

Impact of adding cows to the current EuroGenomics bull reference population on genomic prediction

H. Alkhoder, Z. Liu, F. Reinhardt, and R. Reents

vit – IT solutions for animal production, Heinrich-Schröder-Weg 1, 27283 Verden, Germany

Abstract

Genomic evaluation has been based on a reference population comprising only daughter-proven bulls in most countries until recently. With more and more cows routinely genotyped, adding non-selectively genotyped cows to current bull reference population may improve the accuracy of genomic evaluation. Using approximately 20,000 cows genotyped within the framework of a German research project *Kuh_L*, we wished to quantify the impact of adding cows to the current EuroGenomics bull reference population on genomic prediction. Four types of reference population: only cows from the research project (*Cows*), the current EuroGenomics reference bulls (*Bulls*), a mixed of the bulls and cows from the project (*Mixed*) and a mixed of the bulls and all cows (*Mixed_All*), were compared in accuracy and bias of genomic prediction using phenotype and genotype data from April 2017 monthly genomic evaluation for German Holsteins. Analyzed traits included all routine national and MACE traits and novel health traits on claw health, udder health, reproduction diseases and metabolic disorders. Among all the four types of reference population, the scenario *Cows* had the lowest validation R^2 value due to the limited number of reference cows, this suggested a reference population composed of only cows being not yet accurate enough for routine genomic evaluation. We observed an increase in validation R^2 values for all the traits when the cows were added into the current bull reference population. No single trait was identified with a decrease in the accuracy of genomic prediction. For the novel traits accuracy of prediction was significantly lower than the regular ones, due to the much smaller size of reference population. The regression coefficient of validation bulls' deregressed EBV on their candidate GEBV showed that under the scenarios *Bulls*, *Mixed* and *Mixed_All* genomic prediction remained unbiased, having the cows included in the bull reference population. In summary, adding the non-selectively genotyped cows to the current EuroGenomics bull reference population increased the accuracy of genomic prediction for all the regular traits. No bias of genomic prediction was seen when the reference cows were observed to the current EuroGenomics bull reference population.

Key words: genomic evaluation, cow reference population, genomic validation

Introduction

Genomic reference populations for dairy cattle compose, until today, only daughter-proven bulls in most countries. The U.S. dairy populations belong to few exceptions that include also cows in the reference population (Wiggans et al., 2011). One of the main advantages of including cows in the bull reference population is that a possible negative impact of highly genomically pre-selected bulls on genomic prediction would be reduced,

when the reference cows are not genomically or phenotypically pre-selected, e.g. stemming from a whole-herd genotyping scheme.

Since 2010, German Holstein genomic evaluation has been based on an across-country EuroGenomics bull reference population (Liu et al. 2011), with approximately 36,500 Holstein reference bulls for genomic evaluations in 2017. To maintain and further increase the accuracy of genomic predictions, two research projects had been

conducted for German Holsteins, prior to an industry funded project *Kuh-Vision* (Rensing et al., 2017). In the first research project, *Kuh-L*, about 20,000 cows from 56 large herds were genotyped, 5000 of them genotyped with the Illumina 50K version 2 chip and the remaining with EuroGenomics 10K chip. The cows were born between 2011 and 2013, mostly in 2012. The other research project *GKUHplus* focused on novel health traits, it covered setting up a data recording system to developing routine genetic and genomic evaluation systems for German Holsteins. The novel traits included mastitis, claw health traits, reproduction diseases and metabolic disorders.

On the basis of the two research projects, the initial goal of the industry founded *KuhVision* project was to genotype 100,000 reference cows within three years, started in 2016. To avoid selective genotyping, female animals were genotyped on a whole-herd basis, with >600 herds currently participating in the project. In addition to the conventional traits collected through milk recording schemes, the novel health traits have being also collected in the participating herds.

From the research project *Kuh-L*, genotype and phenotype data were available for almost all the 20,000 cows. The aim of this study was to investigate the impact of adding cows to the current EuroGenomics bull reference population on genomic prediction.

Materials and Methods

Phenotype and genotype data as well as pedigree file were obtained from the official monthly genomic evaluation April 2017 for German Holsteins. There were a total of 34,707 Holstein reference bulls included in the EuroGenomics bull reference population for milk yield. MACE EBV were deregressed for all bulls included in the April 2017 MACE evaluation. Using a similar procedure, national EBV of all cows were deregressed to obtain phenotypes for all the cows with own phenotype data. A total of 314,608 genotyped

animals were evaluated in the monthly genomic evaluation.

Four types of genomic reference population (RP) were investigated: a bull reference population as the current EuroGenomics RP (*Bulls*); a RP of cows from the project *Kuh-L* (*Cows*); a mixed RP including the bulls and only the cows genotyped within the *Kuh-L* project (*Mixed*); and a mixed RP with all the bulls and all available genotyped cows (*Mixed_All*). For the two mixed RP, the reference cows made contributions both to their own phenotypes, deregressed EBV and EDC, and to phenotypes of their sires. To avoid double counting of the cows contributions, the phenotypes of their sires were modified.

A validation population for a genomic validation was made to contain only German national bulls born between 2010 and 2012. In order to avoid a ‘favorable’ influence of their reference daughters, the validation bulls with daughters in the above RP were discarded. Milk yield was chosen to represent regular traits, whereas claw ulcers the novel traits. The total number of validation bulls was 894 for the regular traits and 250 for the novel health traits. Table 1 shows the size of reduced reference populations for the genomic validation study. Because only cows from the *Kuh-L* project had the novel traits recorded at the time of this study, the scenario *Mixed_All* did not exist for the novel traits, and it was, in fact, identical to the scenario *Mixed*.

Table 1. Size of the reduced reference populations for the genomic validation

Reduced RP	Regular traits	Novel traits
Bulls	31,428	2,007
Cows	19,064	16,206
Mixed	50,492	18,210
Mixed_All	81,002	-

Statistical Analysis

A SNP BLUP model (Liu et al. 2011) was applied to the genotypes and deregressed EBV of the reference animals. Interbull’s GEV

Test software was applied to obtain the validation results.

Results & Discussion

SNP effect estimates

Using milk yield as example, we show in Figure 1 SNP effect estimates from the three reduced reference populations.

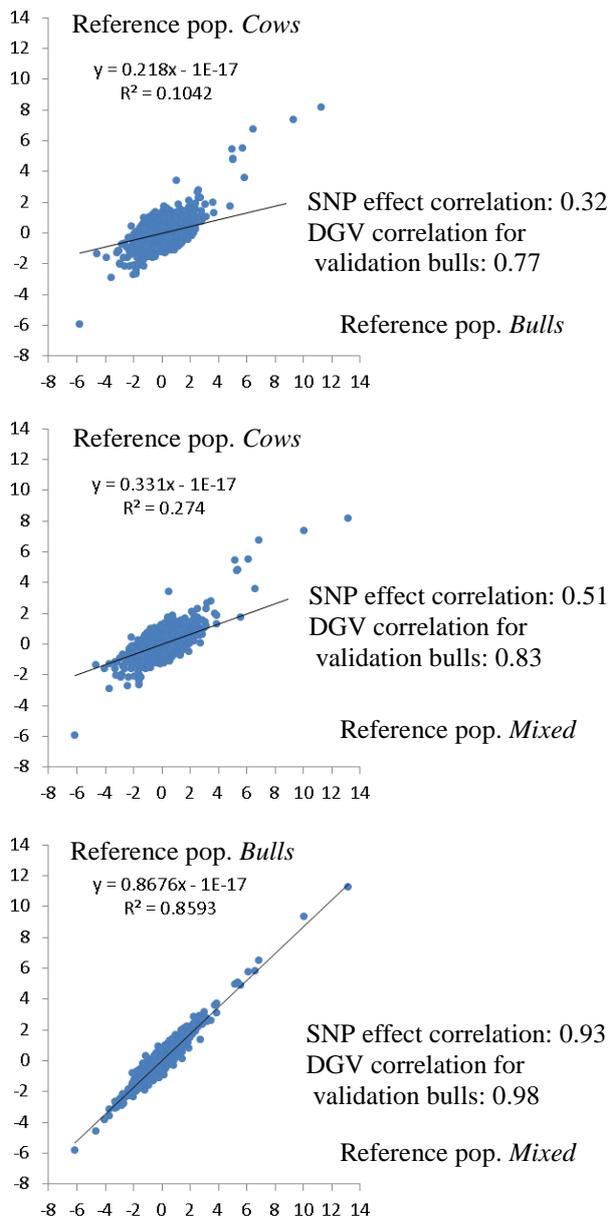


Figure 1. Scatter plots of estimated SNP effects of milk yield for the three reduced reference populations: *Bulls*, *Cows* and *Mixed*.

The SNP effect estimates were standardized using genetic standard deviation of average SNP markers, thus the SNP effect estimates can be directly compared between the scenarios and across the traits. It can be clearly seen that the RP *Cows* has the smallest variation in the SNP effect estimates than the RP *Bulls* or *Mixed* as a result of the smallest size of RP. In addition, the SNP effect estimates of Scenario *Cows* have the lowest correlation with the other two scenarios. DGV of the validation bulls have higher correlations between any two scenarios than the SNP effect estimates. With DGV correlation as high as 0.98 between scenarios *Bulls* or *Mixed*, we expect smaller changes in genomic prediction from the RP *Bulls* to *Mixed* than from *Bulls* to *Cows*.

Due to the much smaller size of RP, variances of SNP effects of the novel health traits were lower than the regular traits under all the scenarios. DGV of the validation bulls were also less correlated than for the regular traits, because of a lower number of daughters of the validation bulls for the novel traits.

Observed validation R^2 values

Figure 2 shows observed R^2 values, multiplied with 100, from the genomic validation for milk production traits and somatic cell scores. The R^2 values were not adjusted for conventional reliability values of the validation bulls, thus they are not comparable between the traits. However, the validation R^2 values can be directly compared between the scenarios of different RP within each trait. Due to the large EuroGenomics reference bull RP, the three scenarios *Bulls*, *Mixed* and *Mixed_All* have high R^2 values. The scenario *Cows* has the lowest R^2 value among all the scenarios, indicating that genomic reference population must not be composed of only cows. Because of the significantly lower accuracy of the scenario *Cows*, subsequent analyses excluded this scenario. As for the four traits in Figure 2, we have seen also similar results for all the other traits.

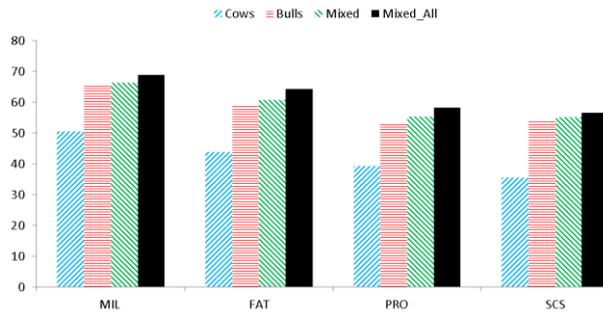


Figure 2. Observed R^2 values (x100) of the genomic validation using four different reference populations.

Change in the accuracy of genomic prediction with cows added

By comparing the scenario RP *Mixed* or *Mixed_All* to the scenario RP *Bulls*, the impact of adding cows on the validation R^2 values were determined. Figures 3 and 4 show changes in validation R^2 values by adding cows to RP for conformation traits and the other traits, respectively.

It can be seen that adding cows to the EuroGenomics bull reference population resulted in an increase in validation R^2 value for all the scenarios. We have seen the increased R^2 value, ranging from 0.3% to 5.5%, by adding c.a. 20,000 cows to the current bull reference population for all national and MACE traits, including the novel health traits as well. Adding all genotyped cows in scenario RP *Mixed_All* led to an even higher R^2 increase for all the traits than scenario RP *Mixed* with only cows from the project *Kuh-L*, suggesting that additionally including many old, possibly more selectively genotyped cows helped further increase the accuracy of genomic prediction. We have not observed any trait with a decreased R^2 value by adding cows to the current bull reference population.

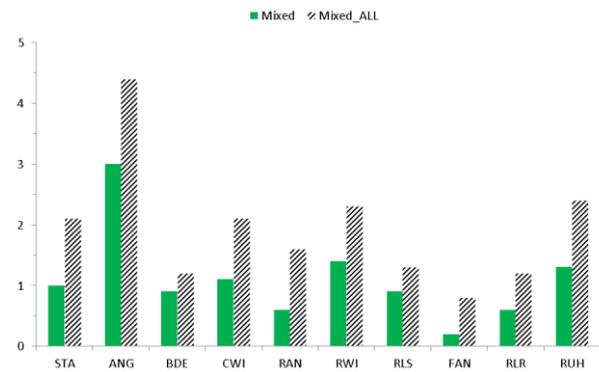


Figure 3. Change in validation R^2 values (x100) for some selected conformation traits by adding cows to the EuroGenomics bull RP.

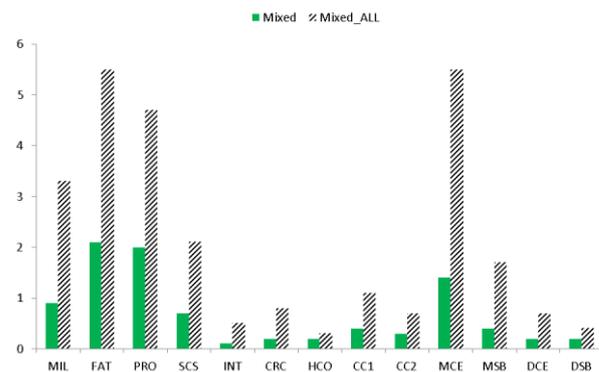


Figure 4. Change in validation R^2 values (x100) for traits other than conformation traits by adding cows to the EuroGenomics bull RP.

Bias of genomic prediction when cows were added to the bull reference population

According to Interbull’s GEBV Test, the bias of genomic prediction was measured by regression coefficient of validation bulls’ deregressed EBV on their candidate GEBV. Figure 5 shows the regression coefficient (b_1) for some selected traits. It can be seen that the b_1 values do not deviate from its expected value of 1 for all the traits. Differences in b_1 values between any two of the three scenarios are rather small, indicating that adding the cows to the bull reference population did not lead to a notable bias in genomic prediction. All the regular national and MACE traits have passed the GEBV Test for scenarios *Bulls*, *Mixed* and *Mixed_All*.



Figure 5. Regression coefficient of validation bulls' deregressed EBV on their candidate GEBV for some selected traits.

Conclusions

Genomic prediction based on the current EurGenomics bull reference population can be improved by adding reference cows that are genotyped non-selectively. Based on the data from two German research projects *Kuh-L* and *GKUHplus*, a comprehensive genomic validation was conducted to investigate the impact of adding reference cows to the current bull reference population. For all national and MACE traits, adding about 20,000 cows to the current reference population with c.a. 35,000 EuroGenomics Holstein bulls always resulted in an increase in the accuracy of genomic prediction. No single trait was seen with a decrease in the validation R^2 value by adding the cows to the bull reference population. A mixed bull and cow reference population gave the highest accuracy for all the traits. The accuracy of the cow reference population was much lower than the other types of reference population, due to the relatively small number of reference cows. For the novel health traits, accuracy of prediction was significantly lower than for the regular traits because of a much smaller reference population. Interbull's GEBV Test criteria have been reached for all the regular traits based on the bull or the mixed reference populations. This study demonstrated clearly the positive impact of adding cows to the current EuroGenomics bull reference population on genomic prediction. Therefore, we recommend using a mixed bull and cow

reference population for routine genomic evaluation of German Holsteins.

Acknowledgements

The German federal Ministry of Education and Research (BMBF) is greatly thanked for financing the research projects. Our acknowledgements extend also to our project partner organizations: German cattle breeding organization FBF, University of Halle (MLU), and University of Kassel.

References

- Liu, Z., F. R. Seefried, R. Reinhardt, S. Rensing, G. Thaller, and R. Reents. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet. Sel. Evol.* 43:19.
- Liu, Z., Aamand, G.P., Fritz, S., and Schrooten, Ch. 2013. Comparison of national genomic predictions of EuroGenomics exchanged young bulls. *Interbull Bulletin.* 47: 38-42.
- Rensing, S., Alkhoder, H., Kubitz, Ch., Schierenbeck, S., and Segelke, D. 2017. Cow reference population – benefit for genomic evaluation system and farmers. *Interbull Bulletin* 51:63-66.
- Wiggans, G.R., Cooper, T.A., VanRaden P.M., and Cole, J.B. 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. *J. Dairy Sci.* 94:6188-6193.