447 Iterative combination of national phenotype, genotype, pedigree, and foreign information. P. M. VanRaden,* *Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD.*

Single step methods can combine all sources of information into accurate rankings for animals with and without genotypes. Equations that require inverting the genomic relationship matrix G work well with limited numbers of animals, but equivalent models without inversion are needed as numbers increase. An equivalent model that includes extra equations to solve for the added contribution of genomic information was applied to national Jersey data. The extra equations solved for $G\gamma = u$ and $A_{22}\phi = -u$, where A_{22} contains pedigree relationships for genotyped animals and u contains genomic estimated breeding values (GEBV) from the previous iteration. Solutions for γ and ϕ were then added when solving for u. Multi-trait across country evaluations (MACE) were deregressed and inserted as extra records containing foreign information. The methods were tested on US Jersey yield data containing 4.4 million lactation records, 4.1 million animals in the pedigree, 16,852 genotyped animals, and 7,072 bulls with foreign MACE records. Heritability was reduced from 0.35 in official evaluations to 0.23 to mimic the effect of cow adjustments. For genotyped young bulls, single-step evaluations were correlated by 0.966 to multi-step evaluations. Both had the same reliability when tested using 4 year truncated data to predict deregressed proofs from the last 4 years, but regressions for single-step evaluations were closer to expected values. The weight on ϕ was reduced to 0.8 in the single step method and polygenic variance was increased to 20% in the multi-step method, both to improve the regressions. Convergence was much slower when the same algorithm was applied to Holstein data, and correlations were poor even after thousands of iterations. The number of Holstein genotypes was 135,724, with 65 million lactation records and 50 million animals in the pedigree. Second order Jacobi iteration was used in this study, but preconditioned conjugate gradient algorithm should be faster. More efficient strategies are needed because algorithms that work well on small or medium-sized data sets may not handle very large populations.

Key Words: single step methods, genomic evaluation, mixed model equations

448 Adaptation of BGF90 package for genomic computations. I. Misztal*[1], A. Aguilar[3], S. Tsuruta[1], and A. Legarra[3], *[1]University of Georgia, Athens, [2]INIA, Las Brujas, Canelones, Uruguay, [3]INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), Castanet-Tolosan, France.*

The BGF90 package is a tool for mixed model analyses. The original package contains programs for renumbering, BLUP, variance components estimation, accuracy approximation and visualization. A renumbering program (RENUMF90) prepares data files for application programs, prunes pedigrees and can support national data sets. BLUP programs are for equations in memory (BLUPF90) and iteration on data (BLUP90IOD). Parameter estimation is via REML (REMLF90 and AIREMLF90) or Bayesian methods (GIBBS*F90), which are able to support large number of traits (20+). Samples from GIBBS* programs can be analyzed by POSTGIBBSF90, and accuracies of predictions can be approximated by ACCF90. Specific programs are available for threshold-linear models. Nearly all programs were updated to support the genomic information and several new programs were added. Program PreGSF90 analyzes the SNP information, provides basic quality control, creates a genomic relationship matrix using a large variety of options, and combines pedigree and genomic relationship matrices for a single-step methodology. Computations with PreGSF90 are optimized for parallel processing; preparing matrices for 30k animals with 50k SNP takes about 1 h. PreGSF90 can be run separately or as part of application programs. PostGSF90 converts GEBV to SNP effects, displays Manhattan plots possibly using moving averages, and estimates variances of SNP effects. Program PredF90 predicts GEBV based only on estimates of SNP effects obtained from PostGSF90. Most of the programs are available online at nce.ads.uga.edu. The package can be used for genomic predictions (including national data sets), parameter estimation (including GBLUP and G-REML), and GWAS. Unequal variances for SNP effects similar to those in BayesA and subsequently "Manhattan" plots can be obtained by iterating on postGSF90 and possibly one of BLUP programs; no deregression is required and complex models may be used. Classical GWAS can be carried out with BLUPF90 fitting one SNP at a time as fixed regression and an animal effect with a genomic (or combined) a relationship matrix. The package has been used for genomic analyzes of models with up to 10 million animals, 18 traits, 40k genotypes and 400k SNP. The updated package simplifies genomic analyses in breeding applications.

Key Words: genomic selection, genome wide association, software

449 Methods to include foreign information in national evaluations. P. M. VanRaden and M. E. Tooker,* *Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD.*

Genomic evaluations (GEBV) with higher reliability often result from including genotypes and phenotypes from foreign bulls in the reference population. Multi-step methods evaluate domestic phenotypes first using only pedigree relationships (EBV), then add foreign data available from multi-trait across country evaluations (MACE), then compute GEBV for genotyped animals, and finally propagate information from GEBV to EBV of non-genotyped relatives. An alternative is to include domestic and foreign phenotypes together so that GEBV and EBV for all animals can be computed in a single step. The MACE EBV could be treated as a correlated trait, but previous research indicates that including these as the same trait with their lower reliability (REL) is sufficient. To include foreign data, the bull's deregressed proof (DRP) was obtained from the MACE EBV as: DRP = PA + (EBV - PA)/REL, where PA is the parent average from MACE. For bulls with both domestic and foreign daughters, domestic EBV was used instead of PA to compute DRP, and domestic daughter equivalents (DE) were subtracted from the total. Remaining DE were added to diagonals of the mixed model equations and were used to compute REL. This strategy included 1 extra record per bull and differed from previous methods that included 1 record for each foreign daughter. For multi-trait models, diagonal matrix D contained the DE for each trait of a bull. The vector of DRP was premultiplied by $D^{.5}T^{-1}D^{.5}$, where T is the genetic covariance matrix among traits, and $D^{.5}T^{-1}D^{.5}$ was added to the mixed model equations. A mean for the DRP was included in the model because the base is not fixed during iteration, only after convergence. The methods were tested using national Holstein data for 25 million cows, MACE data for 88,000 bulls, and a pedigree file of 52 million animals. For bulls with only foreign daughters, correlations between MACE EBV and national EBV after including the foreign data were 0.991 to 0.994 for yield traits, 0.986 for somatic cell score, 0.973 for single-trait productive life, and 0.974 for

daughter pregnancy rate. This simple approach is reasonably accurate for including foreign data in national evaluations.

Key Words: MACE, foreign daughters, genomic evaluation

450 Characteristics and use of the Illumina BovineLD BeadChip. G. R. Wiggans*[1], P. M. VanRaden[1], T. A. Cooper[1], C. P. Van Tassell[2], T. Sonstegard[2], and B. Simpson[3], *[1]Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD, [2]Bovine Functional Genomics Laboratory, ARS, USDA, Beltsville, MD, [3]GeneSeek, Lincoln, NE*

Genotypic information from the 6,909-SNP Illumina BovineLD (LD) Genotyping BeadChip, which replaced the Illumina GoldenGate Bovine3K (3K) Genotyping BeadChip, have been included in US genomic evaluations since November 2011. Of 73 LD single nucleotide polymorphisms (SNP) not used in genomic evaluation, 35 were from the Illumina BovineSNP50 (SNP50) Genotyping BeadChip, and 38 (13 mitochondrial, 9 Y-chromosome, and 16 X-chromosome SNP to improve genome coverage) were from the Illumina BovineHD (HD) Genotyping BeadChip. As of February 2012, the USDA national genotype database for dairy cattle included LD genotypes for 19,515 animals (550 males). Call rate for LD SNP used in genomic evaluation was 99.4%. The 9 Y SNP were highly effective in sex validation (call rate of 98% for males and 0.5% for females). Rate of parent-progeny conflicts on a SNP basis was similar to that for SNP50 SNP. Imputation accuracy averaged 98.9% for Holsteins, 98.3% for Jerseys, and 97.9% for Brown Swiss for LD genotypes compared with 95.9, 94.6, and 93.9% for 3K genotypes. To calculate reliability of genomic evaluations, the fraction of correctly imputed SNP was estimated as a function of the number of low-density SNP that were not missing and the number of animals with SNP50 genotypes. Reliabilities for LD genotypes were about 5 percentage points higher than for 3K genotypes. Using the add-on capability of the LD chip, the GeneSeek Genomic Profiler (GGP) for Dairy Cattle BeadChip was developed with 8,655 SNP. The additional SNP were for proprietary single-gene tests, detection of haplotypes that affect fertility, imputation of microsatellite alleles to facilitate parentage validation, and improved imputation by including more 3K SNP. The GeneSeek Genomic Profiler 80K (GGP-80) also was developed with around 80,000 SNP. The added SNP were SNP50 and HD SNP with the largest effects on primarily the net merit index. Consideration also was given to spacing as well as maintaining around 30,000 SNP50 SNP for imputation accuracy. The GGP and GGP-80 genotypes are expected to further improve accuracy of imputation and genomic evaluation because of the additional SNP.

Key Words: genomic evaluation, beadchip, SNP

451 Partitioning genetic (co)variances leading to alternative derivation of single-step type genomic prediction equations allowing joint estimation of GEBV and SNP effects. N. Gengler*[1], G. Nieuwhof[2], K. Konstantinov[2], and M. Goddard[3,4], *[1]ULg - Gembloux Agro-Bio Tech, Gembloux, Belgium, [2]ADHIS, Bundoora, Australia, [3]DPI, Bundoora, Australia, [4]University of Melbourne, Melbourne, Australia.*

Interest in single-step type procedures to do genomic prediction is growing because of its numerous advantages especially its robustness and its simplicity. Current derivations of single-step equations modify relationships among animals replacing for genotyped animals on an inverted scale, pedigree based by modified, partially genomic based, relationships. From an theoretical standpoint these methods are all based on assumptions and use a hidden underlying hypothesis that modified relationships are obtained as linear combination of strictly genomic and pedigree based relationships, therefore implicitly "weighting" SNP and polygenic effects. Alternative equations were recently proposed deabsorbing the genomic relationships out of the equations. This derivation did not change basic assumptions, but was derived using a matrix of relationship differences. This presentation will show a new and alternative derivation of single-step type genomic prediction equations allowing joint estimation of GEBV and SNP effects based on the partitioning of genetic (co)variances. The method was derived from a random mixed inheritance model where SNP and residual polygenic effects are jointly modeled. The derived equations were modified to allow non-genotyped animals and to estimate directly and jointly GEBV and SNP effects. Equations resemble recently proposed alternative single-step equations but were derived differently and are based on completely different assumptions and avoid certain issues in de-absorbing derivation linked to the matrix of relationship differences by using (co)variances. Several other advantages of the new equations are that weighting of SNP and polygenic effects becomes explicitly and that SNP effects are also estimated. This method makes better use of High-Density SNP panels and can be easily modified to accommodate other genetic effects as major gene effects or copy-number variant based effects. Finally these alternative equations combine advantages of single-step and of explicit SNP effect estimation based methods. Additional research is required to test and validate the proposed method.

Key Words: genomic prediction, single-step method, alternative equations

452 Use of canonical discriminant analysis for detecting selection signatures in cattle. R. Steri, C. Dimauro, S. Sorbolini, G. Marras, M. Cellesi, G. Gaspa, and N. P. P. Macciotta,* *Dipoartimento di AGRARIA, Università di Sassari, Sassari, Italia.*

The development of high throughput SNP platforms for several livestock species allows to study genetic variability both within and between breeds. Several techniques have been used to exploit information derived by these new tools, including principal component analysis (PCA). However, results are not easy to interpret in terms of markers, or genomic regions, linked to phenotypic traits. To explain the biological meaning of multivariate approach, canonical discriminant analysis (CDA) could be proposed. CDA is based on PCA applied on within-between (co)variance ratio matrix of predefined groups instead the (co)variance matrix. The new orthogonal variables maximize the differences among groups on the basis of variances within each groups. The biological meaning of canonical variables (CANi) can be inferred by the canonical coefficients (loadings) that represents the correlations between CAN and the original variables. In the present work, CDA was used to analyze a total of 2,627 bulls of Italian Holstein (IH; 1000), Italian Brown (IB; 755), Italian Simmental (IS; 493) and Piemontese (IP; 379). Thus CDA was used to study genetic differences between 2 dairy, one beef and one dual-purpose cattle breeds. Animals were genotyped with the 50k SNP panel. The analysis was carried out separately for the 29 autosomes. The separation among breeds was always clear. On average, the CAN1 explained about 50% of the total variability and was able to discriminate between IH and all the other breeds. This result, probably, underline the high selection pressure exerted on this population. CAN2 and CAN3, explaining on average 25% each, usually separated dairy breeds (IH and IB) from the IP, whereas the IS tended to be located in an intermediate position. Structure of the new variables shows genomic regions associated with extreme loadings value. Considering only values exceeding the 0.99 quantile as an empirical threshold, we found 743 SNPs across the whole genome that can be considered involved in differences among

breeds. Interesting clustered signals were found near ABCG2 for CAN1, MSTN, LEPR and MC1R for CAN2 and LEP and KIT for CAN3.

Key Words: SNP, canonical discriminant analysis, selection signatures

453 Genome-wide association mapping including phenotypes from relatives without genotypes. H. Wang*[1], I. Misztal[1], I. Aguilar[2], A. Legarra[3], and W. Muir[4], *[1]Department of Animal and Dairy Science, University of Georgia, Athens, [2]Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, Canelones, Uruguay, [3]INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), Castanet-Tolosan, France, [4]Department of Animal Science, Purdue University, West Lafayette, IN.*

The purpose of this study was to extend single-step GBLUP (ssGBLUP) to genome wide association analysis (GWAS). The ssGBLUP is a procedure that calculates breeding values (GEBVs) based on combined pedigree, genomic and phenotypic information. The procedure achieves these goals by blending traditional pedigree relationships with those derived from genetic markers. In this study, GEBVs were converted to marker (SNP) effects. Unequal variances for markers were incorporated by deriving weights from SNP solutions, and incorporating the calculated weights into a new genomic relationship matrix. Improvements on the SNP weights were obtained iteratively either by recomputing the SNP effects only or also by recomputing the GEBVs. Efficiency of the method was examined using simulations for 10 replications with 15,800 subjects across 6 generations, of which 1500 were genotyped with 3000 SNP markers evenly distributed on 2 chromosomes. Heritability was assumed 0.5 all due to 30 QTL effects that were simulated based on Gamma distribution across genome. Comparisons included accuracy of breeding values and cluster of SNP effects of ssGBLUP and BayesB with several options for each procedure. For genomic evaluation, an accuracy of prediction of 0.89 (0.01) was obtained by ssGBLUP after only one iteration, which was slightly higher than BayesB of 0.88 (0.02), but required only a small fraction of time. Power and precision for GWAS applications was evaluated by correlation between true QTL effects and the sum of m adjacent SNP effects, where m varied from 1 to 40. The highest correlations were achieved with m = 8 and were 0.82 (0.02) for ssGBLUP, and 0.83 (0.07) for BayesB with m = 16 according to marker density and extent of linkage disequilibrium in simulated population. Computing time for ssGBLUP took about 2 min while BayesB took about 5 h. Therefore, ssGBLUP with marker weights is 2 orders of magnitude faster than the next best procedure, accurate, and easy to implement for GWAS applications. In particular, ssGBLUP is applicable to GWAS with complex models including multitrait, maternal and random regression.

Key Words: genomic evaluation, genome-wide association mapping, single step procedure

454 Genotyping by sequencing (GBS): A novel, efficient and cost-effective genotyping method in cattle. M. De Donato*[1,2], S. O. Peters[1,3], S. E. Mitchell[4], T. Hussain[1,5], and I. G. Imumorin[1], *[1]Department of Animal Science, Cornell University, Ithaca, NY, [2]IIBCA, Universidad de Oriente, Cumana, Venezuela, [3]Department of Animal Breeding and Genetics, University of Agriculture, Abeokuta, Nigeria, [4]Institute for Genomic Diversity, Cornell University, Ithaca, NY, [5]Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Pakistan.*

High-throughput genotyping methods have increased the analytical power to study complex traits by increasing the resolution and ultimately identifying the changes responsible for phenotypic differences in economic traits. However, high cost has prevented large scale use for animal improvement. In this study, we applied a recently published method for genotyping plants to 7 taurine and indicine breeds of cattle from the US and Africa, to determine the efficiency and feasibility of this method in cattle. Genomic DNA from each animal was individually digested with *Ape*KI and *Pst*I. Each sample was then ligated to adaptors containing one of 96 unique bar codes. Samples were then pooled and sequenced in a single lane on the Illumina HiSeq 2000. *Ape*KI GBS libraries produced more than 1.37 million unique reads, but had low number of SNPs, low call rate and too many reads with multiple locations in the genome, so this data was not analyzed any further. *Pst*I libraries produced about 500,000 unique reads, 93.9% of which were tags with at least 64 bases with no "Ns," with a significantly lower number of non-unique reads. On average, 1.14 million reads were produced per animal. A total of 62,295 SNPs were detected throughout all autosomes with an average distance of 39.9 kb, as well as 1,402 SNPs on the X chromosome at an average distance of 106.1 kb. The average marker density per autosome was highly correlated with size (CC = 0.797, r^2 = 0.635) with more markers per Mb in smaller chromosomes. Average SNP call rate in the genotyped individuals was higher than 0.70 in 81.5% of all loci, and the average minor allele frequency was 0.223 ± 0.001. Average observed heterozygosity per individual ranged from 0.046 to 0.294, with 0.064 as the lowest found in the Nigerian Sokoto Gudali breed (indicine) and the highest of 0.197 in Brangus (indicine × taurine). This technique has shown to be a novel, flexible, cost effective and sufficiently high-throughput, yet requiring no previous knowledge of the population, genome structure or diversity and can provide different levels of marker density depending on the resolution or cost desired.

Key Words: cattle, genotyping, NGS

455 Models' predictive ability of breeding values for a small data set of genotyped animals. F. M. Rezende*[1], J. B. S. Ferraz[1], F. V. Meirelles[1], J. P. Eler[1], and N. Ibañéz-Escriche[2], *[1]Faculdade de Zootecnia e Engenharia de Alimentos-Universidade de São Paulo, Pirassununga, São Paulo, Brazil, [2]Genètica i Millora Animal-IRTA, Lleida, Cataluña, Spain.*

The aim of this study was to compare the breeding values' predictive ability of 3 different models for a small data set composed by 3,149 animals genotyped for 106 SNP markers, for which adjusted phenotype and pedigree information were available. The 106 SNP are causal mutations or are located in transcript or promoter regions of *Bos taurus* genes. A data set composed by 83,404 Nellore beef cattle animals measured for production traits and their pedigree, contained a total of 116,652 animals, were used to estimate fixed and random effects solutions on single traits analysis by MTDFREML software, under animal model. The direct additive effects estimated from that analysis were assumed to be the "true" breeding values for those animals. The individual records for all traits were adjusted for fixed and random effects solutions, except for the direct additive effect. The adjusted phenotypes, composed by the direct additive and residual portions of raw phenotype, were used as dependent variables in tested models. Model 1 included only polygenic effects, model 2 included only markers effects and model 3 included both polygenic and markers effects. These analyses were performed by TM software. The models' predictive ability was verified by Spearman rank correlation coefficient, estimated by PROC CORR from SAS, between the animals' rank based on breeding values estimated on models 1, 2 and 3 and the rank based on "true" breeding value. The correlation coefficients estimated for models 1, 2 and 3 were 0.47, 0.22 and 0.66 for weaning weight, 0.53, 0.36 and 0.83 for post weaning gain, 0.57, 0.31

and 0.94 for scrotal circumference and 0.57, 0.29 and 0.84 for muscle score, respectively. The estimates of rank correlation coefficients lead to the same inferences for all analyzed traits. The reduced number of genetic markers available was not enough to retain a large proportion of additive effects contained in the adjusted phenotypes, as indicated by the lower values of rank correlation for model 2. The outcomes for model 3 suggested that for a small data set and a reduced set of genetic markers, the additive effects were better estimated when markers and polygenic effects were considered together, what suggests that marker assisted selection can be useful for Nellore populations.

Key Words: Marker assisted selection, SNP markers, *Bos indicus* cattle

456 Improving efficiency of inferring genetic architecture parameters in whole genome prediction models. W. Yang* and R. J. Tempelman, *Michigan State University, East Lansing.*

The reliability of whole genome prediction models (WGP) based on using high density single nucleotide polymorphism (SNP) panels critically depends on knowledge and/or reliable estimation of key hyperparameters that partly specify genetic architecture for the traits of interest. These hyperparameters include π, the proportion of SNP not associated with the trait as well as df and s^2. These latter 2 are, respectively, the degrees of freedom and scale parameter for the Student t density, often used to characterize the distribution of SNP effects in BayesB ($\pi > 0$) and BayesA ($\pi = 0$) models. Estimation schemes, however, based on the use of Markov Chain Monte Carlo (MCMC) methods have been plagued by poor mixing, in part because of the high correlation between df and s^2 in current univariate (UNI) sampling approaches. We consider 2 alternative approaches based on Metropolis-Hastings sampling schemes; one based on univariate draws from each of df and scale (UNIMH) and the other based on bivariate draws of the 2 parameters (BIVMH). We tested these 3 sampling methods on 6 replicated data sets, each analyzed at 3 different SNP marker densities with average pairwise LD levels of r^2 = 0.17, 0.25 and 0.32 . The BIVMH and UNIMH methods had significantly higher computational efficiencies for estimating df and scale compared with UNI ($P < 0.001$) in BayesA and BayesB implementations at all LD levels with the BIVMH outperforming UNIMH for s^2 in BayesA only. For BayesA, these efficiencies were 3–7 times greater in BIVMH and UNIMH relative to UNI whereas they were 5–35 times greater in BayesB with the largest gains being attained at higher LD levels. One ominous result is that the effective number of independent samples (ESS) from MCMC on estimating df and s^2decreases substantially with increasing marker densities such that reliable inference from higher density SNP marker panels require not only greater computing time per MCMC cycle but also greater total number of cycles as well. We also demonstrate how sensitive the accuracy of WGP is to misspecification of these key hyperparameters.

Key Words: genomic prediction, Bayesian inference, genetic architecture

457 A multi-compartment model for genomic selection in admixture populations. E. Hay,* S. Smith, and R. Rekaya, *University of Georgia, Athens.*

Currently, genome wide association studies and genomic selection (GS) are often conducted using purebred populations. Estimation and often validation of SNP are carried out using a select elite set of purebred animals (i.e., proven sires). This process was successful when estimated SNP effects were used to predict genomic breeding values on animals of the same breed. But it fails at different degrees when these SNP estimates are used for genomic prediction in other breeds or crossbred animals. Current approaches for dealing with admixed and crossbred populations in GS rely on using different groups of pooled animals in training and validation sets, hence are data dependent and often lead to reduction in accuracies for animals in pure breed populations. In admixture populations or in presence of crossbred animals, pooled databased methods assume SNP effects are the same across breeds or sub-populations. This assumption is seldom true due to several parameters such as minor allele frequency, strength of LD between markers and QTLs, and linkage phase between marker and QTL alleles change across sub-populations. To deal with this problem, we proposed a multi-compartment model where the effect of a SNP could be different between breeds and parameterized as a function of its effect on one of the breeds in pooled population through a one to one mapping function. An admixture population consisting of 2 lines (A and B) of birds was used to test our proposed method. It consisted on 2807 birds (1989 for A and 818 for B) genotyped for around 57 k SNPs. Three analyses were conducted: 1) each line analyzed separately (M1); 2) pooled data (M2); 3) pooled data using our multi-compartment model (M3). For M1, accuracy (correlation between EBVs and GEBVs) was 0.69 and 0.68 for line A and line B, respectively when training and validation were conducted within the same line. These accuracies dropped to 0.15 and 0.21 when training and validation were conducted in different lines. Using M2 (training and validation on pooled data), the accuracy decreased to 0.53. Using our method (M3), the accuracy was 0.59 or 11% increase compared with M2.

Key Words: genomic selection, admixture, SNP

458 Bayesian integration of external information into the single step approach for genomically enhanced prediction of breeding values. J. Vandenplas*[1,2], I. Misztal[3], P. Faux[1], and N. Gengler[1], *[1]University of Liege - Gembloux Agro-Bio Tech, Gembloux, Belgium, [2]National Fund for Scientific Research, Brussels, Belgium, [3]University of Georgia, Animal and Dairy Science Department, Athens.*

An assumption to compute unbiased estimated breeding values (EBV) is that all information, i.e., genomic, pedigree and phenotypic information, has to be considered simultaneously. However, current developments of genomic selection will bias evaluations because only records related to selected animals will be available. The single step genomic evaluation (ssGBLUP) could reduce pre-selection bias by the combination of genomic, pedigree and phenotypic information which are internal for the ssGBLUP. But, in opposition to multi-step methods, external information, i.e., information from outside ssGBLUP, like EBV and associated reliabilities from Multiple Across Country Evaluation which represent a priori known phenotypic information, are not yet integrated into the ssGBLUP. To avoid multi-step methods, the aim of the study was to assess the potential of a Bayesian procedure to integrate a priori known external information into a ssGBLUP by considering simplifications of computational burden, a correct propagation of external information and no multiple considerations of contributions due to relationships. To test the procedure, 2 dairy cattle populations (referenced by "internal" and "external") were simulated as well as milk production for the first lactation of each female in both populations. Internal females were randomly mated with internal and 50 external males. Genotypes of 3000 single-nucleotide polymorphisms for the 50 males were simulated. A ssGBLUP was applied as the internal evaluation. The external evaluation was based on phenotypic and pedigree external information. External information integrated into the ssGBLUP consisted to external EBV and associated reliabilities of the 50 males. Results showed that rank correlations among Bayesian EBV and EBV based on the joint use of

external and internal data and genomic information were higher than 0.99 for the 50 males and internal animals. The respective correlations for the internal evaluation were equal to 0.50 and 0.90. Thereby, the Bayesian procedure can integrate external information into ssGBLUP.

Key Words: Bayesian, genomic, single step

459 Conceptual comparison between standard multiple-trait and structural equation models in animal breeding applications. B. D. Valente,* G. J. M. Rosa, X.-L. Wu, D. Gianola, and K. A. Weigel, *University of Wisconsisn, Madison.*

Structural equation models (SEM) are multivariate specifications capable of conveying causal relationships among traits. Although these models offer insights into how phenotypic traits relate to each other, it is unclear how SEM can improve multiple trait selection. This is a major issue, ultimately defining how SEM can serve animal breeding. Here, we explored concepts involved in SEM, seeking for benefits it could bring to breeding programs, relative to the standard multi-trait models (MTM) commonly used in practice. Genetic effects pertaining to SEM and MTM have distinct meanings. In SEM, these represent genetic effects acting directly on each trait, without mediation by any other instances of the multiple-trait set under study; in MTM they represent overall genetic effects on each trait. Hence, by using a SEM, it is possible to disentangle the overall genetic components into direct and indirect effects. However, in breeding programs one is interested in selecting candidates that produce offspring with best phenotypes, regardless of how traits are causally associated, and overall additive genetic effects are predictive of offspring phenotypes. So, there is no loss of information by using MTM based predictions, even if there are causal associations among traits. Conversely, the extra knowledge provided by causal information may give the ability of predicting effects of external interventions. One may be interested in selecting for a scenario where interventions are performed, e.g., artificially defining the value of a trait, blocking causal associations, or modifying their magnitudes. By knowing SEM genetic effects and mirroring interventions on the causal structure of the model, predictions for these scenarios are possible from data recorded without the interventions. MTM, on the other hand, do not provide information for such predictions. As livestock production involves many interventions, SEM may be then advantageous in many settings.

Key Words: multiple-trait models, selection, structural equation models