Accuracy of prediction for a genomic evaluation in rotational crossbreeding scheme (Montbéliarde x Holstein x Red Danish)

R. Saintilan[1,2], P. Croiseau1*, A. Baur[1,2], I. Croue[1,2], V. Ducrocq[1], J.R. Thomasen[3], E. Karaman[4], D. Boichard[1], H. Leclerc[1,2] and B.C.D. Cuyabano[1]

1 Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France

2 ALLICE, Paris, France

3 Viking Genetics, Ebeltoftvej 16, 8960 Randers SO, Denmark

4 Center for Quantitative Genetics and Genomics, Faculty of Technical Sciences, Aarhus University, 8830 Tjele, Denmark.

**pascal.croiseau@inrae.fr*

Abstract

In recent years, crossbreeding strategy has grown in dairy cattle farms at an international level. Breeders are interested in keeping crossbred cows in their herd both to combine the strengths of the pure breeds, compensate for their weaknesses and benefit from heterosis. However genetic tools are still lacking to manage these crossbred animals. In this study, we evaluate the performances of a genomic evaluation adapted for rotational crossbreeding schemes with real data. This genomic evaluation was applied to a population that includes pure-breed animals from Holstein, Montbéliarde, and Red Danish breeds, as well as crossbreds between these three breeds. The genomic evaluation approach was based on the estimation of SNP specific effect according to the Breed of Origin of the Alleles.

Key words: Genomic evaluation, Crossbreeding, ALLOR, Heterosis

Introduction

Crossbreeding strategy in dairy cattle breeding schemes has significantly grown over the past years, and this trend may be reinforced in the future to face challenges linked to agroecology or to new constraints induced by climate change. Indeed, this strategy is particularly suitable to combine the strengths of the pure-breeds and to compensate for their weaknesses, being an efficient way for breeders to obtain more adaptable and robust animals, resulting in a more sustainable breeding system. Crossbreeding is also relevant to decrease inbreeding or, equivalently, generate heterosis.

Genomic evaluation in pure breed is common, it is developing in terminal crossbreeding for slaughter purposes, but only a few countries routinely evaluate crossbred animals in continuous crossbreeding programs. For most countries, breeders engaged in a crossbreeding process have information limited to purebred bulls and the raw performances of their crossed females to manage their selection and matings. In this context, the literature dedicated to continuous crossbreeding relates, for the most part, to simulated data (Van Raden et al., 2021; Eiriksson et al., 2021; Karaman et al, 2021) and rarely to applications on real data (Sevillano et al., 2017).

In this study, we propose an application of genomic evaluation for continuous crossbreeding programs, through a GBLUP based on Breed of Origin of Alleles (BOA). This approach estimates specific SNP effect depending on its breed of origin. The rationale for this approach is that linkage disequilibrium between SNP and the causal mutation differs according to the breed; additionally, QTL effect may be breed dependent.

The data used in this analysis consisted of the first steps of a three-way crossbreeding scheme including Montbéliarde (Mo), Holstein (Ho) and Red Danish (RD) breeds. This data set was augmented by purebred data to ensure

accurate estimation of within breed SNP effects. The procedure included several steps: imputation and phasing at the 50K level of all the animals; identification of BOA of each allele of crossbred animals; genomic evaluation on five production traits, including a cross-validation procedure. Accuracy of prediction and slope of regression were estimated and compared to a conventional GBLUP ignoring breed information.

Materials and Methods

Genotypes

The data used in this study originated from France and Nordic Countries (NAV). They consisted of 5,238 genotypes of crossbred animals, 20,000, 22,265, and 6,866 genotypes of pure Mo, Ho, and RD animals, respectively. 53,498 autosomal SNP markers were retained from the Illumina 50K chips used routinely in France for genomic selection. Imputation was carried into two steps. In the first step, genotypes of purebreds were imputed with FImpute (Sargolzaei et al., 2014) with the pipeline used in the routine French national evaluation system and they were assumed to be known. Therefore, in the second step, the crossbred genotypes were imputed using FImpute using the purebred Ho, Mo, and RD as a reference, and without using any pedigree information.

BOA of alleles

The frequency (f) of each haplotype of n consecutive SNPs was estimated by counting within each pure breed. BOA was selected if $\frac{f_i}{\sum_{j=1}^{nbreed} f_j}$ was higher than a given threshold (0.9). The whole genome was scanned by a sliding window moving one SNP at a time. The initial value of n was set to 16 (leading to 65,536 theoretical haplotypic combinations). When a haplotype origin remained undetermined, the process was iterated after n was divided by 2. Undetermined origins surrounded by identical breed origins were assigned to this breed. When n=1, the small proportion of finally remaining undetermined origins were allocated according to allelic frequencies. This procedure was implemented in the in-house BreedOrigin fortran software and applied to a target population of 5,238 crossbred animals. The results were compared to expected breed proportions based on pedigree.

Crossbreeding parameters

Fractions of heterosis (H) and recombination losses (R) were calculated for each crossbred animal from proportions of breed origins of their parents based on pedigree. Values of the H and R coefficients were calculated as in Dechow et al. (2007) using the following formulae

$H = 1 - \sum_{i=1}^{nb_{breed}} s_i d_i$ and

$R = 1 - \sum_{i=1}^{nb_{breed}} \frac{(s_i^2 + d_i^2)}{2}$

where si and di are the proportions of sire and dam genes from breed i, respectively. Heterosis and recombination losses were then estimated by regressing H and R in the model of analysis.

Phenotypes

The genomic evaluation was a two-step procedure. In a first step, a polygenic model was used to estimate all non-genetic effects (herd-year, age-year and year-month of calving), heterosis and recombination losses (as regression coefficients) and heterogeneous variances, as in Dezetter et al (2015). Unknown parent groups accounted for breed of origin and breed composition of crossbreds was accounted for by the pedigree. Yield deviations (YD) produced by this model correspond to performances adjusted for fixed and non-genetic random effects. Five production traits were analysed: lactation milk yield, protein yield, protein content, fat yield, and fat content. Two batches of evaluations were performed, the first one accounting for heterosis and recombination losses, and the second one ignoring them.

Crossbreed genomic evaluation accounting for BOA

To account for crossbred animals in a genomic evaluation model, we proposed an extension of the SNP-BLUP model where a SNP effect βi is estimated for each breed as described in the following model:

$$y_i = \sum_{b=1}^{nb\_breed} (p_{i,b}\mu_b) + \sum_{b=1}^{nb\_breed} \left(\sum_{j=1}^{nb\_SNP} (\beta_{i,j,b} X_{i,j,b}) \right) + e_i \quad (1)$$

where yi is the YD of the animal i, μ is a vector of means defined within each breed, pi,b is the proportion of breed b in the genome of individual i, estimated with the BOA approach. Xi,j,b is the allele content of SNP j that originates from the breed b for animal i, centered for the allelic frequency of the SNP in breed b :

$$X_{i,j,b} = (k_{i,j} - n_{i,j,b} f_{j,b}) \quad (2)$$

where ki,j,b and ni,j,b are the number of "2" alleles and the total number of alleles of the SNP j that originates from breed b for the animal i, respectively; fb is the frequency of allele "2" of the SNP j in breed b.

To assess the performance of the genomic evaluation, the data set was divided into a training dataset with both genotypes and phenotypic records used to estimate SNP effects according to their BOA, and a validation data set for which predicted breeding values were computed using the effects obtained with the training dataset and then compared to the observed phenotypes. The validation population consisted of 2000 crossbred cows without progeny.

Results & Discussion

Breed of origin of alleles

Results for breed composition based on BOA methodology and pedigree information are presented in Table 1 and 2 respectively. For genotyped crossbred animals, 94% of the alleles were assigned to a breed. 48% of markers were from Ho origin, 34% from RD origin and 13.6% from Mo origin. In comparison, based on pedigree information, the corresponding origins were 56.3%, 29.8% and 11.8% respectively. Correlation between breed compositions for both methodologies ranged from 0.95 for the RD breed to 0.99 for the Mo breed.

Table 1. - Results of determination of Breed of Origin based on BOA

Breed of Origin	Breed composition with BOA	
	Mean	s-d
Montbéliarde (Mo)	13.60%	20.00%
Holstein (Ho)	48.00%	19.40%
Red Danish (RD)	34.00%	17.80%

Table 2. - Results of determination of Breed of Origin based on Pedigree information

Breed of Origin	Breed composition based on Pedigree	
	Mean	s-d
Montbéliarde (Mo)	11.80%	19.20%
Holstein (Ho)	56.30%	20.20%
Red Danish (RD)	29.80%	17.80%

Across breed genomic evaluation accounting for BOA

Regression coefficients between YD and genomic breeding values and associated slope of regression are presented on Table 3 and 4. On the training population, all the correlations are around 0.80 for milk yield, protein yield, and protein content and around 0.90 for fat yield and fat content. Adjusting for Heterosis weakly affected these correlations. The associated slopes of regression were slightly higher than 1 for all the traits. In the validation population, without integration of the Heterosis effect, correlations ranged from 0.36 for protein content and 0.65 for fat yield trait (table 3). The gain of correlation obtained after adjusting for Heterosis was marginal (+1 point for milk yield, protein yield, and protein content, table 4). Regarding the slopes of regression, they were overestimated both with and without adjustment of the performances for Heterosis (between 1.47 and 1.69).

Table 3: - Regression coefficient and slope of regression of the YD not adjusted for Heterosis for 5 production traits, on the estimated genomic breeding values, in the training and validation population.

	Training Population		Validation Population	
Traits	Correlation	Slope	Correlation	Slope
Milk lactation	0.82	1.19	0.41	1.07
Fat Content	0.78	1.22	0.38	1.15
Protein Content	0.76	1.26	0.36	1.16
Fat yield	0.92	1.15	0.65	1.05
Protein yield	0.93	1.12	0.62	1.00

Table 4: - Regression coefficient and slope of regression of the YD adjusted for Heterosis for 5 production traits, on the estimated genomic breeding values, in the training and validation population.

	Training Population		Validation Population	
Traits	Correlation	Slope	Correlation	Slope
Milk lactation	0.83	1.17	0.40	1.03
Fat Content	0.80	1.20	0.36	1.09
Protein Content	0.77	1.23	0.35	1.11
Fat yield	0.92	1.15	0.65	1.05
Protein yield	0.93	1.12	0.62	1.00

Conclusions

In this study, we propose an application on real data of a genomic evaluation in a rotational crossbreeding scheme based on a SNP-BLUP model accounting BOA.

This approach required to impute and phase genotyping data of crossbred animals in order to predict BOAs for these animals. These steps were tested and an average error of about 1.5% for the predicted BOA was observed (data not shown). To complete this information, we compare in this study the breed composition of crossbred animals measured using the pedigree information of the animals and estimated from the genotyped animals. The very strong correlations obtained between these two approaches are completely consistent with previous tests (Table 1).

Finally, the genomic evaluations carried out have shown an honourable accuracy of prediction of around 0.30 for traits with a heritability of 0.30 and around 0.60 for traits with a heritability of 0.50 (Table 2). The adjustment of the performances for Heterosis did not impact the accuracy of the prediction of the breeding values.

Acknowledgments

This project has received funding from APIS GENE and from the European Union's Horizon 2020 research and innovation program - GenTORE - under grant agreement No. 727213.

References

Dechow, C. D., Rogers, G. W., Cooper, J. B., Phelps, M. I., & Mosholder, A. L. (2007). Milk, Fat, Protein, Somatic Cell Score, and Days Open Among Holstein, Brown Swiss, and Their Crosses. JDS - 90(7):3542-3549.

Dezetter, C., Leclerc, H., Mattalia, S., Barbat, A., Boichard, D. et al. (2015). Inbreeding and crossbreeding parameters for production and fertility traits in Holstein, Montbéliarde, and Normande cows. JDS - 98(7):4904-4913.

Eiriksson, J.H., Karaman, E., Su, G. & Christensen, O. F. (2021) Breed of origin of alleles and genomic predictions for crossbred dairy cows. GSE – 53: 84.

Karaman, E., Su, G., Croue, I., & Lund, M. S. (2021). Genomic prediction using a reference population of multiple pure breeds and admixed individuals. GSE - 53(1):1-15.

Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. BMC genomics - 15(1):1-12.

Sevillano, C. A., Vandenplas, J., Bastiaansen, J. W., Bergsma, R., & Calus, M. P. (2017). Genomic evaluation for a three-way crossbreeding system considering breed-of-origin of alleles. GSE - 49(1):1-14.

Van Raden, P. M., Tooker, M. E., Chud, T. C. S., Norman, H. D., Megonigal Jr et al. (2020). Genomic predictions for crossbred dairy cattle. JDS - 103(2):1620-1631.