

# Importance du phénotypage pour maintenir la précision des prédictions génomiques des caractères mesurés en station

Céline CARILLIER-JACQUIN (1), Alban BOUQUET (2), Pauline BRENAUT (2), Chris SIMON (1)

(1) GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, 31320, Castanet-Tolosan, France

(2) IFIP-Institut du Porc, La Motte au Vicomte, 35651 Le Rheu Cedex, France

[celine.carillier-jacquin@inrae.fr](mailto:celine.carillier-jacquin@inrae.fr)

## Importance of phenotyping to maintain the accuracy of genomic predictions of traits measured in a test station

Genomic evaluation of French maternal lines, set up in 2016, has helped increase genetic progress, especially for reproductive traits. However, computational problems have emerged for genomic evaluation of certain production traits for which phenotyping capacity is limited. This particularly concerns genotyped candidates on breeding farms that have no phenotypes and only a few phenotyped relatives. This data structure seems to pose convergence problems for predicting genomic breeding values. To check this hypothesis, we simulated such a situation, based on a set of actual phenotype data measured for all farm candidates and genotypes. The simulation consisted of deleting phenotypes of the animals measured on-farm in order to reproduce the data structure encountered for the traits recorded at the FGPorc/INRAE test station in Le Rheu. Phenotypes were then progressively added in different scenarios to identify whether prediction accuracy improved and to estimate the number of phenotypes required. The simulations showed that the unbalanced structure between genotypes and phenotypes was responsible for the computational problems that led to low accuracy of genomic predictions. Phenotyping 12% of all pigs phenotyped at 100 kg each year made it possible to solve the computational problems observed and to recover 61% of the maximum expected accuracy. In conclusion, these results highlight the importance of collecting large-scale phenotypes in the context of genomic selection schemes. Further studies will be conducted to study the impact of genotyping animals measured at the station.

## INTRODUCTION

L'évaluation génomique dans les lignées porcines maternelles françaises a été mise en place en 2016 et a permis d'augmenter le progrès génétique, notamment pour les caractères de reproduction. Cependant, des problèmes calculatoires sont apparus pour l'évaluation génomique de certains caractères de production mesurés en station, dont la capacité de phénotypage est limitée. Cela concerne particulièrement les candidats génotypés des élevages de sélection qui ne disposent pas de phénotypes en station et peu d'apparentés phénotypés. La structure de données pour ce type de caractère (i.e. porcs génotypés (candidats) non phénotypés et porcs phénotypés (animaux station) non génotypés) pourrait être responsable des problèmes de convergence observés pour la prédiction des valeurs génomiques. Dans cette étude, nous avons simulé, à partir de phénotypes réels mesurés sur l'ensemble des candidats à la sélection, différents scénarios pour mimer la situation de phénotypage partiel rencontrée dans le cas des caractères mesurés en station et évaluer l'impact d'une augmentation du nombre de porcs phénotypés sur la précision des valeurs génomiques obtenues.

## 1. MATERIEL ET METHODES

### 1.1. Les différents scénarios

Les différents scénarios ont été simulés à partir des phénotypes d'âge à 100 kg disponibles en ferme et en station (UE Porcs Rennes, 35590 Saint-Gilles – doi : 10.1545

4/1.5573932732039927E12) pour les animaux de race Landrace. Pour l'ensemble des scénarios considérés, les génotypes de l'ensemble des porcs Landrace génotypés en routine ont été considérés, soit 9 133 génotypes d'animaux nés entre 2004 et 2018. Afin de reproduire la situation observée pour les caractères mesurés en station, les phénotypes des porcs génotypés (candidats), n'ont été inclus dans aucun des scénarios. Le scénario 1 correspond à la structure actuelle pour les caractères mesurés en station, où environ 1% des porcs sont phénotypés pour l'âge à 100 kg sont phénotypés. Les scénarios suivants (5, 4, 3, 2) correspondent à l'inclusion de différentes proportions de phénotypes correspondant, respectivement, à 82%, 36%, 19%, 12% et 1% des phénotypes totaux (Tableau 1). Le scénario 6, correspondant au jeu de données complet, comprenait 144 312 porcs phénotypés.

**Tableau 1** – Nombre de porcs génotypés et phénotypés selon les différents scénarios

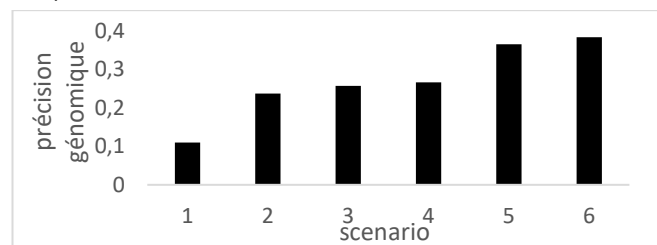
|            | Nombre de phénotypes | % phénotypes totaux | Nombre de génotypes |
|------------|----------------------|---------------------|---------------------|
| Scénario 1 | 1 898                | 1                   | 9 133               |
| Scénario 2 | 16 837               | 12                  |                     |
| Scénario 3 | 27 699               | 19                  |                     |
| Scénario 4 | 52 069               | 36                  |                     |
| Scénario 5 | 118 430              | 82                  |                     |
| Scénario 6 | 144 372              | 100                 |                     |

## 1.2. La validation croisée

Les évaluations génomiques ont été réalisées pour chaque scénario avec la méthode BLUP génomique en une étape (Legarra *et al.*, 2009) en utilisant le programme blupf90 (Misztal *et al.*, 2014). Les individus de validation correspondent aux porcs dont les phénotypes ont été ignorés / masqués dans les évaluations génomiques, afin de prédire leurs valeurs génomiques uniquement à partir de leurs génotypes. Dans cette étude, les animaux de validation correspondent aux 1 449 porcs nés en 2018 dont les phénotypes pour l'âge à 100 kg sont connus par ailleurs. La précision des valeurs génomiques obtenues dans chaque scénario a été calculée comme la corrélation de Pearson entre les valeurs génomiques obtenues dans le scénario considéré pour les individus de validation et leurs phénotypes corrigés des effets fixes, divisée par la racine carrée de l'héritabilité (Zhang *et al.*, 2018). De plus, les pentes de régression des valeurs génomiques en fonction des phénotypes corrigés pour ces mêmes individus ont également été calculées afin d'estimer le biais des valeurs génomiques estimées.

## 2. RESULTATS ET DISCUSSION

Les précisions génomiques obtenues sont comprises entre 0,11 et 0,38 pour les scénarios 1 à 4, respectivement (Figure 1). La faible valeur de précision observée pour le scénario de référence (scénario 1), qui reproduit le cas d'un caractère mesuré en station, confirme notre hypothèse considérant que le faible nombre de phénotypes pouvait expliquer les problèmes de prédictions génomiques observés pour ces caractères. L'ajout de phénotypes dans les évaluations s'accompagne d'une augmentation progressive de la précision des prédictions.



**Figure 1** – précisions des valeurs génomiques obtenues dans les différents scénarios pour les individus de validation

En effet, considérer 12% des animaux phénotypés totaux au lieu de 1% permet d'augmenter les précisions de 0,11 à 0,25, soit plus qu'un doublement de la précision. Ainsi, pour les caractères typiquement mesurés en station comme la qualité de viande, l'intégration de phénotypes mesurés à plus grande

échelle sur des candidats élevés en ferme semble être une solution possible pour améliorer la précision des prédictions génomiques des caractères station.

Le choix des phénotypes des animaux à considérer pour les évaluations génomiques a donc un impact sur les précisions obtenues, comme montré dans la littérature (Lee *et al.*, 2017 ; Liu *et al.*, 2011).

Le niveau de précision des valeurs génomiques obtenues dans les scénarios 2 à 4 est comparable à ce qui est estimé dans les populations porcines (Lopez *et al.*, 2018 ; Song *et al.*, 2017). Dans le cas de l'âge à 100 kg, considérer seulement 12% des animaux dans les évaluations génomiques permet d'obtenir des niveaux acceptables de précisions génomiques.

Les biais de prédiction les plus faibles sont obtenus lorsque les pentes de régression sont proches de 1. Dans cette étude, les biais obtenus sont importants (pentes de régression entre 0,04 et 0,67, Tableau 2), ce qui peut être expliqué par le niveau de précision obtenu. De même que les précisions, les biais sont réduits avec l'ajout de phénotypes d'animaux de ferme.

**Tableau 2** – Pentes de régression des valeurs génomiques en fonction des phénotypes corrigés pour les individus de validation dans les différents scénarios

| Scénarios | 1    | 2    | 3    | 4    | 5    | 6    |
|-----------|------|------|------|------|------|------|
| Pentes    | 0,04 | 0,27 | 0,42 | 0,44 | 0,69 | 0,67 |

## CONCLUSION

Le faible nombre de phénotypes disponibles pour l'évaluation génomique des caractères mesurés exclusivement en station, tels que la qualité de viande jusque récemment, semble être responsable de problèmes de qualité de prédiction limitant la précision des valeurs génomiques estimées. La prise en compte de phénotypes supplémentaires permet d'améliorer la précision des valeurs génomiques. Phénotyper seulement 12% des candidats sur ces caractères permettrait d'augmenter la précision de 11% à 25%.

Le développement actuel du phénotypage de candidats sur l'ensemble des caractères habituellement mesurés en station comme la qualité de viande, l'efficacité alimentaire ou les qualités de carcasse, doit permettre d'améliorer la précision des valeurs génomiques prédites pour les candidats.

Le génotypage d'animaux phénotypés en station pourrait également permettre d'améliorer ces précisions. Des études sont actuellement en cours afin de confirmer et quantifier l'intérêt du génotypage de collatéraux.

## REFERENCES BIBLIOGRAPHIQUES

- Lee S.H., Clark S., Van der Werf J.H.J., 2017. Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. *PLoS One* 12, 12.
- Legarra A., Aguilar I., Misztal I., 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663.
- Liu Z., Seefried F., Reinhardt F., Rensing S., Thaller G., Reents R., 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet Sel Evol.*, 43, 19.
- Lopez B.I., Song C.W., Seo K.S., 2018. Efficiency of genomic selection to improve meat quality in pigs using ZPLAN+. *Indian J. Anim. Res.* 53, 294-298.
- Misztal I., Tsuruta S., Lourenco D., Aguilar I., Legarra A., Vitezica Z., 2014. Manual for BLUPF90 family of programs. University of Georgia, Athens, USA
- Song H., Zhang J., Jiang Y., Gao H., Tang S., Mi S., Yu F., Meng Q., Xiao W., Zhang Q., Ding X., 2017. Genomic prediction for growth and reproduction traits in pig using an admixed reference population. *J. Anim. Sci.* 95, 3415–3424.
- Zhang Z., Xiao Q., Zhang Q., Sun H., Chen J., Li Z., Xue M., Ma P., Yang H., Xu N., Wang Q., Pan Y., 2018. Genomic analysis reveals genes affecting distinct phenotypes among different Chinese and western pig breeds. *Scientific Reports* 8, 13352.