J. Dairy Sci. 99:1–14
http://dx.doi.org/10.3168/jds.2015-10715

Improving reliability of genomic predictions for Jersey sires using bootstrap aggregation sampling

Ashley A. Mikshowsky,*[1] Daniel Gianola,*†‡ and Kent A. Weigel*
*Department of Dairy Science,
†Department of Animal Sciences, and
‡Department of Biostatistics and Medical Informatics and Department of Computer Science, University of Wisconsin, Madison 53706

ABSTRACT

Genomic selection has revolutionized the dairy genetics industry and enhanced the rate of response to selection for most economically important traits. All young bulls are now genotyped using commercially available single nucleotide polymorphism arrays to compute genomic predicted transmitting ability (GPTA) and reliability (REL) values. Decisions regarding the purchasing, marketing, and culling of dairy bulls are based on GPTA until roughly 5 yr of age, when milk-recorded offspring become available. At that time, daughter yield deviations (DYD) can be used to assess the accuracy of the GPTA computed several years earlier. Although agreement between predictions and DYD is often good, the DYD of some bulls differ widely from corresponding GPTA, and published REL are of limited value in identifying such bulls. A method of bootstrap aggregation sampling (bagging) using genomic BLUP (GBLUP) was implemented to predict the GPTA of 379, 379, and 342 young Jersey bulls for protein yield, somatic cell score, and daughter pregnancy rate, respectively. For each trait, 50 bootstrap samples from a reference population consisting of 2011 DYD of 1,738, 1,616, and 1,551 older Jersey bulls were used, and correlations between bagged GBLUP predictions and 2014 DYD were lower than GBLUP predictions derived from the full reference population. Although the bagged GBLUP approach did not improve the predictive correlations, it allowed computation of bootstrap predictive reliabilities across random samples of the reference population. The bootstrap predictive reliabilities could be a useful diagnostic tool for assessing genome-enabled prediction systems or evaluating the composition of a reference population. Our main objective was to determine if bagging GBLUP of young Jersey bulls could lead to measures of reliability that would be a useful alternative to published REL values. The standard deviations of bagged GBLUP predictions were found to weakly improve our ability to identify bulls whose future daughter performance may deviate significantly from early GPTA for protein, but not for somatic cell score or daughter pregnancy rate.

Key words: genomic prediction, reliability, bootstrap sampling, dairy cattle

INTRODUCTION

The transmitting ability of a young dairy animal can be predicted using a large number of molecular markers throughout the genome, most commonly SNP, and genomic selection refers to the use of these predictions to make breeding and culling decisions (Meuwissen et al., 2001). The first official genomic evaluations for Holsteins and Jerseys were released in January 2009 (Wiggans et al., 2011), and virtually all young bulls that are candidates for AI and all young heifers with potential as elite breeding stock are genotyped using low-density (<20,000 SNP) or medium-density (50,000 to 150,000 SNP) arrays. The resulting genomic data are seamlessly integrated into the national genetic improvement program, which is managed by the Council on Dairy Cattle Breeding (**CDCB**; Bowie, MD), and the genomic PTA (**GPTA**) of young genome-tested animals are predicted using data from a reference population of older animals with genotypic and phenotypic data. Various methods can be used for genome-enabled prediction, including genomic BLUP (**GBLUP**), Bayesian regression models, and kernel-based methods (e.g., de los Campos et al., 2013; Gianola and van Kaam, 2008; Moser et al., 2009; VanRaden, 2008). The resulting GPTA take into account information from all genotyped relatives and from nongenotyped females, as they are included in the daughter deviations (**DD**) of their genotyped sires (Legarra et al., 2009). Single-step GBLUP (**ssGBLUP**) can also use information from genotyped animals to improve evaluations of their nongenotyped ancestors. The reliability (**REL**) values corresponding to the GPTA reflect the approximate amount of information (termed

---

Received December 3, 2015.
Accepted January 20, 2016.
[1] Corresponding author: mikshowsky@wisc.edu

as "daughter equivalents") contributed by an animal's parents, progeny, own records, and from the markers.

The effect of genomic selection on reducing the generation interval in the Holstein and Jersey breeds has been dramatic (Hutchison et al., 2014). By 2012, approximately 51% of Holstein and 52% of Jersey breedings in the United States were to genotyped young bulls that had no milk-recorded offspring at the time of insemination. Furthermore, Holstein and Jersey herds that used >75% young, genome-tested bulls had \$58 and \$63 greater expected lifetime net merit in the resulting calves, respectively, than herds that used no semen from young bulls. The use of GPTA for identifying groups of young bulls to be used as service sires in commercial dairy herds, or for identifying groups of young heifers that should be culled or retained as future herd replacements (Weigel et al., 2012), is now commonplace. However, REL values are lower than those typically achieved by progeny testing (VanRaden et al., 2009), and the GPTA for some individual bulls differ widely from actual daughter performance, as measured by daughter yield deviations (**DYD**) for production traits or DD for health traits of these bulls 3 yr later. In addition, the range in REL values for young, genome-tested bulls is relatively small, and published REL have limited utility as indicators of the accuracy or stability of GPTA for specific bulls.

Genomic reliability measures were initially calculated from inversion of the mixed model equation, but this approach was abandoned when data sets became too large. Misztal et al. (2013) developed algorithms to estimate reliabilities from ssGBLUP. An algorithm that used inversion of a matrix containing inverses of the genomic relationship matrix and the pedigree relationship matrix for genotyped animals was found to be fairly accurate and inexpensive for data sets of fewer than 100,000 genotypes.

It is known that the expected accuracy (in the sense of predictive correlation) of genomic prediction is affected by the magnitude of the relationships between testing set (e.g., young bulls) and older animals in the reference population (e.g., Lund et al., 2009; Habier et al., 2010). Habier et al. (2007) showed that the accuracy of genomic predictions will be greatest when many close relationships exist between animals in the reference population (i.e., training set) and the validation population (i.e., testing set). If the relationships between individuals in the reference population and testing set are distant, GBLUP can provide genomic predictions but the accuracy will be lower.

One possible way of computing a more informative estimate of the stability of a bull's GPTA is through the use of bootstrap sampling. Bootstrap aggregation sampling, commonly known as "bagging," is a resampling method that is relatively simple to implement, and it can increase the accuracy of predictions in situations where sampling from training set leads to large variance in the predictor (Breiman, 1996). Bagging involves repeated sampling with replacement from the original reference population to create a set of predictors, which are then averaged across samples to calculate the bagged predictor. By averaging over the bootstrap samples, the variance of the predictor is decreased, and predictive mean-squared error is decreased concomitantly. Gianola et al. (2014) applied this methodology in the context of genome-enabled prediction, computed bagged GBLUP (hereafter, **BGBLUP**) predictors, and showed that 25 to 50 bootstrap samples could provide reasonable predictions and stable measures of predictive mean-squared error for individual selection candidates. Furthermore, if some cows or bulls in the reference population have errors or biases in their phenotypes, perhaps due to preferential treatment of potentially valuable animals, bagging may provide more robust predictions by providing bagged GPTA that are averaged over bootstrap samples that contain or omit certain individuals. Some combinations of bootstrap samples may produce more accurate predictions than those computed from the full reference population. In addition, the observed variation in GPTA between bootstrap samples may provide an indication of which selection candidates may have genomic predictions that could lack stability or deviate significantly from their actual future performance. Therefore, the objective of the present study was to examine whether bagging GBLUP for protein yield, SCS, and daughter pregnancy rate (**DPR**) of young Jersey bulls could provide a useful alternative to published REL values when forecasting which bulls may have GPTA that are more or less accurate predictors of future daughter performance than the GPTA of a typical bull.

MATERIALS AND METHODS

Data

The genotypes of 4,372 Jersey bulls were provided by the Cooperative Dairy DNA Repository (Columbia, MO). Genomic data included 60,671 SNP markers for each bull, and after discarding SNP with more than 20% missing values and those with a minor allele frequency $\leq$5%, 50,819 markers remained for the analysis.

Three phenotypic traits were analyzed: protein yield (kg), SCS $\{\log_2[(\text{cells/mL})/100{,}000] + 3\}$, and DPR (%). The PTA values for all 3 traits, as well as DYD for protein yield and DD for SCS and DPR, were ob-