Use Of Genomic SNP Information In Pig Breeding

B. Nielsen[*], T. Ostersen[*], G. Su[†], O.F. Christensen[†] and M. Henryon[*]

Introduction

The DNA typing technology has now been developed to a stage where it is ready for implementation in calculation of breeding values as well as in the design of breeding schemes. It is expected that in the near future, pig breeding companies will increasingly use genomic information.

Based on this, a combination of SNP information with classical BLUP evaluation seems appealing. In future breeding programs the existing sources of information from pedigree information and performance data should be combined with the genomic information from SNP typing. Methods have been proposed to do so using all the data available (Aguilar et al., 2010; Christensen and Lund, 2010; Van Raden et al., 2009; VanRaden, 2008). The one-step approach suggested by Aguilar et al. (2010) and Christensen and Lund (2010) is expected to be superior to the simple index approach. However, before that approach is confirmed to be ready for implementation in routine genetic evaluation, a more simple animal model approach can be used as a transition.

The objective of this paper is to combine genomic breeding values with traditional BLUP breeding values and to show how genomic information affects the predicted breeding values for both genotyped and non-genotyped animals.

Material and methods

The method presented here for combining genomic breeding values with traditional BLUP breeding values involved three steps. First, a prediction model associating SNP markers and traditional BLUP-breeding values was fitted on a training data set of old animals. Second, genomic breeding values were predicted for the young animals in the test data set using the fitted model from step one. In the third step the genomic breeding values of young animals in the test data set were included in a bivariate animal model in which trait 1 was phenotype recordings for all animals, and trait 2 was estimated genomic breeding values of the young animals.

A genomic SNP model A prediction model associating SNP markers and traditional BLUP-breeding values was first fitted using a linear regression model with a random coefficient for each SNP marker. The overall distribution of all coefficients for the SNP markers was assumed to be Gaussian.

Let x_{ij} be the recorded gene content of animal i at SNP j such that x_{ij} equals -1, 0, and 1 for the homozygote, the heterozygote, and the other homozygote, respectively. Then let y_i be

---

[*] Pig Research Centre, Axelborg, Axeltorv 3, DK-1609 Copenhagen V, Denmark

[†] Aarhus University, Faculty of Agricultural Sciences, Department of Genetics and Biotechnology, Denmark

the observed value of the random variable Y_i denoting the estimated BLUP-breeding value of animal i. The regression model with random coefficients used is then given by

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_{1j} x_{ij} + \varepsilon_i \tag{1}$$

where m is the number of SNP's, β_0 is an intercept, $\beta_{1j} \sim N(\beta_1, \sigma^2_{\beta_1})$ is the coefficient of SNP j, and $\varepsilon_{ij} \sim N(0, \sigma^2_{\varepsilon})$ is the random noise. A Gibbs sampling procedure implemented in the BAMD R-package (Gopal et al., 2009) was used.

Incorporating genomic SNP information into a BLUP evaluation approach To combine genomic SNP information with traditional BLUP evaluations into a joint animal breeding value a bivariate model was used. Consider trait 1 as the traditional trait with phenotypic recordings and trait 2 as the estimated genomic breeding values. Ordering the data by traits then a bivariate model is

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{b}_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{Z}_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix}, \tag{2}$$

where
$\boldsymbol{y}_1$: vector of phenotypes.
$\boldsymbol{y}_2$: vector of estimated genomic breeding values.
$\boldsymbol{b}_1$: vector of fixed effects for the phenotypes.
$\boldsymbol{u}_1$: vector of random effects for the phenotypes.
$\boldsymbol{a}_1$: vector of random animal effects for the phenotypes.
$\boldsymbol{a}_2$: vector of random animal effects for estimated genomic breeding values.
$\boldsymbol{e}_1$: vector of random residuals for the phenotypes.
$\boldsymbol{e}_2$: vector of random residuals for the estimated genomic breeding values.
$\boldsymbol{X}_1$ and $\boldsymbol{Z}_1$ are incidence matrices relating records of the trait to fixed and random animal effects respectively.

It is assumed that

$$var \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{V} & 0 & 0 & 0 & 0 \\ 0 & g_{11}\boldsymbol{A} & g_{12}\boldsymbol{A} & 0 & 0 \\ 0 & g_{21}\boldsymbol{A} & g_{22}\boldsymbol{A} & 0 & 0 \\ 0 & 0 & 0 & r_{11} & r_{12} \\ 0 & 0 & 0 & r_{21} & r_{22} \end{pmatrix}$$

where $\boldsymbol{V}$ is a covariance matrix of the random effects, $\boldsymbol{A}$ is the relationship matrix indicating the additive genetic relationship among individuals, g_{ij} are elements of the additive genetic covariance matrix, $\boldsymbol{G}$ for the animal effects, and r_{ij} are elements of $\boldsymbol{R}$ denoting the covariance matrix for the residual effects. Each element in $\boldsymbol{G}$ is defined as: g_{11} is the additive genetic variance for the direct effect for the trait, $g_{12} = g_{21}$ is the covariance between the additive genetic values of the trait and the genomic breeding values of the trait, and g_{22} is the additive genetic variance for the genomic breeding values of the trait. The elements in $\boldsymbol{R}$ are defined as: r_{11} is the variance for the residual effect of the trait, $r_{12} = r_{21}$ is the the residual covariance between the trait and the genomic breeding values of the trait, and r_{22} is the residual variance for the genomic breeding values of the trait.

The estimates of g_{11} and r_{11} as well as the estimate of b_1 and u_1 were obtained from a single trait animal model for y_1. The estimates of g_{22}, g_{12} and r_{22} were obtained from a cross validation study (results not shown). Here the correlation between genomic breeding breeding values and BLUP breeding values range from 0.59 to 0.72 suggesting that $g_{22}/(g_{22} + r_{22})$ range from 0.35 to 0.52. The covariances $r_{12} = r_{21}$ were assumed to be zero.

Animal and SNP data Blood samples of 970 Landrace boars were used for identification of genomic information using a 6K Illumina chip, from which 2299 SNPs had good quality and minor allele frequency larger then 1%. The boars were selected among the highest ranking boars with the highest number of progeny tested offspring concerning back fat recordings.

The 970 boars were divided into a training data set of 800 old boars born before January 2006 and a test data set of 170 young boars born in 2006 and 2007.

Traditional BLUP-breeding values of log-backfat were obtained using a single trait animal model with 135894 recorded animals and 148976 animals in the pedigree in which 440 animals cover the base population. This model corresponds to trait one in (2). The log-transformation was used to stabilise the the variance across the range of predicted backfat thickness (0.4mm to 18mm). Hence, the heritability was 0.53. The animal model was estimated using VCE5 and BLUP breeding values were obtained by PEST (Neumaier and Groeneveld, 1998).

The training data of the old boars was used to estimate SNP effects in the prediction model (1).

Results and discussion

Figure 1 presents the relationship between genomic breeding values and traditional BLUP breeding values for the 170 young boars in the test data set. Data shows a linear relationship between the two variables with a correlation of 0.628, suggesting that the reliability of genomic breeding values was about 0.40.

Use of genomic SNP information in a BLUP evaluation approach The new estimated breeding values of all 148976 animals predicted using the bivariate model (2) showed small deviations from the traditional estimated breeding values (Figure 2). The mean absolute deviation for the genotyped animals was 0.0015, and as expected this was smaller than the mean absolute deviation between raw estimated genomic breeding values and traditional estimated breeding values, which was 0.0523 for the genotyped animals.

As shown in Figure 2, the deviations of the new breeding values from the traditional breeding values were centered at zero, no matter the traditional breeding values, suggesting that the new breeding values were unbiased. The deviations reflected the contribution from SNP information. It was observed that the deviations appeared in both genotyped and non-genotyped animals, though larger for the genotyped animals than the non-genotyped animals. These results suggested that breeding values predicted from the bivariate model could have higher reliability than traditional breeding values not only for genotyped animals, but also for non-genotyped animals.

Conclusion

Our results show that genomic breeding values predicted using a genomic SNP model can be used as an extra trait extending a single trait BLUP evaluation into a bivariate BLUP evaluation model. The breeding values of all animals were affected through the genetic correlations between the two traits obtained by the pedigree information. The deviations appeared in both genotyped and non-genotyped animals, though larger for the genotyped animals than the non-genotyped animals. The results suggested higher reliability of the breeding values predicted from the bivariate model than traditional breeding values not only for genotyped animals, but also for non-genotyped animals.

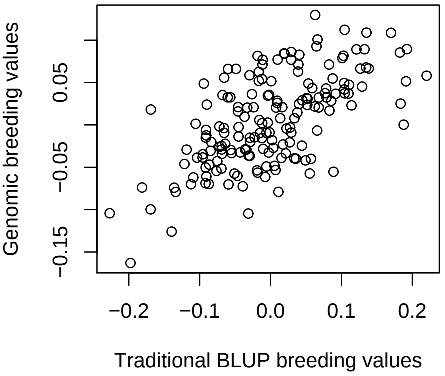

Figure 1: Genomic breeding values v.s. traditional BLUP breeding values of the 170 young boars in the test data set (correlation: 0.628).

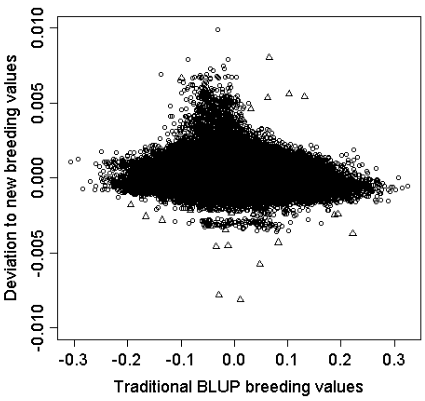

Figure 2: Deviation between traditional and new breeding values predicted by the bivariate model of animals with SNP information (△) and without (○).

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). *Journal of Dairy Science*, 93:743–752.

Christensen, O. and Lund, M. S. (2010). *Genetics Selection Evolution*, 42:2.

Gopal, V., Li, Z., and Casella, G. (2009). R package version 1.2.

Neumaier, A. and Groeneveld, E. (1998). *Genetics Selection Evolution*, 30:3–26.

Van Raden, P., Tassell, C. V., Wiggans, G., Sonstegard, T., Schnabel, R., Taylor, J., and Schenkel, F. (2009). *Journal of Dairy Science*, 92(1):16–24.

VanRaden, P. M. (2008). *Journal of Dairy Science*, 91:4414–4423.