



Are evaluations on young genotyped animals benefiting from the past generations?

D. A. L. Lourenco,*¹ I. Misztal,* S. Tsuruta,* I. Aguilar,† T. J. Lawlor,‡ S. Forni,§ and J. I. Weller#

*Department of Animal and Dairy Science, University of Georgia, Athens 30602

†Instituto Nacional de Investigacion Agropecuaria, Las Brujas 90200, Uruguay

‡Holstein Association USA Inc., Brattleboro, VT 05302

§Genus PIC, Hendersonville, TN 37075

#Agricultural Research Organization, The Volcani Center, Bet Dagan 50250, Israel

ABSTRACT

Data sets of US Holsteins, Israeli Holsteins, and pigs from PIC (a Genus company, Hendersonville, TN) were used to evaluate the effect of different numbers of generations on ability to predict genomic breeding values of young genotyped animals. The influence of including only 2 generations of ancestors (A2) or all ancestors (Af) was also investigated. A total of 34,506 US Holsteins, 1,305 Israeli Holsteins, and 5,236 pigs were genotyped. The evaluations were computed by traditional BLUP and single-step genomic BLUP, and computing performance was assessed for the latter method. For the 2 Holstein data sets, coefficients of determination (R^2) and regression (δ) of deregressed evaluations from a full data set with records up to 2011 on estimated breeding values and genomic estimated breeding values from the truncated data sets were computed. The thresholds for data deletion were set by intervals of 5 yr, based on the average generation interval in dairy cattle. For the PIC data set, correlations between corrected phenotypes and estimated or genomic estimated breeding values were used to evaluate predictive ability on young animals born in 2010 and 2011. The reduced data set contained data up to 2009, and the thresholds were set based on an average generation interval of 3 yr. The number of generations that could be deleted without a reduction in accuracy depended on data structure and trait. For US Holsteins, removing 3 and 4 generations of data did not reduce accuracy of evaluations for final score in Af and A2 scenarios, respectively. For Israeli Holsteins, the accuracies for milk, fat, and protein yields were the highest when only phenotypes recorded in 2000 and later were included and full pedigrees were applied. Of the 135 Israeli bulls with genotypes (validation set) and daughter records only in the complete data set, 38 and 97 were sons of Israeli and foreign bulls, respectively. Although more

phenotypic data increased the prediction accuracy for sons of Israeli bulls, the reverse was true for sons of foreign bulls. Also, more phenotypic data caused large inflation of genomic estimated breeding values for sons of foreign bulls, whereas the opposite was true with the deletion of all but the most recent phenotypic data. Results for protein and fat percentage were different from those for milk, fat, and protein yields; however, relatively, the changes in coefficients of determination and regression were smaller for percentage traits. For PIC data set, removing data from up to 5 generations did not erode predictive ability for genotyped animals for the 2 reproductive traits used in validation. Given the data used in this study, truncating old data reduces computation requirements but does not decrease the accuracy. For small populations that include local and imported animals, truncation may be beneficial for one group of animals and detrimental to another group.

Key words: single-step genomic BLUP, pedigree depth, genomic selection, dairy cattle

INTRODUCTION

Quantitative genetics theory postulates that accuracy of genetic evaluations increases if all known ancestors are included in construction of the relationship matrix (Henderson, 1984), provided that the analysis model corresponds to reality. However, models used in practice are only approximations of “true” models. For example, definitions of traits change over time, accounting for selection may be incomplete, and nonadditive genetic effects are ignored in the model. Also, the contributions of distant generations decay with time. Although parents can explain up to 50% of the genetic variation in an animal, this fraction is divided by 4 with each previous generation. Therefore, the effect of distant ancestors on the accuracy of the youngest animals can be small or even negative. Furthermore, larger data sets require more computing resources.

Mehrabani-Yeganeh et al. (1999) studied the selection response in a simulated population. The accuracy of evaluation for the most recent generation was the same

Received November 26, 2013.

Accepted February 11, 2014.

¹Corresponding author: danilino@uga.edu

regardless of whether all 9 or only the last 2 generations of data were used. The mean simulated breeding value of the selected animals was the same in both scenarios, but mean inbreeding of selected animals was lower for the truncated data set.

In initial predictions with genomic selection, the decay of accuracy for subsequent generations without phenotypes was much slower than with the traditional selection (Meuwissen et al., 2001). Muir (2007) found that the decay of accuracy in genomic selection is much faster under strong selection. In a real population of broiler chickens, that decay was faster than initially expected but still slower than in traditional BLUP (Wolc et al., 2011); the rate of decay changed only slightly for different methods, with lesser decay with bivariate genomic BLUP (**GBLUP**) and BayesC π (Habier et al., 2011) than with univariate GBLUP. If the decay in accuracy is faster than expected, the contributions of older generations may be overestimated with genomic selection.

Recently, Misztal et al. (2013) studied possible biases with unknown parent groups (**UPG**) in a single-step genomic evaluation (**ssGBLUP**). In this method, calculation of unbiased GEBV requires scaling the genomic relationship matrix (**G**) to make this matrix compatible with the numerator relationship matrix for the genotyped animals (**A**₂₂) (Chen et al., 2011; Vitezica et al., 2011). Too-small **G** causes downward bias for the genotyped animals relative to all the animals, and too-large **G** causes upward bias. The additive relationships for the young animals depend on the length of their pedigrees. Because scaling of **G** is for an average of **A**₂₂, GEBV for young animals may be biased up or down depending on the length of the pedigree, with a corresponding decrease in accuracy. A partial solution for this problem is to delete pedigree and phenotypic data of older generations. In this case, missing information from the eliminated pedigrees does not bias evaluations.

The purpose of this study was to evaluate the effect of deleting phenotypic and pedigree data on the accuracy of young genotyped animals in several populations and different traits.

MATERIALS AND METHODS

Three different data sets were analyzed in this study: US Holstein final score data provided by Holstein Association USA Inc. (Brattleboro, VT); Israeli Holstein 305-d milk, fat, and protein yields and fat and protein percentage data provided by Israel Cattle Breeders Association (Caesaria, Israel); and pig reproductive traits from purebred and crossbred lines, provided by PIC (a Genus company, Hendersonville, TN). For all data

sets, variance components were estimated based on the full data using phenotypes and pedigree. Multiple species and a range of population structures were included to give this study a broad application. Animal Care and Use Committee approval was not obtained for this study, because the data were obtained from existing databases.

US Holsteins

Data. Initially, 2 data sets were prepared for US Holsteins. The full data set contained 10,944,571 final score records up to 2011 for 6,586,605 cows born from 1951 to 2009, and a reduced data set (**TR**) included 10,167,064 records up to 2007 for 6,012,441 cows born from 1951 to 2006. Records were deleted from the reduced data set to exclude data of cows born before 5 different thresholds. The thresholds set according to an approximate average generation interval of 5 yr in dairy cattle were T1980, T1985, T1990, T1995, and T2000. Thus, T1980 comprised data of cows born from 1980 to 2006 with records up to 2007, with the same procedure applied for the other 4 thresholds. Two scenarios for constructing the numerator relationship matrix (**A**) were used. The first scenario included relatives of phenotyped animals traced back 2 generations (short pedigree = **A2**); the second scenario included all known relatives of phenotyped animals (deep pedigree = **Af**). The number of animals included in **A** and the number of phenotypes available for each data set are shown in Table 1.

After a general quality control analysis, genotypes on 42,503 SNP markers from the BovineSNP50K Bead-Chip (Illumina Inc., San Diego, CA) were available for 34,506 bulls.

Model. A single-trait animal model was used for evaluation of final score (Tsuruta et al., 2002). The heritability for this trait is 0.31 (Table 2). Unknown parent groups were assigned for missing parents according to year of birth and sex. Traditional evaluations (BLUP) were performed for all data sets, whereas genomic evaluations were not performed for the full data set. Pedigree, genotypes, and phenotypes were analyzed by ssGBLUP (Aguilar et al., 2010). In this method, the inverse of matrix **A** is replaced by the inverse of matrix **H** in the mixed model equations. **H** inverse is as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha\mathbf{G} + \beta\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix},$$

where **G** was constructed as in VanRaden (2008), using current allele frequencies; **A**₂₂⁻¹ is the inverse of pedigree-

Download English Version:

<https://daneshyari.com/en/article/10973960>

Download Persian Version:

<https://daneshyari.com/article/10973960>

[Daneshyari.com](https://daneshyari.com)