32 A computer program for detecting additive, dominance, imprinting, sex-influenced and the overall QTL effects. Y. Duan, J. Garbe, N. London, and Y. Da*, *University of Minnesota*, *St. Paul*.

A computer program was developed to test the overall QTL effect and individual QTL effects including additive, dominance, imprinting, and sex-influenced QTL effect under the F-2 design. The least-squares analysis is used to implement the statistical test for each type of QTL effect. Bi-allelic and multi-allelic markers are allowed, and the program can accommodate an arbitrary number of fixed non-genetic factors. The computer program requires three input files, marker genotypes, phenotypes and pedigree, and a parameter file. Two output files are produced. The first output file prints significant individual QTL effects (additive, dominance, imprinting, and sex-influenced) at a user specified significance level. This file also prints test results of the overall QTL effect, which generally is more significant than any individual QTL effect. The second output file prints significant overall QTL effects at a user specified significance level and also prints test results of each individual QTL effect, which generally is less significant than the overall QTL effect. By default, QTL effects with "suggestive linkage" or "significant linkage" are printed in the output files. The computer program was applied to analyze swine QTL mapping data under an F-2 design.

Key Words: QTL, Computer program, Mapping

33 A mixed model approach to map QTL controlling complex binary disease traits and interacting with environments. Y. Li and H. N. Kadarmideen*, *Statistical Animal Genetics Group, Swiss Federal Institute of Technology*, *ETH Zentrum, Zürich, Switzerland*.

A statistical method using generalized mixed model which fitted QTL effects as random was proposed for mapping quantitative trait locus (QTL) affecting binary traits (e.g. disease) and showing QTL-by-environment interactions (QEI) in multi-family half-sib designs. Estimates of QTL variance and location and power were compared between two random models (I and II) using simulated datasets. Binary dataset was generated from twenty paternal half-sib families each with 200 progeny. The QTL was assumed to be interacting with environment, having different QTL substitution effects among five environments. Phenotypic values of offspring were first generated on the liability scale as a sum of population mean, polygenic effects of sire, environmental effect, the substitution effect of sire QTL allele, interactions (QEI) and residuals. The liability values were transformed into observable 0-1 binary (non-diseased vs diseased) phenotypes using a threshold value corresponding to disease incidence of 50%. The statistical model I did not fit QEI but only random common parent (sire) QTL effect. Model II fitted QTL substitution effects and their interactions with environment (QEI) as random effects. Environmental effects in both models were fitted as fixed effects. QTL variance component in the random model was directly estimated using a restricted maximum likelihood approach. The power of detecting the QTL was 5.5% and 89% in Model I and Model II, respectively. The variance estimate was 0.008 with Model I and 0.050 with Model II. It is concluded that when QEI are present in a population, the model which did not consider QEI was unable to detect the QTL whereas the model which considered QEI significantly improved power and variance estimates. This is the first study to map and estimate breeding values at the QTL which differ across environments in half-sib populations using generalized mixed models and hence would be useful in environment-specific marker assisted genetic evaluation and selection.

Key Words: QTL mapping, QTL-by-environmental interaction, Generalized mixed model

34 A comparison of sire and animal model genetic parameter estimates from herds with high and low within-herd heritabilities. C. D. Dechow*[1] and H. D. Norman[2], [1]*The Pennsylvania State University*, *University Park*, [2]*Animal Improvement Programs Laboratory*, *Beltsville, MD*.

The objectives of this study were to determine if within-herd heritability (WHH) estimated with regression techniques accurately reflects heritability (h^2) differences among herds and to compare sire and animal model genetic parameter estimates among herds varying in WHH. Milk, fat, and protein yield and somatic cell score (SCS) were obtained from states representing the Northeast, Southeast, Midwest and Western regions of the United States. Four random subsets of data were selected across regions. Within-herd heritabilities were estimated with daughter-dam regression (h^2_{DD}) and daughter-sire predicted transmitting ability (PTA) regression (h^2_{DS}) in ASReml with a model that included fixed effects of herd-year-season of calving, age, dam yield, and sire PTA. Dam yield and sire PTA were also treated as random covariables nested within herd to create h^2_{DD} and h^2_{DS} that were regressed toward the subset average. Herds were ranked from lowest to highest WHH for each trait and average WHH rank was calculated. Fifty herds with the highest and lowest average WHH were retained for further analysis in each subset. Milk, fat, protein and SCS were each considered two separate traits (one in the high, another in the low h^2 herds) and analyzed with two trait sire and animal models in ASReml. Fixed effects included herd-year-season of calving, age and lactation number. Random effects included animal, permanent environment and error. Sire and animal model h^2 estimates were greater in herds with high h^2 for all traits. Average h^2_{DD} across subsets for milk yield was 0.41 for high h^2 herds and 0.29 for low h^2 herds, while average milk yield h^2_{DS} were 0.35 and 0.13 from high and low h^2 herds, respectively. Sire model h^2 estimates were more highly correlated with h^2_{DS} (0.93) than with h^2_{DD} (0.85). Animal model h^2 estimates were more highly correlated with h^2_{DD} (0.91) than with h^2_{DS} (0.72). There are significant differences among herds for h^2 and it appears WHH estimated with daughter-dam and daughter-sire PTA regression reflect those differences with some accuracy.

Key Words: Within herd heritability, Daughter-dam, Daughter-sire

35 Modeling extended lactations in Holsteins. C. M. B. Dematawewa*[1], R. E. Pearson[1], and P. M. VanRaden[2], [1]*Virginia Polytechnic Institute and State University*, *Blacksburg*, [2]*Animal Improvement Programs Laboratory, Agricultural Research Services, USDA*, *Beltsville, MD*.

The objective of this study was to develop an equation for predicting average yield of cows still in milk from 1 to 999 days. Test day yields (kg/d) of 903,529 lactations of 305,202 Holstein cows calved between 1997 and 2003 were used. Average daily yield (**Y**) for each 30–d

interval of lactation was calculated for each parity (9 parities), based on cows which were in milk during the month considered. Various lactation models available in literature (i.e. Wood's model and other variants of incomplete Gamma function, inverse polynomials and, mono–, di–, and multi–phasic curves) were tested, before and after a modification made by including a new additive parameter (**k**). Nonlinear regression procedures in SAS were carried out between Y (34 30–d means) and respective days in milk (**DIM**), within and across parities. R–squared value, mean square error, Bayesian information criteria, and autocorrelation of errors were considered as the model selection criteria. Standard models underestimated yields at later stages of lactation for all parities until modified to include the new parameter. The modified Dijkstra model (i.e. $Y= k + a[\exp(b(1–\exp(–c\times DIM))/c – d\times DIM)]$) fitted best (R squared = 0.99), both within and across parities, with the lowest estimates of mean square error and Bayesian information criteria (BIC=59.6). The parameter estimates for a, b, c, d, and k were 12.9088, 0.0183, 0.00806, and 0.00925, and 20.1303, respectively for across parities. Durbin–Watson statistic showed that the first order autocorrelations of errors were negligible for the above model for both within and across parities (DW=1.412, N=34). The modified mono–phasic curve (i.e. $Y = k + ab[1– \tanh^2(b(DIM–c))]$) provided the second best fit for both within and across parities. These results show that modified Dijkstra formula stated above can be effective in modeling the mean yield of cows remaining in milk through lactations well beyond 305 days.

Key Words: Modeling, Lactation curve, Holsteins

36 Improving stability and reliability of test day model evaluation in the Italian Holstein. F. Canavesi*, S. Biffani, and F. Biscarini, *Associazione Nazionale Allevatori Frisona Italiana*, *Cremona, Italy*.

In November 2004 the first Italian genetic evaluation based on Test Day Random Regression Model (TDRRM) was published. The new model assumes that different genetic values can be expressed during the lactation and also across lactations. The published proof is a combination of the first three lactations. Along with the combined proof also breeding values for each lactation are published as well as the persistency Transmitting Ability in the three lactations and for each lactation. The advantage of the Test Day model is that it does not only provide farmers with the expected production superiority of the daughters of a particular bull but also with a description of the way the daughters of that bull behave in expressing their superiority during each lactation and across the first three lactations. Farmers can now choose the bulls that better match their herd management. The overall variability of such a complex system is larger than the variability observed with the former repeatability model based on lactation records. In order to improve the stability of the proofs, an extensive research program has been set up to investigate the scale of expression of the proofs and the definition of the fixed effects to be included in the model. In February 2006 few changes were introduced: the scale of expression of proofs was defined by the standard deviation of third parity cows in the genetic base population. The genetic base population is now defined by cows born in the triennium 1997-1999 and it is a rolling base that will be updated each year in the August evaluation. These changes improved stability over time by 15-20%. Differences across years of production and days open classes, that are not yet accounted for in the genetic evaluation model, were also investigated. Changes in the structure of fixed effects will be introduced in the August evaluation and this will further increase the stability and reliability of the genetic evaluation system.

Key Words: Proof stability, Test day model, Random regression

37 Use of phenotypic information to ascertain paternity. R. L. Sapp*, R. Rekaya, W. Zhang, and J. K. Bertrand, *The University of Georgia*, *Athens*.

A new method of implementing mixed model equations without constructing the inverse of the relationship matrix (WO-A) was developed and compared to the classical best linear unbiased prediction implementation where the inverse of the relationship matrix was constructed (W-A). Both methods were compared in a univariate and multiple-trait situation using simulated data. For the univariate case, the Pearson correlation between estimates of WO-A and W-A was 1.00 for fixed and random effects. Similarly, parameter estimates were exactly the same using W-A and WO-A for three correlated traits. These results indicate that W-A could be implemented without the need to construct or to invert the relationship matrix. The proposed method should not be considered as an alternative to the classical implementation except in situations where memory is a limiting factor, for iteration on data or in genetic evaluation with uncertain paternity. For the latter, the proposed method is, perhaps, the only viable procedure. The second objective was to investigate the implementation of a genetic evaluation using WO-A in presence of uncertain paternity. The scenarios investigated included either one or two records of a single trait with varying heritability and one record of three correlated traits. The average probability of the true sire being identified as such (PSA) was computed for all scenarios. As expected, PSA increased with increasing heritability using one trait. However, the results of the current study suggested that the probability of identifying the true sire was the highest when three correlated traits were used to compute PSA and the lowest when only one record was used. For traits such as birth weight and weaning weight where only one measurement is taken, the multiple-trait scenario could result in more animals being assigned the true sire than if birth or weaning weight was used separately. Further research is needed to determine the performance of this methodology in field data as well as the potential implementation of this methodology in conjunction with molecular information.

Key Words: Genetic evaluation, Uncertain paternity

38 Ascertaining paternity using phenotypic and molecular information. R. L. Sapp*, R. Rekaya, and J. K. Bertrand, *The University of Georgia*, *Athens*.

The objective of the current study was to evaluate the effectiveness of paternity assignments for 15 markers prior to and after inclusion of phenotypic information. Fifteen markers with the number of alleles per marker ranging from 3 to 7 were simulated for every animal in the pedigree file. A linear mixed model which included a fixed effect and random effects of additive breeding values and residual terms was used to generate and analyze data for three correlated traits. The marker information was used to assign paternity to all animals in the pedigree file, regardless of paternity status, assuming the mother's genotype was unknown. A paternity index (PI) was calculated for each of the 15 markers using the candidate sire (CS) and offspring's genotypes. The probability of paternity (W) was computed using the product of PI for the 15 markers. For animals with uncertain paternity, PI and W were calculated for each CS. The molecular information, on average, resolved 94% of the uncertain paternity cases. For the remaining 6% of animals with uncertain paternity, more than one CS was not excluded based on the marker information. The standardized probability of paternity for each non-excluded CS was used as prior information together with the phenotypic information to discriminate between those non-excluded sires. For the true sires of animals with more

than one non-excluded CS, the prior probability based on marker information was 0.58. Likewise, the prior probability for non-true CS was approximately 0.42. When combined with the phenotypic information, the posterior probability of paternity for the true sires increased to nearly 0.80. Thus, the true sire was correctly assigned 80% of the time when molecular information was not able to unambiguously assign paternity. The results of the current study indicated that when molecular information was available, but not able to determine paternity, phenotypic information could be used to ascertain paternity. Moreover, the method proposed in the current study was able to increase the probability of assigning the true sire using the molecular and phenotypic information.

Key Words: Paternity index, Paternity testing

39 The combination of genetic test information and phenotypic records for the prediction of breeding values. M. L. Spangler*, R. Rekaya, and J. K. Bertrand, *The University of Georgia*, *Athens*.

The use of marker assisted selection in the beef cattle industry to date has been comprised of using traditional EPDs in tandem with molecular test information. The combination of these two sources of information into one quantitative genetic value has clear advantages to livestock producers and genetic testing companies. The instances were molecular information is of the most benefit are those in which the trait of interest is difficult or expensive to measure or the heritability is low. In the current study a multiple trait simulation was carried out to create a beef cattle data set using genetic parameter estimates from the literature in order to identify the best procedure for combining both sources of information and to assess the added benefit of the procedure. To reach these objectives the following simulation/analysis steps were implemented: 1) all phenotypic records are observed without knowledge of a causal gene (A1) 2) all animals in the data file have phenotypic records for the trait of interest and molecular information is available on all animals in the pedigree (A2) and 3) varying percentages of missing records for the trait of interest and complete molecular information (A3). In the latest scenario the percentage of missing records for marbling score was kept very high in order to mimic real production systems. The data set for A3 included six correlated traits with complete records for five traits and only 5% of the records were observed for marbling score for which molecular information was available for a causative gene that explained 10% of the genetic variation. The preliminary results showed a better prediction of the true breeding values of the top 10% of sires (candidates for selection) using A3 compared to the case where genomic information was not considered. In fact, the rank correlation was increased by more than 21%. This study is currently being extended to situations with varying allelic frequencies for the causative gene and the genetic progress over several generations of selection is being assessed.

Key Words: Beef cattle, Best linear unbiased prediction, Gene-assisted selection

40 Genetic evaluations for mixed breed populations. P. M. VanRaden*, M. E. Tooker, J. B. Cole, G. R. Wiggans, and J. H. Megonigal, Jr., *Animal Improvement Programs Laboratory, USDA*, *Beltsville, MD*.

New programs to include data from crossbred animals were tested for genetic evaluation of US dairy cattle. An all-breed animal model including an adjustment for general heterosis was applied to usable records back to 1960. Yield records were adjusted to 36 months of age instead of mature equivalent. Total numbers of cows with records were >20 million for yield traits, for productive life (PL), and for daughter pregnancy rate (DPR), and >10 million for somatic cell score (SCS). About 1% of recent cows are first generation (F1) crossbreds, and the percentage is growing. Unknown parent groups were separate by breed, pedigree path (dams of cows, sires of cows, and parents of sires), US or foreign origin, and birth year. Groups were formed when they included at least 500 animals within time periods and at least 2000 animals across all years. Heterogeneous variance adjustments were estimated separately within herd, year, and sire breed. Convergence of the animal model was fairly rapid, indicating sufficient connections among the breeds and crossbred groups. Genetic differences among breeds were similar to previous estimates, and genetic rankings were very highly correlated with previous within-breed evaluations. Correlations for high reliability bulls exceeded .999 for Holsteins and exceeded .98 within other breeds. Changes were largest in the smallest breeds. Predicted transmitting abilities (PTA) changed more for bulls with fewer daughters, and for cows. Additional group-mates of another breed can add accuracy but also could cause bias if managed differently or not modeled properly. Evaluations of crossbred animals are more accurate in all-breed than within-breed analyses because the relationship matrix can link to reliable sire PTA for breeds in both the maternal and paternal ancestry. Joint evaluation of all breeds and crossbred animals does not greatly change rankings for animals with mostly purebred group-mates and relatives, but provides routine calculation of breed differences. Evaluations may be converted back to original within-breed bases for display to the public.

Key Words: Genetic evaluations, Heterosis, Crossbreeding

41 A new statistical model and method of multiple breeds evaluation. L. Zhang*[1,2], E. J. Pollak[2], and R. L. Quaas[2], [1]*Inner Mongolia Agricultural University*, *Huhhot, China*, [2]*Cornell University*, *Ithaca, NY*.

Crossbreeding and multiple breed genetic evaluations of beef cattle have triggered an urgent need for a reasonable and effective model to analyze simultaneously the mixture of multi-genome for crossbreeding and mixture of the multi-population for multi breeds to facilitate the estimation of the breeding values for both of the mixtures. Since the different breeds differ from each other in genetic background and belong to different population, it is reasonable to assume the existence of a probability distribution of for each breed. The mixtures of the gene pool and the populations taken from all breeds will have a new probability distribution which is a mixture of distributions of every breed. So, the model of multiple breed evaluation should be a new normal distribution in which the mixture of multiple normal distributions, multiple traits with common known/unknown covariance matrices, and different means within 3 standard deviations from common mean, when the proportions of the distributions of every breed were known a priori. By this definition, all the available linear models and algorithms are still worked without any new difficulties or limitations involved. In this mixture linear model, the effects of the breed and the combination of crossbreed were included in the fixed effects; the effects of genomic mixture among different breeds belong to random effects. When the coordinates represent measurements that are subject to random fluctuations of differing magnitudes, it is often desirable to weight coordinates subject to a great deal of variability less heavily than those that are not highly variable. A statistical distance that accounts for differences in variation and, in the presence of correlation

was developed. The statistical distances among the breeds, populations and individuals are the genetic or breeding differences which can be used to multiple breeds genetic evaluations.

Key Words: Mixture distribution, Statistical distance, Multiple breed

42 Use of Principal Components and Factor Analysis to factorize genetic correlation matrices of multivariate phenotypes. N. P. P. Macciotta*, N. Bacciu, C. Dimauro, and A. Cappio-Borlino, *Dipartimento di Scienze Zootecniche, Università di Sassari*, *Sassari, Italia*.

Principal Component Analysis (PCA) has been used by several authors to address issues of genetic analysis of multivariate complex phenotypes, i.e. large number of parameters to be estimated, computational constraints and sampling error. PCA is able to describe the maximum amount of original variation with the minimum number of orthogonal new variables. However, it is scarcely flexible and results are not always easy to interpret. In order to cope with these shortcomings, the multivariate Factor Analysis (FA) could be proposed. FA, via the rotation technique, is able to extract latent variables with defined relationships with original variables. In the present work the ability of PCA and FA to factorize genetic correlation matrices (published data) of milk yield, logSCC for dairy cattle and weights at different ages for beef cattle is compared. The PCA yields a first component related to all original variables, able to explain most of the original variance (82%, 50%, 82% for milk yield, logSCC and body weight, respectively). Other principal components highlight the contribution of specific original variables to the total variance in decreasing order. On the other hand, FA extracts common latent factors related to specific groups of variables and that have a clear biological meaning: in the case of milk yield, for example, the fist latent common factor is able to explain 55% of the original variance, it is related to the milk tests of the second part of lactation and can be considered as an index of lactation persistency; the second factor (42% of explained variance) is related with tests of the first part of lactation and can be considered as an index of level of production in early lactation. Similar results are obtained for body weights and, even if less defined, for logSCC. Results of the present work suggest that a combined used of PCA and FA could be usefully for a deep understanding of latent structure of correlations in multivariate complex phenotypes.

Key Words: Multivariate phenotypes, Principal component analysis, Factor analysis

Food Safety: Ruminants as Reservoirs for Shiga Toxin-Producing Escherichia coli

43 Shiga toxin-producing *Escherichia coli*: The big picture. C. L. Gyles*, *University of Guelph*, *Guelph, Ontario, Canada*.

Shiga toxin-producing *Escherichia coli* (STEC) represent a diverse group of *E. coli* that have one thing in common, namely, the ability to produce Shiga toxin (Stx). STEC behave as normal flora in ruminants, which are a major reservoir. In humans, STEC can cause disease of varying severity. Serotyping is used as an initial basis for differentiating STEC, and virulence characteristics that are serotype-related have allowed the concept of sero-pathotype to be developed. Serotype O157:H7 is the serotype that is most frequently implicated in outbreaks of disease and in severe disease in North America and several other regions. This serotype has therefore been the subject of the most intense investigation both with respect to its relation to its reservoir host, ruminants, and its accidental host, humans. Other serotypes are also implicated in outbreaks and in severe disease. STEC pathogenesis involves two phases: colonization of the intestine, and production of toxin. Several approaches have been used in attempts to identify the bases for virulence of STEC. These have resulted in recognition of a large number of putative virulence factors, but only a small number are clearly significant contributors to virulence. Most of the highly pathogenic STEC have the ability to produce a characteristic attaching and effacing lesion in the intestine. This lesion is the result of a bacterial type III secretion system that injects certain effector proteins into the host intestinal epithelial cell. Profound changes in the architecture and metabolism of the host cell occur and contribute to the diarrha that develops. The severe complications that develop in a certain percentage of affected humans are attributable to the Shiga toxin, which exists as two major types, Stx1 and Stx2. STEC may produce one or both of these types of Stx. Production of Stx2 is associated with disease of greater severity, but strains that produce Stx1 may also cause severe disease. The genes for Stx are encoded on temperate bacteriophages in the chromosome of the bacteria and production and release of the toxin are highly dependent on induction of lysis of the phages.

Key Words: Serotype, Virulence, Shiga toxin

44 Prevalence and pathogenicity of Shiga toxin-producing *Escherichia coli* in beef cattle and their products. H. S. Hussein*, *University of Nevada*, *Reno*.

In the past 25 years, many human illness outbreaks have been traced worldwide to consumption of undercooked ground beef and other beef products contaminated with Shiga toxin-producing *Escherichia coli* (STEC). Because of the global nature of food supply, the safety concerns with beef will continue and the challenges facing the beef industry will increase. To be prepared to address these concerns and challenges, it is critical to assess the beef cattle role in human infection with STEC. Because most STEC outbreaks in the US were traced to beef containing *E. coli* O157:H7, the epidemiological studies have focused on prevalence of this serotype in beef cattle. Worldwide, however, additional STEC serotypes (e.g., members of the O26, O91, O103, O111, O118, O145, and O166 serogroups) have been isolated from beef and caused human illnesses ranging from bloody diarrhea and hemorrhagic colitis to the life-threatening hemolytic uremic syndrome (HUS). To provide a global assessment of the STEC problem, published reports on beef and beef cattle in the past three decades were evaluated. Prevalence rates of *E. coli* O157 ranged from 0.1 to 54.2% in ground beef, from 0.1 to 4.4% in sausage, from 1.1 to 36.0% in various retail cuts, and from 0.01 to 43.4% in whole carcasses. The corresponding prevalence rates of non-O157 STEC were 2.4 to 30.0%, 17.0 to 49.2%, 11.4 to 49.6%, and 1.7 to 58.0%, respectively. Of the 161 STEC serotypes isolated from beef products, 43 were detected in HUS patients and 36 are known to cause other human illnesses. With