260 Genome wide selection: Potential and pitfalls. B. Hayes* and M. Goddard, *University of Melbourne*, *Attwood, Victoria, Australia*.

Genomic selection exploits dense marker information to calculate genomic breeding values (GEBV) as the sum of the effects of all quantitative trait loci (QTL) across the genome, thereby potentially exploiting all the genetic variance for a trait. The QTL effects, either inferred from haplotypes or individual single nucleotide polymorphism (SNP) markers, are first estimated in a large reference population with phenotypic information. In subsequent generations, only marker information is required to calculate GEBV. Computer simulation suggested that accuracies of GEBV calculated for animals with marker information only could be as high as 0.85. Results from analysis of real data from the Australian Holstein Friesian population suggest that to achieve this level of accuracy, particularly for young animals with no performance records, a number of challenges must be met. In order for single SNP or haplotype effects used in calculating the GEBV to persist across the population and across generations, the SNP or haplotypes must be in high linkage disequilibrium (LD) with the QTL. This requires adequate marker density and methods for estimating QTL effects which distinguish between contributions from LD as opposed to pedigree tracing or linkage. Another challenge is recent reports in both livestock and human whole genome association studies suggest even the largest QTL effects on some traits may only account for 1-2% of the phenotypic variance. The implications of such a distribution of QTL effects is that a large number animals will be required in the reference population. Finally, ensuring long term genetic gain with genomic selection will be challenging. Particularly if the marker density is not high enough to ensure all QTL are captured, long term gain from genomic selection may be less than from a traditional BLUP based breeding. One solution to this problem is to include a polygenic component in the GEBV. Breeding programs based on genomic selection do however offer the prospect of reducing the rate of inbreeding, as greater emphasis will be placed on individual selection as opposed to family selection, particularly for traits of low heritability.

Key Words: Genome Wide Selection, Genomic Breeding Values, GEBV Reliability

261 Reliability of genomic predictions for North American dairy bulls. P. M. VanRaden*[1], C. P. Van Tassell[1,2], G. R. Wiggans[1], T. S. Sonstegard[2], R. D. Schnabel[3], and F. Schenkel[4], [1]*USDA Animal Improvement Programs Laboratory*, *Beltsville, MD*, [2]*Bovine Functional Genomics Laboratory*, *Beltsville, MD*, [3]*University of Missouri*, *Columbia*, [4]*University of Guelph*, *Guelph, ON Canada*.

Genetic progress will increase when breeders examine genotypes instead of only pedigrees and phenotypes. Genotypes for 39,835 markers and August 2003 genetic evaluations for 2609 Holstein bulls born before 1998 were used to predict January 2008 daughter deviations for 510 bulls born 2001-2. Genotypes were from the Illumina Bovine SNP50™ chip and semen contributed by U.S. and Canadian AI organizations to the Cooperative Dairy DNA Repository. Genomic predictions for 5 yield traits, 5 fitness traits, 16 conformation traits, and net merit were computed by a linear model with an assumed normal distribution for marker effects and also a nonlinear model with a heavier tailed prior to account for major genes. The official parent average from 2003 and a 2003 parent average computed from only the subset of genotyped ancestors were combined with the genomic predictions by selection index. The combined predictions were significantly ($P < .0001$) more accurate than official parent averages for all 27 traits. Squared correlations were 0.03 to 0.22 higher with linear genomic predictions included than those from parent average alone. Nonlinear genomic predictions had R-square similar to linear except for a few traits such as fat percentage, with maximum additional increase of 0.08. Squared correlations were converted to realized reliabilities by dividing by the average reliability of 2008 daughter deviations and by adding the difference between published and observed reliabilities of the 2003 parent averages. When averaged across all traits, combined genomic predictions had realized reliabilities 18% higher than reliabilities of parent average (48% vs. 30%), and gains in information were equivalent to 9 daughter records. Reliability increased more by doubling the number of bulls genotyped than the number of markers genotyped. Genomic selection can decrease generation interval and greatly increase accuracy by tracing the inheritance of minor genes.

Key Words: Genotype, Genomic Selection, Marker Assisted Selection

262 Data optimization techniques for large phenotypic and molecular data sets. R. Rekaya*, *University of Georgia*, *Athens*.

Recent advances in high throughput technologies have lead to the generation of massive amounts of genomic/molecular data in addition to the already extensive phenotypic recording. Over the years several creative statistical and computational optimization techniques have been developed to successfully analyze phenotypic data; achieving some success in the area of quantitative genomics. However, for the ultimate objective of marker assisted or genomic selection to be realized, serious challenges must still be met in order to optimally combine these two sources of information. The complexity of this task stems in part from the huge imbalance of genomic and phenotypic information in selection candidates. In fact, genomic information, although extensive at an individual level, is only available on a small number of animals. By contrast, the phenotypic information is limited at the individual level, but it is extensively recorded on a large number of animals. This situation creates two major problems: 1) an NxP problem in which the number of parameters to be estimated is far greater than the degrees of freedom available to estimate them and 2) inferring the genotypes of untyped individuals and the selective sampling of animals to be genotyped. Theoretically both problems are simple to solve; however, they are computationally intractable. Consequently, some dimension reduction techniques and/or feature selection procedures need to be developed and implemented prior to analysis. In this study a machine learning procedure, often implemented to select optimal subsets of features, was developed to deal with both problems. Ant colony algorithm (ACA) was developed to efficiently search large sample spaces for optimal solutions, making it ideal for applications on high dimension data sets. This algorithm is based on the natural process of communication between real ants, and utilizes simple units to perform complex tasks. When used in applications to disease diagnosis, features identified by ACA yielded increases in prediction accuracy from 13.94 to 43.97% over filter methods. For optimal genotyping strategies, ACA yielded at least 5% superiority compared to any other method.

Key Words: Ant Colony, Feature Selection

263 The next steps in genomic selection: An industry perspective. J. P. Chesnais*[1], F. Schenkel[2], and N. Caron[1], [1]*Semex Alliance*, *Guelph, ON,Canada*, [2]*University of Guelph*, *Guelph, ON,Canada*.

Recent results from genomic selection projects indicate that significant increases in the accuracy of evaluation of young dairy animals can be obtained by combining genomic data from high-throughput SNP panels with traditional pedigree data. In a recent project based on a 50K SNP panel, the reliability of prediction increased by an average of 18 points over parent average accross all traits. As expected, the increase is proportionally larger for traits with low heritability such as fertility or longevity. These results are still far from those of several simulation studies. In theory, the proportion of genetic variance explained by the SNP markers does not have to be very high to justify moving from a traditional progeny-testing scheme to one where animals are selected at a younger age based on their SNP genotype. The ensuing reduction in the generation interval more than compensates for the decrease in selection accuracy. In practice, however, dairy producers are accustomed to proven bull reliabilities that are significantly higher than those currently obtained with genomic selection. Dairy cattle breeding companies are therefore likely to first use genomic selection to pre-select bull dams and young bulls rather than as a replacement for organized progeny testing. Over time, as the number of genotyped animals and SNP density increase, and methods of data analysis are refined, reliability may improve. An increasing number of young genotyped bulls will then be used as sires of sons and to commercially breed cows. The use of genomic selection is not necessarily synonymous with an increase in homozygosity. First, large SNP panels can be used to provide better estimates of homozygosity than current measures of inbreeding. Second, genomic selection may lead to the selection of more genetically diverse candidates than BLUP selection based only on phenotypes. Finally, the use of optimum genetic contribution methods may become more common among breeding companies, which would prevent the selection of only very few animals with top genomic breeding values as parents of future generations.

Key Words: Genomic Selection, Dairy Cattle, Genetics