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, 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 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, genotyped-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. Mieza 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, BGGBLUP) 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...
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 obtained from the CDCB for the August 2011 and August 2014 sire summaries. Jersey bulls with at least 50 US daughters for a given trait in August 2011 were used as the reference population, whereas Jersey bulls with no daughters for a given trait in August 2011 and at least 50 US daughters for that trait in August 2014 were used as the testing set. After editing, the reference population consisted of 1,738 bulls for protein yield, 1,616 bulls for SCS, and 1,551 bulls for DPR. The corresponding testing sets contained 379 bulls for protein, 379 bulls for SCS, and 342 bulls for DPR.

The August 2011 CDCB GPTA and the August 2014 DD values for SCS were on 2 different scales. The original data had been adjusted by adding 3 to compute the reported PTA, but no adjustments had been made for DD. Rescaling was done by subtracting the mean from each individual value and dividing by the standard deviation.

### GBLUP Prediction

Genomic predictions were calculated using GBLUP, as implemented in the Bayesian generalized linear regression package of the R software (R Core Development Team, 2015). The GBLUP method uses a genomic relationship matrix (\( G \)) that incorporates genome-based estimates of the additive genetic similarity between individuals. The basic GBLUP model for predicting direct genomic breeding values (DGV) is

\[
y = 1\mu + Zg + e,
\]

where \( y \) is a vector of observations, \( 1 \) is a vector of ones, \( \mu \) is the overall mean, \( Z \) is an incidence matrix relating observations to the random effects, \( g \) is a vector of random genomic breeding values, and \( e \) is a vector of random residual effects. It was assumed that \( g \sim N(0, G\sigma_g^2) \), where \( \sigma_g^2 \) is the additive genomic variance and \( G \) is the marker-based genomic relationship matrix, and \( e \sim N(0, D\sigma_e^2) \), where \( \sigma_e^2 \) is the residual variance and \( D \) is a diagonal matrix with elements \( d_{ii} = 1/w_i \). The \( w_i \) are weights that account for differences in the amount of information in the DYD or DD between individual bulls in the reference population. The variance components were estimated within the Bayesian generalized linear regression package from the data.

In this study, the \( y_i \) represented the DYD for protein yield or DD for SCS or DPR of bull \( i \) in August 2011, whereas the \( w_i \) were set equal to the square root of the number of US daughters of bull \( i \) for that trait in August 2011. Bulls in the testing set had no daughters in August 2011, and the \( w_i \) for these bulls were set equal to 1. A weight of 1 (rather than 0) was used for computational reasons. Although the effect of this approximation is negligible in the case of bulls with many progeny, this would not be the case if females had been included in the reference population.

### Bagging GBLUP

Bootstrap samples were created using the August 2011 reference populations for protein yield, SCS, and DPR. Bulls were sampled randomly from the reference population, with replacement, until each bootstrapped sample from the population had the same number of bulls as the original reference population. Genomic predictions for each trait were computed for all bulls in the testing set using GBLUP. Fifty bootstrap samples from the reference population were created for each trait, as recommended by Gianola et al. (2014).

### Bootstrap Reliabilities

The realized bootstrap squared prediction error (BPE) for bull \( i \) (\( BPE_i \)) based on 50 bootstrap samples for each trait were computed as follows:

\[
BPE_i = \sum_{j=1,50} (DYD_i - GPTA_{ij})^2/50,
\]

where \( DYD_i \) for protein yield (or \( DD_i \) for SCS or DPR) represented the standardized average of actual daughter performance in August 2014 for bull \( i \), and \( GPTA_{ij} \) represented the GBLUP prediction for a given trait for bull \( i \) in bootstrap sample \( j \) as computed from August 2011 data of bulls in the reference population. The realized bootstrap prediction reliability (\( BPRL_i \)) was then calculated as

\[
BPRL_i = 1 - \left[ BPE_i/\text{diag}(V_{Test}) \right],
\]

where \( V_{Test} = G_{Test} \sigma_g^2 + I_{Test} \sigma_e^2 \), \( G_{Test} \) is the genomic relationship matrix between bulls in the testing set, \( I_{Test} \) is an identity matrix of the same dimension, and \( \sigma_g^2 \) and \( \sigma_e^2 \) are the genetic and residual variance components for a given trait. The variance components were calculated from parameters used in August 2014 CDCB genetic evaluations for each trait; values of \( \sigma_g^2 \) and \( \sigma_e^2 \) were estimated.
were 282 and 514 for protein yield, 0.144 and 0.782 for SCS, and 9.12 and 182 for DPR, respectively.

**Relationship to Reference Population**

As noted earlier, the expected accuracy of genomic prediction for a given selection candidate can be influenced by the animal’s degree of relatedness to older animals in the reference population. In this study, the relationship of each bull in the testing set to the reference population was computed as the average of the bull’s 10 largest off-diagonal elements of the G matrix corresponding to relatives in reference population, as suggested by Clark et al. (2012). For larger populations, the top 25 or top 50 relationships might be more robust than the top 10 when considering outcross animals such as foreign bulls or older research cows.

**RESULTS**

**Accuracy of Published Reliabilities**

The absolute August 2011 GBLUP prediction error (calculated as the absolute value of the difference between the official August 2011 CDCB GPTA and the August 2014 DYD or DD) for each Jersey bull in the testing set is compared in Figure 1 with the official REL value published by the CDCB in 2011. It is clear that the published REL values associated with August 2011 GPTA of these bulls give little indication of the magnitude of differences that could be expected between the published GPTA and the subsequent daughter performance of individual bulls. The relationship between REL values and the difference between prediction and future progeny performance lacked statistical significance for all 3 traits analyzed.

**Accuracy of GBLUP and BGBLUP Predictions**

Table 1 shows the correlations of August 2014 DYD for protein yield and DD for SCS and DPR with GBLUP and BGBLUP predictions for the same traits based on August 2011 data of bulls in the reference population. The GBLUP predictions were based on one realization of the full reference population, whereas the BGBLUP predictions represented averages of genomic predictions from 50 bootstrap samples for each trait. Correlations for GBLUP from the full reference populations were 0.660, 0.481, and 0.400 for protein yield, SCS, and DPR, respectively, and were slightly higher than correlations of 0.646, 0.445, and 0.383 for BGBLUP. Similarly, the mean-squared errors (MSE) associated with predicting August 2014 DYD (or DD) by applying GBLUP or BGBLUP to August 2011 data of bulls in the reference population are shown in Table 1. The GBLUP MSE values of 41.8, 0.033, and 2.01 for protein yield, SCS, and DPR, respectively, were slightly smaller than the MSE of 42.6, 0.034, and 2.09 for BGBLUP. As expected, correlations and MSE values varied between bootstrap samples. However, in this study, the maximum BGBLUP correlation for any individual sample was always less than that of the GBLUP correlation from the full reference population, whereas the minimum BGBLUP MSE for any individual sample was always greater than that of the GBLUP MSE.

**Relationship Between Bootstrap Standard Deviations and Accuracy of BGBLUP Predictions**

The bootstrap standard deviations (SD) of GBLUP predictions for individual bulls were used as a measure of the variation in GBLUP predictions that one could expect due to random sampling from the reference population. A large bootstrap SD for a particular bull would indicate that his predictions are highly influenced by the sample of older animals represented in the reference population for a given bootstrap sample. For example, such bulls might have close relatives in the reference population whose phenotypic data contain errors or reflect bias due to preferential treatment. The relationship between the bootstrap SD of August 2011 GBLUP predictions and the absolute August 2011 GBLUP prediction error is shown in Figure 2. For protein, there was a weakly significant positive correlation (P = 0.057) between bootstrap SD and the accuracy of 2011 GPTA predictions. No clear relationship was observed for SCS and DPR, although the number of Jersey bulls with large bootstrap SD was limited.

As seen in Figure 2, a few bulls had very large values of bootstrap SD. Each trait had different bulls with extreme bootstrap SD, but there were sires and dams in common between the traits. For protein, the 11 bulls with the highest bootstrap SD all had very low values for the average of the top 10 relationships. Ten of these bulls shared the same sire, and the accuracy of their GBLUP prediction for each bootstrap sample was very dependent on whether or not their sire was present in the sample. Their GBLUP predictions decreased when their sire was not included in the sample, as shown in Figure 3. For both SCS and DPR, 2 of the 3 bulls with the highest bootstrap SD were also paternal half-brothers and had low values for the average of the top 10 relationships. Their GBLUP predictions tended to increase when their sire was absent from the bootstrap sample (Figure 3). These observations demonstrate the effect that a single close relative in the training set can
have on prediction accuracy for DGV, especially if a particular bull has very few close relatives overall in the reference population. The effect would be slightly less if blended GPTA were used, as discussed later.

**Bootstrap Squared Prediction Error and Presence of Sire in Bootstrap Sample**

Bulls in the testing set are expected to have higher prediction accuracy if their sire is present in the reference population. The influence of the sire’s presence in the reference population for a particular bull was observed by separating bootstrap samples into those samples containing his sire and those omitting his sire, and then computing the BPE for each. The absence of a bull’s sire from the reference sample led to higher BPE values than if the sire was present. The average difference between BPE for samples excluding the sire and samples including the sire was 5.60, 0.0036, and 0.14 for protein, SCS, and DPR, respectively. The absence

![Graphs showing relationship between genomic BLUP prediction error and CDCB reliability for protein, SCS, and DPR](image)

**Figure 1.** Relationship between the absolute August 2011 genomic BLUP (GBLUP) prediction error for protein (n = 379), SCS (n = 379), and daughter pregnancy rate (DPR; n = 342) with August 2011 Council on Dairy Cattle Breeding (CDCB) reliability.
of the sire from the reference sample also increased the range of BPE values. Samples including the sire had BPE ranges of 341.50, 0.361, and 19.69 for protein, SCS, and DPR, respectively. Samples excluding the sire had BPE ranges of 435.05, 0.365, and 20.89 for protein, SCS, and DPR, respectively, indicating smaller variation of prediction errors in samples containing the sire than those omitting the sire.

**Bootstrap SD and Accuracy of Prediction Based on Relationship to the Reference Population**

As noted earlier, selection candidates with more close relatives in the reference population are expected to have greater accuracy of genomic predictions. In this study, the average of the 10 largest relationships of each bull in the testing set with sires in the reference population was used as the measure of relatedness. The August 2011 GBLUP prediction errors for protein, SCS, and DPR were compared with this measure of relatedness in Figure 4. As the average of the top 10 relationships with the reference population increased, the August 2011 GBLUP prediction error increased slightly, but the relationship was not statistically significant.

The bootstrap SD of August 2011 GBLUP predictions for protein, SCS, and DPR was compared with the average of the top 10 relationships to bulls in the reference population in Figure 5. A very significant strong trend was observed: as the average of the top 10 relationships increased, the bootstrap SD of GBLUP predictions decreased. This indicates that bulls with more close relatives in the reference population had significantly less variability in GBLUP predictions over bootstrap samples.

**Accuracy of Bootstrap Prediction Reliabilities**

Bootstrap prediction reliabilities (BPREL) were calculated as described previously, and the absolute value of the August 2011 GBLUP prediction error for each Jersey bull in the testing set was compared in Figure 6 with the BPREL value computed from repeated bootstrap sampling. As expected, the relationship was strong and negative, because the August 2014 DYD (or DD) values are used when computing $\text{BPE}_i$ (the bootstrap prediction error for each bull), which is used in the numerator when computing $\text{BPREL}_i$. Nonetheless, BPREL can give a useful retrospective measure of reliability that can be used to diagnose potential problems in the reference population or genomic prediction system.

Last, as shown in Figure 7, the relationship between REL values published by CDCB in August 2011 and the realized BPREL in August 2014 was weak, due in part to the limited range in published REL of young bulls with genomic data but no progeny.

**DISCUSSION**

In the present study involving production and fitness traits of Jersey bulls, BGBLUP did not improve the accuracy (predictive correlation) of genomic predictions over that of GBLUP applied to the original reference population. In fact, GBLUP predictions from the full reference population were closer to subsequent DYD or DD than the BGBLUP predictions derived from repeated bootstrap sampling. This contradicts the results of Abdollahi-Arpanahi et al. (2015), who reported greater prediction accuracy of BGBLUP than GBLUP for growth, carcass, and egg production traits in broiler chickens. Our results are not unexpected, because bagging is known to be more effective when the prediction model is unstable (Breiman, 1996), and pedigree or genomic BLUP are low-variance predictors. However, our primary goal was not to improve overall prediction accuracy but rather to identify a metric for identifying young dairy bulls for which the early GPTA for economically important traits might be exceptionally good or exceptionally poor predictors of future daughter performance. Published REL values associated with...
the GPTA of these bulls were not indicative of the magnitude of future deviations from actual daughter performance, and hence their utility appears to be limited. Our hypothesis—that the bootstrap SD of GBLUP predictions over repeated samples from the reference population would be informative for identifying bulls whose GPTA was more strongly or loosely related to future DYD or DD—was weakly supported for protein, but not for SCS or DPR in the present study. We had expected that differences between GPTA and DYD or DD would be smaller for bulls that demonstrated less variation in GPTA between bootstrap samples (i.e., GPTA that were robust to the choice of reference animals). We detected a weak positive relationship between these differences and bootstrap variation for protein, but such relationships were not observed for SCS or DPR. The

![Figure 2](image_url)

**Figure 2.** Relationship between the absolute August 2011 genomic BLUP (GBLUP) prediction error with the SD of August 2011 bootstrap genomic BLUP (GBLUP) predictions for Jersey bulls for protein (n = 379), SCS (n = 379), and daughter pregnancy rate (DPR; n = 342).
Figure 3. August 2011 genomic BLUP (GBLUP) predictions for each bootstrap sample from the reference population for paternal half-brothers with large bootstrap SD in Figure 2. Bootstrap samples in which the bulls' sire is missing are denoted by ● (n = 10 for protein; n = 2 for SCS; n = 2 for daughter pregnancy rate).
weak relationship for protein could indicate bias due to preferential treatment introduced by certain bulls in the reference population, which would not be expected for SCS or DPR. The BPE calculated from bootstrap samples excluding the sire were higher on average than BPE from samples including the sire for protein, SCS, and DPR, as expected. The range of BPE values for samples excluding the sire were also greater than for samples including the sire. Therefore, the presence of a young bull’s sire in the reference sample decreased the squared prediction error and the variability of that error.

As the magnitude of genomic relationships of bulls in the testing set with sires in the reference population increased, the bootstrap SD over repeated sampling decreased. This indicates that GPTA for bulls with more

Figure 4. Relationship of the August 2011 genomic BLUP (GBLUP) prediction error for Jersey bulls and average of their 10 largest genomic relationships with bulls in the reference population [n = 379 for protein; n = 379 for SCS; n = 342 for daughter pregnancy rate (DPR)].
As discussed previously, relationships between the reference population and bulls in the testing set can affect prediction accuracy. The more close relatives a bull has in the reference population, the more accurate his prediction will be. Erbe (2013) found that the presence of sires, full-siblings, and half-siblings in the reference population, especially those that are relatively close in age to the bulls being predicted, is imperative to attain the highest prediction accuracy. If 1 or 2 generations are missing from the reference population, there will be a decrease in prediction accuracy.

Figure 5. Relationship of bootstrap SD of August 2011 genomic BLUP (GBLUP) predictions for protein, SCS, and daughter pregnancy rate (DPR) of Jersey bulls with the average of their 10 largest genomic relationships with bulls in the reference population (n = 379 for protein; n = 379 for SCS; n = 342 for DPR).
be a significant loss in prediction accuracy (Erbe et al., 2011). This is starting to become a concern as the generation interval decreases. There are fewer progeny-tested bulls who are closely related to the young bulls whose values are being predicted. The reference population is beginning to comprise bulls that are 2 or 3 generations older than the current selection candidates, and these bulls contribute much less to the accuracy of genomic predictions than closer ancestors.

Although close relationships between the reference and testing populations are desired, relationships among individuals in the reference population should be minimized. Equivalently, the reference population should span as much genetic variation as possible, because the average prediction reliability is increased when there are more unrelated individuals within the reference population who have alleles that are not identical-by-state (Pszczola et al., 2012).

**Figure 6.** Relationship of the absolute August 2011 genomic BLUP (GBLUP) prediction error for protein (n = 379), SCS (n = 379), and daughter pregnancy rate (DPR; n = 342) with bootstrap prediction reliability (BPREL).
Whereas the current study used DGV, many national evaluations use selection index blending to include information from animals' nongenotyped ancestors into the published GPTA, as in Lund et al. (2009), Reinhardt et al. (2009), and VanRaden et al. (2009), and some evaluations use ssGBLUP. This resulting blended GPTA uses phenotypic data from daughters to improve progeny predictions if a sire is nongenotyped. Therefore, these blended GPTA include more information than the DGV used in this study, yet they are not as accurate as the values that would be computed if the ancestor genotypes were known. Most sires today are genotyped, so the difference between blended GPTA and those computed from DGV is minimal.

A very large reference population is needed to predict genomic values with the highest expected accuracy (Goddard, 2009; VanRaden et al., 2009; Brøndum et al., 2011). This has been accomplished for Holsteins (Wiggans et al., 2011), but for smaller breed populations, such as Jerseys, it is harder to achieve because

---

**Figure 7.** Relationship of August 2011 Council on Dairy Cattle Breeding (CDCB) reliability for protein (n = 379), SCS (n = 379), and daughter pregnancy rate (DPR; n = 342) with bootstrap prediction reliability (BPREL).
fewer progeny-tested bulls are available for the reference population. Now that there are many more cows genotyped per year than bulls, one potential way to increase the size of the reference population is to include cows. However, there are 2 main concerns about the addition of cows to the reference population: (1) the possibility of introducing bias due to preferential treatment of cows with high perceived value, and (2) complications associated with including data of females that are genotyped selectively. The diagnostic tools described herein, particularly those based on analysis of prediction errors associated with subsets of bootstrap samples that include or excluded various relatives, could be useful in determining the conditions under which data from genotyped cows should be added to the reference population.

The computation for BGBLUP, compared with GBLUP, is very demanding in terms of time. For 50 bootstrap samples, BGBLUP takes approximately 50 times as long to run as GBLUP, due to the creation of new bootstrap reference samples and calculation of genomic predictions using each of these samples. Therefore, this approach would best be used by breeding companies to assess the variation or stability of predictions for high-value animals such as sires of sons, bulls and cows involved in in vitro fertilization, or boars and sows in the top of the nucleus breeding pyramid. It could be used as a way of verifying or double-checking predictions before making purchasing or breeding decisions.

CONCLUSIONS

Although bagging GBLUP using bootstrap sampling did not improve the accuracy of genomic predictions in Jerseys, it allowed computation of bootstrap prediction REL across random samples of the reference population. The latter can be used to construct useful diagnostic tools for assessing genomic prediction systems or for evaluating the composition of a genomic reference population. Published REL values seem to have limited utility for identifying young Jersey bulls whose future daughter performance may deviate significantly from early GPTA, but despite the conceptual advantages of bootstrap SD of GBLUP predictions, this new metric did not significantly improve our ability to identify such bulls in the present study. The bootstrap SD of GBLUP was a weak indicator of the magnitude of prediction errors for protein, but not for SCS or DPR. Now that genomic reference populations for major breeds such as Holstein and Jersey have become very large, we can hypothesize that customized subsets of reference animals can be constructed to enhance the predictive ability for specific selection candidates. Future research should seek to identify methods for diagnosing errors, biases, or nuances in the reference data that could lead to more stable or robust predictions for young selection candidates.

ACKNOWLEDGMENTS

This project was supported by the USDA National Institute of Food and Agriculture (Washington, DC), accession no. 1006669, as Hatch grant no. MSN179976 from the Wisconsin Agricultural Experiment Station. A. Mikshowsky acknowledges partial financial support from Accelerated Genetics (Baraboo, WI), and K. Weigel acknowledges partial financial support from the National Association of Animal Breeders (Columbia, MO).

REFERENCES


