Proceedings of the World Congress on Genetics Applied to Livestock Production, 11. 499

Accounting for breed origin of alleles increase accuracy of genomic prediction in admixed populations

S. M. Hosseini-Vardanjani[1], M. M. Shariati[1], L. Janss[2], O. F. Christensen[2], M. T. Anche[2], M. S. Lund[2]*

Corresponding Author: M. S. Lund

mogens.lund@mbg.au.dk

[1] University of Mashhad, Faculty of Agriculture, Department of Animal Science, Mashhad, Iran

[2] University of Aarhus, Faculty of Science and Technology, Department of Molecular Biology and Genetics, Denmark

Summary

In admixed populations, markers may be associated to different QTL depending on the origin of a given genomic segment. The goal of this study was to investigate if taking into account of the breed origin of alleles in a breed of origin genomic model (BOGM) can improve genomic predictions compared to a traditional genomic model (TGM) in admixture populations. Real genotype data of Danish Holstein and Jersey breeds were used as base populations to simulate F1 crosses. This was followed by simulating 5 discrete generations of random mating to achieve a highly admixed population. A single trait with heritability of 0.25 and 100 QTL was simulated on the genome. Three different scenarios were considered, where the QTL effects of two breeds were sampled from a multivariate normal distribution with correlation 1.0, 0.5, or 0.1. Accuracy and bias of models which were measured as correlation and regression coefficient between true and estimated genomic breeding values, respectively, were validated using each first three generations as training, and subsequent generations as test populations. The BOGM had higher accuracy than TGM with a correlation of 0.5 (0.455 to 0.719 *vs* 0.285 to 0.594) and the correlation of 0.1 (0.419 to 0.749 *vs* 0.172 to 0.576). In the scenario with identical QTL effects in the two breeds, the accuracy of TGM was higher than BOGM (0.548 to 0.610 *vs* 0.456 to 0.557). Prediction bias was only observed when the first generation was used for training and subsequent generations as test. In conclusion, accuracy of genomic prediction in admixed population can be increased by taking into account the breed of origin of alleles.

Key words: *Genomic prediction, admixture populations, breed-specific effects, validation*

Introduction

Genomic selection has been widely implemented in purebred dairy cattle

populations with homogeneous genetic structure such as Holstein breed. However, some populations are more admixed in nature and systematic crossing schemes are expected to increase in the future. For the Nordic Red breeds Makgahlela et al. (2013) found no improvements by taking the breed composition into account. This is likely because the models used only the genome wide composition, and failed to model the local genomic differences arising from linkage disequilibrium across breeds and linkage disequilibrium phase between markers and QTL which are different between breeds (De Roos et al., 2008). For crossbreeding schemes it has been shown, that taking account of breed of origin in genomic prediction is crucial under some conditions (Ibánẽz-Escriche et al., 2009; Xiang et al., 2016).

With an increased use of crossbreeding, applying genomic prediction in admixed or crossbred populations is of great importance. However, while methods exist for systems with terminal crosses, there is an unmet need for tools to perform predictions in admixed populations. Therefore, the aim of this study was to compare accuracy of a specific model which accounts for breed origin of alleles with current genomic prediction models for genomic selection in a simulated admixed population.

Materials and Method

A simulation was carried out using available real genotypes of two breeds: Danish Holstein and Jersey. Animals were genotyped using the Illumina 50K SNP panel. Common SNP markers between the two breeds on the first 5 chromosomes were used for simulation. Nine hundred animals from each breed were sampled as a base population. In each breed it was assumed that 800 of them were females and the rest males. The mating was performed randomly between the males of one breed to the females of the other breed with a mating ratio of 1:8. Of the 8 mating, 2 mating were assumed to produce 2 offspring each, and the rest of 6 mating give 1 offspring each which led to 2000 individuals in total. In subsequent generations, we have set 200 male and 1800 females and, a mating ratio of 1:9. From 9 mating, 1 mating was assumed to generate 2 offspring and the rest of 8 mating generate 1 offspring which led to 2000 individuals in each generation of admixed population (i.e. generation 1 to 5). A single trait with heritability of 0.25 and 100 QTL was simulated on the genome. Additive effects of QTLs were sampled from a multivariate normal distribution. Three different scenarios were defined where the correlation between QTL effects of the two founding breeds were 1.0, 0.5, or 0.1. Phenotypes of the trait were simulated by adding a standard normal residual effect to the genotypic value of each animal such that a heritability of 0.25 was achieved. All simulated scenarios were replicated 10 times.

Two Bayesian models were used to predict genomic breeding values. The first we term here the traditional genomic model (TGM) was BayesB (Meuwissen et al., 2001) as follows:

,

where is the phenotype of animal , is the general mean, is the genotype of at marker with element 0, 1 or 2 corresponding to genotypes of animal at SNP marker of 11, 12 or 22 respectively, is the allele substitution effect of locus and, is an index variable for locus j that is 1 with probability 1- π and zero with probability π to be included in the model. The second model was a specific genomic model to take into account the breed origin of alleles (BOGM):

,

where is an index for the number of breeds contributing to admixed population corresponding to 1 and 2 in present study, is a matrix for the number of each allele of marker originating from breed , and is the vector of breed-specific allele effects of SNP marker in breed . The training set for genomic prediction models included the first 3 generations of the admixed population, and predictions were evaluated in the same or following generations. Prediction accuracy and bias of models was calculated based on correlation and regression coefficient between true and estimated genomic breeding values, respectively.

Results

Figure 1 shows boxplots of the fivefold cross validation for each training generation using the traditional genomic model (TGM) and the breed of origin genomic model (BOGM) at different scenarios. BOGM outperformed TGM when the correlation between QTL was less than one, while the TGM shows higher accuracy when QTL effects were similar across breeds independent of which generation was used for training.

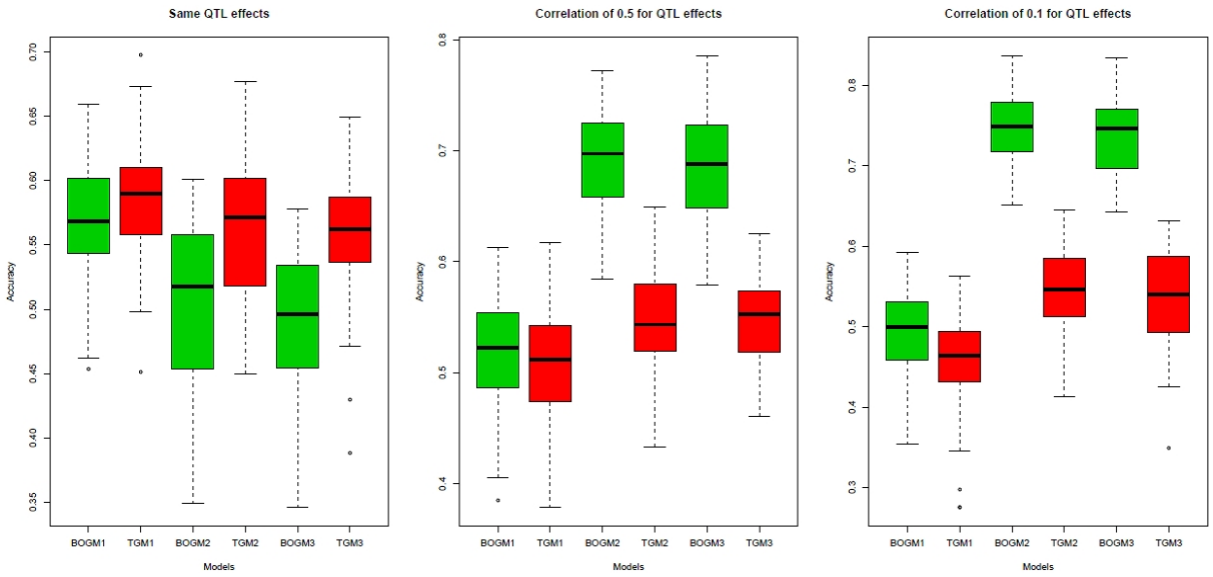

Figure 1- Accuracy of fivefold cross validation within generation with BOGM and TGM for first 3 generations (1, 2, and 3) of the admixed population

Prediction accuracies and biases of models based on different training populations for predicting subsequent generations are shown in Table 1. Accuracies show almost same differences to those in cross validation for the two models. Except for the situation where

generation one is used as training population, two models did not result in major changes in prediction bias when generation 2 and 3 were used for training, and those that were different remained small and always better state for BOGM.

Table 1- Accuracy and bias of genomic prediction from different models with different training and test populations

Scenarios		1				0.5				0.1			
Populations		Accuracy		Bias		Accuracy		Bias		Accuracy		Bias	
TrG	TeG	BOGM	TGM	BOGM	TGM	BOGM	TGM	BOGM	TGM	BOGM	TGM	BOGM	TGM
1	2	0.476	0.610	0.823	0.922	0.504	0.339	1.104	0.831	0.460	0.208	1.191	0.643
1	3	0.456	0.548	0.793	0.898	0.455	0.285	1.093	0.764	0.419	0.172	1.179	0.567
2	3	0.557	0.608	0.980	1.010	0.719	0.594	1.010	1.033	0.749	0.576	0.993	1.018
2	4	0.494	0.563	0.907	0.985	0.685	0.535	1.009	0.988	0.714	0.510	0.982	0.960
3	4	0.541	0.607	0.977	0.988	0.713	0.588	1.004	1.020	0.748	0.569	0.986	1.012
3	5	0.480	0.561	0.918	0.971	0.675	0.527	0.990	0.975	0.723	0.516	0.992	0.988

TrG: Training generation, TeG: Test generation, BOGM: Breed of Origin Genomic Model, TGM: Traditional Genomic Model

Discussion

In this study we showed the potential increase in accuracy of genomic prediction in admixed populations by taking into account the breed of origin of markers. This is important because some populations are highly admixed and because there is an increased focus on the use of crossbreeding of multiple dairy breeds for milk production in combination with sexed semen and usage of a terminal beef sire. Many cows will therefore be admixed and an increased need exist for genomic predictions for both purebred and admixed animals using phenotypes from both types of cows.

Low to medium correlations between QTL in the different breeds lead to significantly higher accuracy of BOGM compared to TGM. Linkage disequilibrium, linkage phase, and allele frequencies are different between breeds, which result in different estimated marker effects in each breed. Predicting genetic merit of admixed or crossbred animals should therefore be improved by accounting for which breed the marker alleles have descended from and the effect in that breed. With identical QTL effects across breeds the accuracy is highest for a model that assumes the same effect. These results are in line with Lund et al. (2014) who showed that combining closely related breeds in a multi breed reference population increase genomic prediction accuracies while distantly related breeds do not result in a benefit. In total, if there is a relatively high overall positive correlation between marker effects in the breeds of origin, it is difficult to separate these effects using BOGM at least with the same number of records used in the training for TGM because of a greater number of effects to be predicted in BOGM. Ibánẽz-Escriche et al. (2009) showed the model that accounts for breed-specific marker effects can outperform models in which marker effects are assumed to be same across breeds only when there was a large training data size, and low relatedness between breeds.

It is worth noting that the necessary condition for the application of BOGM in real genomic selection is that the purebred origin of alleles in the admixed population be

estimated accurately. If not, uncertainties in assigning breed origin of alleles will most likely impact the accuracy of subsequent genomic predictions. However, vast progress has been accomplished to develop methods for inference breed origin of alleles where they originate from. In a recent study, Sevillano et al. (2016) applied a method to trace breed origin of alleles in crossbreds using long-range phasing with or without need for tracking pedigree relationships of crossbreds and assigned a proportion of 94.6 and 93 percent of alleles to origin, respectively.

Acknowledgements

The first author wishes to acknowledge Ministry of Science, Research and Technology of Iran, and center for Genomic Selection in Animals and Plants (GenSAP) funded by Innovation Fund Denmark under grant 0603-00519B for financially supporting his visit to the University of Aarhus.

List of References

De Roos, A., Hayes, B. J., Spelman, R., & Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics, 179*(3), 1503-1512.

Ibánẽz-Escriche, N., Fernando, R. L., Toosi, A., & Dekkers, J. C. (2009). Genomic selection of purebreds for crossbred performance. *Genet Sel Evol., 41*(1), 12.

Lund, M. S., Su, G., Janss, L., Guldbrandtsen, B., & Brøndum, R. F. (2014). Genomic evaluation of cattle in a multi-breed context. Livestock Science, 166, 101-110.

Makgahlela, M. L., Mäntysaari, E. A., Strandén, I., Koivula, M., Nielsen, U., Sillanpää, M., & Juga, J. (2013). Across breed multi trait random regression genomic predictions in the Nordic Red dairy cattle. *J Anim Breed Genet., 130*(1), 10-19.

Meuwissen, T., Hayes, B., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819.

Sevillano, C. A., Vandenplas, J., Bastiaansen, J. W., & Calus, M. P. (2016). Empirical determination of breed-of-origin of alleles in three-breed cross pigs. *Genet Sel Evol., 48*(1), 55.

Xiang, T., Nielsen, B., Su, G., Legarra, A., & Christensen, F. (2016). Application of single-step genomic evaluation for crossbred performance in pig. *J. Anim. Sci.,* 94(3):936-948.