Distribution and location of genetic effects for dairy traits

J.B. Cole[1], P.M. VanRaden[1], J.R. O'Connell[3], C.P. Van Tassell[1,2], T.S. Sonstegard[2], R.D. Schnabel[4], J.F. Taylor[4] & G.R. Wiggans[1]

[1]Animal Improvement Programs and [2]Bovine Functional Genomics Laboratories, USDA Agricultural Research Service, Beltsville, MD, USA 20705-2350
[3]University of Maryland School of Medicine, Baltimore, MD, USA
[4]University of Missouri, Columbia, MO, USA

Summary

A nonlinear model with heavy tails allows markers with large effects to be regressed away from the mean and regresses markers with small effects towards zero, reflecting the biology of traits such as DGAT. These models provide better fits to the data than models that allow only a finite number of loci to have non-zero effects. No advantage was gained by including interactions between heavy tails and the number of non-zero marker effects. Linear model predictions were intermediate to the heavy-tailed and finite loci models.

Markers on BTA 18 centered on SNP ARS-BFGL-NGS-109285 had large effects on economic merit, longevity, calving ease, and conformation. Relationships among those traits may be attributable to a gene product or regulatory element associated with calf birth weight. The presence of significant marker effects resulted in greater-than-expected proportions of explained genetic variance.

***Key words**: Genomic data, Non-linear model, Linear and non-linear predictions, X chromosome, Marker effects.*

Introduction

The high-speed genotyping of large numbers of single nucleotide polymorphisms (SNP) has recently become affordable for dairy cattle, which has allowed development of genomic selection programs. A number of questions must be answered before genomic evaluation becomes routine, including the optimal choice of models of gene action, the size, location, and distribution of marker effects, and the proper treatment of X chromosome effects.

Marker locations and effects can be used to assess alternative models of gene action, and to identify chromosomal segments of interest for functional genomic study. Different prior assumptions about the distribution of marker effects correspond to varying models of gene action. Markers with large effects on traits of economic importance may be used to identify regions of the genome that merit further study.

This paper describes the effect of different priors on the reliability of genomic predictions, reports the location and size of a marker on BTA 18 that is associated with dystocia, conformation, economic merit, and longevity, and describes how to account for effects on the X chromosome.

Genomic data

Genotypes for 39 314 SNP of 5 360 Holsteins were examined. The selected SNP were from the Illumina Bovine SNP50™ chip (Van Tassell *et al*., 2008) and had minor allele frequencies greater than 5% in Holsteins. Genotyping and DNA extraction was done at six locations: Bovine Functional Genomics Laboratory, University of Missouri, University of Alberta, Geneseek, GIVF, and Illumina. Cooperating AI organizations in North America contributed the DNA.

Prior assumptions about the distribution of marker effects were tested using historical data from August 2003 for 3 576 bulls born before 1999 to predict current data for 1 759 bulls born 1999-2002. Advantages of various models were tested using weighted regressions of current daughter deviations on traditional and genomic evaluations computed from 2003 data.

Reported marker locations with largest effects used official April 2008 evaluations for 5 285 proven bulls and 75 cows with records.

The genomewide association method of Aulchenko *et al*. (2007) was also used to calculate marker effects for data visualization.

Linear and nonlinear predictions

Predictions were computed by linear and nonlinear genomic models (VanRaden, 2008). For linear predictions, the traditional additive genetic relationship matrix is replaced by a genomic relationship matrix and is equivalent to assigning equal genetic variance to all markers. Differing assumptions about numbers and sizes of QTL effects can result in better or poorer predictions, but may be appropriate in some cases, such as when a single gene is known to have a large effect on a trait. Three different nonlinear models were considered: an infinitesimal alleles model with a heavy-tailed prior in which smaller effects are regressed further toward 0 and markers with larger effects are regressed less to account for a non-normal prior distribution of marker effects (model A), a finite locus model with a normal distribution of marker effects (model B), and a finite locus model with heavy-tails (model AB). Infinitesimal alleles models assume that all loci have non-zero effects, and finite loci models assume that only a fixed number of alleles have effects. Models A and AB are analogous to the Bayes A and B methods of Meuwissen *et al*. (2001), respectively.

Model A had little advantage in R^2 over the linear model except for fat and protein percentages with increases of 8% and 7%, respectively (Table 1). Gains obtained in simulation averaged 3% but were mostly smaller with real data, indicating that most traits are influenced by more loci than the 100 QTL used in simulation (VanRaden, 2008).

Model B provided similar or poorer fits than Model A for all traits assuming 5 000, 10 000, or 20 000 loci with non-zero effects. Model AB produced better R^2 for the percentage traits than model B, which provided results similar to those of the linear model, but provided similar results to model A.

With the exception of fat and protein percentages, for which there are known genes of large effect (Grisart *et al*., 2004; Zinder *et al*., 2005), models assuming that all markers have some effect rather than that most have no effect provided better R^2.Slight decreases in R^2 were noted for most traits in model A when the variance beyond 2 SD was increased (data not shown).

Largest effects

The largest marker effects were for fat percentage on BTA 14 flanking the DGAT1 gene (Grisart *et al*., 2004), with lesser effects on both milk and fat yield. Large marker effects for protein percentage were also present on BTA 6 flanking the ABCG2 gene (Cohen-Zinder *et al*., 2005). This demonstrates that the genomic predictions work by tracking the inheritance of causal mutations. Markers on BTA 18 centered on ARS-BFGL-NGS-109285 had the largest effects for several traits: productive life, sire calving ease, daughter calving ease, rump width, stature, strength, and body depth (Figure 1). Another marker on BTA 18 had the largest effect of any on net merit, in the region previously identified by Ashwell *et al*. (2004) as having a large effect on daughter pregnancy rate. This marker had a greater effect on economic merit than DGAT (Figure 2).

Correlations (r) among chromosome 18-specific EBV (data not shown) reveal favorable correlations among longevity and economic merit (r = 0.88); undesirable correlations among conformation and calving ease traits (r = 0.78 to 0.95); and unfavorable correlations among economic merit and longevity and conformation and calving ease (r = -0.44 to -0.72). Selection for extreme conformation (larger body size) has resulting in larger calves and increased rates of dystocia which are largely attributable to feto-pelvic incompatibility (Meijering, 1984). The increase in calf size is apparently not offset by increasing internal pelvic size of cows.

The expected proportion of genetic variance for each trait accounted for by SNP on a chromosome was calculated based on chromosome lengths assuming that all markers had equal effects. Chromosome 18 was expected to account for 2.2% of genetic variance for each trait, but actually accounted for 2.9% (economic merit) to 7.6% (sire calving ease) of genetic variation.

Table 1. Squared correlations for sire predicted transmitting abilities from predictions using different numbers of QTL and prior distributions of QTL sizes.

	Model[1]			
Trait	Linear	A	B	AB
Net Merit	28.2	28.4	27.6	27.6
Milk	47.2	48.5	46.7	47.3
Fat	41.8	44.2	41.5	43.6
Protein	47.5	47.0	46.8	46.6
Fat %	55.3	63.3	57.5	63.9
Protein %	51.4	57.7	51.4	56.6
Longevity	25.6	27.4	25.4	26.4
Somatic cell	37.3	38.3	37.3	37.6
Days open	29.5	29.0	29.4	29.2

[1]Linear = linear model; A = heavy tails model with a priorof 1.12; B = finite alleles model with 20 000 markers; AB = finite alleles model with 20000 markers and a heavy tails prior of 1.08.

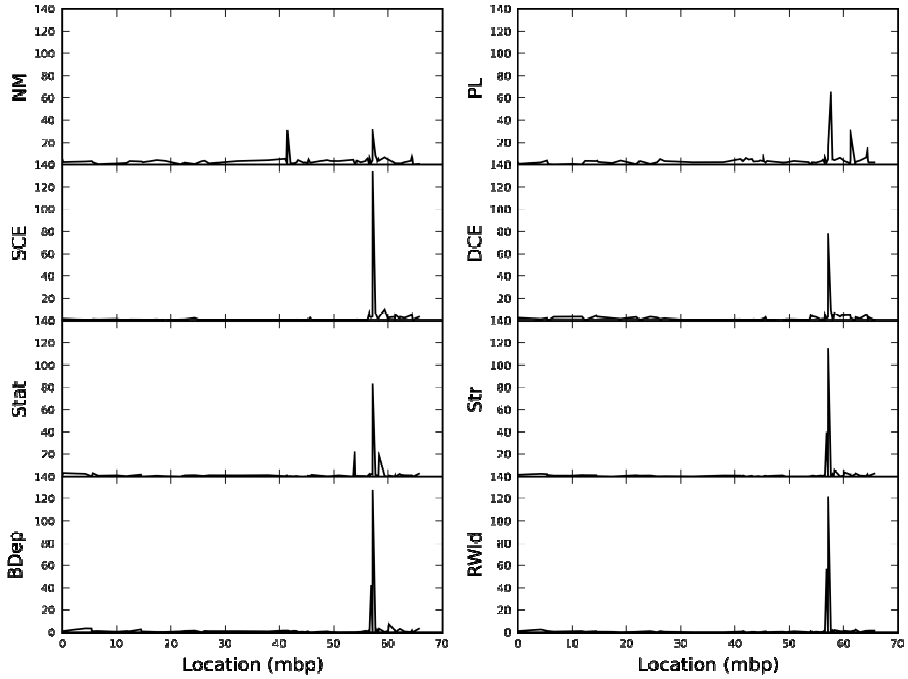

Figure 1. Size (in SD) and location of marker effects on BTA 18 affecting net merit (NM), longevity (PL), sire (SCE) and maternal calving ease (DCE), stature (Stat), strength (Str), body depth (BDep), and rump width (Rwid).

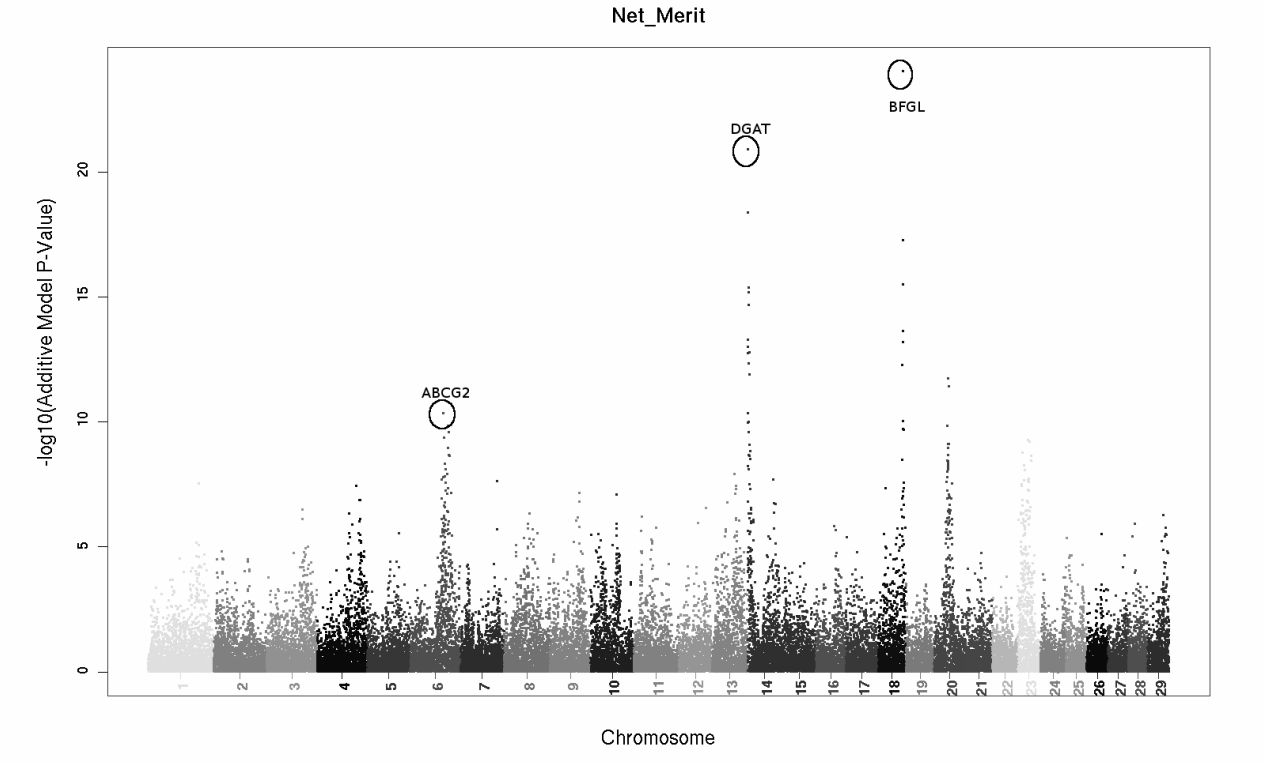

Figure 2. Genome-wide significance of marker effects associated with net merit.

Marker effects for most other traits were evenly distributed across all chromosomes, with only a few regions having larger effects. This may explain why the infinitesimal model and standard quantitative genetic theories have worked well. The distribution of marker effects indicates that favorable alleles will not become homozygous quickly, and genetic variation will remain even after intense selection. Thus, dairy cattle breeders may expect genetic progress to continue for many generations.

X chromosome

The X chromosome of a bull is inherited by all of his daughters but by none of his sons. Thus, two estimates of his genetic merit can be provided: EBV for his daughters is the sum of all marker effects, whereas EBV for his sons excludes effects of 487 markers on the X chromosome. Thirty five markers located on the pseudo-autosomal region were included in the autosomal sum. There are fewer identified SNP on the X chromosome, and markers are spaced more widely than on the autosomes.

Females can also have differing EBV for daughters than sons. Effects on the X are doubled for producing sons because the X transmitted to sons will be transmitted to 50% of the granddaughters instead of the 25% expected for autosomes. Differences between EBV from daughters and average of sons' EBV from 796 genotyped sires that had at least 10 evaluated sons were used to test if effects for net merit on the X chromosome were real. Effects were small but had significant (P<0.0001) associations with differences between the genetic merits of a bull's sons and his daughters.

Acknowledgements

This project was supported by National Research Initiative Grant nos. 2006-35205-16888 and 2006-35205-16701 from the USDA Cooperative State Research, Education, and Extension Service and by the National Association of Animal Breeders. The computational assistance provided by M. E. Tooker is gratefully acknowledged.

List of References

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

Aulchenko, Y. S., D.-J. De Koning & C. Haley. 2007. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177: 577–585.

Cohen-Zinder, M., E. Seroussi, D.M. Larkin, J.J. Loor, A. Everts-van der Wind, J.H. Lee, J.K. Drackley, M.R. Band, A.G. Hernandez, M. Shani, H.A. Lewin, J.I. Weller & M. Ron. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 15: 936–944.

Grisart B., F. Farnir, L. Karim, N. Cambisano, J.J. Kim, A. Kvasz, M. Mni, P. Simon, J. M. Frere, W. Coppieters & M. Georges. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. U.S.A. 101: 2398–403.

Meijering, A. 1984. Dystocia and stillbirth in cattle – a review of causes, relationships, and implications. Livest. Prod. Sci. 11: 143–177.

Meuwissen, T.H.E., B.J. Hayes & M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

VanRaden, P.M. 2007. Genomic measures of relationship and inbreeding. Interbull Bull. 37: 33–36.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91: (Submitted).

Van Tassell C. P., T.P.L. Smith, L.K. Matukumalli, J.F. Taylor, R.D. Schnabel, C.T. Lawley, C.D. Haudenschild, S.S. Moore, W.C. Warren & T.S. Sonstegard. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods 5: 247–252.