

# 801. Genomic prediction of ham weight loss during dry-curing in heavy pigs

V. Bonfatti<sup>1</sup>, S. Faggion and P. Carnier

*Department of Comparative Biomedicine and Food Science, University of Padova, viale dell'Università 16, 35020 Legnaro, Italy; valentina.bonfatti@unipd.it*

## Abstract

Excessive ham weight losses during dry-curing (WL) represent a loss of marketable product, hindering the quality of dry-cured hams. Bayesian regression models were developed for observed (OB) and infrared-predicted WL (IR), from the EBV of 640 and 956 crossbred pigs, respectively. The accuracies of direct genomic values of the purebred sires (DGV-PB) of crossbred pigs were estimated in a leave-one-family-out validation scheme. Accuracies were satisfactory and very similar for both traits. The correlation between the DGV-PB for IR and sire EBV for OB was slightly lower (0.318) than the correlation between the DGV-PB for OB and the sire EBV for OB (0.383). While genomic prediction of OB and IR can be equally suggested to be incorporated in future selection programs aiming at reducing WL, the use of IR allows a reduction in phenotyping costs, favouring the establishment of larger reference populations, with accuracies comparable to those achievable using OB phenotype.

## Introduction

Genetic selection of pigs aiming at reducing excessive ham weight losses (WL) during dry-curing is challenging because measures of WL: (1) require the establishment of ham individual traceability systems throughout the dry-curing process, largely limiting the availability of phenotypes; and (2) can be obtained only after the completion of the entire dry-curing process, resulting in long generation intervals, with consequent penalizing effects on the response to selection. Recently, on-site infrared spectroscopy has demonstrated its effectiveness as a large-scale phenotyping tool for WL, by enabling cost-effective, high-throughput, early predictions of WL, using spectral data acquired at the slaughterhouse on green hams, combined with carcass and green ham quality traits (Bonfatti and Carnier, 2020). In addition, it provides infrared predictions of WL (IR) that are highly genetically correlated with the measured trait (OB) and characterized by a higher heritability than the corresponding measured trait (Bonfatti and Carnier, 2020). The incorporation of IR into pig genomic selection programs aiming at reducing WL might be a promising alternative to the genomic prediction of OB, as it would allow periodical phenotype collection at virtually no additional costs and the creation of larger reference populations. The objective of this study was to compare the accuracy of genomic prediction models for OB and IR and evaluate the ability of genomic predictions for IR to predict the genetic merit for OB.

## Materials & methods

**Phenotypes.** A measure of WL was obtained on the left ham at the end of dry-curing (12 mo) from 1,888 crossbred (CB) pigs produced in the sib-testing program of the C21 Goland sire line (Gorzagri, Fonzaso, Italy). Age at slaughter and body weight were those typical of heavy pigs (270 days, and 167 kg, respectively). Of these pigs, 1,624 had also spectra records and were used to develop infrared prediction models for WL, as detailed by Bonfatti and Carnier (2020). Briefly, the prediction model included the following explanatory variables: (1) visible-infrared spectra, collected at the slaughterhouse on the transversal section of the subcutaneous fat of the raw ham using a LabSpec<sup>®</sup>5000 (ASD Inc., Boulder, Colorado, USA) equipped with a fibre optic contact probe; (2) sex; (3) carcass traits (carcass weight, backfat depth, lean meat content, weight of the hams); (4) green ham quality traits (ham subcutaneous fat depth, and linear scores for ham round shape, subcutaneous fat thickness, and ham marbling).

To obtain the IR values for the 1,624 pigs with available records of OB, a set of prediction models was developed excluding, at each round, all the samples from a given slaughter batch from the training set and applying the obtained model to the samples of the batch that was left out. The process was repeated until all the slaughter batches (n=25) were used in prediction. Finally, a model trained on the entire dataset (n=1,624) was applied to obtain records of IR in a larger dataset, extracted from the historical database of the genetic evaluation program of the Goland C21 sire line. The larger dataset included records of carcass traits, green ham quality traits and spectral variables from 8,048 CB pigs. Hence, a total of 9,672 records of IR were available for this study.

**Pseudo-phenotypes.** Genomic prediction models were developed following a multiple-step genomic evaluation approach, where the estimated breeding values (EBV) of the CB pigs were used as pseudo-phenotypes. The EBV for OB and IR were obtained using BLUPF90 (Misztal *et al.*, 2002). The complete pedigree information was available for all CB pigs with phenotypic records and all their purebred (PB) C21 Goland sires. The model included the fixed effect of sex (female and castrated male) and slaughter batch and the random additive genetic effect of the animal.

**Genotypes.** Genotypes were available for 1,029 CB pigs, spanning four generations. Of these, 640 had a record of OB and were offspring of 57 sires, and 956 had a record of IR and were offspring of 104 sires. Sire families with records of both OB and IR were 53. Genotypes of the sires were also available. Genotypes at 8,826 SNP were obtained using the GGP Porcine LD chip (Genomic Profiler for Porcine LD, GeneSeek Inc., a Neogen Co., Lincoln, NE) according to the manufacturer's protocol, and subsequently imputed to the GeneSeek® Genomic Profiler™ Porcine 50K (Neogen, Lansing, MI, USA) using Fimpute v. 2.2 (Sargolzaei *et al.*, 2014). All genotypes were then subjected to quality control using PLINK v. 1.9 software (<http://www.cog-genomics.org/plink2>). After editing, a total of 29,559 SNPs were retained for the genomic prediction models.

**Genomic prediction model.** Bayesian Ridge Regression was fitted to the CB data using the BGLR package (De los Campos and Rodriguez, 2021) in the R software (R Core Team, 2021). A total of 500,000 iterations were generated to estimate the parameters of the models, with a burn-in period of 100,000 iterations, and a thinning interval of 100 samples. To simulate a scenario under which a reference population including only CB pigs with phenotypic information is used to provide genomic predictions of the genetic merit of PB candidates, the estimated SNP substitution effects were used to calculate the direct genomic values of the PB sires (DGV-PB) of the CB pigs.

In particular, models were trained on the EBV of CB pigs and tested on the EBV of PB sires (EBV-PB) according to the following scenarios:

1. The model was trained on the EBV of CB pigs for OB and tested on the EBV-PB for OB;
2. The model was trained on the EBV of CB pigs for IR and tested on the EBV-PB for IR;
3. The model was trained on the EBV of CB pigs for IR and tested on the EBV-PB for OB.

For each scenario, a leave-one-family-out training-validation scheme was performed: models were trained excluding, at each round, the members of a sire family, a Bayesian Ridge Regression model was fitted to the training set, and the estimated SNP substitution effects obtained at each round were used to calculate the DGV-PB of the sire that was excluded from training. The correlation between EBV-PB and DGV-PB was used as a measure of model accuracy. The bias was calculated as the regression coefficient of the EBV-PB on DGV-PB.

## Results

On average, OB was  $27.8 \pm 2.4\%$ , whereas IR averaged  $26.5\%$ , and exhibited a lower variability ( $SD=1.8\%$ ) compared to the corresponding measured trait. The accuracies of the genomic predictions obtained for the investigated traits are summarized in Table 1. Values of accuracies were very similar for OB and IR (0.383 and 0.396, in scenarios 1 and 2, respectively), whereas a lower accuracy (0.318) was obtained in scenario 3. All the correlation coefficients were significantly different from 0. Regression coefficients were highly variable across cross-validation folds and not significantly different from 1.

## Discussion

The leave-one-family-out training-validation scheme allowed to mimic as far as possible the circumstances under which models are going to be applied in practice, where selection is based on the DGV estimated for PB breeding candidates and estimates are based on a reference population of CB finishing pigs belonging to families related to, but different from, those of the PB candidates.

Scenario 1 had the aim to evaluate the accuracy attainable by models exploiting OB phenotyping, whereas scenario 2 allowed testing the ability of models trained on IR pseudo-phenotype to predict the EBV-PB for the same trait. Results obtained in the two scenarios were comparable, probably due to the larger training set and higher heritability of IR compared to OB (0.39 vs 0.31, data not reported in tables).

The correlation between DGV-PB and EBV-PB in scenario 3 corresponds to the accuracy of models in predicting the genetic merit of PB sire from a reference population based on CB pigs with no close relationships with the sires, thus representing the most realistic scenario. The accuracy under this scenario was lower than the accuracy achievable if the model was trained on OB (scenario 1), but comparable to that value for practical applications. In pigs the accuracy of the DGV tends to decline rapidly over generations, and periodical phenotype collections, especially in animals related to selection candidates, are likely to be needed to update genomic prediction models (van Grevenhof and van der Werf, 2015). Routine phenotyping for OB is currently unfeasible, and the periodical phenotype collection is expensive. In addition, when the target is the prediction of PB genetic merit for CB animals, an increased number of training animals with phenotypes is needed. Under these circumstances, the use of IR in place of OB can allow a reduction of the phenotyping costs, favouring the construction of larger reference populations, a greater flexibility in the collection of phenotypes, with only a minimal impact on model accuracy compared to the use of OB phenotypes.

Considering the limited phenotyping costs for IR and the relatively short generation length achievable using IR phenotypes, the genetic gain attainable by a genomic selection program based on IR phenotyping

**Table 1.** Accuracy of direct genomic values of sires (DGV-PB) for observed ham weight loss (OB) or its infrared prediction (IR) in predicting the EBV of sires (EBV-PB) for OB or IR.<sup>1</sup>

Scenario	DGV-PB	EBV-PB	n	Accuracy		Bias	
				r	P-value	B	P-value <sup>2</sup>
1	OB	OB	57	0.383±0.125	<0.01	1.190±0.387	0.626
2	IR	IR	104	0.396±0.091	<0.001	0.838±0.192	0.400
3	IR	OB	104	0.318±0.094	<0.001	0.739±0.218	0.234

<sup>1</sup> n: number of animals used to calculate the accuracy; r: correlation coefficient between the DGV-PB and the EBV-PB; β: regression coefficients of EBV-PB on DGV-PB.

<sup>2</sup> H<sub>0</sub>: β = 1.0

of the reference population should be compared to that achievable by a traditional selective program based on routine IR phenotyping.

## References

- Bonfatti V., and Carnier P. (2020) Prediction of dry-cured ham weight loss and prospects of use in a pig breeding program. *Animal* 14, 1128-1138, <https://doi.org/10.1017/s1751731120000026>.
- De los Campos G., and Rodriguez P.P. (2021) BGLR: Bayesian generalized linear regression. Available at: <https://cran.r-project.org/web/packages/BGLR/index.html>.
- Misztal I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., *et al.* BLUPF90 and related programs (BGF90) (2002) Proc. of the 7<sup>th</sup> World Congress on Genetics Applied to Livestock Production, Montpellier, France.
- R Core Team 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>.
- Sargolzaei M., Chesnais J.P., Schenkel F.S. (2014) A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15, 478, <https://doi.org/10.1186/1471-2164-15-478>.
- Van Grevenhof I., Van der Werf J. (2015) Design of reference populations for genomic selection in crossbreeding programs. *Gen. Sel. Evol.* 47, 14, <https://doi.org/10.1186/s12711-015-0104-x>.