

## **Faisabilité d'une évaluation génomique pour les animaux issus d'un croisement trois voies : Montbéliarde X Holstein X Rouge Danoise**

SAINTILAN R. (1,2), CUYABANO B.C.D. (1), BAUR A. (1,2), CROUE I. (1,2), DUCROCQ V. (1), THOMASEN J.R. (3), KARAMAN E. (4), BOICHARD D. (1), LECLERC H. (1,2), CROISEAU P. (1)

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

(2) ELIANCE, 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.

### **RESUME**

Ces dernières années, la stratégie de croisement s'est développée dans les élevages bovins laitiers au niveau international. Les éleveurs sont intéressés par le fait de garder des vaches croisées dans leur troupeau à la fois pour allier les atouts des races pures, compenser leurs faiblesses et bénéficier des effets d'hétérosis. Cependant, les outils génétiques ne sont pas adaptés pour gérer ces animaux croisés. Dans cette étude, nous évaluons les performances d'une évaluation génomique adaptée aux schémas de croisements rotationnels avec des données réelles. Cette évaluation génomique a été appliquée à une population comprenant des animaux purs de races Holstein, Montbéliarde et Rouge Danoise, ainsi que des animaux croisés issus de ces trois races en utilisant une méthode basée sur l'estimation de l'effet spécifique du SNP en fonction de la race pure d'origine des allèles.

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

SAINTILAN R. (1,2), CUYABANO B.C.D. (1), BAUR A. (1,2), CROUE I. (1,2), DUCROCQ V. (1), THOMASEN J.R. (3), KARAMAN E. (4), BOICHARD D. (1), LECLERC H. (1,2), CROISEAU P. (1)

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

(2) ELIANCE, 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.

### **SUMMARY**

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.

## INTRODUCTION

La stratégie de croisement dans les élevages bovins laitiers s'est considérablement développée ces dernières années, et cette tendance pourrait se renforcer à l'avenir pour faire face aux défis liés à l'agroécologie ou aux nouvelles contraintes induites par le changement climatique. Cette stratégie est particulièrement adaptée pour combiner les forces des races pures et compenser leurs faiblesses, ce qui est un moyen efficace pour les éleveurs d'obtenir des animaux plus adaptables et robustes, aboutissant à un système d'élevage plus durable. Le croisement est également pertinent pour diminuer la consanguinité ou, de manière équivalente, générer de l'hétérosis.

L'évaluation génomique en race pure est courante, et est en cours de développement pour les animaux issus de croisement terminal, mais seuls quelques pays évaluent systématiquement les animaux croisés dans des programmes de croisement continu. Pour la plupart des pays, les éleveurs engagés dans une démarche de croisement disposent d'informations limitées aux taureaux de race pure et aux performances brutes de leurs femelles croisées pour gérer leur sélection et leurs accouplements. Dans ce contexte, la littérature consacrée au croisement continu porte, pour la plupart, sur des données simulées (Eiriksson et al., 2021 ; Karaman et al., 2021) et rarement sur des applications sur données réelles (Sevillano et al., 2017). Dans cette étude, nous proposons une application de l'évaluation génomique pour les programmes de croisement continu, au travers de la méthodologie du GBLUP basée sur la détermination de l'origine raciale des allèles pour les animaux croisés (BOA). Cette approche consiste à estimer l'effet spécifique de chaque marqueurs SNP en fonction de sa race d'origine. La justification de cette approche est que le déséquilibre de liaison entre le SNP et la mutation causale diffère selon la race ; de plus, l'effet QTL peut dépendre de la race.

## 1. MATERIEL ET METHODES

Les données utilisées dans cette analyse sont issues des premières étapes d'un schéma de croisement à trois voies comprenant les races Montbéliarde (Mo), Holstein (Ho) et Rouge Danoise (RD). Cet ensemble de données a été complété par des données de chacune des trois races pures pour assurer une estimation précise des effets SNP intra-race. La procédure comprenait plusieurs étapes : imputation et phasage au niveau 50K des génotypes de tous les animaux ; identification du BOA pour les allèles de tous les animaux croisés ; évaluation génomique des cinq caractères de production laitière et mesure de la précision des prédictions génomique et des pentes de régression par une procédure de validation croisée.

### 1.1 GENOTYPES

Les données utilisées dans cette étude proviennent de la France et des pays nordiques (Suède, Danemark et Finlande). Elles consistent en 5 238 génotypes d'animaux croisés réalisés dans le cadre du projet GENTORE, majoritairement dans les pays nordiques où le croisement est plus développé, ainsi qu'en 20 000, 22 265 et 6 866 génotypes d'animaux purs Mo, Ho et RD, respectivement. Les génotypes des animaux RD ont été réalisés par les partenaires des pays nordiques et sont des typages moyenne densité. Actuellement en France, 53 498 marqueurs SNP autosomiques sont utilisés en routine pour la sélection génomique des différentes races bovines. Afin de disposer de l'ensemble des informations pour les 53 498 marqueurs pour l'ensemble des individus génotypés sur un panel de puces potentiellement très large, une procédure d'imputation a été réalisée. Elle consiste à estimer le

génotype le plus probable pour les marqueurs manquants d'un génotypage sur la base des fréquences alléliques de la population et sur la base d'information d'apparentement. L'imputation a été effectuée en deux étapes. Dans un premier temps, les génotypes des animaux de races pures ont été imputés avec FLmpute (Sargolzaei et al., 2014) avec le pipeline utilisé pour les évaluations officielles françaises. Cette procédure est réalisée de manière hebdomadaire. La deuxième étape consistait en l'imputation des génotypes des animaux croisés à l'aide de FLmpute en utilisant les génotypes imputés et phasés des animaux de races pures Ho, Mo et RD comme référence, et sans utiliser aucune information sur le pedigree.

### 1.2 ORIGINE RACIALE DES ALLELES (BOA)

Pour déterminer l'origine raciale des allèles portés par les animaux croisés, on s'intéresse aux fréquences alléliques raciales d'haplotypes de grande taille, un haplotype étant une séquence d'allèles de marqueurs présents sur un même chromosome. Ainsi, la fréquence ( $f$ ) de chaque haplotype de  $n$  SNP consécutifs a été estimée par comptage au sein de chacune des trois races pures définissant ainsi les fréquences de référence. Ensuite, l'origine raciale de chacun des haplotypes des animaux croisés a été déterminée en recherchant pour quelles races pures les haplotypes portés par les animaux croisés sont les plus fréquents, permettant d'estimer une composition raciale pour chaque animal croisé.

Le BOA d'un haplotype d'un animal croisé est dit de la race  $i$  si  $\frac{f_i}{\sum_{j=1}^{n_{\text{breed}}} f_j}$  était supérieur à un seuil de 0,90 (seuil par défaut).

L'ensemble des haplotypes du génome a ainsi été analysé par fenêtre coulissante en se déplaçant d'un SNP à la fois. La valeur initiale de la taille des haplotypes ( $n$ ) a été fixée à 16 (conduisant à 65 536 combinaisons haplotypiques théoriques). Lorsqu'une origine d'haplotype restait indéterminée, le processus était répété après que la taille de l'haplotype ait été divisée par 2. Lorsque l'origine d'un haplotype restait indéterminée mais que les haplotypes flanquants étaient attribués à une race, alors on attribue à l'haplotype inconnu la race des haplotypes flanquants. Lorsque  $n = 1$ , la faible proportion d'origines finalement restée indéterminée a été attribuée en fonction des fréquences alléliques à la race ayant la valeur la plus élevée. Cette procédure a été implémentée dans un logiciel interne développé par INRAE en fortran (BreedOrigin) et appliquée à la population cible des 5 238 animaux croisés. Pour les animaux croisés avec parents connus, les proportions de chaque race calculées à partir des BOA des SNP ont été ensuite comparées à celles obtenues sur la base des informations du pedigree.

### 1.3 HETEROSIS ET CROISEMENT

Comme décrit par Dezetter et al., 2015, les fractions d'hétérosis (traduisant la supériorité phénotypique de l'animal croisé par rapport à ses parents ( $H$ )) pour chaque animal croisé sont estimées à partir des proportions d'origines raciales de leurs parents en fonction du pedigree. Les valeurs du coefficient  $H$  ont été calculées comme décrit par Dechow et al. (2007) en utilisant les formules suivantes :

$$H = 1 - \sum_{i=1}^{n_{\text{breed}}} s_i d_i$$

Où  $s_i$  et  $d_i$  sont les proportions de gènes de père et de mère de la race  $i$ , respectivement. L'hétérosis a ensuite été pris en compte dans le modèle d'analyse des performances des animaux croisés.

### 1.4 PHENOTYPES

Cinq caractères de production laitières ont été analysés dans cette première étude : production de lait standardisé

sur la durée de la première lactation (LAIT), quantité de protéines (MP), teneur en protéines (TP), quantité de matières grasses (MG) et teneur en matières grasses (TB). L'évaluation génomique est une procédure en deux étapes, une évaluation polygénique afin de calculer les performances corrigées et l'évaluation génomique en tant que telle.

Dans un premier temps, un modèle polygénique a été utilisé pour estimer tous les effets non génétiques (troupeau-année, âge-année et année-mois de vêlage), les effets d'hétérosis (régression selon H) et en prenant compte des variances résiduelles hétérogènes, tel que décrit par Dezetter et al (2015). Les groupes de parents inconnus pris en compte dans l'évaluation étaient fonction de la race d'origine et la composition raciale des croisements d'après les informations du pedigree

Les performances des femelles de races pures et croisées, exprimées en écarts de rendement (YD), correspondent à des performances corrigées des effets fixes pris en compte dans le modèle précédent ainsi que des effets aléatoires non génétiques. Par caractère et par animal, un YD est disponible avec un poids associé de 1.

Les données de 22 571 femelles génotypées de races pures et 4 380 femelles génotypées croisées avec YD ont été utilisées pour constituer la population de référence des évaluations génomiques.

Deux séries d'évaluations ont été réalisées, la première en ne tenant pas compte des effets d'hétérosis, et la seconde en tenant compte des effets d'hétérosis. On obtient ainsi pour chaque animal, deux YD par caractère, l'un avec correction pour H et l'autre non corrigé pour H.

## 1.5 EVALUATION GENOMIQUE POUR LES ANIMAUX CROISES EN TENANT COMPTE DU BOA

Pour tenir compte des animaux croisés dans un modèle d'évaluation génomique, nous avons proposé une extension du modèle SNP-BLUP où un effet SNP  $\beta_i$  est estimé pour chaque race comme décrit dans le modèle suivant :

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

Où  $y_i$  est le YD de l'animal  $i$ ,  $\mu_b$  est un vecteur de moyennes définies au sein de chaque race,  $p_{i,b}$  est la proportion de la race  $b$  dans le génome de l'individu  $i$ , estimée par l'approche BOA.  $X_{i,j,b}$  est le contenu allélique du SNP  $j$  qui provient de la race  $b$  pour l'animal  $i$ , centré sur la fréquence allélique du SNP dans la race  $b$  :

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

Où  $k_{i,j,b}$  et  $n_{i,j,b}$  sont le nombre d'allèles alternatifs (recodifié en "2") et le nombre total d'allèles du SNP  $j$  qui provient de la race  $b$  pour l'animal  $i$ , respectivement ;  $f_{j,b}$  est la fréquence de l'allèle « 2 » du SNP  $j$  dans la race  $b$ .

Pour évaluer l'efficacité de la méthode d'évaluation génomique testée, l'ensemble de données des animaux de la population de référence a été divisé en un ensemble de données dites d'apprentissage avec des animaux ayant à la fois des informations de génotypes et des YD pris en compte pour estimer les effets SNP en fonction de leur BOA, et un ensemble de données dites de validation consistant en une liste de femelles croisées avec génotypes dont les YD n'ont pas été prises en compte et pour lesquelles les prédictions génomiques ont été calculées en utilisant les effets estimés de marqueurs sur la population d'apprentissage. Les prédictions génomiques sont ensuite comparées aux phénotypes observés (YD).

Deux types de résultats sont analysés, (1) la corrélation entre la prédiction génomique de l'animal et son YD par caractère, (2), la pente de la droite de régression entre prédiction génomique et YD. Dans notre étude, la population

de validation était constituée des 2000 vaches croisées les plus jeunes sans descendance.

## 2. RESULTATS

### 2.1 RACE D'ORIGINE DES ALLELES

Les tableaux 1 et 2 présentent la proportion d'allèles provenant des différentes races, et l'écart type associé, estimée en exploitant une information génomique (tableau 1) ou généalogique (tableau 2) Pour les animaux génotypés croisés, 94% des allèles ont été attribués à une race, 48 % des marqueurs étaient d'origine Ho, 34 % d'origine RD et 13,6 % d'origine Mo. En comparaison, sur la base des informations généalogiques, les origines correspondantes étaient respectivement de 56,3%, 29,8% et 11,8%. La corrélation entre les compositions des races pour les deux méthodologies est de 0,95, 0,97, 0,99 pour les races Rouge Danoise, Holstein et Montbéliarde, respectivement.

| Race d'origine     | Composition raciale avec BOA |            |
|--------------------|------------------------------|------------|
|                    | Moyenne                      | Ecart-type |
| Montbéliarde (Mo)  | 13,60%                       | 20,00%     |
| Holstein (Ho)      | 48,00%                       | 19,40%     |
| Rouge Danoise (RD) | 34,00%                       | 17,80%     |

**Tableau 1.** Résultats de la détermination de la race d'origine sur la base des informations BOA

| Race d'origine     | Composition raciale sur Pedigree |            |
|--------------------|----------------------------------|------------|
|                    | Moyenne                          | Ecart-type |
| Montbéliarde (Mo)  | 11,80%                           | 19,20%     |
| Holstein (Ho)      | 56,30%                           | 20,20%     |
| Rouge Danoise (RD) | 29,80%                           | 17,80%     |

**Tableau 2.** Résultats de la détermination de la race d'origine sur la base des informations généalogiques

### 2.2 EVALUATION GENOMIQUE DES ANIMAUX CROISES

Les corrélations entre YD et les prédictions génomiques et les pentes de régression associées sont présentées dans les tableaux 3 et 4. Sur la population d'apprentissage, toutes les corrélations sont d'environ 0,80 pour les caractères LAIT, TB et TP et d'environ 0,90 pour les caractères MG et MP. Dans la population de validation, sans intégration de l'effet d'hétérosis (tableau 3), les corrélations varient de 0,36 pour le TP et de 0,65 pour la MG.

| Caractère | Apprentissage |       | Validation  |       |
|-----------|---------------|-------|-------------|-------|
|           | Corrélation   | Pente | Corrélation | Pente |
| LAIT      | 0,82          | 1,19  | 0,41        | 1,07  |
| TB        | 0,78          | 1,22  | 0,38        | 1,15  |
| TP        | 0,76          | 1,26  | 0,36        | 1,16  |
| MG        | 0,92          | 1,15  | 0,65        | 1,05  |
| MP        | 0,93          | 1,12  | 0,62        | 1,00  |

**Tableau 3 :** Corrélations et pentes de régression des YD non ajustées pour les effets d'hétérosis avec les prédictions génomiques pour cinq caractères de production laitière.

| Caractère | Apprentissage |       | Validation  |       |
|-----------|---------------|-------|-------------|-------|
|           | Corrélation   | Pente | Corrélation | Pente |
| LAIT      | 0,83          | 1,17  | 0,40        | 1,03  |
| TB        | 0,80          | 1,20  | 0,36        | 1,09  |
| TP        | 0,77          | 1,23  | 0,35        | 1,11  |
| MG        | 0,92          | 1,15  | 0,65        | 1,05  |
| MP        | 0,93          | 1,12  | 0,62        | 1,00  |

**Tableau 4 :** Corrélations et pentes de régression des YD ajustées pour les effets d'hétérosis avec les prédictions génomiques pour cinq caractères de production laitière.

L'ajustement des performances pour l'hétérosis a faiblement affecté ces corrélations (tableau 4). Le gain de corrélation obtenu dans la population d'apprentissage après ajustement pour l'hétérosis est marginal (+1 point pour le LAIT et le TP et +2 points pour le TB). Pour la population de validation, les pentes de régression sont relativement proches de 1 pour le LAIT, la MP et la MG et légèrement surestimées pour les taux, que ce soit avec ou sans ajustement des performances pour les effets d'hétérosis.

### 3. CONCLUSION

Dans cette étude, nous proposons une application sur données réelles d'une évaluation génomique dans un schéma de croisement rotationnel basée sur un modèle SNP-BLUP prenant en compte l'origine raciale des allèles pour les animaux croisés (BOA).

Cette approche nécessite d'imputer et de déterminer l'origine paternelle ou maternelle de chacun des allèles des données de génotypage des animaux croisés afin de prédire les BOA de ces animaux. Ces étapes ont été testées et un écart moyen aux valeurs obtenues sur pedigree d'environ 1,5 % pour la BOA prédite a été observé. Les compositions raciales des animaux croisés mesurées à partir des informations généalogiques des animaux et estimées à partir des données de génotypage montrent de très fortes corrélations entre elles. Les évaluations génomiques réalisées sur ce jeu de données réelles ont montré une précision de prédiction honorable de l'ordre de 0,30 pour les caractères à héritabilité de 0,30 et de l'ordre de 0,60 pour les caractères à héritabilité de 0,50. L'ajustement des performances pour les effets d'hétérosis n'a pas eu d'impact sur la précision de la prédiction des valeurs d'élevage. Cette première étude encourageante sera complétée par des analyses sur d'autres types de croisement et d'autres caractères avant d'envisager un déploiement en production.

*Cette étude a été financée par APIS GENE et par le programme « European Union's Horizon 2020 research and innovation program - GenTORE - under grant agreement No. 727213 ».*

**Dechow, C. D., Rogers, G. W., Cooper, J. B., Phelps, M. I., & Mosholder, A. L. 2007.** JDS - 90(7) :3542-3549.  
**Dezetter, C., Leclerc, H., Mattalia, S., Barbat, A., Boichard, D. et al. 2015.** JDS - 98(7) :4904-4913.  
**Eiriksson, J.H., Karaman, E., Su, G. & Christensen, O. F. 2021.** GSE – 53 : 84  
**Karaman, E., Su, G., Croue, I., & Lund, M. S. 2021.** GSE - 53(1) :1-15.  
**Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. 2014.** BMC genomics - 15(1):1-12.  
**Sevillano, C. A., Vandenplas, J., Bastiaansen, J. W., Bergsma, R., & Calus, M. P. 2017.** GSE - 49(1):1-14.