Pilot genomic evaluations of French crossbred dairy cattle

Audrey A.A. Martin[1,2], Didier Boichard[2], Sandra Dominique[3], Manon Guillerm[3], Sophie Mattalia[2,3], Sébastien Fritz[1,2], Pascal Croiseau[2]

[1] *Eliance, 149 Rue de Bercy, 75595, Paris cedex 12, France*

[2] *Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France*

[3] *Institut de l'Élevage, 149 Rue de Bercy, 75595, Paris cedex 12, France*

Corresponding author: audrey.martin@eliance.fr

Abstract

The use of crossbreeding in dairy cattle has been increasing, reflecting breeders' interest in new strategies to address current challenges in the dairy industry. Crossbred animals benefit from both parental breeds abilities and from heterosis, making them attractive to producers. However, genetic tools commonly used in purebred populations, such as pedigree or genomic estimated breeding values (GEBV), are not yet available for crossbred populations, limiting herd management options and the prediction of future phenotypes. The TransEvagenoc project aimed to develop and test a genomic evaluation tool adapted to current crossbred populations. Genotypes from 4,004 individuals resulting from crosses among the three main French dairy breeds (Normande, Montbéliarde, Holstein) were used. Breed-of-origin of alleles (BOA) and heterosis percentages were estimated based on reference populations of 5,000 individuals per pure breed. SNP effects estimated in pure breeds were then applied to crossbred genotypes according to BOA, and breed differences and heterosis effects were then added to estimate breeding values. Evaluations were performed for milk, fat and protein yields (MY, FY, PY), somatic cell score (SCS) and interval from calving to first service (ICFS). Methodological choices, such as GEBV standardisation, and results preparation were discussed in focus groups (FG) with industry stakeholders to better align to the field needs. Correlations between breeding values and first-lactation phenotypes were 0.35, 0.27, 0.28, 0.31 and 0.08 for MY, FY, PY, SCS and ICFS, respectively. Based on focus group discussions, GEBV were standardised using a representative crossbred population. This tool provides a first genomic evaluation framework for the French crossbred dairy cattle encouraging the collect of additional crossbred genotypes and phenotypes.

Key words: Dairy - Crossbreeding – Genomic evaluation.

Introduction

Historically, French cattle populations are composed of pure breeds. However, dairy crossbreeding has recently gained interest due to its adaptation to current environmental challenges, thanks to a more balanced trait profile and the added benefit of heterosis.

Nowadays, we are observing an increase in the number of inseminations resulting in dairy crossbred cattle in France (IDELE, 2024). In 2024, the proportion of dairy cross inseminations reached 6.1% of whole inseminations. In addition, the increase in the use of crossbreeding for the first insemination (up by 1.5%) indicates a direct breeding choice, rather than positioning dairy crossbreeding as an alternative following potential failure of purebred insemination.

This increase happened while no adapted genetic tools, such as genomic indices, were available for crossbred females, restricting crossbred replacement heifers' selection and mating plans.

Research groups and breeding organisations around the world have developed advanced and appropriate methods to perform accurate genomic evaluations for crossbred animals

(examples: Brøndum et al., 2011; Madalena et al., 1990; Wiggans et al., 2021). However, due to the lack of adequate management tools, there have been little efforts to phenotype and genotype crossbreds over the past decades. Therefore, none of the proposed advanced methods can be practically applied to the French crossbred population at the moment.

It is important to propose an intermediate approach, applicable at a small population scale and with outputs practically usable for breeders. This would promote performance recording and genotyping for crossbreds, bringing it to an adequate scale for more advanced and accurate methods. The TransEvagenoc project aimed to develop and test a pipeline based on a proposed intermediate approach (Clasen et al., 2023; Eiríksson et al., 2022, 2021), and to produce the first genetic evaluation for dairy crossbreds. The number of available performances being limited in crossbreds and in order to avoid any change in the pure breed evaluations, the crossbred individuals were evaluated by their average breed composition and the SNP effects estimated in pure breeds, as proposed by Eiríksson et al (2021).

Material and Methods

The crossbred evaluation is composed of two steps: the determination of the Breed of Origin of Alleles (BOA), and the actual crossbred evaluation. In-house programs were developed to perform the different steps. The pipeline requires both purebred and crossbred information.

Datasets

The pilot evaluation was conducted using crosses from any combination of the three main French dairy breeds: Holstein, Montbéliarde, and Normande. This initial selection of crossbred animals was performed by tracing back pedigree origins using Pedig (Boichard, 2002). Additionally, all individuals in the evaluation had to be genotyped (54,205 SNP) and phenotyped.

Five traits were considered: milk yield (MY), fat yield (FY), protein yield (PY), somatic cell score (SCS), and calving to first service interval (CTFS). To retain as many animals as possible, the pilot evaluation was performed using first-lactation phenotypes only. A total of 4,004 crossbred individuals were selected.

Methods description

Breed of Origin of Alleles: BreedOrigin

The BOA was determined using the in-house program BreedOrigin. Its performances in terms of accuracy and computational efficiency is described in the benchmark study of Liu et al. (2026). Its principle is simple: BreedOrigin analyses haplotypes. In a first step, the haplotype frequencies are estimated within each pure breed with a sufficiently large number of animals from each pure breed.

For the pilot evaluation, a sample of 5,000 individuals was selected to represent each pure breed. The first 3,000 individuals were the most widely used bulls (based on their number of offspring in the population), while the remaining 2,000 individuals were randomly selected from the most recent generation (born from 2023 to 2025) to capture the latest haplotypes present in the population.

The haplotypes of crossbreds were then compared, position by position, to those in purebreeds. A BOA was assigned when the haplotypic frequency was substantially higher in one breed; otherwise, the haplotype origin remained unknown. In subsequent iterations, unassigned segments flanked by segments assigned to the same breed were attributed to that breed. Otherwise, no BOA was assigned at this step. This iterative procedure progressively filled the gap, starting with haplotypes of 16 markers, then 8, 4, 2, and finally 1 marker.

Among other outputs, BreedOrigin provided crossbred genotypes including BOA information, together with the breed proportions and heterosis level of each crossbred individual.

Genetic evaluation: CrossSG

CrossSG program implements an extension of the GBLUP model designed for the genomic evaluation of crossbred animals. This approach accounts for the BOA when applying purebred marker effects to crossbred genotypes according to their BOA.

For the pilot model, and for each trait, the genetic value of each crossbred individual was estimated as follows:

$$\boldsymbol{y_i} = \sum_{\boldsymbol{b=1}}^{\boldsymbol{\#_breed}} \left((\boldsymbol{p_{i,b}}\mu_{\boldsymbol{b}}) + \sum_{\boldsymbol{j=1}}^{\boldsymbol{\#_{SNP}}} (\boldsymbol{\beta_{i,j,b}}\boldsymbol{X_{i,j,b}}) \right)$$

where y_i is the genetic value of individual i; μ is the genetic difference of breed b; $p_{i,b}$ is the proportion of breed b of individual i; $\beta_{i,j,b}$ is the breed of origin b of marker j of individual i; and $X_{i,j,b}$ is the effect of marker j, carried by individual i, centred on the allele frequency of marker j within breed b, such that :

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

where $k_{i,j,b}$ is the number of copies of the second allele at marker j originating from breed b in individual i; $n_{i,j,b}$ is the total number of alleles at marker j originating from breed b in individual i; and $f_{j,b}$ is the frequency of the second allele at marker j in breed b.

A heterosis term, ph_iH, could be added to this breeding value, where ph_i is the fraction of heterosis of individual i calculated by BreedOrigin as the proportion of loci with alleles originating from two different breeds; and H is the heterosis effect associated with the considered trait.

The marker effects and allele frequencies used in this pilot were obtained from the national purebred evaluations carried out by GenEval. The breed genetic differences (μ_b; expressed with the Holstein as standard) and heterosis effects (H) for all traits were obtained from Dezetter et al. (2015).

Discussion with the dairy breeding industry

Focus groups with the different stakeholders of the dairy cattle breeding industry were organised to discuss the future implementation and practical use of the crossbred evaluation.

The traits to be evaluated required careful consideration due to the difficulty of using multiple breeds, resulting in diverse measuring protocols, different trait definitions, or simply in the absence or presence of some traits in specific breeds.

Several scenarios were proposed for index standardisation. For ease of use, fairness across breeds, and technical considerations, the indices will be expressed on a crossbred scale in this pilot evaluation. The subset of the crossbred population used for the standardisation was selected to be representative of the breed composition observed in the overall current crossbred population. Additionally, raw indices will be shared with breeding organisations so that they can standardised them on their breed of choice.

Results & Discussion

Predictive ability

Including heterosis had little effect on EBV rankings (Pearson correlations > 0.99 for all traits) and slightly increased the correlation between EBV and phenotypes (Table 1).

Predictive ability, assessed as the Pearson correlation between genomic indices and phenotypes, was moderate but encouraging for the first crossbred indices. Our results showed that genomic prediction for crossbred dairy cattle is feasible, which is consistent with the conclusion of Steyn et al. (2021).

Table 1: Pearson correlations between EBV, EBV including heterosis (EBV+Het), and phenotype (pheno).

	Pearson correlation	
	pheno, EBV	pheno, EBV+Het
MY	0.35	0.36
PY	0.27	0.27
FY	0.28	0.28
SCS	0.31	0.31
CTFS	0.08	0.09

MY: milk yield; PY: protein yield; FY: fat yield; SCS: somatic cell score; and CTFS: calving to first service.

Out of the five traits, only CTFS showed a slightly different pattern for the results due to its low heritability and reversed optimum, i.e. lower values for the interval are desirable. Therefore, only results for MY and CTFS are detailed in this paper.

Genetic progress

The indices were standardised on a subset of the crossbred population. The genetic progress for MY and CTFS is shown in Figures 1.a and 1.b, respectively.

Before interpreting these figures, the overall breed proportions within the crossbred population across birth years were assessed to disentangle genetic progress from potential temporal changes in the representation of different crossbred types. For example, an increasing proportion of Holstein ancestry in the crossbred population could have mimicked the trend of genetic progress. As breed composition remained relatively constant across birth years, the observed trends can be attributed to genetic progress rather than changes in breed composition.

Accordingly, MY showed a clear upward genetic trend across birth years, indicating ongoing genetic improvement in the crossbred population, reflecting the genetic improvement achieved in the parental purebred populations (Figure 1a). In contrast, CTFS exhibited a weaker trend and greater year-to-year variability, consistent with its lower heritability (Figure 1b). The limited number of animals born after 2018 reflects the availability of crossbred genotypes, most of which originated from the Evagenoc project (2019–2022).

As shown in Figure 2, animals with higher genomic indices performed better in the first lactation. Animals with higher MY indices had better production and those with higher CTFS indices had shorter calving-to-first-service intervals, although the difference between index classes was less pronounced.

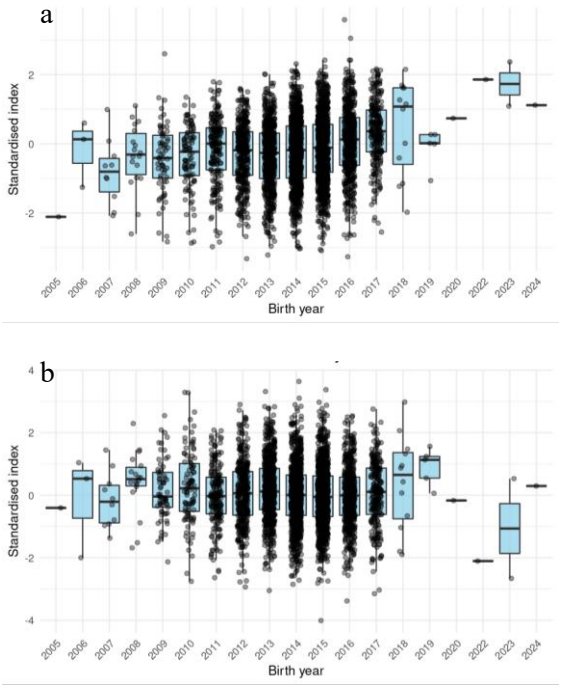


Figure 1. Genetic trend in crossbreds: Distribution of the genomic indices by birth year. a. Milk yield (MY) b. Calving to first service interval (CTFS) - note that lower values are desirable.

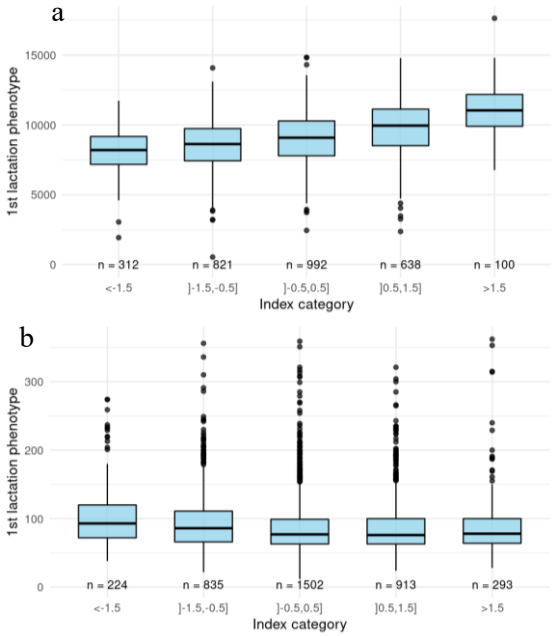


Figure 2. Distribution of first-lactation phenotypes across categories of genomic indices including heterosis effect. Boxplots represent the distribution of first-lactation phenotypes within each category. n corresponds to the category sample size. a. milk yield (MY). b. calving to first service (CTFS).

Evaluation strategy

Previous studies have demonstrated that genomic evaluations for crossbred dairy cattle can be implemented by combining genomic information from the parental breeds to the crossbred information (VanRaden et al., 2020; Wiggans et al., 2021). From a methodological perspective, the main challenge for a common evaluation comes from the difference of population sizes: purebred populations are much larger than crossbred populations and their information strongly dominate in the estimation of the SNP effects, leaving no opportunity for the crossbred information to impact estimates. From a practical perspective, implementing a common evaluation would also affect the organisation of national purebred evaluations and require substantial computational resources.

Because French crossbred resources are not yet enough developed, we used an intermediate method to obtain a genomic index for crossbred animals, as suggested by Eiriksson et al (2021). The observed associations between the indices and first-lactation performance suggest that this approach provides a useful tool until a dedicated crossbred genomic evaluation can be proposed.

Conclusions

Two programs, BreedOrigin and CrossSG, were developed within the TransEvagenoc project to implement genomic evaluations in the current French crossbred dairy population. They cope with the limited phenotypic and genomic data available in French dairy crossbreds, hampering the use of more advanced and adequate crossbred evaluation methods. This intermediate approach enabled the production and diffusion of the first genomic indices for crossbred dairy cattle, released at the end of 2025. Although only three breeds are presented in this paper, the routine evaluation includes the six breeds the most used in dairy crossbreeding in France.

By providing practical genomic evaluations for crossbred animals, this approach supports the choice of breeders using crossbreeding in France while encouraging the collection of additional phenotypic and genotypic data, paving the way for the future implementation of more comprehensive crossbred evaluation methods.

Acknowledgments

The work presented was supported by the Evagenoc and TransEvagenoc projects, funded by Apis-Gene, and by the contribution of several breeding companies that coordinated the genotyping of crossbred animals.

References

Boichard, D., 2002. Pedig: a fortran package for pedigree analysis suited for large populations. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*.

Brøndum, R.F., Rius-Vilarrasa, E., Strandén, I., Su, G., Guldbrandtsen, B., Fikse, W.F., Lund, M.S., 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. *J. Dairy Sci*. 94, 4700–4707. https://doi.org/10.3168/jds.2010-3765

Clasen, J. B., W. F. Fikse, G. Su, E. Karaman. 2023. Multibreed genomic prediction using summary statistics and a breed-origin of-alleles approach. *Heredity* 131, 33–42. https://doi.org/10.1038/s41437-023-00619-4.

Dezetter, C., Leclerc, H., Mattalia, S., Barbat, A., Boichard, D., Ducrocq, V., 2015. Inbreeding and crossbreeding parameters for production and fertility traits in Holstein, Montbéliarde, and Normande cows. *J. Dairy Sci.* 98, 4904–4913. https://doi.org/10.3168/jds.2014-8386

Eiríksson, J. H., E. Karaman, G. Su, and O. F. Christensen. 2021. Breed of origin of alleles and genomic predictions for crossbred dairy cows. *Genet. Sel. Evol.* 53, 84.

https://doi.org/10.1186/s12711-021-00678-3.

Eiríksson, J. H., K. Byskov, G. Su, J. R. Thomasen, and O. F. Christensen. 2022. Genomic predictions for crossbred dairy cows by combining solutions from purebred evaluation based on breed origin of alleles. *J. Dairy Sci.* 105, 5178-5191. https://doi.org/10.3168/jds.2021-21644.

IDELE, 2024. Statistiques des inséminations sur femelles laitières : le croisement laitier en 2024. Bilan des inséminations animales bovines en 2024. Rapport Résultats No. 0025203011.

Liu, H., Christensen, O.F., Eiríksson, J.H., Boichard, D., Milkevych, V., Thomasen, J.R., Karaman, E., 2026. Accuracy of local ancestry inference and its impact on genomic prediction in admixed dairy cattle populations. *J. Dairy Sci.* 109, 2743–2758. https://doi.org/10.3168/jds.2025-27019

Madalena, F.E., Teodoro, R.L., Lemos, A.M., Monteiro, J.B.N., Barbosa, R.T., 1990. Evaluation of Strategies for Crossbreeding of Dairy Cattle in Brazil. *J. Dairy Sci*. 73, 1887–1901. https://doi.org/10.3168/jds.S0022-0302(90)78869-8

Steyn, Y., Gonzalez-Pena, D., Bernal Rubio, Y.L., Vukasinovic, N., DeNise, S.K., Lourenco, D.A.L., Misztal, I., 2021. Indirect genomic predictions for milk yield in crossbred Holstein-Jersey dairy cattle. *J. Dairy Sci.* 104, 5728–5737. https://doi.org/10.3168/jds.2020-19451

VanRaden, P.M., Tooker, M.E., Chud, T.C.S., Norman, H.D., Megonigal, J.H., Haagen, I.W., Wiggans, G.R., 2020. Genomic predictions for crossbred dairy cattle. *J. Dairy Sci.* 103, 1620–1631. https://doi.org/10.3168/jds.2019-16634

Wiggans, G., VanRaden, P., Null, D., Nicolazzi, E., Jansen, G., Megonigal, J., 2021. Genomic evaluation of crossbred dairy cattle in the United States – an update. *Interbull Bulletin 56:17-21.*