Use of Gene Markers for Genetic Selection of Dairy Cattle

John B. Cole, Ph.D.
Animal Improvement Programs Laboratory
Agricultural Research Service, USDA
10300 Baltimore Avenue
Beltsville, MD 20705-2350
Telephone: +1-301-504-8334
Fax: +1-301-504-8092
E-mail: john.cole@ars.usda.gov

Keywords: best linear unbiased prediction, dairy cattle, genetic evaluation, genomic selection, marker assisted selection, quantitative genetics, single nucleotide polymorphisms

ABSTRACT

Marker-assisted selection (MAS) is a tool for the genetic improvement of livestock based on linkage relationships between markers and quantitative trait loci (QTL). The use of MAS in livestock populations has had only limited success historically because of high genotyping costs and difficulties associated with marker validation. High-speed genotyping of large numbers of single nucleotide polymorphisms (SNP) became affordable for dairy cattle late in 2007, which has allowed development of genomic selection programs. In addition to increasing rates of genetic improvement and reducing costs of progeny testing, genomic evaluations produce estimates of the contributions of individual markers to additive genetic merit. Most traits of economic importance in dairy cattle are controlled by many genes, each of small effect, which means that a large amount of data is needed to estimate the effects accurately. Dense markers are necessary to ensure that associations between markers and QTL persists across families, and permit accurate imputation of parental haplotypes.

INTRODUCTION

Marker-assisted selection exploits linkage relationships between markers and QTL to provide genomic information for traits of interest. High-speed genotyping of large numbers of DNA markers became affordable for dairy cattle late in 2007, which permitted the development of genomic selection programs as originally described by Nejati-Javaremi et al. (1997). In addition to increasing rates of genetic improvement and reducing costs of progeny testing (Schaeffer, 2006), genomic evaluations produce estimates of the contributions of individual markers to additive genetic merit. Earlier studies involved at most a few hundred markers, while current studies include tens- or hundreds-of-thousands of markers, the results of which must be evaluated statistically rather than by individual inspection. The objective of this chapter is to describe the implementation of MAS in dairy cattle populations using systems which combine genotypic, phenotypic, and pedigree data to increase the accuracy of estimates of genetic merit and decrease generation interval.

BODY OF TEXT

GENETIC MARKERS

From Sax's (1923) demonstration almost a century ago that it is possible to identify the effect of individual loci on quantitative traits, geneticists have searched for markers linked to traits of scientific interest or economic importance. Animal blood groups were used for parentage verification for many years, as well as for some early QTL mapping attempts. Restriction fragment length polymorphisms, first described in the early 1970s, were successfully used to identify many QTL in plants, but were often found to be monomorphic in animals. Following the discovery of the polymerase chain reaction in 1986 and the rapid development of related technologies, microsatellite markers became the marker of choice in livestock studies (e.g., Ashwell et al., 2004). However, microsatellites are relatively slow and expensive to genotype, limiting the number available for QTL studies. Interested readers are referred to Weller (2001) for a detailed discussion of early work with genetic markers.

The ideal genetic marker is distributed uniformly across the genome, may be rapidly genotyped, and is inexpensive. Following the completion of a draft assembly of the bovine genome in 2009, tens-of-thousands of SNP were discovered. These biallelic (having only 2 states) markers can be used to track the inheritance of short chromosomal segments. If enough data are available then the effect of each SNP on a phenotype of interest may be predicted with high accuracy. The dairy cattle population is ideally suited for this analysis because millions of phenotypes are available for traits of interest, pedigrees extend back many generations, individual animals are valuable, and historical collections of DNA are available.

GENOMIC SELECTION

Development of SNP Chips

After the bovine genome was sequenced, an international consortium of government, university, and industry cooperators was assembled to work with Illumina (San Diego, CA) to develop a set of SNP to be included on a genotyping chip. The resulting set of 54,001 SNP was included in the original release

of the Illumina BovineSNP50 BeadChip (Matukumalli et al., 2009), which became publicly available in December 2007. Genotypes from the Beltsville Agriculture Research Center (MD), University of Missouri (Columbia), and University of Alberta (Edmonton, AB, Canada) were used to identify SNP suitable for inclusion in genomic evaluation. Markers were excluded for many reasons, such as low call rates (genotypes frequently could not be determined), low minor allele frequencies (variants occurred too rarely), or high correlations with adjacent SNP (added little new information). Procedures for using SNP to detect errors in pedigrees also were developed. In many cases where parents are shown by genotyping to be incorrect, the correct parents can be identified.

Several additional SNP chips currently are in use, including: a second version of the BovineSNP50, 3K and 6K low-density (LD) chips, and a high-density (HD) chip from Illumina; the GeneSeek Genomic Profiler, a customized version of the 6K Illumina LD chip, developed by GeneSeek (Lincoln, NE); and 10K and 25K chips from Affymetrix (Santa Clara , CA). Most bulls that produce semen for sale by major artificial insemination (AI) firms have BovineSNP50 genotypes, while most cows are genotyped using an LD chip. Low-density genotypes may be imputed to 50K and to HD genotypes with a high degree of accuracy (VanRaden et al., 2011). Contrary to expectations, recent research has found greater reliability gains from the use of additional predictor animals than from additional markers provided by HD genotpyes.

Estimation of Genomic Predicted Transmitting Abilities

For many years, dairy cattle genetic improvement programs have been based on the mixed model methodology developed by Henderson (1984). In a simple mixed model analysis, a phenotype is modeled as a function of fixed (e.g., sex) and random (e.g., genetic) effects:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where **y** is a vector of phenotypes; **X** and **Z** are design matrices relating observations to levels of fixed and random effects, respectively; **b** is a vector of values for fixed effects; **u** is a vector of random

88 animal breeding values; and **e** is a vector of residual error effects. Under an infinitesimal alleles model
89 of inheritance Var(**u**) = $\mathbf{A}\sigma^2_a$, where **A** is the numerator relationship matrix that describes the genetic
90 (co)variances among the animals in the population, and is the additive genetic variance σ^2_a.
91
92 Allele effects are estimated using phenotypic and genotypic information from a predictor population of
93 animals with high-reliability predicted transmitting abilities (PTA) for the traits of interest. Those PTA
94 are used to predict the marker effects and do not include genomic information, avoiding part-whole
95 confounding, and may be used to calculate genomic evaluations for animals without traditional
96 evaluations (VanRaden, 2008). When the first official US genomic evaluations were released in January
97 2009, the predictor population included Holstein bulls born primarily between 1995 and 1997, and the
98 resulting SNP effects were used to produce genomic PTA for younger animals. Models with heavy-
99 tailed distributions of SNP effects accommodate traits on which one or a few genes have large effects
100 (Cole et al., 2009), such as fat yield (Figure 1). Marker effects must be periodically re-estimated, so
101 phenotypes must be collected continuously.
102
103 Final genomic predictions combine 3 terms by selection index: 1) direct genomic prediction; 2) parent
104 average (PA) or PTA computed from genotyped ancestors using traditional relationships; and 3)
105 published PA or PTA. For each animal, a 3 × 3 matrix is set up with reliabilities for the 3 terms on the
106 diagonals and functions of those 3 reliabilities on the off-diagonals. Reliabilities of direct genomic
107 predictions are approximated using the genomic relation of each animal to the predictor population and
108 the reliability of the predictor evaluation. Misztal et al. (2009) have developed an alternative approach
109 for combining all information in a single-step. Their approach is based on replacing **A** with a matrix
110 including both pedigree and genotypic information, which addresses the problem of genomic pre-
111 selection of young bulls, a source of bias in genomic evaluations.
112 **Adoption by the Dairy Industry**

Genomic selection was rapidly adopted by the US dairy cattle industry, with AI firms providing much of the DNA and money needed to genotype thousands of animals. Additional funding for genotyping came from research grants and individual breed organizations. In return for their support, the AI organizations received the exclusive right to have males genomically evaluated until May 2013. Much of the initial genotyping was performed by the Bovine Functional Genomics Laboratory (ARS, USDA, Beltsville, MD), with some assistance from the commercial genotyping laboratories GeneSeek (Lincoln, NE) and Genetics & IVF Institute (Fairfax, VA). Several other commercial laboratories now provide sample processing and genotyping services, including DNA LandMarks (Saint-Jean-sur-Richelieu, QC, Canada), Genetic Visions (Middleton, WI), and Expression Analysis (Durham, NC).

While the initial genotyping was underway, VanRaden (2008) used simulated data to develop and test a genomic evaluation system that was able to process real genotypes as soon as they became available in December 2007. The first unofficial USDA evaluations based on SNP genotypes were released to the industry in April 2008. Genomic evaluations became official for Holsteins and Jerseys in January 2009 and for Brown Swiss in August 2009. Many other countries also have adopted genomic evaluations systems (Loberg and Dürr, 2009). including Australia, Britain, Canada, Italy, New Zealand, and Switzerland. Canada, Great Britain Italy, and the US currently exchange genotypes, increasing the size of each country's predictor population. A cooperative endeavor called EuroGenomics was established in 2009 by 5 European breeder-owned companies that represent Belgium, Denmark, Finland, France, Germany, the Netherlands, Spain, and Sweden.

CONCLUSIONS

The adoption of genetic evaluations including genomic information by the global dairy genetics industry has been very rapid. Results shows that the genomic predictions for young animals are very accurate, maintaining selection intensity, and that young bulls are being used extensively, reducing the generation interval. As the number of genotyped cows continues to increase, the demand for improved

mating tools will increase. Small breeds which do not have large populations for predicting SNP effects will have a challenge in maintaining commercial viability against breeds that do.

The Animal Improvement Programs Laboratory (ARS, USDA, Beltsville, MD) provides an extensive collection of information including genetic evaluation results, peer-reviewed publications, and educational materials on its website (http://aipl.arsusda.gov/). The International Bull Evaluation Service (Uppsala, Sweden) publishes lots of information related to genetic evaluation programs in member countries on its website (http://www.interbull.org/) and in the Interbull Bulletin.

REFERENCES

Ashwell, M.S.; Heyen, D.W.; Sonstegard, T.S.; Van Tassell, C.P., Da, Y.; VanRaden, P.M.; Ron, M.; Weller, J.I.; Lewin, H.A. Detection of quantitative trait loci affecting milk production, health, and reproduction traits in Holstein cattle. J. Dairy Sci. **2004,** 87, 468–475.

Cole, J.B.; VanRaden, P.M.; O'Connell, J.R.; Van Tassell, C.P.; Sonstegard, T.S.; Schnabel, R.D.; Taylor, J.F.; Wiggans, G.R. Distribution and location of genetic effects for dairy traits. J. Dairy Sci. **2009,** 92, 2931–2946.

Henderson, C.R. *Applications of Linear Models in Animal Breeding*; Univ. Guelph; Ontario, Canada, 1984.

Loberg, A.; Dürr, J. Interbull survey on the use of genomic information. Interbull Bull. **2009,** 39, 3-14.

Matukumalli, L.K.; Lawley, C.T.; Schnabel, R.D.; Taylor, J.F.; Allan, M.F.; Heaton, M.P.; O'Connell, J.R.; Moore, S.S.; Smith, T.P.L.; Sonstegard, T.S.; Van Tassell, C.P. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE **2009,** 4, e5350, doi:10.1371/journal.pone.0005350.

Misztal, I.; Legarra, A.; Aguilar, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. **2009,** 92, 4648–4655.

Nejati-Javaremi, A.; Smith, C.; Gibson, J.P. Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Animal Sci. 1997, 75, 1738–1745.

Sax, K. The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics **1923,** 8, 552–560.

Schaeffer, L.R. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. **2006,** 123, 218–223.

VanRaden, P.M. Efficient methods to compute genomic predictions. J. Dairy Sci. **2008,** 91, 4414–4423.

VanRaden, P.M.; O'Connell, J.R.; Wiggans, G.R.; Weigel, K.A. Genomic evaluations with many more genotypes. Genet. Sel. Evol. **2011,** 43, 10.

Weller, J.I. *Quantitative Trait Loci Analysis in Animals*; CABI Publishing; New York, NY, USA, 2001.

FIGURE CAPTIONS

Figure 1. Marker effects in additive genetic standard deviations (SD) of 45,187 single nucleotide polymorphisms for fat yield in US Holsteins (P = pseudo-autosomal region of the X chromosome; X = remainder of the X chromosome).