national trends. Hot carcass weight (HCW), 12th rib fat thickness (BF), longissimus dorsi area (REA), USDA yield grade (YG) and USDA marbling score (MS) were analyzed using records from the Alabama Beef Connection (ABC) and Alabama Pasture to Rail Program (P2R). The ABC database contained 5,160 records from 2003 to 2006 on cattle sold as feeder calves and fed primarily in the Midwest and High Plains regions of the United States. The P2R database contained 4,997 records from 1994 to 2005 of co-mingled retained ownership cattle fed in the High Plains region of the United States. Data were analyzed using a general linear model in SAS. Fixed effects included year, breed of sire and region fed. A covariate of harvest date was included for all traits. The HCW of Alabama feeder cattle have not followed audit trends. Hot carcass weights have tended to increase nationally (344 to 361 kg). In P2R cattle, HCW has significantly decreased from 352 kg in 1994 to 335 kg in 2005. The ABC cattke show a similar trend (392 kg in 2003 to 351 kg in 2006, $P<0.05$). In both data sets, REA remained stable across all years (P2R 87.04 $cm^2$; ABC 88.11 $cm^2$, audit 84.5 $cm^2$). The MS trend was significantly positive across years in both datasets. The P2R data from 1994 moved from a MS of 411 to 484 in 2006 ($P<0.05$). The ABC data from 2003 to 2006 moved from a MS of 484 to 490 ($P<0.05$). Back fat and YG were the most variable carcass traits for Alabama across years with positive and negative trends ($P<0.05$). Feedlot, market conditions and weather probably affected these traits as much as genetic predisposition. Alabama cattle are generally YG 2 cattle (P2R 2.54 vs. ABC 2.56; audit 3.0). This is primarily due to 3.23 $cm^2$ more REA than required for the associated HCW. Alabama results do not agree with audit findings of increased HCW and REA over time.

**Key Words:** Beef Cattle, Carcass Characteristics, Beef Quality

## Breeding and Genetics - Livestock and Poultry: Analyses and Methods I

**409 Using epidemiological models and genetic selection to identify theoretical opportunities to reduce disease impact.** G. D. Snowder*, *USDA, ARS, U.S. Meat Animal Research Center*, *Clay Center, NE*.

Selection for disease resistance is a contemporary topic with developing approaches for genetic improvement. Merging the sciences of genetic selection and epidemiology is essential to identify selection schemes to enhance disease resistance. Epidemiological models can identify theoretical opportunities for genetic selection to reduce the impact of a disease. Potential selection venues may be more appropriately comprehended by compartmentalization of disease components using epidemiological models. This study considers the standard Susceptible, Infected, Recovered (SIR) epidemiological model and five other common epidemiological models (MSIR, SEIR, SIS, Carrier State, and SIR Vector) with genetic selection alternatives. Theoretical modeling of genetic selection effects on epidemiological models were used to: predict the economic effect of selection, estimate the optimal number of resistant animals to prevent an epidemic, and determine genetic selection alternatives. Selection alternatives to genetic disease resistance include lowering the probability of being infected, tolerance for the pathogen, longer latency period, less severe clinical expression, faster recovery rate, and compensatory rebound. These selection alternatives can result in favorable changes to the differential equations for susceptibility, infected, and recovery rates. Potentially undesirable consequences due to selection can be predicted, such as an increase in the size of sub clinical populations harboring and shedding pathogens. When applied to actual data for bovine respiratory disease, this approach identifies the complexity of genetic resistance to this disease while detecting potential opportunities for genetic selection. When a disease such as bovine respiratory disease is caused by different pathogens (bacterial, viral, mycoplasmal, etc.) with different pathways of infection, the probability of reducing the disease prevalence with genetic selection is diminished.

**Key Words:** Animal Breeding, Cattle, Disease Resistance

**410 Assessment of different selective phenotyping design strategies for genetical genomics studies with outbred F2 populations.** F. F. Cardoso*[1,2], J. P. Steibel[1], G. J. M. Rosa[3], C. W. Ernst[1], R. O. Bates[1], and R. J. Tempelman[1], [1]*Michigan State University*, *East Lansing*, [2]*Embrapa Pecuária Sul*, *Bagé, RS, Brazil*, [3]*University of Wisconsin*, *Madison*.

Quantitative genetic analysis of transcriptional profiling experiments is emerging as a promising approach to discover candidate genes underlying variation of complex biological traits. However, adoption of these genetical genomics approaches is currently limited by the high cost of microarrays. We studied variants of three recently proposed design strategies to optimally select subsets of individuals for transcriptional profiling including maximizing genetic dissimilarity between selected individuals, maximizing the number of recombination events in selected individuals, and selecting phenotypic extremes within genotypes of a previously identified quantitative trait locus (QTL). We also investigated two other options, namely purely random selection and profiling animals with the highest and lowest phenotypic values within each family-gender subclass. A simulation study was conducted based on linkage map and marker genotypes provided from a dataset on Chromosome 6 for 510 F2 animals from an actual pig resource population. Comparisons between methods were based on a biallelic QTL with pleiotropic effects on a phenotypic trait and a particular expression profile. The model included an overall mean, fixed additive QTL and sex effects and random polygenic and family effects. Bivariate (gene expression with phenotypic data) mixed model analyses were conducted for subset selection intensities of 80/510, 160/510 and 240/510. All methods were deemed to be similar for the mean absolute distance of the estimated QTL to the true QTL location. Precision and bias of estimates of QTL effects was further assessed by their Mean Square Error (MSE). The genetic dissimilarity and extremes within genotype methods had the smallest MSE and maximum sensitivity, outperforming all other selection strategies, particularly at the smallest proportion of selected samples (80/510).

**Key Words:** Genetical Genomics, Selective Phenotyping, QTL

**411 Different methods of selecting animals for genotyping to maximize the amount of genetic information known in the population.** M. L. Spangler*[1], R. L. Sapp[2], J. K. Bertrand[1], M. D. MacNeil[2], and R. Rekaya[1], [1]*University of Georgia*, *Athens*, [2]*USDA-ARS Fort Keogh Livestock and Range Research Laboratory*, *Miles City, MT*.

It is possible to predict genotypes of some individuals based on genotypes of relatives. Different methods of sampling individuals to be genotyped from populations were evaluated using simulation. Simulated pedigrees included 5,000 animals and were assigned genotypes based on assumed allelic frequencies (favorable/unfavorable) of 0.3/0.7, 0.5/0.5, and 0.8/0.2. A real beef cattle pedigree which included 29,101 animals was used to test selected methods using simulated genotypes with allelic frequencies of 0.3/0.7 and 0.5/0.5. For the simulated pedigrees, known and unknown allelic frequencies were assumed. The methods used included random sampling, selection of males, and selection of both sexes based on the diagonal element of the inverse of the relationship matrix (A-1) and absorption of either the A or A-1 matrix. For random sampling, scenarios included selection of 5 and 15% of the animals while all other methods presented concentrated on the selection of 5% of the animals for genotyping. The methods were evaluated based on the percentage of alleles correctly assigned after peeling (AKP), the probability of assigning true alleles (AKG) and the average probability of correctly assigning the true genotype (APTG). As expected random sampling was the least desirable method. The most desirable method in the simulated pedigrees was selecting both males and females based on their diagonal element of A-1. Increases in AKP and AKG ranged from 26.58 to 29.11% and 2.76 to 6.08%, respectively, when males and females (equal to 5% of all animals) were selected based on their diagonal element of A-1 compared with selecting 15% of the animals at random. In the case of a real beef cattle pedigree, selection of males only or males and females yielded similar results and both selection methods were superior to random selection.

**Key Words:** Genotype Sampling, Marker-Assisted Selection, Simulation

**412 Effect of raw data normalisation on detection of differentially expressed genes in cDNA microarray experiments.** C. Dimauro, N. P. P. Macciotta*, and A. Cappio-Borlino, *Dipartimento di Scienze Zootecniche, Università di Sassari*, *Sassari, Italia*.

cDNA microarrays allow for the monitoring of the expression of thousands genes in parallel in different tissues or conditions. However, the interpretation of microarray experiments requires large manipulation of raw fluorescence data that may seriously affect the reliability and repeatability of final results. Two data processing main steps can be identified: normalization, i.e. the pre-scaling of data to correct for technical biases; choice of a suitable statistical model to detect differentially expressed genes, controlling for false discovery rate. Whereas the impact of the latter has been widely investigated, the influence of normalization techniques has not been fulfilled to date. In this work, four different combinations of normalization techniques have been carried out on published fluorescence data generated in an experiment aimed at defining the temporal gene profiling of liver from periparturient dairy cows (GEO accession n.GSE2692). The normalizations were aimed at removing bad spots, correcting for background intensity (2 corrections), bias and dye effect. Differentially expressed genes were detected, for all combinations, with a gene-specific linear mixed model that included the fixed effect of treatment, and random effects of array and cow, plus the random residual. The range of differentially expressed genes ranged from about 2500 to 4000 among the different combinations, with a statistical significance threshold of 0.05, increasing dramatically to a range of about 300 to 500 when a Bonferroni corrected statistical test was applied. The percentage of genes commonly detected in the different combinations is between 20-50% and 2-8% for unadjusted and adjusted tests respectively. Moreover a great variation in the results can be still observed whether mean or median fluorescence intensity data was used. All these results highlight a relevant impact of normalization techniques on detection of differentially expressed genes in cDNA microarrays experiments and strongly suggest that consistent comparisons should be made by using the same normalization procedures.

**Key Words:** cDNA Microarrays, Normalisation, Differentially Expressed Genes

**413 Methods to explain genomic estimates of breeding value.** P. M. VanRaden and M. E. Tooker*, *Animal Improvement Programs Laboratory, USDA*, *Beltsville, MD*.

Genetic markers allow animal breeders to locate, estimate, and trace inheritance of many unknown genes that affect quantitative traits. Traditional models use pedigree data to compute expected proportions of genes identical by descent (assumed the same for all traits). Newer genomic models use thousands of marker genotypes to obtain actual fractions of DNA shared by any two individuals. Full sibs, for example, may actually share 45% or 55% of their DNA rather than the 50% expected in the traditional relationship matrix. The actual percentage of shared genes has a standard deviation (SD) equal to 50% divided by the square root of twice the number of independent loci affecting the trait. This SD does not decrease below 3.5% even with very large numbers of loci on 30 chromosome pairs because loci on the same chromosome are linked rather than independent. Accounting for these small differences in the relationship matrix and tracing individual genes can increase reliability, especially if the number of genotyped individuals is large. Reliability from parent average is <50% and is the upper limit for individuals without phenotypic data or progeny in traditional models. With genotypes and phenotypes for full sibs, genomic models can increase the reliability to 62% with 100 full sibs or 95% with 1000 full sibs. If no sibs have both genotypes and phenotypes but 100 full sibs of each parent do, reliability can increase to 57%. Less gain is provided from each distant relative, but the number of distant relatives may be very large. “Unrelated” individuals actually share many unknown common ancestors born prior to the known pedigree file and thus can provide additional information. More markers are required to estimate and trace genetic effects for distant rather than close relatives because the shared DNA segments are shorter. More markers are also required when more loci affect a trait. Another useful concept is the proportion of genes in common that affect a particular trait, called a quantitative trait loci (QTL) relationship matrix. Genetic evaluations should be more accurate if genomic relationships replace the traditional relationships computed from pedigrees.

**Key Words:** Relationship Matrix, QTL, Genomics

**414 Efficient estimation of breeding values from dense genomic data.** P. M. VanRaden*, *Animal Improvement Programs Laboratory, USDA*, *Beltsville, MD*.

Genomic, phenotypic, and pedigree data can be combined to produce estimated breeding values (EBV) with higher reliability. If coefficient

matrix Z includes genotypes for many loci and marker effects (u) are normally distributed with equal variance at each, estimation of u by mixed model equations or EBV by selection index equations that include a genomic relationship matrix (G) are equivalent models. Matrix G is analogous to traditional relationship matrix A and is obtained by subtracting allele frequencies from coefficients of Z and then dividing the revised Z Z' by the number of marker effects (m). Equations that include either Z'Z or Z Z' are dense and can be solved by several methods tested on simulated data. Off-diagonals count individuals that inherited two different alleles (in Z'Z) or alleles shared by two individuals (in Z Z'). Algorithms that estimate marker effects using Z'Z and then sum to obtain EBV are more efficient than those that use Z Z' unless m greatly exceeds the number of genotyped individuals (n). With direct inversion to obtain reliabilities, computing times increase by $n^3$ for EBV or $m^3$ for marker effects. With iteration to estimate u, computing times increase with the number of iterations (i) times $m^2$. The algorithm known as iteration on data reduces memory, and a simple trick can increase speed. For each individual, its genotypes (left-hand sides) are multiplied by previous round estimates and this sum minus the diagonal coefficient is used to adjust right-hand sides instead of summing off-diagonals times previous solutions again for each effect. Computing time is linear with number of effects in the model (not quadratic as in many previous algorithms) and linear with total number of genotypes, increasing with i times n times m. More iterations and under-relaxation were required for convergence as m increased. The methods include only phenotypes (or daughter deviations) for genotyped individuals, but future algorithms ideally should also include phenotypes of un-genotyped individuals, perhaps by absorbing equations for marker effects into equations for EBV.

**Key Words:** Iteration on Data, Genetic Markers, Algorithm

**415 Recursive algorithm to compute inbreeding coefficients assuming non-zero inbreeding of unknown parents.** I. Aguilar* and I. Misztal, *University of Georgia*, *Athens*.

The objectives of this study were to investigate a recursive algorithm to calculate inbreeding coefficients using rules from the tabular method and to expand it to consider animals with missing parent information using VanRaden's method. For each animal $x$ present in the pedigree, the inbreeding coefficient $F_x$ is calculated as $0.5R_{sd}$, where $R_{sd}$ is the numerator relationship between sire and dam of the animal $x$. Computation of $R_{sd}$ is recursive and involves tracing the ancestors. Three cases are considered. In the first case, if $x=0$ or $y=0$, $R_{xy}=0$. In the second case, if $x=y$, $R_{xy}=1+F_x$. Finally, $R_{xy}=0.5(R_{sy}+R_{dy})$, being $x$ younger than $y$; and $s,d$ sire and dam of $x$. The modification was done in the first case, where $x=0$ or $y=0$ denote unknown parents. Let a negative code denote the year of birth of their progeny, and $b_i$ be an average inbreeding of all animals born in year $i$. Then $R_{xy}=2b_i$, where $b_i$ is based on $x$ and $y$ code: if $x<0$ and $y>0$, $b_i=b_{-x}$; if $x>0$ and $y<0$, $b_i=b_{-y}$; and if $x<0$ and $y<0$, $b_i=max(b_{-x},b_{-y})$. The algorithm is iterative, in the first round $b_i$ is zero, and in the others $b_i$ is the average of inbreeding of animals with year of birth $i$, considering only the inbreeding coefficient of animals with known parents. Testing involved 17 million US Holsteins. Solutions were compared with two algorithms. Convergence was reached in 6 rounds. The computing time per round was 4 minutes, 2 times lower than the implementation of VanRaden algorithm based on tabular method, and two times higher than the algorithm by Meuwissen and Luo, which assumes zero inbreeding of unknown parents. The presented algorithm can also be used for other purposes, including computing of relationship of groups of animals (e.g., sires) using the full pedigree, or creation of nonadditive relationships.

**Key Words:** Inbreeding Coefficient, Unknown Parent, Recursive Algorithm

**416 A social competitive model with the categorical expression.** I. Misztal and R. Rekaya*, *University of Georgia*, *Athens*.

A model by Muir and Schinckel used to model social competition among animals assumes that the animal competitive effects are expressed on a continuous scale. This may not be realistic and could lead to theoretical problems such as variance inflation. A model is proposed where these effects could be expressed in a few discrete categories (strongly dominant, .., independent,.., passive) or as binary (dominant, passive). Let $y_{ij}$ be a record generated under a set of environmental effects i, and let $d_j$ and $c_j$ be the direct and competitive effects of animal j, respectively. Further, let $\alpha(k,x_i)$ be the effect of animal with dominance status k on its mates in the same pen in a set of environmental effects represented by $x_i$. Let $p_j$ be a social dominance category of animal j. The model could be represented as:

$$y_{ij}= other + d_j+ \Sigma\, \alpha(p_k,x_i) +e_{ijk}$$

where *other* are effects other than animals, and the summation are over all the remaining animals in the pen. If only two categories are considered, the model simplifies to:

$$y_{ijk}= other +d_j-\alpha(1,x_i)q_j +e_{ijk}$$

where $q_j$ is 1 if animal j is dominant, and 0 otherwise. Additionally, the dominance category of an animal can be described through a liability model:

$$l_{ij}= other +c_j+e_{ij}$$

where $l_{ijk}$, is an unobserved liability. Assignment of liabilities to categories can involve either fixed or variable thresholds. Thresholds that are independent of pen effect can result in multiple occurrences of "dominant" animals per pen. If there is only one dominant animal per pen, thresholds need to be adjusted for every pen to result in the desired decomposition of categories. If the dominance status of all animals is known, implementation of the proposed models can be achieved through a modified linear-threshold model. If the dominance status is not known, an additional step is needed where the dominance status is inferred using the observed data via a Bayesian MCMC approach. The social dominance model that assumes a categorical expression can allow for a more realistic expression of social dominance for animals housed in pens.

**Key Words:** Social Dominance, Competition

**417 Comparison of two methods for computing approximated accuracies for growth traits in random regression models.** J. P. Sanchez*[1,2], I. Misztal[1], and J. K. Bertrand[1], [1]*University of Georgia*, *Athens*, [2]*University of Leon*, *Leon, Spain*.

The random regression model (RR) fitted direct genetic and permanent environmental effects, maternal genetic and maternal environmental effects, and the residual variance, all using linear splines. Knots were set at 1, 205 and 365 d. Approximate accuracies were obtained by a method specifically designed for RR by Tier and Meyer (2004) (M1),

and by a method developed for multiple-trait models by Strabel and Misztal (2001) (accf90 program; M2). In both methods the only fixed effect considered was the contemporary group (CG). D1 was a simulated data set containing 30,250 animals, 5,000(250) base dams(sires) and 5 generations (G) of 5,000 animals, 3 measurements (M) per animal, distributed in 15 CG (GxM combination). The sex rate in non base animals was 1 M:4 F. D2 was a data set of 1,812,871 records of Gelbvieh animals in 199,168 CG. Initial analyses involved D1, where the exact accuracies were also computed by inversion (M3). Regression coefficient, intercept and R2 of direct accuracies when M1 was regressed on M3 at 205 days (Weaning Weight) for males were 0.95, 0.03 and 0.99. The same quantities of M2 on M3 were 0.96, 0.04 and 0.99, and of M1 on M2 were 0.98, 0.00 and 0.98. The corresponding numbers for females were 1.06, -0.01, 0.99 (M1 on M3), 1.09, 0.00, 0.99 (M2 on M3), and 0.98, 0.00, 1.00 (M1 on M2). For maternal effects both methods showed similar performance and errors. Both M1 and M2 overestimated accuracies for base dams with many offspring. Using data set D2 without distinction of sex and retaining only animals with computed accuracy >= .6 for M1, similar statistics for M1 on M2 were 1.05, -0.06, 0.97 (1.13, -0.14, .96) for direct (maternal) effects. Computing requirements for D2 were 8 CPU min. and 878 Mb. of RAM with M1 and 7 min. and 326 Mb. with M2. A multiple trait accuracy algorithm is useful for computing accuracy of a RR linear spline model when there is at most one observation per trait and no interest on other ages than those defining traits.

**Key Words:** Accuracy, Beef Cattle, Linear Splines

**418 Equivalent mixed model equations for genomic selection.** D. J. Garrick*, *Colorado State University*, *Fort Collins*.

Henderson's mixed model equations are used for genetic evaluation from pedigree and performance information. Equations are solved for factors influencing a trait (eg direct genetic, maternal genetic, maternal permanent environment). Assumptions to obtain BLUP include the var-cov matrix between effects on one animal (eg G0) and the var-cov matrix of additive effects between animals (numerator relationship or A matrix). These var-cov matrices are commonly full rank, and their inverses are used in computations. Widespread implementation has occurred because A-1 is sparse and can be easily made directly from pedigrees. Genetic merit is assumed to result from an infinitesimal number of genes of small effects. An incidence matrix is used to relate genetic merit to phenotypes. This indicator matrix contains one or more unit elements that identify the animals own additive and any other effects. The number of equations increases each evaluation, in proportion to the number of new animals. Genomics has delivered an apparently very different approach to selection. Genetic merit can be considered the finite sum of perhaps tens of thousands of effects, physically located at some place on the genome whose transmission can be traced through genetic markers or haplotypes. Genomic selection might involve mixed model equations that ignore animal effects but include haplotype effects. Pedigree relationships are not necessarily required, the dense markers being used to trace identity by descent (IBD) at each locus and these IBD probabilities being used to construct incidence matrices. Such equations would not increase according to the number of new animals added over time, only the number of new markers or haplotypes. Total genomic merit of candidates would be obtained by summing up many relevant haplotype effects.

An equivalent model can be written that does not explicitly fit haplotype effects but total genomic effects for each animal. This demonstrates the similarity between total genomic selection and conventional A matrix evaluation. The animal-based formulation may be computationally attractive in the short-term when there are more haplotype effects than animals with markers.

**Key Words:** Genetic Evaluation, Equivalent Models

**419 Detection and use of single gene effects in large animal populations.** N. Gengler*[1,2], S. Abras[1], M. Szydlowski[1], and R. Renaville[1], [1]*Gembloux Agricultural University*, *Gembloux, Belgium*, [2]*National Fund for Scientific Research*, *Brussels, Belgium*.

Unbiased estimation of single gene effects can only be achieved by estimating them simultaneously with other environmental and polygenic effects in mixed inheritance models. As in large animal populations the vast majority of animals are however not genotyped, missing genotypes have to be estimated. Currently used methods as iterative peeling or MCMC are unpractical for large datasets. Recently an alternative method to estimate missing gene content, defined as the number of copies of a particular allele was developed. Unknown gene content is approximated from known genotypes based on the additive relationships between animals. In this study the proposed method was tested for the detection of candidate gene effects for bovine transmembrane GHR on first lactation milk, fat and protein test-day yields in Holsteins. The GHR gene was estimated to show moderate to small gene substitution effects of 295 g/day for milk, -8.14 g/day for fat yield and -1.83 g/day for protein yield for a phenylalanine replacement by a tyrosine (frequency 23.3%). Only 961 mostly recent sires out of 2,755,041 animals were genotyped. The accuracy of the procedure was then estimated by doing 15 simulations using gene dropping and adjustment of the observed 12,858,741 records using the estimated parameters. The new method to estimate missing gene content resulted to be functional and accurate as relative bias in the estimation of allele frequency was very low (0.2%) as were the biases for moderate allele substitution effects (milk: 3.7%; fat yield: 3.3%). Biases were larger for traits with smaller substitution effects (protein yield: 55.3%). The new method has the potential to allow even in very large animal population with few genotyped animals reliable estimation and use of moderate to large single gene effects.

**Key Words:** Single Gene Effects, Large Population, Estimation of Gene Content