J. Dairy Sci. 97:1774–1781
http://dx.doi.org/10.3168/jds.2013-7054

Accuracy of direct genomic values for functional traits in Brown Swiss cattle

M. Kramer,*[1] M. Erbe,* F. R. Seefried,† B. Gredler,† B. Bapst,† A. Bieber,‡ and H. Simianer*
*Department of Animal Science, Animal Breeding and Genetics Group, Georg-August-Universität Göttingen, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany
†Qualitas AG, Chamerstraße 56, 6300 Zug, Switzerland
‡FiBL–Research Institute of Organic Agriculture, Ackerstrasse 21, 5070 Frick, Switzerland

ABSTRACT

In this study, direct genomic values for the functional traits general temperament, milking temperament, aggressiveness, rank order in herd, milking speed, udder depth, position of labia, and days to first heat in Brown Swiss dairy cattle were estimated based on ~777,000 (777K) single nucleotide polymorphism (SNP) information from 1,126 animals. Accuracy of direct genomic values was assessed by a 5-fold cross-validation with 10 replicates. Correlations between deregressed proofs and direct genomic values were 0.63 for general temperament, 0.73 for milking temperament, 0.69 for aggressiveness, 0.65 for rank order in herd, 0.69 for milking speed, 0.71 for udder depth, 0.66 for position of labia, and 0.74 for days to first heat. Using the information of ~54,000 (54K) SNP led to only marginal deviations in the observed accuracy. Trying to predict the 20% youngest bulls led to correlations of 0.55, 0.77, 0.73, 0.55, 0.64, 0.59, 0.67, and 0.77, respectively, for the traits listed above. Using a novel method to estimate the accuracy of a direct genomic value (defined as correlation between direct genomic value and true breeding value and accounting for the correlation between direct genomic values and conventional breeding values) revealed accuracies of 0.37, 0.20, 0.19, 0.27, 0.48, 0.45, 0.36, and 0.12, respectively, for the traits listed above. These values are much smaller but probably also more realistic than accuracies based on correlations, given the heritabilities and samples sizes in this study. Annotation of the largest estimated SNP effects revealed 2 candidate genes affecting the traits general temperament and days to first heat.

Key words: direct genomic value prediction, accuracy of direct genomic value, functional trait, gene annotation

INTRODUCTION

Genomic selection is well established in dairy cattle breeding, and accuracies of direct genomic values (**DGV**) for traditional production or conformation traits are high (e.g., Segelke et al., 2012). To maximize the benefit, genomic selection should also be used for scarcely recorded functional traits, which are becoming increasingly relevant in dairy cattle breeding. For this reason, one aim of our study was to apply the estimation of DGV to a set of functional traits. High-density chips with information of more than 777,000 (**777K**) SNP are increasingly available in cattle breeding. We thus compared results from an estimation of DGV with 777K data with the results of estimation of DGV with ~54,000 (**54K**) SNP data. To assess the accuracy of DGV, it is easy to calculate the correlation between DGV and conventional breeding value (EBV), but this correlation is different from the correlation between DGV and true breeding value (**TBV**), which is normally used in practical breeding to reflect the accuracy of a breeding value. To account for this, the correlation between DGV and EBV is often divided by the accuracy of the EBV. As stated by Amer and Banos (2010), this correction leads to an overestimation of accuracy; therefore, we applied a new method proposed by Wellmann et al. (2013) to estimate more realistic accuracies. To date, only few genome-wide association studies for functional traits have been conducted in Brown Swiss (**BS**) cattle (e.g., Guo et al., 2012). Another aim of our work was thus to screen the genome for SNP with large effects on the observed traits to find genes that are associated with these SNP and could influence the respective trait.

MATERIALS AND METHODS

Genotypic Data

The 777K genotypes used in this study were derived by imputation from a data set of 880 BS animals genotyped with the Illumina BovineHD chip (727 cows, 153 bulls) and from a data set of 6,016 animals genotyped with the Illumina Bovine SNP50 chip (548

Received May 23, 2013.
Accepted November 25, 2013.
[1] Corresponding author: mkramer@gwdg.de

cows, 5,468 bulls; both chips from Illumina Inc., San Diego, CA). Before imputation, both data sets were checked for quality separately: SNP with minor allele frequency <0.5%, a call rate <90%, missing position, or position on the sex chromosomes were excluded. In addition, within the 777K data set, mitochondrial SNP and 1 SNP of each of 55 pairs of SNP with identical position but different denomination, respectively, were excluded. Animals with a call rate <90% were excluded from both data sets. After quality checks, imputation was done by using FImpute (Sargolzaei et al., 2011) from 39,004 to 627,306 SNP, using a combination of family and population imputation. This method led to a correlation of 0.99 between true genotype and imputed genotype if both parents were genotyped with the 777K chip and to a correlation of 0.97 between true genotype and imputed genotype if no close relatives were genotyped with the 777K chip (Gredler et al., 2013). We used a subset of 1,126 animals (930 cows from Switzerland with phenotype information for the traits observed and 196 bulls with performance records of at least one daughter) for estimation of DGV and SNP effects in our study.

Phenotypic Data

Deregressed proofs (**DRPF**) for the traits general temperament, milking temperament, aggressiveness, rank order in herd, milking speed, udder depth, position of labia, and days to first heat of the 1,126 BS animals were used as quasi phenotypes. For detailed information on the trait definitions, see Kramer et al. (2013). Although some of these traits are routinely evaluated and multiple-trait across-country evaluation (MACE) proofs would have been available, we decided to use our own breeding values to evaluate the possibilities of genomic estimation based on genotypes and phenotypes of a mixed data set of cows and bulls. The DRPF for individual i within a given trait were derived as described in Garrick et al. (2009):

$$DRPF_i = \frac{EBV_i}{r^2_{EBV\,i}},$$

where $DRPF_i$ is the DRPF of individual i, EBV_i is the EBV for individual i, and r^2_{EBV} is the squared accuracy of the EBV_i for individual i. In our study, EBV of cows with several phenotypic measurements as well as EBV of bulls with a different number of daughters (mean: 3.22; range: 1–31) were used. Distribution of birth year of animals is shown in Figure 1. Ninety-five percent of bulls were born between the years 1988 and 2004, and 95% of the cows were born between the years 1998 and 2007. Accuracies of conventional breeding values (r_{EBV}) were in a wide range and, on average, higher in cows than in bulls for most traits (Table 1).

Model

The DGV were estimated using the model $\mathbf{y} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e}$, where $\mathbf{y}$ is a vector of DRPF, $\mathbf{1}$ is a vector of 1s, μ is the overall mean, $\mathbf{g}$ is a vector of additive genetic effects, $\mathbf{e}$ is a vector of residual effects, and $\mathbf{Z}$ is a design matrix allocating genetic effects to the observations. For random elements, the distributions $\mathbf{e} \sim N(0, \mathbf{R}\sigma^2_e)$ and $\mathbf{g} \sim N(0, \mathbf{G}\sigma^2_g)$ were assumed, where $\mathbf{R}$ is a diagonal matrix with $1/w_i$ on the diagonal and w_i is a weighting factor for the residual variance, and $\mathbf{G}$ is a genomic relationship matrix. Additive genetic variance

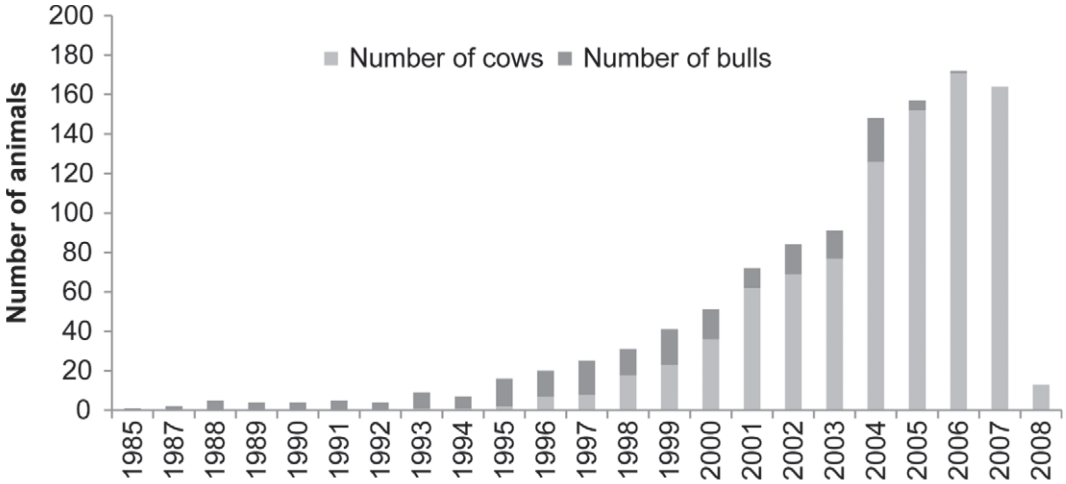

Figure 1. Distribution of birth year for genotyped cows and bulls.