

## Genomic evaluation using combined reference populations from Montbéliarde and French Simmental breeds

C. Hozé<sup>\*,†</sup>, S. Fritz<sup>\*,†</sup>, F. Phocas<sup>\*</sup>, D. Boichard<sup>\*</sup>, V. Ducrocq<sup>\*</sup> and P. Croiseau<sup>\*</sup>.

<sup>\*</sup>INRA UMR1313 GABI, Jouy en Josas, France, <sup>†</sup>UNCEIA, Paris, France

**ABSTRACT:** The French-Simmental and Montbéliarde breeds are currently evaluated jointly for production traits in international evaluations. Therefore, we investigated the feasibility of using the large reference population available in Montbéliarde breed to implement a genomic evaluation in French-Simmental. Data consisted in 229 Simmental and 1,758 Montbéliarde progeny tested bulls genotyped for a common set of 37,364 SNP. Genomic evaluations were performed using GBLUP and BayesC $\pi$ . SNP effects were either estimated based on Montbéliarde, Simmental or a joint reference population. Montbéliarde population was also used to preselect SNP before Simmental evaluation. Accuracy of evaluation was assessed based on the youngest Simmental bulls. Average correlations were 0.28 with SNP effects estimated in Montbéliarde, 0.39 with effects estimated in Simmental, 0.42 after pre-selection, and 0.48 with multi-breed reference population. This study demonstrates that the Montbéliarde reference population is beneficial to implement genomic evaluation in Simmental.

**Keywords:** dairy cattle; genomic selection ; multi-breed evaluation

### Introduction

In France, a very large number of animals from the main dairy breeds have been genotyped with the Illumina Bovine SNP50 BeadChip® (50K) for genomic selection (GS) and it is now possible to predict breeding values of animals at birth with high accuracy in Holstein, Montbéliarde and Normande. For the French Simmental breed and more generally for all regional breeds, a limited number of progeny-tested bulls are available and assembling a large enough reference population is a real challenge.

One strategy proposed to implement genomic evaluations in small breeds was the use of multi-breed reference population (Hayes et al., 2009, Erbe et al., 2012). Some increase in correlation was observed when combining Holstein and Jersey populations (Erbe et al., 2012) and major French breeds (Hozé et al., 2014). However, the gain in reliability was limited, depended on the trait considered and on the size of reference populations (Hozé et al 2014).

Larger increases with multi-breed reference populations were observed when populations were closely related (Brondum et al., 2011) or when a denser chip was used (Hozé et al., 2014). One possible explanation is the better conservation of linkage disequilibrium (LD) in breeds with common origin whereas conserved LD is only observed at short distance in unrelated breeds.

In France, the French Simmental breed has a limited population and therefore cannot benefit alone from efficient genomic evaluations. In contrast, the Montbéliarde breed, which also belongs to the Simmental and Fleckvieh breed group, is a national breed and has more than 1,700 progeny-tested bulls genotyped. The two breeds take part separately to the international genetic evaluations for production traits implemented since 1996 and involving ten Simmental European populations. Such an evaluation leads to estimated breeding values on the French Simmental scale for Montbéliarde bulls. Accordingly, we can envisage using Montbéliarde bulls for French Simmental genomic evaluations.

The aim of this paper is to assess the accuracy of predictions of genomic evaluation in French Simmental using four different strategies involving Montbéliarde and Simmental reference populations separately or jointly.

### Materials and Methods

**Data.** The French Simmental reference population consisted of 229 progeny tested-bulls genotyped on the BovineHD BeadChip® (HD). Among them, 123 were German bulls selected on their marginal contribution to the French Simmental population and whose genotype was exchanged with LfL. In the Montbéliarde breed, 1,758 progeny tested bulls genotyped on the BovineSNP50 BeadChip® (50K) were available. Most of them originate from the French genomic evaluations. A quality control procedure described in Hozé et al. (2014) was first performed on both datasets. Then, markers present on both 50K and HD chip were extracted leading to a common set of 37,364 SNP for the two breeds.

The phenotypes used were deregressed proofs (DP) and associated equivalent daughter contribution (EDC) from Interbull evaluation expressed on the French Simmental scale. The traits studied here were milk yield, fat yield, protein yield and somatic cell score (SCS).

**Evaluation model.** Breeding values (EBV) were estimated using three approaches : a pedigree-based BLUP, GBLUP (VanRaden, 2008) and BayesC $\pi$  approaches (Habier et al., 2010). Evaluations were performed using GS3 software (Legarra et al., 2013). BayesC $\pi$  was used for single-breed and multi-breed genomic evaluation. For this latter case, a breed effect was included in the model to account for the genetic mean of the breeds and avoid selecting SNP to estimate this difference.

Four scenarios were compared in this study: a) SNP effects were estimated in the Simmental training population and then applied to the Simmental validation population; b) SNP effects were estimated in Montbéliarde and applied to the Simmental validation population; c) SNP effects were estimated in the joint reference population of Montbéliarde and Simmental bulls; d) a preselection of SNP was performed with the Montbéliarde data and SNP effects were estimated in Simmental breed.

The latter approach, derived from the study of Croiseau et al. (2011) allows a reduction of the number of effects to estimate in the Simmental breed. SNP preselection was based on results of a QTL detection in the Montbéliarde breed. QTL detection was based on a Linkage Disequilibrium Linkage Analysis (LDLA, Druet et al., 2008) study. Phenotypes used were daughter yield deviation of the national Montbéliarde evaluation as they were assumed to be more reliable for the Montbéliarde breed, than Montbéliarde deregressed proofs expressed on the Simmental scale. First, we defined a QTL as the SNP with a LRT higher than 5 and with the highest value within a window of 50 SNP. Then, we retained this SNP and the 50 SNP around the LRT peak for the Simmental analysis.

**Accuracy of evaluation.** The accuracy of evaluation was assessed using a validation study. The 46 youngest bulls of the Simmental reference population formed the validation set. Since phenotypes used here were deregressed proof expressed on the Simmental scale, we did not assess the accuracy of evaluations in Montbéliarde. For animals in the validation set, phenotypes were erased to mimic candidates to selection. SNP effects were estimated on the training population and then applied to the validation population to compute GEBV (Genomic Enhanced Breeding Value). Then, the accuracy of genomic evaluation was measured through the weighted correlation between GEBV and DP of validation bulls.

Among the 46 bulls in the validation population, 28 had their sire in the reference population and 18 did not. We also computed the correlation separately for the two subpopulations to assess the impact of connectedness to reference population on the accuracy of genomic evaluation.

## Results and Discussion

**French Simmental breed considered alone.** Accuracies of pedigree-based BLUP, GBLUP and BayesCp were computed for the four traits using Simmental reference population alone. Results are presented in Table 1.

It can be observed that even for a breed with fewer than 200 animals in the reference population, using genomic information allowed a gain in correlation of 0.06 with GBLUP and 0.08 with BayesCp. Increases in correlation were trait dependent: a large gain was observed for protein yield (+0.23) while a decrease (-0.04) was observed for fat yield.

**Table 1: Correlations between deregressed proofs and estimated breeding values for the four traits using pedigree-based BLUP, GBLUP and BayesCp based on a training population of 181 Simmental bulls**

	Milk	Fat Yield	Protein Yield	SCS <sup>1</sup>	Mean
BLUP	0.46	0.41	0.31	0.12	<b>0.33</b>
GBLUP	0.47	0.34	0.54	0.21	<b>0.39</b>
BayesCp	0.48	0.36	0.54	0.17	<b>0.39</b>

<sup>1</sup> Somatic Cell Score

Whatever the method, correlations were clearly lower for somatic cell score (SCS) than for other traits. SCS is also the only traits where GBLUP performed better than BayesCp. This is at least partially due to the lower heritability and a more polygenic genetic determinism of this trait.

When computing correlations for bulls that did or did not have their sire in the training population, we observed a mean difference between the two groups of bulls of 0.18. Though the limited number of individuals and probably large sampling errors associated with these values, part of the difference between groups could be explained by the lack of polygenic component in our model and spurious association between SNP and causative mutations due to pedigree relationship (Solberg et al., 2009, Liu et al., 2011).

**Using Montbéliarde breed to estimate SNP effect.** SNP effects estimated in the Montbéliarde breed were used to predict GEBV in Simmental. Results are presented in Table 2. The mean correlation achieved with this strategy (0.28) was lower than the one achieved with pedigree-based BLUP (0.32) and with effects estimated in Simmental for bulls without their sire in the reference population (0.31). The Montbéliarde breed has been highly selected for milk production, therefore large QTL influencing production traits which may have been fixed in this breed may still have an influence in Simmental. For SCS, which has been less selected than production, estimating SNP effects in Montbéliarde improved accuracy compared to the single-breed situation.

**Table 2: Correlations between deregressed proofs and estimated breeding values for the four traits using BayesCp with SNP effects estimated in Montbéliarde breed or SNP effects estimated in the French Simmental breed after a SNP pre-selection in the Montbéliarde breed**

	Milk	Fat Yield	Protein Yield	SCS	Mean
EstMO <sup>1</sup>	0.28	0.36	0.28	0.19	<b>0.28</b>
PreselMO <sup>2</sup>	0.54	0.38	0.58	0.16	<b>0.42</b>
Nb SNP <sup>3</sup>	11,627	12,690	12,698	7,927	<b>11,235</b>

<sup>1</sup> SNP effect estimated in Montbéliarde breed

<sup>2</sup> SNP preselected in Montbéliarde breed

<sup>3</sup> Number of SNP conserved after preselection

### Using the Montbéliarde breed to preselect SNP.

When the number of effects to estimate was reduced by preselection based on Montbéliarde data, correlations achieved were up to 0.06 higher than without preselection (Table 2). However, for somatic cell score the preselection led to a decrease in correlation (-0.03). This result could be explained by the relatively low number of SNP preselected for this trait compared to production traits (Table 2).

**Pooling Montbéliarde and Simmental reference populations.** Combining Simmental and Montbéliarde reference populations led to a large increase in accuracy compared to single-breed genomic evaluation (Table 3). The gain in correlation was on average 0.09 and ranged from 0.05 to 0.13 depending on the trait. Larger increases in correlation were observed for SCS and protein yield. It is consistent with the relatively high correlation observed for these two traits using SNP effects estimated in Montbéliarde breed.

**Table 3: Correlations between deregressed proofs and estimated breeding values for the four traits using BayesC $\pi$  and a joint Montbéliarde/Simmental reference population**

	Milk	Fat Yield	Protein Yield	SCS	Mean
BayesC $\pi$	0.53	0.49	0.61	0.27	<b>0.48</b>

The benefit of multi-breed genomic evaluation was higher for bulls that did have their sire in reference population (+0.14) than for those that did not (+0.05). The larger reference population in the multi-breed GS allows a better estimation of SNP effect and avoids detecting SNP associated with pedigree relationship (Hozé et al., 2014).

### Conclusion

The Montbéliarde and French Simmental breeds belong to the same breed family and are jointly evaluated in international evaluations. Despite the lack of connectedness observed between the two breeds (Fouilloux et al., 2006), using both Montbéliarde and French Simmental reference populations for the French Simmental genomic selection allowed a 0.09 increase in the accuracy of prediction compared to the use of the French Simmental breed alone. This study was performed with a small validation population and results need to be confirmed on a larger population. However, accuracies achieved with multi-breed GS are promising and encourage the development of a joint routine genomic evaluation.

A noticeable limit for this implementation is the use of deregressed proofs from Interbull evaluation as phenotypes. Indeed, up to now, production traits are the only traits evaluated by Interbull for Simmental. The extension of international evaluations to other traits, in particular to type traits which has been shown to be feasible (Regaldo et al., 2006), is required for the development of genomic se-

lection in French Simmental. An alternative is to use daughter yield deviation from national evaluation as phenotypes for genomic evaluation but its consequences need to be investigated further.

### Acknowledgment

GEMBAL project is funded by the Agence Nationale de la Recherche (ANR-10-GENM-0014), APISGENE, Races de France and INRA “AIP Bioressources”. Montbéliarde genotypes originated from the Cartofine-ANR-05-GENANIMAL-007 project funded by ANR (French National Research Agency) and ApisGene, and from genomic selection activity generated by the French cattle breeding companies, with LABOGENA as main genotyping lab. The Bayerische Landesanstalt für Landwirtschaft (LfL) 123 Simmental genotypes.

### Literature cited

- Brondum, R. F., Rius-Vilarrasa, E., Strandén, I. et al. (2011). *J Dairy Sci* 94:4700-4707.
- Croiseau, P., Legarra, A., Guillaume, F. et al. (2011). *Genet Res* 93:409-417.
- Druet, T., Fritz, S., Boussaha, M. et al. (2008). *Genetics* 178:2227-2235.
- Erbe, M., Hayes, B. J., Matukumalli L. K. et al. (2012). *J. Dairy Sci* 95:4114-4129.
- Fouilloux, M.N., Minery, S., Mattalia, S., and Laloë D. (2006). *Interbull Bulletin* 35: 129-135
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2010). *BMC bioinformatics* 12:186.
- Hayes, B., Bowman, P., Chamberlain, A. et al. (2009). *Genet Sel Evol* 41:51.
- Hozé, C., Fritz, S., Phocas, F. et al. (2014) *J Dairy Sci.* (in press)
- Legarra, A., Ricard, A., and O. Filangi (2013). <http://snp.toulouse.inra.fr/~alegarra>.
- Liu, Z., Seefried, F. R., Reinhardt, F. et al. (2011). *Genet Sel Evol* 43(19):9.
- Regaldo, D., Minery, S., and Mattalia S. (2006). *Interbull Bulletin* 35: 136-140
- Solberg, T. R., Sonesson, A.K., Woolliams, J.A. et al. (2009). *Genet Sel Evol* 41:53.
- VanRaden P. M., (2008). *J Dairy Sci* 91: 4414-4423