J. Dairy Sci. 90:4821–4829
doi:10.3168/jds.2007-0158

Breeding Value Estimation for Fat Percentage Using Dense Markers on *Bos taurus* Autosome 14

A. P. W. de Roos,*[1] C. Schrooten,* E. Mullaart,* M. P. L. Calus,† and R. F. Veerkamp†
*HG, 6802 EB Arnhem, the Netherlands
†Animal Breeding and Genomics Centre, Animal Sciences Group, Wageningen University and Research Centre, 8200 AB Lelystad, the Netherlands

ABSTRACT

Prediction of breeding values using whole-genome dense marker maps for genomic selection has become feasible with the advances in DNA chip technology and the discovery of thousands of single nucleotide polymorphisms in genome-sequencing projects. The objective of this study was to compare the accuracy of predicted breeding values from genomic selection (GS), selection without genetic marker information (BLUP), and gene-assisted selection (GEN) on real dairy cattle data for 1 chromosome. Estimated breeding values of 1,300 bulls for fat percentage, based on daughter performance records, were obtained from the national genetic evaluation and used as phenotypic data. All bulls were genotyped for 32 genetic markers on chromosome 14, of which 1 marker was the causative mutation in a gene with a large effect on fat percentage. In GS, the data were analyzed with a multiple quantitative trait loci (QTL) model with haplotype effects for each marker bracket and a polygenic effect. Identical-by-descent probabilities based on linkage and linkage disequilibrium information were used to model the covariances between haplotypes. A Bayesian method using Gibbs sampling was used to predict the presence of a putative QTL and the effects of the haplotypes in each marker bracket. In BLUP, the haplotype effects were removed from the model, whereas in GEN, the haplotype effects were replaced by the effect of the genotype at the known causative mutation. The breeding values from the national genetic evaluation were treated as true breeding values because of their high accuracy and were used to compute the accuracy of prediction for GS, BLUP, and GEN. The allele substitution effect for the causative mutation, obtained from GEN, was 0.35% fat. The accuracy of the predicted breeding values for GS (0.75) was as high as for GEN (0.75) and higher than for BLUP (0.51). When some markers close to the QTL were omitted from the model, the accuracy of prediction was only slightly lower, around 0.72. The removal of all markers within 8 cM from the QTL reduced the accuracy to 0.64, which was still much higher than BLUP. It is concluded that, when applied to 1 chromosome and if genetic markers close to the QTL are available, the presented model for GS is as accurate as GEN.
Key words: genomic selection, genetic marker, fat percentage

INTRODUCTION

Molecular genetic selection can lead to much higher genetic gains than conventional quantitative genetic selection, especially for traits with low heritability, phenotypes that are difficult to record, unfavorable genetic correlations, and genotype × environmental interactions (Meuwissen and Goddard, 1996; Dekkers and Hospital, 2002). Animal breeding programs have been using molecular genetic information for many years, but its effect has been less than initially expected (Dekkers, 2004). One of the reasons is the difficulty to find the causal mutations in QTL, or genetic markers that are in high population-wide linkage disequilibrium (**LD**) with a QTL. Many genetic markers that are in population-wide linkage equilibrium or low LD with a QTL have been found, but these are much more difficult to use in molecular genetic selection, because the linkage phase between the marker and the QTL needs to be estimated for each family (Dekkers, 2004).

Advances in DNA chip technology and the discovery of many thousands of single nucleotide polymorphisms (**SNP**) in genome sequencing projects have provided new opportunities to find markers in LD with QTL and to use them for selection (Andersson and Georges, 2004). Haley and Visscher (1998) predicted that the development of cheap and high-density marker maps would move the selection based on polygenes plus individual loci to effective total genomic selection (**GS**). This would greatly improve selection before phenotypic information from the animal or its progeny is available (for example, selection among young bulls before prog-

Received March 1, 2007.
Accepted June 22, 2007.
[1]Corresponding author: roos.s@hg.nl

eny testing). Furthermore, it can enable selection among young animals or embryos, which may dramatically reduce the generation interval. Meuwissen et al. (2001) presented a method to predict breeding values using genome-wide dense marker maps. Using Bayesian statistics, the effects of 50,000 simulated haplotypes were estimated from only 2,200 phenotypic records. After that, the total genetic value of an animal was predicted with an accuracy of 0.85 by summing the estimated effects of the haplotypes of the animal for each marker bracket. This method, GS, attempts to explain all genetic variation by genetic markers without selection of markers that contribute to the genetic variance. It was concluded that GS can substantially increase the rate of genetic gain, especially if combined with reproductive techniques to shorten the generation interval. It can be argued that prediction of breeding values is not the same as making selection decisions, but because GS is the accepted name for the method proposed by Meuwissen et al. (2001), this term is used throughout the paper.

Meuwissen et al. (2001) used the flanking markers of a putative QTL to define a haplotype, which means that all marker brackets that carry the same marker alleles are assumed to have the same effect, whereas in reality they may carry different QTL alleles. Furthermore, they did not include a matrix of identical-by-descent (**IBD**) probabilities between marker brackets, which means that covariances among different haplotypes were assumed to be zero. These assumptions were relaxed in the multiple-QTL mapping method presented by Meuwissen and Goddard (2004), which used the IBD probability matrix among haplotypes as described by Meuwissen and Goddard (2001). The multiple-QTL mapping method has been applied in QTL mapping studies (Olsen et al., 2005) but can also be applied as a method for GS by using a dense marker map with whole-genome coverage.

In the present literature, GS has not been applied to real data. The objective of this study was to validate the method of Meuwissen and Goddard (2004) for prediction of genomic breeding values on a dairy cattle data set for 1 chromosome and to compare the accuracy of prediction to a method without marker information and to a method in which the causative mutation underlying an important QTL is known. Furthermore, the effect of omitting markers close to the QTL on the accuracy of the prediction was analyzed.

MATERIALS AND METHODS

Materials

Data were obtained from a QTL mapping study using a granddaughter design comprising 1,300 progeny-tested Holstein-Friesian bulls born from 1973 to 1994 (Farnir et al., 2002). Twenty-seven grandsires had at least 10 sons, which summed up to 1,135 sons in total, and, on average, 42 sons per grandsire for validation, as explained later. Estimated breeding values for fat percentage, obtained from the official August 2006 genetic evaluation for the Netherlands and Flanders, were used as phenotypic records. All bulls were genotyped for 32 markers on *Bos taurus* autosome 14. The marker set comprised 13 microsatellite markers and 19 SNP markers (Figure 1). The percentage of heterozygous animals was from 41 to 81 (64 on average) for the microsatellite markers and from 0 (for marker 14) to 65 (43 on average) for the SNP markers. Marker 7 was the K232A substitution in the acyl coenzyme A:diacylglycerol acyltransferase 1 (**DGAT1**) gene, which was shown to have a large effect on fat percentage (Grisart et al., 2002; Winter et al., 2002). Eleven grandsires were heterozygous for this marker, whereas 12 grandsires were homozygous for the A allele that was associated with low fat percentage, and 4 grandsires were homozygous for the K allele that was associated with high fat percentage.

The map was constructed based on the bovine composite map (www.livestockgenomics.csiro.au/perl/gbrowse.cgi/bosmap/). The centimorgan position of the markers that were not placed on the composite map were calculated by interpolation using their base pair positions on the National Center for Biotechnology Information bovine sequence map (version 3.1, www.ncbi.nlm.nih.gov) and the base pair position of neighboring markers on the composite map. Figure 1 shows the positions of the SNP and microsatellite markers relative to the causative mutation in the DGAT1 gene. Haplotypes were constructed from the marker genotypes by comparing the genotype of an animal to that of its sire (dams were not genotyped). This was informative in situations when the animal or its sire was homozygous. If both animal and sire were heterozygous but the animal had genotyped offspring, the linkage phase with the closest informative marker was assumed the same as in the majority of the offspring. For example, consider an animal with genotype A/a at locus 1 and B/b at locus 2 and its sire with genotypes A/a and B/B, respectively. For locus 1, it is unclear whether the A or a allele was inherited from the sire, whereas at locus 2, allele B was inherited from the sire. To infer the phase at locus 1, the genotypes of the progeny of the animal were considered: for progeny that were homozygous at both loci, their haplotypes could be determined. If the majority of the progeny of this animal inherited haplotype AB or ab, allele A was assumed paternal, whereas if the majority inherited haplotype aB or Ab, allele a was assumed paternal. Markers with