Reliabilities of genomic estimated breeding values in Danish Jersey

J.R. Thomasen[*], B. Guldbrandtsen[†], G. Su[†], R. F. Brøndum[†] and M.S. Lund[†]

Introduction

Within the last two years genomic selection has been used for selection of breeding candidates of dairy cattle in many countries.

In Genomic Selection (GS) marker effects are estimated in a genotyped reference population where registrations of phenotypic data are available. The marker alleles observed in a candidate are used to construct a predictor for the breeding value of the candidate. This is called the direct genomic value (DGV). GS can be performed as soon as the DNA is available. This opens up for more accurate selection of young breeding candidates. Schaeffer (2006) argued that in a realistic scenario genetic progress would be doubled.

In order to apply GS to practical breeding programs, it is important to know the reliability of the genomic predictions. As the reliabilities of DGV are highly dependent on the number of bulls in the reference population, the heritability of the trait and the genetic structure of the population (Hayes et al., 2008a), it is important to evaluate the reliabilities of the genomic predictions in the population from which breeding candidates are selected. Most studies until now on real data have reported results from Holstein populations (Hayes et al., (2008a); VanRaaden et al., (2009); Su et al., (2010)), whereas only Harris and Johnson (2010) and Hayes et al. (2009) have reported results from Jersey populations.

The aim of this study is to investigate the reliability of DGV for economically important traits in the Danish Jersey population. Three different methods for investigating the reliability of DGV are studied.

Material and methods

Data. The Jersey bulls analyzed were born from 1984 to 2004 and came from 106 paternal half-sibs families (1-71 bulls each) and were genotyped using the Illuminia Bovine SNP50 BeadChip (Illumina, San Diego, CA). After marker data quality checking 1,002 bulls and 33,524 SNP markers were available.

Two datasets were used in this study. The first was published EBV from September 2009. The other was published EBV from June 2006, which was used as reference dataset to predict DGV of the bulls having daughter records after June 2006.

[*] VikingGenetics, Ebeltoftvej 16, Assentoft, DK-8960 Randers SØ, Denmark
[†] Aarhus University, Faculty of Agricultural Sciences, Department of Genetics and Biotechnology, DK-8830 Tjele, Box 50, Denmark

Statistical analyses. All SNP were used as individual predictors. Conventional EBV were used as response variables weighted by a function of reliability of EBV as 1/(1 – reliability of EBV). The following model was used to fit EBV data:

$$\boldsymbol{y} = \boldsymbol{1}\mu + \sum_{i=1}^{m} \boldsymbol{X}_i \boldsymbol{q}_i v_i + \boldsymbol{e}$$

where $\boldsymbol{y}$ is the vector of published conventional EBV, μ is the intercept, $\boldsymbol{1}$ is a vector of ones, m is the number of SNP markers, $\boldsymbol{X}$ is the design matrix allocating genotypes to the animals, $\boldsymbol{q}_i$ is the vector of scaled SNP effects (scaled by standard deviation) of marker i with $\boldsymbol{q}_i \sim N(\boldsymbol{0}, \boldsymbol{I})$, v_i ($v_i > 0$) is a scaling factor for SNP effects of marker i, and **e** is the vector of residual with $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_e^2)$. The effects of SNP alleles of marker i are the products of v_i and $\boldsymbol{q}_i$. Scaling factors v_i were assumed to have a common prior distribution across the markers of chromosome segment effects,

$$v_i \sim TN(0, \sigma_v^2), \quad v_i > 0$$

DGV for individual k was defined as the sum of predicted effects of SNP over all markers,

$$\mathrm{DGV_k} = \hat{\mu} + \sum_{i=1}^{m} \mathbf{x}_{i(k.)} \mathbf{q}_i v_i$$

A more detailed description of the Bayesian model is given in Villumsen et al. (2009) and Su et al. (2010)

Reliability of DGV. Three different methods were used to investigate the reliability of the DGV. 1) Five-fold cross validation 2) Prediction of DGV for the last 3-year of bulls with EBV. 3) Model estimated reliabilities.

Five-fold cross validation: The reference bulls were divided into five nearly equally sized subsets (184 to 205) according to birth year. Half-sib families having sons in more than one subset were moved to the same subset. Cross-validations were performed by omitting EBV in one subset (test data) from the whole dataset per fold of validation. The reliability of DGV was estimated as the within-year squared correlation between EBV and DGV in the five test data sets. In order to reduce strong dependency between reference data and test data, the bulls in the test data which had sons in the reference data were removed from the calculations (63 bulls).

Last 3-year validations: The 860 bulls with official EBV in June 2006 were used as reference bulls. 138 Bulls born in 2002-2004 with official EBV in September 2009 were assigned as test bulls. The reliability of DGV was estimated as the squared correlation between EBV and DGV for the 138 test bulls.

Model estimated reliabilities: Model estimated reliabilities were obtained from prediction error variance (PEV) following Su et al. (2010). The PEV for a DGV was calculated as the variance of the posterior samples of each DGV. In the context, this measure of reliability was denoted as expected reliability. The expected reliabilities were calculated for the 138 young bulls in the "Last-3 year validations" analysis.

The IBay package v1.46 (Janss, 2009) was used to estimate the marker SNP effects. The Gibbs sampler was run as a single chain with a length of 50,000 iterations. Samples from the first 10,000 iterations were discarded as burn-in. Every 5th sample of the remaining 40,000 was saved to estimate the parameters of the realized posterior distributions.

Results and discussion

Table 1 shows calculated reliabilities of DGV for the traits protein, udder-health, fertility, udder conformation and longevity. The analyzed traits are the main index traits with the highest economic weights in the Nordic total merit index (Nordic Genetic Evaluation, 2010).

Reliabilities calculated from the five-fold validation method range from 0.24 to 0.50 with an average of 0.30. Reliabilities calculated from the last three years of bulls with EBV are on average 0.27, which is marginally lower than the five-fold validation method. The variation between traits are much smaller (0.24 – 0.29) in the last 3-year validation. Expected reliabilities give the highest reliabilities and range from 0.34 to 0.48 with an average of 0.41.

Heritabilities (Danish Cattle Federation, 2006) for the component traits that are included with economic weight in the main index are also given in table 1. Although the heritabilities vary considerably between traits there is not a strong connection between heritability and calculated reliability.

Table 1: Reliabilities of DGV in Danish Jersey predicted from different validations methods

	Five-fold	Last 3 years	Expected reliabilities	Heritability
No of ref bulls	1002	864		
No of test bulls	1002	138	138	
Protein	0.25	0.26	0.39	0.23-0.38
Udder-health	0.50	0.25	0.48	0.01-0.03
Fertility	0.24	0.29	0.47	0.01-0.04
Udder-conf.	0.28	0.29	0.34	0.17-0.42
Longevity	0.24	0.24	0.35	0.12
Mean	0.30	0.27	0.41	

The measures of reliabilities depend on the validation methods. The reliabilities obtained from the 5-fold validation are similar to those from the last 3-year validation, except for udder-health for which the 5-fold validation gives a much higher reliability than the last 3-year validation. However, expected reliabilities are much higher than those from the cross validations. The possible reasons for the big difference could be: 1) expected reliabilities calculated from the model may overestimate the reliabilities (Su et al., 2010); 2) reliability from the cross validations may be underestimated because the comparison is based on EBV which contains error, and the test bulls have been selected instead of using a random sample. The true reliability of DGV for the Danish Jersey may be in the area between 0.27 and 0.41.

The calculated reliabilities in this study are lower than the reliabilities reported for the Nordic Holsteins with an average of 0.42 (Su et al., 2009), but the analysis used a much bigger reference population consisting of 3,330 bulls. Harris and Johnson (2010) reported an average reliability of 0.54 for 4 traits in the New Zealand Jersey population with 1738 reference bulls using a mixed linear model. The level of the reliabilities in Danish Jersey is expected to increase with increasing size of future reference population, according to the demonstration about reference populations by Hayes et al. (2008b).

The reliability is expected to increase by increasing the size of future reference data and by blending with information from the conventional pedigree index.

Conclusion

Averaged over the 5 traits, the reliability of genomic prediction using the current reference data is in the range between 0.27 and 0.41, despite the small size of the reference data. In this study no connection was found between the heritability of the trait and the calculated reliability. For most traits "five-fold" and "last 3 years" validation predicted similar reliabilities, while "expected reliabilities" predicted higher reliabilities.

Acknowledgments

This work was performed in the project "Genomic Selection – from function to efficient utilization in cattle breeding (grant no. 3412-08-02253)", founded by Danish Directorate for Food Fisheries and Agri Business, VikingGenetics, Nordic Genetic Evaluation and Aarhus University.

References

Danish Cattle Federation (2006). http://www.lr.dk/kvaeg/diverse/principles.pdf

Harris, B. L and Johnson, D. L. (2010). *J. Anim. Sci*., 93:1243–1252.

Hayes, B. L., Bowman, P. J, Chamberlain, A. J. *et al.* (2008a). *J. Anim. Sci*., 92:433–443.

Hayes, B. L., Daetwyler, H. D., Bowman P. *et al.* (2008b). *Proc. Assoc Advmt. Anim. Breed. Genet*., 18:34-37

Janns, L. (2008). iBAY manual version 1.44. http:/www.lucjannss.com

Meuwissen, T. H. E., Hayes, B. L. and Goddard, M. E. (2001). *Genet.,*157:1819–1829.

Nordic Genetic Evaluation (2010). http://www.nordicebv.info/Forside.htm

Su, G., Guldbrandtsen, B., Gregersen, V.R. *et al.* (2010). *J. Anim. Sci*., 93:1175-1183.

Schaeffer, L.R. (2006*). J. Anim. Breed. Genet.*, 123:218–223.

VanRaden, P. M., Van Tassel, C. P., Wiggans, G. R. *et al.* (2008). *J. Anim. Sci*., 92:16–24.

Villumsen, T., Janss, L., Lund, M. (2009). *J. Anim. Breed. Genet.,* 126:3-13