictive power for genomic evaluation of dairy cattle.

**Key Words**: genomic selection, granddaughter design, quantitative trait variant

465 Impact of SNP selection on genomic prediction for different reference population sizes. D. A. L. Lourenço*1, I. R. Menezes2,1, B. O. Fragomeni3, H. L. Bradford1, S. Tsuruta1, and I. Misztal1, 1University of Georgia, Athens, GA, 2University of Sao Paulo, Pirassununga, SP, Brazil.

Methods for SNP selection can improve prediction accuracy over genomic BLUP, but in practice, the improvement is trait and population specific. This study investigates the importance of SNP selection in populations with 2000 to 25,000 genotyped animals. Populations were simulated with effective population sizes (Ne) of 20 or 100, and assuming that 10, 50, or 500 QTL were affecting a trait with heritability of 0.3. Pedigree information was available for 6 generations; phenotypes were recorded for the 4 middle generations. Animals from the last 3 generations were genotyped for 45,000 SNP. Single-step genomic BLUP (ssGBLUP) and weighted ssGBLUP (WssGBLUP) were used to estimate genomic EBV (GEBV). For WssGBLUP, 2 iterations of weights were calculated and were used to derive SNP variances and to construct a weighted genomic relationship matrix (G). Improved prediction accuracies are expected in WssGBLUP because more weight is placed on important SNP. Prediction accuracies were calculated for 1000 genotyped animals in the last generation. Reference populations included 2000, 5000 and 25,000 genotyped animals. The latter genotyped set was used to assess the dimensionality of genomic information (number of effective SNP or effective chromosome segments - Ne). This was calculated as the number of the largest eigenvalues explaining 98% of the variation in the genomic relationship matrix with and without the weights. For the data sets with Ne = 20 and 10 QTL, the accuracy gain from WssGBLUP was 12, 9, and 4 points for 2000, 5000, and 25,000 genotyped animals, respectively. With Ne = 100, this gain was 8, 10, and 7 points, respectively. For both Ne of 20 and 100, the gain assuming 50 QTL was halved, and no gain was observed assuming 500 QTL. The number of effective SNP was about 4-fold less in weighted G (~1512) than in unweighted G (~5790), explaining the greater gain in accuracy with fewer genotyped animals. The impact of SNP selection decreases with increasing size of the reference population and number of QTL.

In large populations, the detection of chromosome segments is more difficult, requiring more genotyped animals.

**Key Words**: accuracy, variable selection, weighted ssGBLUP


The objective was to determine the effect of using core animals from different generations in single-step genomic BLUP with the Algorithm for Proven and Young (APY). Effective population size and number of independent chromosome segments (ICS) are limited in livestock populations indicating limited dimensionality of genomic information. The APY takes advantage of this dimensionality and assumes that breeding values (BV) for noncore animals are functions of the BV for core animals. The core animals represent the same information as the ICS. Simulations comprised a moderately heritable trait for 95,010 animals and 50,000 genotypes for animals across 5 generations. Genotypes consisted of 25,500 SNP distributed across 15 chromosomes. Core animals were defined based on individual generations, equal representation across generations, and at random. For a sufficiently large core size, core definitions had the same accuracies ($r^2 = 0.90 \pm 0.01$) and biases ($\beta = 1.02 \pm 0.01$) for young animals, even if the core animals had imperfect genotypes because of imputation. Using the youngest generations as core caused an increase in the number of rounds to convergence indicating some numerical instability with these core definitions. When 80% of genotyped animals had unknown parents, accuracy and bias were significantly better ($P \leq 0.05$) for random and across-generation core definitions ($r^2 = 0.71 \pm 0.01; \beta = 0.75 \pm 0.01$) than for single generation core definitions ($r^2 = 0.61 \pm 0.01; \beta = 0.53 \pm 0.01$). This difference could result from improved relationship estimates between animals in different generations, because all generations were represented in the core partition that was directly inverted in APY. Thus, any subset of genotyped animals can be used to approximate the ICS when pedigrees are complete, but core animals should represent all generations when pedigrees are incomplete.

**Key Words**: APY, genomic selection, single-step genomic BLUP

467 Including causative variants into single-step genomic BLUP. B. D. Fragomeni*1, D. A. L. Lourenco1, Y. Masuda1, A. Legarra2, and I. Misztal1, 1University of Georgia, Athens, GA, 2INRA, Castanet-Tolosan, France.

The purpose of this study was determining, by simulation, whether (single-step) GBLUP is useful for genomic analyses when causative Quantitative Trait Nucleotides (QTNs) are known. Simulations included 180k animals in 11 generations. Simulated population mimicked a cattle population with weak selection intensity (Ne ~200). Phenotypes were available for animals in generations 6–10. Genotypes were available for 24k parents and 5k young animals in generation 11, and included 60k regular SNPs in 10 chromosomes, with genetic variance fully accounted for by 100 or 1,000 biallelic QTN, with effected sampled from a gamma distribution with shape parameter equal 0.4. LD ($r^2$) between SNPs
Impact of pedigree truncation on accuracy and convergence of sGBLUP in a population with long pedigree when only a fraction of animals are phenotyped. I. Pocrnic*1, D. A. L. Lourenco1, H. L. Bradford1, C. Y. Chen2, and I. Misztal1, 1Department of Animal and Dairy Science, University of Georgia, Athens, GA, 2Genus PIC, Hendersonville, TN.

In a genomic evaluation, it is desirable to have low computing cost while retaining high accuracy of evaluation for young animals. When the population is large but only few animals have phenotypes, especially for low heritability traits, the convergence rate of BLUP or single-step genomic BLUP (ssGBLUP) can be very slow. While eliminating old pedigrees can seriously affect (G)EBV for old animals, usually only younger animals are candidates for selection. This study investigates the effect of pedigree truncation on convergence rate and accuracy of prediction for young animals. The data consisted of 216k, 221k, 722k, and 579k phenotypes on 4 traits (T1, T2, T3, T4) from a purebred pig line. Heritabilities were <0.1 for T1 and T2, and >0.2 for T3 to T4. A total of 2.4 million animals born from 1971 to 2016 were included in the complete pedigree. Genotypes were available for 33,502 animals and consisted of 60,003 SNP. A bivariate animal model was fit for T1–2, and T3–4, separately. Computations were done by BLUP or ssGBLUP, and were conducted with complete pedigree or different levels of pedigree depth (Pn), where n = 1, 2, 3, 4, 5. Pedigree depth n was defined as n ancestral generations from the animals with phenotypes. The number of pedigree animals for T1–2 (T3–4) varied from 226k (760k) for P1 to 228k (767k) for P5. Genomic relationship matrix was inverted either by genomic relationship matrix, or 1% of the identity matrix. Rank of GRM was between the rank of unweighted GRM and that computed with causative SNP only. Single-step GBLUP can account for causative SNP when variances of causative QTN are known.

Key Words: genomic relationship matrix, genomic prediction, causative variant

470 SSGP: SNP-set based genomic prediction to incorporate biological information. J. Jiang*4, J. O’ConnellJ, P. VanRaden3, and L. Ma1, 1Department of Animal and Avian Sciences, University of Maryland, College Park, MD, 2University of Maryland School of Medicine, Baltimore, MD, 3Animal Genomics and Improvement Laboratory, ARS-USDA, Beltsville, MD.

Genomic prediction has emerged as an effective approach in plant and animal breeding and in precision medicine. Including biological information into the genomic model can be of great advantage. Due to the statistical and computational challenges in large genomics studies,
however, a fast and flexible method to incorporate such external information is still lacking. Here, we proposed a linear mixed model that can incorporate biological information in a flexible way and developed a fast variational Bayes-based software package named SSGP. In our model, whole genome markers can be split into groups in a user-defined manner, and each group of markers is given a common effect variance. Since previous functional genomics studies have accumulated much evidence on which genes, genomic regions or pathways are more/less important for a trait of interest, we can divide genome-wide SNPs into several groups based on their levels of importance and then use the predefined SNP sets in SSGP. Additionally, each marker has a pre-specified weight for which the rule can be flexibly assigned, e.g., based on minor allele frequency or LD pattern. The model was implemented with the parameter expanded variational Bayesian method. For testing purpose, we analyzed a large cattle data set consisting of ~24k bulls (20k in training set and 4k in validation set) and ~760k whole-genome SNP markers. By simply grouping markers based on proximity (markers were divided into continuous, non-overlapping chunks, each containing 1k SNPs) and considering only additive effects, SSGP already performed better than Bayes A in all 5 milk traits analyzed, with an increase of up to 8 percent points in prediction accuracy. Meantime, it took only ~5h for each trait with 20 threads. We also analyzed many simulation data sets and the WTCCC heterogeneous stock mice data set for which the results of many existing methods had been reported. Generally, SSGP could achieve similar prediction performance compared with the best approaches reported, though only proximity was used for grouping SNPs. Collectively, the method and software show great potential to increase accuracy in genomic prediction, particularly in the future when more useful biological information is becoming available.

**Key Words:** genomic prediction, SNP set, biological information