# Can microbial data improve prediction of breeding values of efficiency traits in pigs fed conventional or fiber diets?

Vanille Déru, Franscesco Tiezzi, Céline Carillier-Jacquin, Benoit Blanchet, Laurent Cauquil, Olivier Zemb, Alban Bouquet, Christian Maltecca, H Gilbert

# Can microbial data improve prediction of breeding values of efficiency traits in pigs fed conventional or fiber diets?

**V.Déru[1,2], F. Tiezzi[3,4], C. Carillier-Jacquin[1], B. Blanchet[5], L. Cauquil[1], O. Zemb[1], A. Bouquet[6], C. Maltecca[3] and H. Gilbert[1*]**

[1] GenPhySE, INRAE, 31320 Castanet-Tolosan, France; [2] France Génétique Porc, 35651 Le Rheu, France; [3] Department of Animal Science, North Carolina State University, Raleigh, NC, USA; [4] Department of Agriculture, Food, Environment and Forestry, University of Florence, 50144, Firenze, Italy; [5] UE3P, INRAE, 35590 Saint-Gilles, France; [6] IFIP-Institut du Porc, 35650 Le Rheu, France; [*]helene.gilbert@inrae.fr

## Abstract

Recently, digestive efficiency (DE) was proposed as a trait of interest to improve feed efficiency (FE) in pigs, especially when they are fed with alternative feeding resources. Both are influenced by the host genetics, and also by the gut microbiota composition. The goal of this study was to quantify the impact of faecal microbial information on the prediction accuracies of genomic estimated breeding values (GEBVs) of FE and DE traits for pigs fed conventional or fiber diets. For DE traits, gains in prediction accuracy of GEBVs were increased by about 18% when microbial information was included in linear mixed models. In addition, these gains of prediction accuracy were very similar in both diets. For FE traits, no improvement was observed. Thus, the addition of microbial information in breeding programs is promising to better estimate GEBVs for DE traits.

## Introduction

Feed costs in pig production currently represent between 60 to 70% of the total cost of pork production (Patience *et al.*, 2015). Feeding pigs with by-products from the agri-food and biofuel industry is an option to reduce feed costs and feed-food competition. In this context, feed efficiency (FE) remains the major objective of selection, but recent research suggested that digestive efficiency (DE) can contribute to FE (Harris et *al.*, 2012; Mauch *et al.*, 2018), and more especially with alternative feeding stuff (Déru *et al.*, 2021a). The DE is influenced by host genetics (Déru et al., 2021a), and it has significant genetic correlations with some faecal microbiota traits (Déru *et al.*, 2021b). In addition, faecal microbiota can explain more than 44% of the phenotypic variation for DE traits (Verschuren *et al.*, 2020; Déru *et al.*, 2021c), and a moderate part, around 20%, for FE traits variance (Camarinha-Silva *et al*., 2017, Aliakbari *et al*. 2021). The goal of the present study was to evaluate if adding microbial information in the prediction models of breeding values can improve their prediction accuracy for DE and FE traits.

## Materials & Methods

***Animals and diets.*** Data comprised Large White male pigs reared at INRAE UE3P - France Génétique Porc phenotyping station. Couples of full-sibs were raised in post-weaning facilities until nine weeks of age. When moved to growing-finishing pens, one of the siblings was fed with a two-phase conventional (CO) European diet and the other one a two-phase high-fiber (HF) diet. Diets differed in net energy (NE) (9.6 MJ/kg of NE for the CO diet and 8.2 MJ/kg of NE for the HF diet), and neutral detergent fiber (from 13.90 to 15.12 % for the CO diet; from 23.82 to 24.46 % for the HF diet), as described in Déru *et al.* (2020).

***Data.*** Individual average daily gain (ADG) and daily feed intake (DFI) were measured between 35 and 115 kg live weight, with GenStar single place electronic feeders (Genstar, Skiold Acemo, Pontivy, France). Feed conversion ratio (FCR) was calculated as DFI/ADG. Residual feed intake (RFI) was determined as a multiple linear regression of DFI on ADG, average metabolic body weight, lean meat percentage and carcass yield (Déru *et al.*, 2020). Microbiota composition and DE were obtained from a unique faecal spot sampling at 16 weeks of age. The DE was predicted as digestibility coefficients (DC) of energy, nitrogen and organic matter using near infrared spectrometry (Déru *et al.*, 2021a). Microbiota composition was obtained from sequencing the V3-V4 regions of the 16S rRNA (Déru *et al.*, 2021b), and 14,366 operational taxonomic units (OTU) were identified. Sequence reads were rarefied at 10,000 sequences per sample.
Genotyping was carried out using 70K SNP GeneSeek GGP Porcine chip. Quality control (QC) was carried out: the call rate per individual, i.e. the percentage of genotype present per individual, with a threshold set at 95%, and the SNP call rate, i.e. the percentage of genotype by SNP with a threshold set at 95%. The SNPs with minor allele frequencies lower than 5%, with a significant deviation from Hardy-Weinberg equilibrium ($P < 0.000001$) were also deleted. After QC, 48,919 markers were available for further analyses.
Genomic estimated breeding values (GEBVs) for FCR and RFI (FE traits), ADG, DFI, and DE traits were obtained for pigs that had all microbiota, genomic and digestive efficiency data, i.e. 1,082 pigs raised in 32 batches.

***Statistical analyses.*** The microbial relationship matrix **M** was computed after filtering out the OTUs present in less than five samples and with an average abundance lower than 0.001% (2,399 OTUs kept), as $\mathbf{M} = \mathbf{TT}'/k$, where **T** is the $n$ x $k$ matrix containing the centred and standardized log abundancies of each OTU, $n$ the number of pigs and $k$ the number of OTUs (Ross *et al.*, 2013). The genomic relationship matrix **G** was constructed using the first VanRaden method (2008). Predictions were obtained with a Bayesian model fitted with the BGLR package in R (Pérez *et al.*, 2014). Two models were constructed, $\mathbf{y}=\mathbf{X\beta}+\mathbf{Zu}+\mathbf{Wm}+\mathbf{e}$ (model M+G) and $\mathbf{y}=\mathbf{X\beta}+\mathbf{Zu}+\mathbf{e}$ (model G), with **y** the vector of phenotypes for a given trait, **X** the incidence matrix relating observations to fixed effects, **β** the vector of fixed effects, **Z** the incidence matrix for the genetic effects, $\mathbf{u} \sim N(0, \mathbf{G}\sigma^2_u)$ the vector of additive genetic effects, **W** the incidence matrix for the microbiota effects, $\mathbf{m} \sim N(0, \mathbf{M}\sigma^2_m)$ the vector of microbiota effects, and $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$ the random residuals.

***Gains of prediction accuracies.*** The advantage of including the microbial information to predict GEBVs was quantified via changes in prediction accuracies of GEBVs from model G to model M+G. A forward cross-validation approach was used, considering the first 25 batches of the population (876 pigs) as a reference population and the last seven batches (205 "candidate" pigs) as a validation population. The reference GEBV, $GEBV_{C,M+G,}$ were obtained with model M+G on the complete population (i.e. phenotypes available for the reference and validation sets). Then, to assess the gains in prediction accuracy with models M and M+G compared to this reference, the criteria $\rho_M$ and $\rho_{M+G}$ were computed as the Pearson correlation coefficients computed between $GEBV_{C,M+G}$ and GEBVs obtained with model G ($GEBV_{P,G}$) or M+G ($GEBV_{P,M+G}$) with a partial dataset, i.e. where phenotypes were available for the reference population and not available for the validation population (Legarra and Reverter, 2018). First, gains of prediction accuracies were obtained for the full validation set. Next, to test for differences in prediction accuracies between diets, $\rho_M$ and $\rho_{M+G}$ were computed separately for the CO (96 pigs) and the HF diet (109 pigs).

**Results**

The gains of prediction accuracy combining the microbiota and genomic effects were evaluated using the predictions of the M+G model for the complete dataset as reference GEBVs. Indeed, assuming that microbiota composition affects the trait variance components, it should be the model providing the most accurate GEBVs (David and Ricard, 2019). It should be noted that with the complete dataset, correlations between GEBVs obtained with model G and those with model M+G were high, ranging from 0.97 to 0.99 for ADG, DFI and FE traits, and from 0.80 to 0.85 for DC.

**Table 1. Correlation coefficients $\rho$ [95% confidence intervals] between estimated breeding values of the validation population obtained with the complete ($GEBV_{C,M+G}$), and with the partial datasets predicted without ($GEBV_{P,G}$) and with ($GEBV_{P,M+G}$) microbial information, considering both diets in the validation population, and separating the validation population according to the diet.**

| | $\rho(GEBV_{P,G}, GEBV_{C,M+G})$ | | | $\rho(GEBV_{P,M+G}, GEBV_{C,M+G})$ | | |
|---|---|---|---|---|---|---|
| Traits | Both diets | CO diet | HF diet | Both diets | CO diet | HF diet |
| FCR | 0.60 [0.51;0.68] | 0.62 [0.47;0.73] | 0.58 [0.45;0.70] | 0.59 [0.49;0.67] | 0.59 [0.44;0.71] | 0.58 [0.44;0.69] |
| RFI | 0.62 [0.53;0.70] | 0.72 [0.60;0.80] | 0.56 [0.42;0.68] | 0.61 [0.52;0.69] | 0.70 [0.58;0.79] | 0.55 [0.40;0.67] |
| DFI | 0.60 [0.51;0.68] | 0.72 [0.61;0.80] | 0.57 [0.43;0.68] | 0.62 [0.51;0.69] | 0.71 [0.60;0.80] | 0.56 [0.41;0.68] |
| ADG | 0.70 [0.62;0.76] | 0.70 [0.58;0.79] | 0.70 [0.58;0.78] | 0.70 [0.62;0.76] | 0.69 [0.57;0.78] | 0.71 [0.60;0.79] |
| DC of energy | 0.55 [0.45;0.64] | 0.56 [0.40;0.68] | 0.55 [0.40;0.67] | 0.67 [0.58;0.74] | 0.65 [0.52;0.76] | 0.69 [0.57;0.77] |
| DC of OM | 0.53 [0.42;0.62] | 0.55 [0.39;0.67] | 0.52 [0.37;0.64] | 0.64 [0.55;0.71] | 0.65 [0.51;0.75] | 0.66 [0.54;0.75] |
| DC of nitrogen | 0.51 [0.40;0.61] | 0.48 [0.31;0.62] | 0.64 [0.39;0.66] | 0.62 [0.53;0.70] | 0.61 [0.47;0.72] | 0.64 [0.51;0.74] |

FCR = feed conversion ratio; DFI = daily feed intake; ADG = average daily gain; RFI = residual feed intake; DC = digestibility coefficient; OM = organic matter; CO = conventional; HF = high fiber

For FE traits, DFI and ADG, $\rho$ were not improved with the model M+G in comparison to the model G (Table 1). In addition, no difference was observed between $\rho$ computed separately for pigs fed the CO and the HF diets, except for DFI and RFI that had slightly higher gains of prediction accuracy in the CO than in the HF diet (for instance 0.72 vs 0.57 for DFI with CO and HF, respectively) with both models. For DE traits, ratios of prediction accuracies tended to be higher in model M+G (~0.65) than in model G (~0.53). There was no difference in prediction accuracy ratios based on the diet for DC, except for DC of nitrogen with model G that had lower $\rho$ in CO than in HF diet. However, because 95% confidence intervals were large due to limited number of pigs in these sub-groups, these differences between diets were not significant.

**Discussion**

Based on these first results, the joint analysis of microbiota and genomic information does not appear to increase prediction accuracies of GEBVs for FE and growth traits. However, the addition of microbial information to genomic models seemed to improve prediction accuracies for DE traits. A previous study highlighted that a higher proportion of phenotypic variance was explained by gut microbiota for DE traits (≈ 50-60%) than for FE traits (≈ 20%) in this dataset (Déru *et al.*, 2021c). This could explain why adding microbial information in the linear mixed models has more impact for the former traits. However, we also previously found higher proportions of variance due to microbial information in the HF diet than in the CO diet for DE traits, which did not seem to affect the ratios of prediction accuracies obtained when microbial and genomic information were both included in the model. In addition, higher prediction accuracies were observed in the HF diet compared to the CO diet for DC of nitrogen. Previously, heterogeneity of residual variances was observed between diets for this trait (Déru *et al.*, 2021a), which could not be accounted for in the BGLR models, and could explain the higher prediction accuracies in the HF diet. Finally, the GEBV dispersions, appreciated via the regression of GEBVs with the complete dataset on GEBVs of the partial dataset with model G or model M+G (not presented), were also very similar in the two diets.

We can thus conclude that there is an advantage of adding microbial information in linear mixed models to improve the accuracy of breeding values only for DE traits. Furthermore, the accuracy of GEBV predictions was not impaired with the alternative high fiber diet.

**References**


Aliakbari A., Zemb O., Barilly C., Billon Y. and Gilbert H. (2021). Proc. of the 72nd EAAP, Davos, Switzerland.

Camarinha-Silva A., Maushammer M., Wellmann R., Vital M., Preuss S. et al. (2017) Genetics 206(3):1637-1644. https://doi.org/10.1534/genetics.117.200782

Déru V., Bouquet A., Hassenfratz C., Blanchet B., Carillier-Jacquin C. et al. (2020) Animal 14:2236–45. https://doi.org/10.1017/S1751731120001275

Déru V., Bouquet A., Labussière E., Ganier P., Blanchet B. et al. (2021a) J Animal Breed Genet 138:246-58. https://doi.org/10.1111/jbg.12506

Déru V., Bouquet A., Zemb O., Blanchet B., De Almeida M.L. et al. (2021b) Preprint bioRxiv. https://doi.org/10.1101/2021.11.15.468583

Déru V., Tiezzi F., Carillier-Jacquin C., Blanchet B., Cauquil L. et al. (2021c) Proc. of the 72nd EAAP, Davos, Switzerland.

Harris A.J., Patience J.F., Lonergan S.M., Dekkers J.M.C., Gabler N.K. (2012) J. Anim. Sci. 90(4):164-166. https://doi.org/10.2527/jas.53855

Mauch E.D., Young J.M., Serão N.V.L., Hsu W.L., Patience J.F. et al. (2018) J. Anim. Sci. 96:1221-1236. https://doi.org/10.1093/jas/sky065

Patience J.F., Rossoni-Serão M.C. and Gutiérrez N.A. (2015) J Anim Sci Biotechnol 6(1):33. https://doi.org/10.1186/s40104-015-0031-2

Pérez P. and De los Campos G. (2014) Genetics 198(2):483-495. https://doi.org/10.1534/genetics.114.164442

Ross E.M., Moate P.J, Marett L., Cocks B.G. and Hayes B.J. (2013) J Dairy Sci 96(9):6030-6046. https://doi.org/10.3168/jds.2013-6766

VanRaden P.M. (2008) J Dairy Sci 91(11):4414-23. https://doi.org/10.3168/jds.2007-0980

Verschuren L.M.G., Schokker D., Bergsma R., Jansman A.J.M., Molist F. et al. (2020) J Anim Breed Genet 137(1):23-35. https://doi.org/10.1111/jbg.12433