A Continuous Genomic Evaluation System for German Holsteins

H. Alkhoder, Z. Liu, Th. Bach, E. Pasman and F. Reinhardt
vit, Heideweg 1, D-27283 Verden, Germany

Abstract

In March 2014, a continuous genomic evaluation system was introduced in German Holsteins, in addition to routine monthly genomic evaluation. Genomic selection on embryos can lead to a shortened generation interval and substantial savings for breeders. The objectives of this study were to compare genomic breeding values (GEBV) of the continuous with the routine genomic evaluation systems and to investigate the effect of genotype imputation on genomic predictions of embryos. A total of 2440 animals genotyped with Illumina 50K, LD and EuroG10K chips were available between March and April 2014 for comparing genomic evaluations between the two systems. Average differences for all 44 evaluated traits ranged from -4% to 1% genetic standard deviations for direct genomic values (DGV) and from -10% to 5% genetic standard deviations for GEBV. Depending on the traits, correlations of DGV between the two systems were above 0.995, 0.992, and 0.988 for animals genotyped with 50K, EuroG10K and LD chips, respectively. GEBV correlations exceeded 0.98 for all the traits across all the chip types. Due to the genotype imputing step, DGV variances of embryos increased, on average, by 2% to 7% of total additive genetic variance and about half of the variance increase was attributed to Mendelian sampling. For embryos with lower call rates in original genotypes, the increase in DGV variances reached 12% of total genetic variance. The validation study showed that the continuous genomic evaluation system gave highly consistent results as the routine genomic evaluation. Further developments are needed to minimise differences in statistical methods between the two genomic evaluation systems.

Key words: continuous genomic evaluation, genomic breeding value, genotype imputation

Introduction

Since the introduction of genomic evaluation for German Holsteins in August 2010, three major genomic evaluations, including the step of estimation of SNP marker effects, are conducted every year, following each conventional MACE evaluation by Interbull. Between any two consecutive genomic evaluations, breeders are provided with monthly genomic evaluations. As a consequence of continuous genotyping and genomic selection, the German breeding organisations demanded even more frequent genomic evaluations in order to reduce the costs of raising candidates on farm and improve the efficiency of genomic selection programmes. Therefore, a continuous genomic evaluation system was developed, which enables genomic evaluation conducted just in time of reception of genotypes. Thanks to the development of genotyping techniques, genomic selection can now be applied to as early as embryos. The major benefits of the genomic selection at the early life stage of embryos were a shortened generation interval and cost savings for breeders. Because only a very small amount of DNA can be extracted from embryos for genotyping, call rates of genotypes were lower in embryos than real animals, usually between 85% and 95%. The objectives of this study were to investigate the effect of genotype imputation on genomic breeding values of embryos, and to compare genomic predictions of the continuous system with the official routine genomic evaluation.

Materials and Methods

In April 2014 monthly genomic evaluation for German Holsteins, a total of 113,910 genotyped animals were included: 62,486 animals genotyped with the standard 50K Illumina bovineSNP50 version 2 (including embryos), 34,115 animals with Illumina bovineSNP50 chip version 1, 14,221 with EuroG10K chip and 3,081 with Illumina LD chip. A total of 2440 genotyped animals were common between the two evaluations: 985

animals (including 67 embryos) genotyped with 50K chips, 1337 animals with EuroG10K and 134 animals with LD chips. Average call rate for the embryos was about 0.85. A pedigree of the genotyped animals contained 338,761 animals in total.

Genotypes of the lower density chips were imputed with findhap version 2 (VanRaden *et al*., 2011) and missing genotypes of animals genotyped with the standard 50K chips, e.g. for embryos, were filled with the imputation. This genotype imputing step was particularly important for the embryos with a much lower genotype call rate than real animals. In order to investigate thoroughly the effect of genotype imputing on GEBV of embryos, embryos in all preceding genomic evaluations were considered (see Table 1). There were 479 embryos belonging to 174 families from 67 sires and 134 dams. Only 372 embryos had both parents genotyped with the 50K chips.

Table 1. Number of embryos and families selected for this study.

Family size	1	2-3	4-5	>5
Nb families	58	71	26	19
Nb embryos	58	164	112	145

A total of 45,613 SNP markers on the standard 50K v2 chip were used in SNP effect estimation based on a BLUP SNP model with a trait-specific residual polygenic variance (Liu *et al*. 2011). A genomic reference population used for the two genomic evaluations consisted of 27,175 Holstein bulls from EuroGenomics countries. The same SNP markers were used also in the genotype imputation and routine monthly genomic evaluation. A selection index method was used to combine DGV and pedigree index or conventional EBV for the genotyped animals for all the 44 evaluated traits (Liu *et al*. 2011). The resulting DGV and combined GEBV from the monthly evaluation were treated as reference values for validating the continuous genomic evaluation system.

Results & Discussion

To meet the demands of more frequent genomic evaluations, a new continuous, just-in-time system was developed and introduced in March 2014 for genomic evaluation of German Holsteins. New computer source programs were written for the continuous evaluation in Java and SQL, replacing some in Fortran 90 and SAS in the routine monthly genomic evaluation. SNP effect estimates and allele frequencies were obtained from the routine genomic evaluation for calculating DGV in the continuous system. In addition, relevant population parameters for calculating relative breeding values and total merit index were obtained from the corresponding routine genomic evaluation. Pedigree indices and associated reliabilities of the genotyped animals were automatically calculated in vit's own database. When parents were genotyped, parental GEBV, instead of EBV, were used in the calculation of pedigree index, which represented a major difference to the routine genomic evaluation. DGV and pedigree index were combined using the selection index method, as in the routine genomic evaluations, to obtain GEBV for all the 44 traits. Reliability values of GEBV and DGV were calculated in the same way as in the routine genomic evaluation. The calculation method for DGV was kept unchanged for the components of German total merit index (RZG), sub-indices for all trait groups. However, RZG of DGV was computed in the continuous system with a fixed formula for the genotyped candidates which, in contrast, was computed with the selection index in the routine evaluation. The stepwise calculation of the RZG and its sub-indices in the routine evaluations was replaced with a fixed formula for GEBV by treating the indices as if they were the individual evaluated traits, because those indices were available in vit's database system.

Original genotypes of the standard 50K chips, versions 1 or 2, were directly used in the continuous system without being imputed in a prior step. Therefore, animals with the 50K chips were evaluated by the new continuous system just in time receiving the genotypes on vit's ftp servers. For genotypes of the SNP chips other than the standard ones, an additional step of genotype imputing was conducted on a weekly basis, although technically feasible on a daily basis, as currently requested by the German industry. Due to the lower call rates of original genotypes, the imputing step was also applied

to embryos that were mostly genotyped with the standard 50K chips. The imputing software findhap version 2 (VanRaden *et al*., 2011) was used for both the continuous and the routine genomic evaluations. About 2 hours were needed to conduct the imputing with 113,910 genotyped animals on 40 processors in parallel. The RAM usage amounted to c.a. 100Gb in total.

Besides the lower call rates of the original genotypes, embryos had also approximately 3% error rate in their original genotypes, which was considerably higher than that of real animals. Missing genotypes of the embryos, as in case of real animals, were filled via the imputing. We found that error rate of the imputed genotypes decreased for the embryos. However, there was only a small increase in genotypes consistencies with their parents. A linear relationship was observed between the call rate and genotype consistencies with parents: the higher was the call rate and the better genotype consistencies between the embryos and their parents. Genotype call rate after imputing reached 99.9% for all the embryos.

Genotype providers or owners were notified via email as soon as results of the continuous genomic evaluations were available. Same as for routine genomic evaluations, a complete summary of genomic prediction results, among other publication forms, was immediately made available online in PDF or other formats, labelled as interim evaluation results, because the continuous evaluation was not considered to be official yet.

Table 2 shows average differences and correlations of DGV or GEBV of the 2440 common animals between the two systems.

Table 2. Average differences between the continuous and routine monthly genomic evaluations (routine – continuous), expressed as percentages of genetic standard deviation, and correlations for DGV and GEBV of five selected traits.

		Difference (% σ_g)		Correlation	
Trait	Chip	DGV	GEBV	DGV	GEBV
BCS	50K	0.58	-2.00	0.999	0.990
	10K	0.25	-1.17	0.994	0.987
	LD	0.63	-1.53	0.992	0.984
PRO	50K	-4.92	-11.49	0.998	0.996
	10K	-4.15	-10.90	0.991	0.991
	LD	-5.14	-9.05	0.988	0.987
FAT	50K	-1.97	-9.44	0.995	0.997
	10K	-1.79	-8.56	0.990	0.992
	LD	-3.84	-8.94	0.990	0.989
MIL	50K	-4.13	-6.63	0.997	0.996
	10K	-3.47	-6.42	0.992	0.991
	LD	-3.87	-5.26	0.987	0.987
DLO	50K	-0.43	-8.05	0.999	0.982
	10K	-0.82	-6.20	0.993	0.981
	LD	-0.75	-3.25	0.993	0.975

Mean differences between the two systems for the production traits (MIL, FAT and PRO) were relatively small and in the range of -2% σ_g to -5% σ_g for DGV and of -3% σ_g to -11% σ_g for GEBV. The slightly larger differences in DGV for 50K than 10K animals may be explained by the fact that original genotypes of 50K animals were used in the continuous just-in-time system but imputed genotypes in the routine system for computing DGV. For the non-standard chips LD and 10K, the difference in DGV came from two slightly different imputation reference populations, where for the later monthly genomic evaluation the

reference population size for imputation was about 2,000 animals more than that for the continuous system. Additionally, pedigree file for the earlier continuous system was less complete than the later monthly evaluation, in particular for the newly genotyped animals. The DGV differences between the two systems for other traits e.g. body condition score (BCS) or longevity (DLO) were very small because very few SNPs had large effects in comparison to the three milk production traits and wrongly imputed genotypes had, therefore, less impact.

The GEBV differences were greater than the DGV differences because pedigree index was calculated differently between the two systems. The larger differences in GEBV than DGV also indicated that pedigree of the newly genotyped animals was less complete at the time of genomic evaluation with the continuous system than at a later time of the routine monthly evaluation.

The DGV correlations between the just-in-time and routine evaluations were higher than 0.995 for 50K animals and above 0.990 for 10K animals and ranged from 0.988 to 0.995 for LD animals (see Table 2) for the five selected traits. GEBV correlations were slightly lower than DGV ones: ranging from 0.982 to 0.997 for the 50K animals and from 0.975 to 0.992 for the 10K and LD animals.

Figure 1 shows scatter plots of the two sets of DGV stratified by the chip type for somatic cell scores. It can be clearly seen that DGV of common animals with the standard 50K chip are more similar than that of animals with the non-standard 10K or LD chips.

Overall, slightly better validation results were obtained when comparing the continuous with the routine genomic evaluation using new data from May 2014 (unpublished data) than from April 2014.

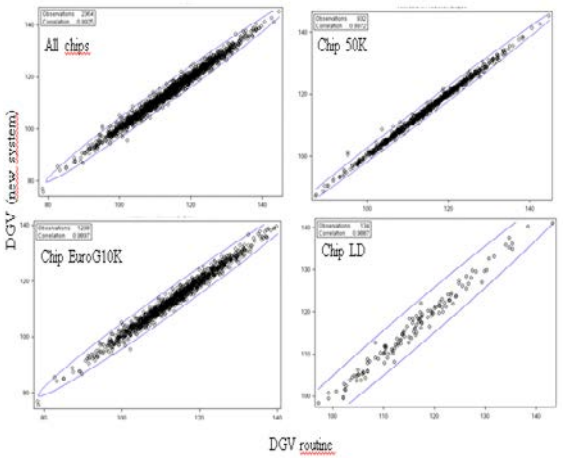

Figure 1. Scatter plots of DGV of the continuous and routine genomic evaluations for somatic cell score, stratified by chip type.

Call rates of original genotypes ranged from 0.41 to 0.99, with mean of 0.85 and standard deviation of 0.11, for the embryos. Missing genotypes of the embryos were filled via genotype imputation, which led to a higher call rate up to 99.9% for the embryos. Figure 2 shows the increase of DGV variance, caused by genotype imputing, for the 372 embryos for 21 selected traits. The DGV variance increase ranged from 2% to 7% of total additive genetic variance, depending on the analysed traits. For embryos with lower call rate of original genotypes, less than 90%, the DGV variance increase was evidently greater than those with higher call rate of original genotypes. This clearly shows that higher genotype call rates or fewer missing genotypes, resulted from the imputation step, made embryos, particularly those full-sibs, more variable in DGV. The differentiated DGV helped breeders make easier selection decision on the embryos.

Figure 3 shows the increase of DGV variance attributed to Mendelian sampling in all the 21 traits for the embryos. On average, half of the DGV variance increase was caused by the Mendelian sampling, as expected. This

indeed reflected the fact that the genotype imputing utilized information from both family pedigree and population haplotypes.

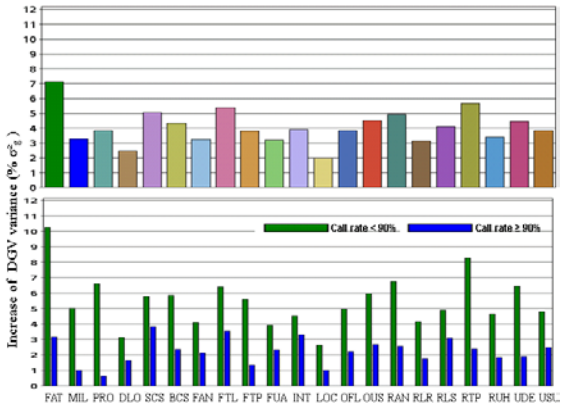

Figure 2. Increase of DGV variance due to genotype imputing for the embryos.

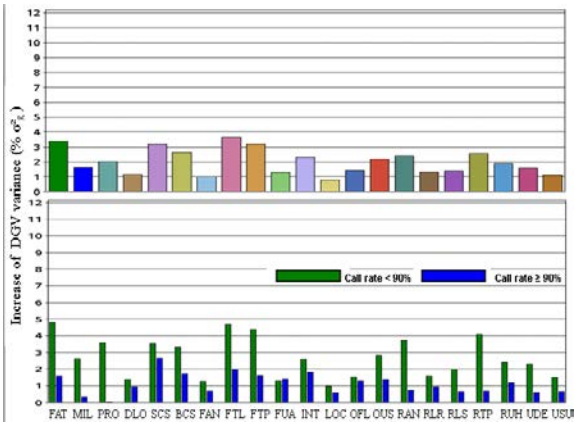

Figure 3. Increase of Mendelian sampling DGV variance due to genotype imputing for the embryos.

Conclusions

In order to reduce the costs and enhance the efficiency of genomic selection, German dairy industry demanded more frequent genomic evaluation than on a monthly routine basis. Therefore, a continuous genomic evaluation system was developed and implemented for German Holstein in March 2014. Animals with genotypes of the standard 50K chips received immediate genomic evaluation results just in time of genotype reception; whereas genotypes of non-standard lower density chips must first be imputed to the standard 50K basis for the just-in-time evaluation. To validate the accuracy of the new, continuous genomic evaluation system, 2440 common animals evaluated in both systems were selected and their DGV and GEBV were compared. Overall, DGV or GEBV were highly correlated between the two evaluations with very small differences, despite the fact that pedigree indices were calculated differently. Accuracy of the continuous genomic evaluation depended also on the pedigree completeness at the time of receiving genotype. Because genomic selection has been extended to embryos, which usually had lower genotype call rates than real animals, the impact of genotype imputing on DGV of the embryos was investigated. Genotype imputing increased the variance of DGV significantly for the embryos, which allowed easier and more accurate genomic selection decision on embryos, also among full-sibs. In summary, the continuous genomic evaluation system was proven to be highly consistent with the routine monthly system. Similar validation studies should be conducted routinely to further examine performance of the new system. Special attention must be paid to the second-generation candidates whose sires have no own phenotypes either, because pedigree indices are calculated differently for this group of animals. Further optimization is needed to minimize the differences in statistical methods between the two genomic evaluation systems.

References

Liu, Z., Seefried, F.R., Reinhardt, F., Rensing, S., Thaller, G. & Reents, R. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. 2011. *Genet. Sel. Evol. 43:19.*

VanRaden, P.M., O'Connell, J.R., Wiggans, G.R. & Weigel, K.A. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol. 43:10.*