Use of Mendelian Sampling Terms in Genomic Models

J. Chen[*], F. Ytournel[*], M. Erbe[*], E.C.G. Pimentel[*] and H. Simianer[*]

Introduction

In breeding programs, animals are usually selected based on breeding values that are estimated using mixed model methodology with best linear unbiased prediction (BLUP) properties. This method has not only accelerated genetic gain but also increased inbreeding rate per generation. Consequently, it increases the risk of losing genetic variation (Sørensen *et al.*, 2005). In some livestock (e.g., Holstein), inbreeding is expected to be a problem in the future, given that few prominent sires are very intensively used. The breeding value of an individual can be decomposed into three components: half the breeding value of its sire, half the breeding value of its dam, and the Mendelian sampling term. Mendelian sampling terms play a major role in determining contributions under truncation selection (Avendano *et al.*, 2004). With conventional methods it is not possible to predict Mendelian sampling terms of individuals when there are no available records on their progeny. Avendano *et al.* (2005) predicted the accuracy of Mendelian sampling terms using selection index methods. Simianer and Pimentel (2009) proposed the application of a quantitative transmission disequilibrium test (QTDT) to the estimated Mendelian sampling terms for QTL fine mapping. The increasing availability of large numbers of single nucleotide polymorphisms (SNPs) distributed across the genome may lead to further development of genetic evaluation methods and increase the accuracy of predicted breeding values (Meuwissen *et al.*, 2001; VanRaden *et al.*, 2008; Misztal *et al.*, 2009). The purpose of this study was to investigate the use of Mendelian sampling terms in genomic models with respect to maintenance of diversity and accuracy of prediction.

Material and methods

Simulated data. The population structure was simulated by first creating a founder generation with 500 males and 500 females, which were mated at random to produce the next generation. This process was repeated for 1000 generations keeping the population size constant at 1000. In generation 1001 a bottleneck was simulated and within the subsequent 10 generations the population size was progressively reduced, down to 100 individuals in generation 1010. In generation 1011 the population size was increased to 500 individuals, which were again mated at random to produce the next generation. This was repeated for 14 more generations with population size constant at 500 per generation. The data sets used in the analyses were the last 5 generations, for which breeding values and phenotypes were created. Breeding values were computed as the sum of QTL effects, drawn from a gamma distribution with shape parameter 0.3 and scale parameter 3.0. Phenotypes were computed by adding a random noise yielding the desired heritability. From the last 5 generations, the first 4 were assigned to the training set and the fifth to the prediction set. The genome consisted

[*] Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August University, 37075 Goettingen, Germany

of 10 chromosomes with a total of 3,000 SNP marker loci (300 per chromosome). Mutation rate was 10^{-5}. The last (5th) generation consisted of 100 groups with 5 full-sibs each.

Statistical models. The reference model (model 1) is a conventional animal model, using complete phenotypic and pedigree information of individuals in the training set. The model equation is $\mathbf{y} = \mu + \mathbf{Za} + \mathbf{e}$ where $\mathbf{y}$ is the vector of phenotypes, μ is the overall mean, $\mathbf{a}$ is the vector of random polygenic effects, $\mathbf{Z}$ is an incidence matrix and $\mathbf{e}$ is a random residual term. Assumptions are: $\mathbf{a} \sim \mathbf{N(0, A\sigma^2_a)}$ and $\mathbf{e} \sim \mathbf{N(0, I\sigma^2_e)}$ where **A** is the numerator relationship matrix, $\boldsymbol{\sigma}^2_{\mathbf{a}}$ is the additive genetic variance and $\boldsymbol{\sigma}^2_{\mathbf{e}}$ is the residual variance. Breeding values from model 1 were computed as parent average for animals in the prediction set.

The alternative model (model 2) is a two-step procedure for estimating breeding values. In the first step, breeding values are estimated using model 1 in the training set and estimated Mendelian sampling terms (**m**) are derived using the estimated breeding values.

In the second step, genomic Mendelian sampling effects (**q**) are estimated using the mixed model $\hat{\mathbf{m}} = \boldsymbol{\mu} + \mathbf{Wq} + \boldsymbol{\varepsilon}$ where $\hat{\mathbf{m}}$ is the vector of estimated Mendelian sampling terms for animal in the training set and $\boldsymbol{\varepsilon}$ is a random error. W_i is a matrix of Mendelian sampling genotypes of animals in the training set, computed as $\mathbf{W_i} = g_i - \left[\frac{1}{2}(g_S + g_D)\right]$, where g_i, g_S, g_D are the marker genotypes for the individual, its sire and dam, respectively. We assumed $\mathbf{q} \sim \mathbf{N(0, I\lambda)}$, with $\boldsymbol{\lambda}$ defined as in VanRaden (2008), but here divided by 2. Values of g are set to -1, 0 and 1 for the homozygote 11, heterozygote 12 and other homozygote 22.

Estimated breeding values for individuals in the prediction set were then obtained as:

$$\hat{\mathbf{a}}_\mathbf{i} = \left[\frac{\mathbf{1}}{\mathbf{2}}(\hat{\mathbf{a}}_\mathbf{S} + \hat{\mathbf{a}}_\mathbf{D})\right] + \mathbf{W}\hat{\mathbf{q}}$$

Where $\hat{\mathbf{a}}_\mathbf{i}$, $\hat{\mathbf{a}}_\mathbf{S}$, $\hat{\mathbf{a}}_\mathbf{D}$ are the estimated breeding value for the individual, its sire and dam, respectively.

Comparison between models. One hundred replicates were simulated for each of three values of heritability (0.1, 0.3, 0.5). For each of them, breeding values were predicted using the two models and animals were assigned to a selection set. Three proportions of selected animals were considered (top 10, top 20 and top 30%). In each scenario, the average additive relationship between the selected animals was computed, the number of full-sib groups represented in the selected set was counted and correlations between predicted and true breeding values were calculated. Analyses were performed using the software R.

Results and discussion

Average additive relationships within the selected animals for each simulated heritability and proportion of selected animals are presented in figure 1. Average additive relationship

decreased considerably for increased proportion of selected animals. For the three simulated heritabilities, average additive relationship within selected animals was lower when breeding values were predicted using model2. The difference in relationship between models 1 and 2 was lower for lower values of heritability but the trend was consistent across different values of heritability. It was found that the use of model 2 yielded less related selected animals, which would incur in lower levels of inbreeding rate per generation. This is consistent with the results of Daetwyler *et al.* (2007).

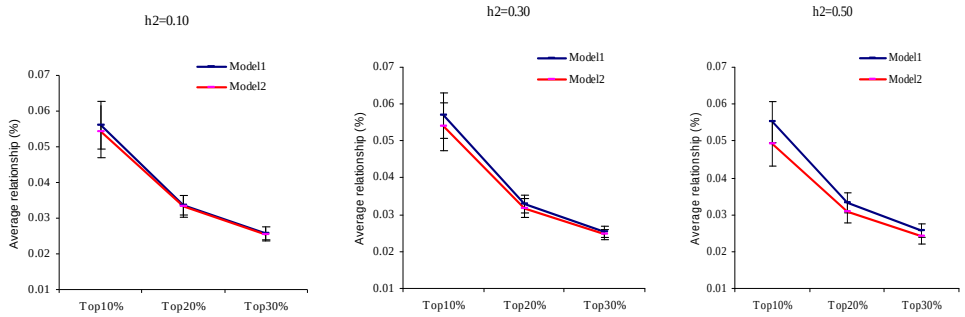

Figure 1: Average additive relationship for each model and scenario

Numbers of full-sib groups within selected animals for each scenario are presented in figure 2. As expected, for both models, the number of groups increased considerably with increased proportion of animals in the selection set. For all simulated heritabilities and proportions of selected animals, the number of full-sib groups in the selection set was higher when breeding values were predicted with model 2 than with model 1. This indicates that the use of model 2 will lead to the maintenance of a larger amount of genetic diversity. As in the comparison between additive relationships, the difference was also more pronounced for larger values of heritability. It shows that the higher the heritability the higher the impact of the use of the Mendelian sampling term.

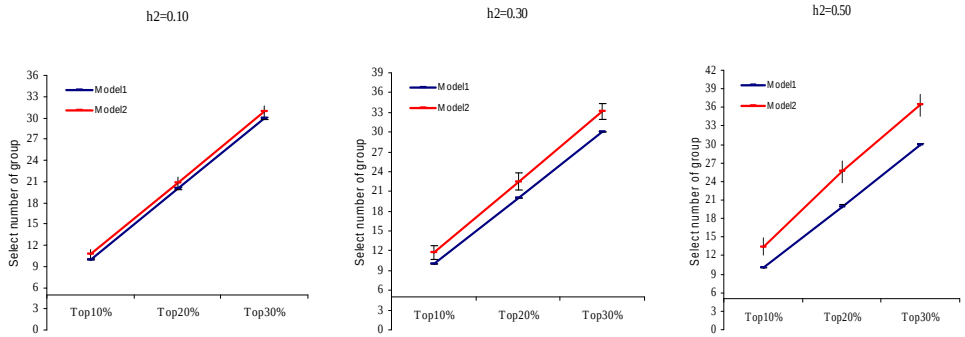

Figure 2: Number of full-sib groups within selected animals for each scenario

The correlation coefficients between true and estimated breeding values from each model and simulated heritability are presented in Table 1. Correlations are higher for model 2, particularly for the highest level of heritability.

Table 1: Correlation coefficients between estimated and true breeding values

Heritability	Model 1	Model 2
0.1	0.2896±0.082	0.2929±0.081
0.3	0.4509±0.058	0.4707±0.057
0.5	0.5332±0.057	0.5828±0.053

Conclusion

These results clearly show that the model including the Mendelian Sampling term may effectively increase the number of family groups and reduce the average additive relationship within selection candidates, without reducing the accuracy of predicted breeding values. Furthermore, for levels of heritability within the range considered here, the use of model 2 may also reduce the rate of inbreeding per generation. The advantage of model 2 is more pronounced if selection operates on highly heritable traits. Use of the Mendelian sampling term in genomic applications could be useful for more efficient selection strategies and a better control of genetic diversity in livestock. This study may contribute with some information for the sustainable development of livestock production.

Acknowledgement

This research was funded by the Germany Federal Ministry of Education and Research BMBF) within the AgroClustEr “Synbreed – Synergistic plant and animal breeding”.

References

Avendaño, S., Woolliams, J. A., and Villanueva, B. (2004). *Genet. Res.,* 83:55-64.
Avendaño, S., Woolliams, J. A., and Villanueva, B. (2005). *J. Anim. Breed. Genet*., 122:302–308.
Daetwyler, H. D., Villanueva, B., Bijma, P., and Woolliams, J. A. (2007). *J. Anim. Breed. Genet*., 124:369-376.
Meuwissen, T. H. E., Hayes, B. J., and Goddard., M. E. (2001). *Genetics,* 157:1819–1829.
Misztal, I., Legarra, A., and Aguilar, I., (2009). *J. Dairy Sci*., 92:4648–4655.
Simianer, H., and Pimentel, E.C.G. (2009). *J. Anim. Breed. Genet*., 126:432-442.
Sørensen, A. C., Sørensen, M. K., and Berg, P. (2005). *J. Dairy Sci*., 88:1865-1872.
VanRaden, P. M. (2008). *J. Dairy Sci*., 91:4414–4423.