GENOMIC SELECTION IN FLECKVIEH WITH LOW DENSITY SNP PANELS

Gábor MÉSZÁROS [1], Elisabeth KIRNBAUER [1], Hermann SCHWARZENBACHER [2], Johann SÖLKNER [1]

ABSTRACT

The aim of this study was to compare accuracies of genomic breeding values (GEBV) based on genotypes from 2123 Austrian Fleckvieh bulls using different SNP panels. Highly reliable estimated breeding values (EBV) and deregressed breeding values for fat content, fat yield, protein content, longevity, non-return rate after 56 days, somatic cell score and dressing percentage were used as phenotypes. The initial genotype data originated from the Illumina Bovine SNP50 BeadChip (50k), from which after quality check 41082 SNP remained. These were further reduced to obtain the 3k (2890 SNP) and the 7k (6565 SNP) SNP panels. GBLUP and BayesB methodologies were used to obtain GEBVs. Correlations between the EBVs, deregressed EBVs and GEBVs increased with SNP panel density. For EBVs the correlations ranged from 0.44–0.71 for the 50k panel in all cases. In case of deregressed EBVs the correlations were 0.40–0.64 for most of the traits, but only 0.15–0.26 for longevity and non-return rate. Correlations from the 7k SNP panel were very close to those from the 50k set, differing only by 0.04–0.09 in EBVs and 0–0.06 in deregressed EBVs.

Key words: cattle / breeds / Fleckvieh / genomic selection / genomic breeding value / accuracy

1 INTRODUCTION

Genomic selection became a standard method to estimate the breeding values in livestock, most often used in dairy cattle. The main goal of genomic selection, as described by Meuwissen *et al.* (2001) is to exploit linkage disequilibrium between quantitative trait loci (QTL) and high density markers (SNP) across the genome for breeding value estimation in genetic improvement in livestock (Habier *et al.*, 2009). Goddard and Hayes (2007) describe the process in three steps, assuming that one treats the markers as if they were QTL and estimate the effects of the marker alleles or genotypes: 1. Use the markers to deduce the genotype of each animal at each QTL; 2. Estimate the effects of each QTL genotype on the trait; 3. Sum all the QTL effects for selection candidates to obtain their genomic EBV (GEBV). While conventional progeny testing programs are still ongoing, GEBVs are heavily influencing the selection of sires.

The most widespread is the Illumina Bovine SNP50 BeadChip consisting from 54001 SNPs, but both higher (777k) and lower (3k, 7k) density SNP chips are available. When implementing a genotyping strategy in large populations, one should consider the tradeoff between the price for genotyping and the accuracy of prediction. While the costs are lower for low density chips the accuracy of the GEBV estimation is higher for denser chips because of the better coverage of the genome by SNP markers. It has to be noted however, that most results indicate an increase in accuracies 1–2% when using HD vs. 50k data.

In this study our aim is to predict GEBV using 3k, 7k and 50k SNP panels and compare accuracy of the prediction using correlations between GEBV and the phenotype (EBVs and deregressed EBVs).

---

1 Univ. of Natural Resources and Life Sciences (BOKU), Division of Livestock Sciences, Gregor-Mendel Str. 33, A-1180 Vienna, Austria
2 ZuchtData EDV-Dienstleistungen GmbH, Dresdner Straße 89/19, A-1200, Vienna, Austria

2   MATERIALS AND METHODS

The dataset consisted from 2123 Fleckvieh bulls, genotyped by Illumina Bovine SNP50 BeadChip. From the initial 54001 SNP makers only 41082 remained the rest was deleted not fulfilling at least one of the criteria: minimal call rate for marker 0.95, minimum minor allele frequency 0.02, and check for Hardy-Weinberg equilibrium. The 3k (2890 SNPs) and the 7k (6565 SNPs) data sets were derived from the quality checked data, keeping the SNPs present in the official 3k and 7k map files.

The data set was divided into reference and test set, the first one with 1637 animals, the latter consisting from 486 animals born after 2002. The strategy to select the youngest animals to the test set was in order to circumvent estimation of GEBVs of sires with their sons in the reference set. This would lead to seemingly high accuracies, but is in fact caused by the data structure, observing the phenotype of the son in the training set.

Estimated breeding values and deregressed breeding values with high accuracies were used as phenotypes, based on progeny testing results from the joint routine genetic evaluation in Austria and Germany. The traits were fat content, fat yield, protein content, longevity, non-return rate after 56 days, somatic cell score and dressing percentage. The deregression procedure of the EBVs removed the contribution of relatives other than daughters to the breeding value (Garrick *et al.*, 2009).

GBLUP and BayesB methodologies were applied, using the bayesgg program provided by Theo Meuwissen. While the BayesB applies a Bayesian mixture model where only a certain proportion of the SNPs are considered to have an effect, the GBLUP considers all SNPs. In BayesB we experimented with different proportions of important SNPs but the results were extremely similar, regardless of the phenotype. From this reason we present only the results with 10% important SNP in this paper. EBVs were used in both training and test sets when correlations between GEBVs and EBVs are reported. For

obtaining the correlations between deregressed breeding values and GEBVs the deregressed breeding values were used in both training and test sets.

3   RESULTS AND DISCUSSION

Correlations between the conventional EBVs and GEBVs for the 486 test animals, in all traits and all SNP panels are shown in table 1, and those between deregressed EBVs and GEBVs are shown in table 2. These correlations in the reference animals in the 50k set were close to 1 in all cases for both GBLUP and BayesB. For BayesB however the correlations in the reference sets were in the range of 0.79–1.00 in case of the 3k panel and 0.91–1.00 in case of the 7k panel. In both cases the lower values were associated with low heritable traits such as longevity and non-return rate.

In case of the conventional breeding values (table 1) the pattern was similar for all traits, with increasing accuracy for SNP panels with higher density. Accuracies for the 50k panel were the highest and similar for all traits (around 0.5–0.6) with a notable exception of fat content using the BayesB model, reaching accuracy of 0.7. Accuracies from BayesB were higher for fat production, fat content and protein content, with the exception of the 50k set, which yielded similar results compared to GBLUP. There were only minor differences in accuracies between the low and moderate heritable traits when using EBVs as phenotypes.

Several simulation studies had shown that high accuracies could be achieved when using SNP data to estimate genomic breeding values. In the milestone paper Meuwissen *et al.* (2001) used ~50.000 markers to calculate genomic breeding values with 0.73 accuracy using BLUP and 0.85 with BayesB. Calus *et al.* (2008) also reported accuracies of 0.83 (traits with $h^2 = 0.5$) and 0.66 (traits with $h^2 = 0.1$).

Correlations from studies using real, instead of sim-

Table 1: Correlations between conventional EBVs and GEBVs for the test set using different SNP panels

	3k		7k		50k	
	BayesB	GBLUP	BayesB	GBLUP	BayesB	GBLUP
Dressing percentage	0.355	0.341	0.431	0.430	0.520	0.521
Protein content	0.404	0.351	0.461	0.405	0.452	0.475
Fat production	0.303	0.253	0.448	0.393	0.491	0.440
Fat content	0.549	0.488	0.671	0.550	0.706	0.555
Longevity	0.317	0.337	0.437	0.450	0.496	0.511
Non-return rate	0.492	0.427	0.527	0.518	0.578	0.582
Somatic cell score	0.430	0.378	0.449	0.437	0.511	0.511

Table 2: Correlations between deregresed EBVs and GEBVs for the test set using different SNP panels

	3k		7k		50k	
	BayesB	GBLUP	BayesB	GBLUP	BayesB	GBLUP
Dressing percentage	0.275	0.277	0.431	0.278	0.441	0.250
Protein content	0.364	0.214	0.408	0.313	0.473	0.403
Fat production	0.294	0.266	0.438	0.395	0.464	0.430
Fat content	0.505	0.446	0.634	0.503	0.640	0.503
Longevity	0.113	0.144	0.190	0.183	0.150	0.178
Non-return rate	0.182	0.143	0.231	0.234	0.250	0.260
Somatic cell score	0.370	0.377	0.387	0.398	0.417	0.446

ulated data were lower though. Possible reason could be that while simulation studies report correlations between true EBVs and GEBVs, while real data are based on estimated breeding values. De Roos *et al.* (2011) reported accuracy of 0.63 for low heritable traits of fertility index, non-return rate and longevity, but correlations milk production traits were higher with an average of 0.76 in the Dutch Holstein population. Similarly VanRaden *et al.* (2009) achieved accuracies of 0.71 in the North American Holstein bulls.

Our results using the 50k set were consistently the highest among the SNP panels, but were somewhat lower compared to those presented in the literature, even when looking at the non-simulation results. One of the reasons could be the considerably higher number of genotypes in the cited studies. In similar sized study in Austrian Simmental cattle Gredler *et al.* (2009) compared several model types, from which GBLUP is a common methodology with our work. Their accuracies for fat content, protein yield, non-return rate and somatic cell score were in the range of 0.42–0.47. The BayesC method used by Gredler *et al.* (2009) resulted into accuracies 0.43–0.60, which is well in line with our outcomes of 0.45–0.58 from BayesB, with an outlier of 0.71 for the fat content.

Accuracies of GEBVs based on deregressed EBVs (table 2) were less straightforward compared to those using EBVs. In majority of cases the accuracies based on denser chip panels gave higher accuracies. In case of dressing percentage however, the GBLUP results had the same low accuracy with all 3 chip panels. In case of protein content, fat production and content the results were higher using BayesB, in case of somatic cell score GBLUP appeared to give somewhat higher results. When using deregressed breeding values as phenotypes for the low heritable traits longevity and non-return rate, the accuracies from both BayesB and GBLUP were low, regardless of SNP panel density.

These results agree with those of Moser *et al.* (2010) who used different subsets of SNP data to calculate cor-relations between the GEBV and the phenotype (deregressed breeding values). In general the production traits were more accurate (0.52–0.64) than for survival (0.19–0.20). They noted that the heritability of the respective trait might have an influence on the correlation. In our case the results were slightly lower and more diverse, in the range of 0.40 to 0.64 for the traits with higher heritability and 0.15–0.26 for low heritable traits of longevity and non-return rate.

In all cases the results from the 7k SNP panel were very close to those from the 50k, with only marginal differences in accuracies in the test sets. The differences were in the range of 0.04–0.09 when of EBVs were used as phenotypes, and 0–0.06 in case of deregressed EBVs in all traits. The results did not fully support the conclusions of Moser *et al.* (2010), who suggested that with 3000 evenly placed SNPs 90% of the accuracy from the 50k panel could be achieved. Our results show that correlations from the 3k SNP panel are lower by 0.1–0.2, depending on the trait. At the same time the results from the 7k chip allow us to speculate about their large scale usage, with eventual imputation to 50k or higher (Dassonneville *et al.*, 2011).

4 CONCLUSIONS

Genotype data from 2123 Fleckvieh bulls were used to assess accuracies of genomic breeding values based on conventional breeding values and deregressed proofs for seven traits. After quality check 41082 SNPs remained from the initial Illumina Bovine SNP50 BeadChip data, from which the 3k and 7k SNP panel data were extracted based on the official map files. GBLUP and BayesB methodologies were used to obtain genomic breeding values.

When comparing GEBVs with the corresponding EBVs of bulls, the results were stable, on similar level of 0.5-0.6 for 50k panels for all traits. Accuracies were increasing with panel size. In case of deregressed breeding

values the low heritable traits showed lower accuracies of GEBVs with all three SNP panels. For fat content, fat production and protein content the BayesB method resulted into consistently higher accuracies.

5   ACKNOWLEDGEMENTS

6   REFERENCES

Calus M.P.L., Meuwissen T.H.E., De Roos A.P., Veerkamp R.F. 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. Genetics, 178: 553–561

Dassonneville R., Brøndum R.F, Druet T., Fritz S., Guillaume F., Guldbrandtsen B., Lund M.S., Ducrocq V., Su G. 2011. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. Journal of Dairy Science, 94, 7: 3679–3686

De Roos A.P.W., Schrooten C., Druet T. 2011. Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. Journal of Dairy Science, 94: 4708–4714

Garrick D.J., Taylor J.F., Fernando R.L. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution, 41: 55

Goddard M.E., Hayes B.J. 2007. Genomic selection. Journal of Animal Breeding and. Genetics, 124: 323–330

Gredler B., Nirea K.G., Solberg T.R, Egger-Danner C., Meuwissen T., Sölkner J. 2009. A comparison of methods for genomic selection in Austrian dual purpose Simmental cattle Association for the Advancement of Animal Breeding and Genetics. In: Proceedings of the 18th Conference, Barossa Valley, South Australia, 28 September – 1 October 2009: 568

Habier D., Fernando R.L., Dekkers J.C.M. 2009. Genomic Selection Using Low-Density Marker Panels. Genetics, 182: 343–353

Meuwissen T. H. E., Hayes B.J., Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819–1829

Moser G., Khatkar M.S., Hayes B.J., Raadsma H.W. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genetics Selection Evolution, 42: 37