Received: 22 December 2023 | Revised: 21 June 2024 | Accepted: 24 June 2024

DOI: 10.1111/jbg.12887

Journal of Animal Breeding and Genetics WILEY

ORIGINAL ARTICLE

# Multiple-trait genomic prediction for swine meat quality traits using gut microbiome features as a correlated trait

**Francesco Tiezzi**[1] | **Clint Schwab**[2] | **Caleb Shull**[3] | **Christian Maltecca**[1,4]

[1]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Florence, Italy

[2]AcuFast LLC, Navasota, Texas, USA

[3]The Maschhoffs LLC, Carlyle, Illinois, USA

[4]Department of Animal Science, North Carolina State University, Raleigh, North Carolina, USA

**Correspondence**
Francesco Tiezzi, Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Firenze 50144, Italy.
Email: francesco.tiezzi2@unifi.it

**Funding information**
National Pork Board Association 'The pork checkoff'; The Maschhoffs LLC; North Carolina State University

## Abstract

Traits such as meat quality and composition are becoming valuable in modern pork production; however, they are difficult to include in genetic evaluations because of the high phenotyping costs. Combining genomic information with multiple-trait indirect selection with cheaper indicator traits is an alternative for continued cost-effective genetic improvement. Additionally, gut microbiome information is becoming more affordable to measure using targeted rRNA sequencing, and its applications in animal breeding are becoming relevant. In this paper, we investigated the usefulness of microbial information as a correlated trait in selecting meat quality in swine. This study incorporated phenotypic data encompassing marbling, colour, tenderness, loin muscle and backfat depth, along with the characterization of gut (rectal) microbiota through 16S rRNA sequencing at three distinct time points of the animal's growth curve. Genetic progress estimation and cross-validation were employed to evaluate the utility of utilizing host genomic and gut microbiota information for selecting expensive-to-record traits in crossbred individuals. Initial steps involved variance components estimation using multiple-trait models on a training dataset, where the top 25 associated operational taxonomic units (OTU) for each meat quality trait and time point were included. The second step compared the predictive ability of multiple-trait models incorporating different numbers of OTU with single-trait models in a validation set. Results demonstrated the advantage of including genomic information for some traits, while in some instances, gut microbial information proved advantageous, namely, for marbling and pH. The study suggests further investigation into the shared genetic architecture between microbial features and traits, considering microbial data's compositional and high-dimensional nature. This research proposes a straightforward method to enhance swine breeding programs for improving costly-to-record traits like meat quality by incorporating gut microbiome information.

**KEYWORDS**

carcass composition, correlated response to selection, genomic selection, gut microbiome, meat quality

## 1 | INTRODUCTION

Traits valuable in modern pork production describe the growth and muscle deposition pattern within the first few months of an animal's life. These traits relate to whole-body growth, muscle and fat deposition, saleable meat production and feed efficiency. In addition, traits that describe the composition of the carcass (e.g. higher vs. lower value cuts) and quality of the meat (e.g. tenderness, marbling) are gaining importance. However, these traits are difficult to include in genetic evaluations because of the high costs of phenotyping. Including genomic information in the prediction models could help identify the genetic material that better meets the breeding needs, and genomic selection has been proven to dramatically increase the efficiency of selection. However, the advantages of genomic selection cannot eliminate the need to collect correlated indicator traits (CIT) in constructing selection indexes, particularly when the aim is to achieve genetic progress in a breeding goal trait (BGT) that is not regularly measured. Both BGT and CIT, which can be assessed in a significantly larger population segment, can be integrated into a multiple-trait (genomic) best linear unbiased prediction (BLUP) model. Consequently, the prediction for the selection of candidates relies on information from CIT and previously collected data for the BGT (Calus & Veerkamp, 2011). Numerous instances demonstrate the utilization of high-throughput, cost-effective traits that are easier to measure on selection candidates (or the general validation population) across diverse livestock species (Bonfatti et al., 2017; Cecchinato et al., 2020; Godinho et al., 2018; Putz et al., 2015; See & Knauer, 2019; van den Berg et al., 2020).

The CIT have historically been considered traits that can be routinely collected with little economic and logistical effort. Recently, measures of gut microbiome composition have become cheaper to collect using targeted rRNA sequencing. Consequently, the applications of this technology in animal breeding are becoming more appealing. The gut microbiome composition is easily assessed using a rectal swab and has been proven to be associated with most of the swine BGT such as growth (Bergamaschi, Tiezzi, et al., 2020; Ramayo-Caldas et al., 2016), feed efficiency (Camarinha-Silva et al., 2017; Quan et al., 2018) and meat quality (Bergamaschi, Tiezzi, et al., 2020; Fang et al., 2017; Khanal et al., 2019).

Simultaneously, there is substantial evidence supporting the partial influence of the host genotype on the gut microbiota, as indicated by moderate heritability estimates for gut microbiota composition (Bergamaschi, Maltecca, et al., 2020; Crespo-Piazuelo et al., 2019; Martinez Boggio et al., 2022; Ramayo-Caldas et al., 2020). Additionally, the phenotypic associations observed between BGT and ruminal or gut microbiome composition extend to the genetic level. Genetic correlations between BGT and ruminal microbiome composition have been estimated in cattle (Martínez-Álvaro et al., 2022; Roehe et al., 2016; Saborío-Montero et al., 2020, 2021), as well as with gut microbiome composition in swine (Aliakbari et al., 2021). The association could be due to shared genetic architecture but also recursive effects at the phenotypic level (Saborío-Montero et al., 2021; Tiezzi et al., 2021; Valente et al., 2013). In the case, the former occurs and considering the cost-effectiveness of gut microbiome sampling and sequencing, this compositional trait could function as a CIT in a selection index. Evaluating the gut microbiota composition, or its diversity measure, on selection candidates could offer timely and valuable information in the evaluation models. This new data source could be integrated with (host) genomic information in the genetic evaluation process, allowing for comparing the relevance between gut microbiome CIT and host genotypes. The optimal allocation of resources, between host genotyping and gut microbiome sequencing, would depend on the specific trait, its genetic architecture, and the shared genetic architecture with the microbiome composition. Once the relevance of the microbiome information is assessed, a potential implementation could entail the rectal sampling of the gut content at different stages of life of the animal, a rapid sequencing procedure and definition of features that describe the microbiome composition and the inclusion of the most relevant features in a multiple-trait model. In the case of a progeny or sib test, the selection candidates (or their offspring, or siblings) would not have records for the BGT available and would be substituted in the index by microbiota composition information. This would imply collecting rectal swabs from the testing candidates or their siblings at one or more timepoints, having them processed rapidly for both the wet-lab and dry-lab parts of the process, and their resulting information incorporated into multiple-trait models. Given the burden at the operational level, a value proposal is needed. Questions about the number of samples and time of sampling would have to be addressed properly. A graphical representation of this potential implementation in the genetic evaluation model is reported in Figure 1. In the absence of CIT (left side panel), the information about the BGT flows from the families in the reference population (training set) to the candidates (validation set) through the (genomic) relationship matrix. In the presence of one or several CIT (right side panel), the information can still follow that same path but will also flow through the other traits, such as the gut microbiome composition. Since the individuals in the validation set are phenotyped for the microbiome composition, the model can inform the predictions of the breeding goal for the validation set using the microbiome information weighted by the variance–covariance structure among the traits.

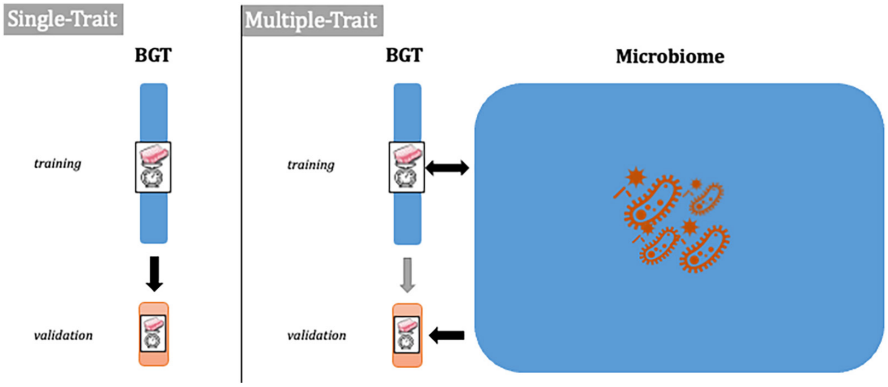


**FIGURE 1** Schematic representation of a breeding value prediction model. The left-side panel pictures a single-trait model, where the information on the breeding goal trait (BGT) flows from the training to the validation set. The right-side panel shows a multiple-trait model where the microbial information enters as correlated indicator trait. Here, breeding goal trait is recorded on the training set while the microbial correlated trait(s) are recorded on both the training and validation sets. In this model, the information can flow directly from the training to the validation set through the pedigree-based or genomic-based relationship matrices, but also indirectly through the correlated traits.

Given the association of microbiota composition at the phenotypic and genetic value with performance traits, incorporating gut microbiota composition into breeding models could potentially offer significant advantages in selection programs. Its use in a multiple-trait model would be similar in principle to other methods already proposed. For example, Weishaar et al. (2020) proposed a method to extract the genetic component of the microbial composition and its use in trait prediction. While different from the modelling standpoint, this method is similar at the conceptual level, since the genetic contribution to the microbial composition is the baseline. In both cases, cross-validation appears to be the most effective way to test the inclusion of the gut microbial information as CIT.

The main objective of this study was to assess the relevance of gut microbial information for the selection of crossbred pigs for meat quality and carcass composition traits. This was achieved by (1) estimating the expected genetic progress (using a deterministic approach) obtainable by selecting for the most relevant microbial features as opposed to direct selection and (2) assessing the accuracy of multiple-trait (genomic or pedigree) predictions where microbial features are included as correlated traits as opposed to single-trait models.

## 2 | MATERIALS AND METHODS

### 2.1 | Animals and data

Phenotypic records presented in this study came from a commercial farm operated by The Maschhoffs LLC (now Acuity Ag Solutions, Carlyle, IL, USA). All methods and procedures were in accordance with the Animal Care and Use policies of North Carolina State University and the National Pork Board. The experimental protocol for faecal sample collection received approval number 15027 from Institutional Animal Care and Use Committee. All pigs were harvested in commercial facilities under the supervision of USDA Food Safety and Inspection Service.

### 2.2 | Data structure

Animals used in this study included commercial crossbred individuals belonging to 28 paternal half-sib families raised at The Maschhoffs LLC (now part of AcuFast LLC, Navasota, TX, USA). Crossbred pigs were from a three-way cross involving purebred Duroc boars and crossbred 'Yorkshire × Landrace' or 'Landrace × Yorkshire' sows. The trial included 6680 individuals evenly distributed over the paternal half-sib families. Matings that produced the animals involved in this study were carried out from November to December 2014 and data collection was from March to November 2015.

The pigs were born in a commercial sow facility; their weight was taken within 24 h from birth, then they were weaned at 18.64 ± 1.09 days old and were moved to a nursery-finishing facility. Pigs were kept in 334 single-sire single-sex pens with 20 pigs per pen. The test period began the day that pigs were moved to the nursery-finishing facility. All pigs were fed a standard pelleted feed based on sex and live weight during the nursery, growth and finishing period. Diet details and nutritional values are provided in supporting information in Khanal et al. (2019). The pigs received a standard vaccination and medication routine (supporting information

in Khanal et al., 2019). End of test was reached when the average weight of pigs in each pen reached 138 kg, pigs' age was 196.4 ± 7.80 days. At the end of the trial, three to five pigs from each pen (single-sire, single-sex group) were selected, as detailed by Wilson et al. (2016). The selected pigs per pen represented an average pig for body weight, along with pigs approximately 1 and 2 SD above and below the pen average. Out of the selected individuals, some showed poor quality in the microbial information, so only 1180 individuals showed complete records for genomic information, faecal microbiome and phenotypic measures.

## 2.3 | Genomic information

Genomic information was obtained using the Porcine SNP60 v2 BeadChip (Illumina, Inc.). Standard quality control procedures were performed on the genomic data, removing SNPs with call rate lower than 0.90 and/or minor allele frequency lower than 0.05. A total of 42,529 SNPs remained after this procedure.

## 2.4 | Phenotypic information on the breeding goals

Meat quality, carcass composition and growth traits were also recorded on the selected pigs and will be used as breeding goal traits (BGT) in this study. Meat quality traits included intramuscular fat (IMF), Warner–Bratzler slice shear force (SSF), Minolta L* (M.L), Minolta a* (M.a), Minolta b* (M.b), pH (PH), subjective colour (SCOL), subjective firmness (SFIR) and subjective marbling (SMAR). Carcass composition traits included loin weight (LOI), belly weight (BEL), ham weight (HAM), loin depth (LD), backfat depth (BF) and carcass average daily gain (CDG). All phenotypic data collection details were described by Wilson et al. (2016) and Khanal et al. (2019). Briefly, both BF and LD were measured after slaughter using a Fat-O-Meater probe (SFK Technology A/S, Herlev, Denmark) at approximately the location of the 10th rib. The CDG was calculated by dividing the difference between the hot carcass weight and birth weight by the pig's age at slaughter. The measurement of PH, M.L (measuring luminosity), M.a (measuring redness) and M.b (measuring yellowness) was done on the ventral side of the longissimus dorsi muscle. LOI, BEL and HAM were measured after carcass dissection. Subjective measures of pork quality, including SCOL (5 categories), SFIRM (1–5 scale), and SMAR (6 classes), together with SSF and IMF, were determined on a loin muscle sample.

## 2.5 | Microbiome information

Faecal samples for 16S rRNA sequencing were collected at three time points, as described in Khanal et al. (2021). While rectal swabs were collected from all the ~6000 pigs, sequencing was performed only on 1252 individuals chosen to be representative of the within-pen variation (see section on data structure). Sample collection was conducted at three stages: weaning (**S1**, average 18.64 ± 1.09 days), 15 weeks post-weaning or on-test (**S2**, average 118.2 ± 1.18 days) and off-test (**S3**, average 186 ± 13.69 days). After quality control, there were 1205, 1295 and 1273 samples at S1, S2 and S3 respectively. Distribution of samples across families, time points and sex are provided in Khanal et al. (2021).

DNA extraction, purification, Illumina library preparation and sequencing were done as described by Lu et al. (2018). Briefly, total DNA (gDNA) was extracted from each rectal swab by mechanical disruption in phenol/chloroform. Bead-beating was performed on the Mini-BeadBeater-96 (MBB-96; BioSpec, OK, USA) for 4 min at room temperature; samples were centrifuged at 3220 × *g*. The DNA was then purified using a QIAquick 96 PCR purification kit (Qiagen, MD, USA). Minor modifications were performed in the purification process as per the manufacturer's instruction. The modification included the addition of sodium acetate (3 M, pH 5.5) to Buffer PM to a final concentration of 185 mM, combination of crude DNA with 4 volumes of Buffer PM, and elution of DNA in 100 μL Buffer EB. All sequencing was performed at DNA Sequencing Innovation Laboratory at the Center of Genome Sciences and Systems Biology at Washington University in St. Louis. Phased, bi-directional amplification of the v4 region (515–806) of the 16S rRNA gene was employed to generate indexed libraries for Illumina sequencing as described in Faith et al. (2013). Sequencing was performed on an Illumina MiSeq instrument (Illumina, Inc. San Diego, USA), generating 250 bp paired-end reads.

Sequencing of 16S rRNA gene and quality control of data were done as described by Lu et al. (2018). Briefly, the pairs of 16S rRNA gene sequences obtained from Illumina sequencing were combined into a single sequence using FLASH v1.2.11 Magoc and Salzberg (2011). The sequences with a mean quality score below Q35 were filtered out using PRINSEQ v0.20.4 (Schmieder & Edwards, 2011). Then, the forward-oriented sequences were matched with primer sequences and trimmed off. Mismatch was allowed up to 1 base pair. Sequences were subsequently demultiplexed using QIIME v1.9 (Caporaso et al., 2010). QIIME was used to cluster the nucleotide sequences into operational taxonomic units (OTU), as explained by Maltecca et al. (2021). A

modified version of GreenGenes (Ley et al., 2006; Schloss & Handelsman, 2006) was used as reference database. Ninety per cent of the input sequences were matched to the reference database. The remaining 10% that did not match the reference database were then clustered de novo with UCLUST (Schloss & Handelsman, 2006) to generate new reference OTU. Then, the 90% of reads that were matched with the reference database were again assigned to the new reference OTU that were derived from the de novo cluster. The OTU were removed if showing a total observation count of less than 1200 counts or appearing in less than 0.05% of the samples. The resulting 3001 OTU table was rarefied to 10,000 counts per sample, obtaining relative abundance measures. Furthermore, The OTU were considered as microbial features (MF) if showing a zero count in less than 80% of the samples. After editing and merging with the rest of the data, a total of 605, 1382 and 1452 MF from each timepoint (S1, S2 and S3) were available on 1180 individuals for this study. On these, zero-value imputation was performed using the function *cmultrepl* from the R package "Zcompositions" (Palarea-Albaladejo & Martín-Fernández, 2015), and centre-log (CLR) transformation was performed using the function *clr* from the R package "compositions" (van den Boogaart & Tolosana-Delgado, 2008), both implemented in the R software (R Core Team, 2020). For further details on the microbial data, see Maltecca et al. (2021).

## 2.6 | Statistical model employed for the selection of relevant microbial features

The first statistical analysis step involved selecting the relevant MF that could be used as a correlated trait for the multiple-trait genomic prediction of BGT. The whole dataset was split for discovery and validation. The 28 sires were ordered for their date of birth, the progeny of the 21 sires born earlier were allocated to the training set (TRN, $N = 979$), while the offspring of the seven sires born later were allocated to the validation set (VAL, $N = 273$).

The selection of the relevant MF was carried out on the TRN dataset. The set of MF was specific for each BGT and each time point of microbial sample collection. First, variance components and breeding values for the 15 BGT and the 3439 MF were obtained using single-trait models implemented on each of the traits individually. The model was defined as follows:

$$y = \mathbf{X}\beta + \mathbf{Z_p}\mathrm{p} + \mathbf{Z_a}\mathrm{a} + e \quad (1)$$

where $y$ is a vector of phenotypes for the investigated trait; $\beta$ is a vector of solutions for fixed effects, including dam line (two levels) and gender by contemporary group (12 levels); p is the vector of solutions for the random effect of the pen, with $\mathrm{p} \sim N(0, \mathbf{I}\sigma^2_{\mathrm{p}})$, where $\sigma^2_{\mathrm{p}}$ is the estimated pen variance; a is the vector of solutions for the random additive genetic effect of the animal, with $\mathrm{a} \sim N(0, \mathbf{G}\sigma^2_{\mathrm{a}})$, where $\mathbf{G}$ is the SNP-derived genomic relationship matrix built using VanRaden's method 1 (VanRaden, 2008) and $\sigma^2_{\mathrm{a}}$ is the estimated additive genetic variance; $\mathbf{X}$ is the incidence matrix of fixed effects, $\mathbf{Z_p}$, and $\mathbf{Z_a}$ are the corresponding incidence matrices for the random effects and $\mathbf{e}$ is the vector of random residuals, with $\mathrm{e} \sim N(0, \mathbf{I}\sigma^2_{\mathrm{e}})$, where $\sigma^2_{\mathrm{e}}$ is the residual variance.

This first analysis was implemented using the software AIREMLf90 (Misztal et al., 2002), setting the first 50 iterations as expectation–maximization REML (EM-REML) and the maximum number of iterations to 1000. The model produced estimates of variance components and genomic breeding values (GEBV). The EBV were used in the following linear model to select the most relevant MF:

$$R_{p|m} = \frac{\sigma_{mp}}{\sigma^2_m} \quad (2)$$

where $R_{p|m}$ is the estimated response to selection in the BGT $p$ when selection is performed on the MF $m$, $\sigma_{mp}$ is the covariance between vector EBV for the BGT of interest $p$ and the vector of EBV for the MF $m$, $\sigma^2_m$ is the variance of the vector of EBV for the MF $m$. The response to selection, defined as the ratio between the genetic covariance and the genetic variance, is taken from Schneeberger et al. (1992).

All MF were tested against all BGT using the Formula (2). For each BGT, the MF from each timepoint were ranked based on the generated response, and the 25 that generated the largest values were selected. In total, 75 MF were selected for each BGT.

## 2.7 | Multiple-trait variance component estimation

The subsequent step involved implementing 26-trait models, including the BGT of interest and the 25 MF selected for each time point. Again, this step was carried out on the TRN dataset.

For ease of implementation, the variance component estimation was carried out using Gibbs sampling. The model specifications were as for the single-trait models, with the fixed effects of dam line and gender by contemporary group as well as a co-variance structure of the vector of random effects was defined as appropriate for a multiple-trait model. Here, a full variance–covariance structure was defined for the additive genetic and residual terms, estimating the 26 variances and corresponding

covariances. Therefore, the vector of solutions for the additive genetic effects was defined as:

$$\begin{bmatrix} a_1 \\ \vdots \\ a_{26} \end{bmatrix} \sim N\left(0, \mathbf{G}\begin{bmatrix} \sigma^2_{a.1} & \cdots & \sigma_{a.1-26} \\ \vdots & \ddots & \vdots \\ \sigma_{a.26-1} & \cdots & \sigma^2_{a.26} \end{bmatrix}\right)$$

where $a_x$ is the vector of solutions for the additive genetic effect for trait $x$, with $x$ going from 1 (BGT) to 26, therefore, including all the chosen MF for that stage of sampling, **G** is again the SNP-derived genomic relationship matrix, $\sigma^2_{a.x}$ is the additive genetic variance for trait $x$ and $\sigma_{a.1-26}$ is the additive genetic covariance among the traits (1 and 26, in this case).

For the random effect of pen, all the covariance values were fixed to 0:

$$\begin{bmatrix} p \\ \vdots \\ p_{26} \end{bmatrix} \sim N\left(0, \mathbf{I}\begin{bmatrix} \sigma^2_{p.1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_{p.26} \end{bmatrix}\right)$$

where $p_x$ is the vector of solutions for the pen effect for trait $x$, **I** is again an identity matrix, $\sigma^2_{p.x}$ is the pen variance for trait $x$. This was done to facilitate the convergence of the models since the low number of individuals per pen did not allow for estimating all the covariance parameters. The vectors of residuals in the multiple-trait models were then defined as:

$$\begin{bmatrix} e_1 \\ \vdots \\ e_{26} \end{bmatrix} \sim N\left(0, \mathbf{I}\begin{bmatrix} \sigma^2_{e.1} & \cdots & \sigma_{e.1-26} \\ \vdots & \ddots & \vdots \\ \sigma_{e.26-1} & \cdots & \sigma^2_{e.26} \end{bmatrix}\right)$$

where $e_x$ is the vector of residuals for trait $x$, **I** is an identity matrix, $\sigma^2_{e.x}$ is the residual variance for trait $x$ and $\sigma_{e.1-26}$ is the residual covariance among the traits. This variance component estimation step was also implemented in the GIBBS3F90 program (v. 1.83) from the BLUPF90 family of programs (Misztal et al., 2002). The multiple-trait models were run for 600,000 iterations, with the first 100,000 samples discarded and every 50th sample saved, leaving a total of 10,000 samples for subsequent inference. Convergence of the variance components estimates was assessed via visual inspection of the trace plots and Geweke's test using the R package 'boa' (Plummer et al., 2006). The posterior mean and the 95% empirical confidence intervals were used as estimates and their errors.

## 2.8 | Relative efficiency of selection

After obtaining (co)variance components for the different phenotypic traits and microbial sampling time points, selection response calculation for each BGT $p$ was carried out using $SR_p = \frac{b^T \mathbf{C_{12}}}{\sqrt{b^T \mathbf{C_{11}} b}}$, where b is a vector of index weights obtained as $b = \mathbf{C_{11}^{-1}C_{12}}v$, where $\mathbf{C_{11}}$ is a $m \times m$ genetic variance–covariance matrix of $m$ traits in the selection criterion (Schneeberger et al., 1992); $\mathbf{C_{12}}$ is a $m \times n$ genetic covariance matrix of $m$ traits in the selection criterion and $n$ traits in the breeding objective and v is a vector of relative emphasis for the traits in the breeding objective, which was always set to 100 for the trait studied. The relative efficiency of indirect selection (REIS) versus direct selection for the BGT was obtained using:

$$REIS = \frac{SR_{Scenario}}{SR_{Direct}}$$

where $SR_{Scenario}$ is the indirect selection response for the BGT based on microbial information from a given timepoint (S1, S2, S3) and $SR_{Direct}$ represents the selection response for the BGT given direct selection for the trait of interest.

## 2.9 | Genomic prediction of crossbred traits

Genomic predictions were validated on the VAL dataset to compare the different microbiome sampling scenarios to predict unobserved phenotypes of interest. The cross-validation aimed at comparing the predictive ability of a single-trait model (ST), where predictions were based only on the measures of the breeding goal on the TRN set individuals, with a multiple-trait model (MT) where the microbiome composition was assessed on both the TRN and VAL sets.

Therefore, for the ST scenario, the breeding values were estimated using the TRN dataset, and prediction accuracy was calculated as the Pearson correlation between the estimated breeding values and the (masked) phenotypes adjusted for the fixed effects (see model 1, estimates obtained with the full dataset). Then, different sources and number of microbial features for the MT scenario were included. The top 1, 2, 5, 10 and all 25 features were subsequently included in five independent runs of the MT model. The procedure was repeated for the microbial features coming from the S1, S2 and S3 timepoints. The microbial information was included for all individuals in TRN and VAL sets.

In addition, the relevance of host genomic information in producing predictions was assessed. Under the same cross-validation scheme, the genomic relationship matrix **G** was replaced with a pedigree-derived relationship matrix **A**, built on a pedigree traced back nine generations.

Overall, the cross-validation was repeated for each BGT for the ST model as well as the three MT models.

Since one measure of predictive ability was obtained for each run, bootstrapping was employed to obtain confidence intervals. The correlation between the estimated breeding values and the (masked) phenotypes was repeated 1000 times, each time resampling from the validation set with replacement, in order to have bootstrapped sets with the same number of records but different representations of each record. Empirical 95% confidence intervals were then calculated and reported in the figures.

## 3 | RESULTS

The descriptive statistics for the BGT over the TRN and VAL datasets are reported in Table 1. For all traits considered, there was no significant difference between the two sets.

The most relevant microbial features (as identified by the operational taxonomic units) for each BGT considered are reported in Tables 2–4 depending on the microbiome sampling stage (S1, S2 or S3 respectively). The genetic correlation estimates are presented in Table S1, together with their significance ('0' value included within the 95% empirical confidence intervals). The total number of genera identified for the 15 BGT was 33, 21, and 17 for each of the three time points. At timepoint S1, the MF belonged primarily to the family *Ruminococcaceae* (72 features), *Lachnospiraceae* (46 features), followed by the genera *Prevotellaceae* (37 features), *Lactobacillaceae* (12 features), *Peptostreptococcaceae* (10 features) and *Porphyromonadaceae* (10 features). Other genera showed a number of MF associated with the BGT, such as *Staphylococcaceae* (9), *Streptococcaceae* (7), and *Fusobacteriaceae* (*8*). Many MF (85) were unassigned at the family level.

At timepoint S2, again, the families *Lachnospiraceae* (46 features) and *Ruminococcaceae* (72 features), followed by *Prevotellaceae* (34 features), and then *Lactobacillaceae* (11 features), *Porphyromonadaceae* (10 features) and *Spirochaetaceae* (9 features). Again, 121 MF were unassigned at the family level. At timepoint S3, the families *Lachnospiraceae* and *Ruminococcaceae* showed the largest number of MF associated (82 and 80 respectively), followed by the genera *Prevotellaceae* (13 features), *Peptostreptococcaceae* (9 features), *Bacteroidaceae* (9 features) and *Clostridiaceae* (8 features). At the same time, 152 MF were unassigned at the family level.

### 3.1 | Variance components and relative efficiency of selection

Figure 2 shows the efficiency of microbiome-enabled selection estimates relative to direct selection (also in Table S2). In this plot, a REIS of the value of 1 indicates that the selection performed on the MF is expected to yield a gain equivalent to the direct selection on the BGT itself (without phenotyping of the individuals in the validation set). A REIS below the value of 1 indicates that microbiome as a correlated trait does not provide relevant

**TABLE 1** Descriptive statistics for the training (TRN) and validation (VAL) sets.

| | | Training | | | Validation | | |
|---|---|---|---|---|---|---|---|
| **Trait** | **Acronym** | ***N*** | **Mean** | **SD** | ***N*** | **Mean** | **SD** |
| Intra-Muscular fat | IMF | 960 | 2.7 | 0.94 | 264 | 2.7 | 0.88 |
| Belly weight | BEL | 976 | 16.0 | 2.44 | 272 | 15.8 | 2.28 |
| Ham weight | HAM | 979 | 25.2 | 2.36 | 273 | 25.3 | 2.30 |
| Loin weight | LOI | 979 | 20.1 | 1.90 | 273 | 20.0 | 1.81 |
| Minolta a | M.a | 966 | 3.8 | 1.11 | 272 | 3.8 | 1.05 |
| Minolta b | M.b | 966 | −0.2 | 0.88 | 272 | −0.1 | 0.98 |
| Slice shear force | SSF | 960 | 158.7 | 36.53 | 264 | 161.4 | 36.24 |
| Subjective colour | SCOL | 962 | 2.7 | 0.47 | 272 | 2.7 | 0.47 |
| Subjective firmness | SFIR | 962 | 3.1 | 0.99 | 272 | 2.9 | 0.98 |
| Subjective marbling | SMAR | 962 | 3.1 | 0.83 | 272 | 3.0 | 0.83 |
| Minolta L | M.L | 966 | 45.3 | 3.14 | 272 | 45.4 | 2.99 |
| pH | PH | 911 | 5.6 | 0.19 | 257 | 5.6 | 0.17 |
| Carcass daily gain | CDG | 970 | 521.5 | 54.31 | 272 | 522.3 | 50.62 |
| Fat depth | BF | 972 | 22.6 | 5.13 | 271 | 21.6 | 4.46 |
| Loin depth | LD | 972 | 68.8 | 7.43 | 271 | 67.3 | 7.45 |

**TABLE 2** Number of operational taxonomic units, as organized by family, used for each trait.

| Family | BEL | BF | CDG | HAM | IMF | LD | LOI | M.L | M.a | M.b | SCOL | SFIR | SMAR | SSF | pH | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Acidaminococcaceae | | | | | | | | | | | | | 1 | | | 1 |
| Aerococcaceae | | | | 1 | | | | | 1 | | | | | | | 2 |
| Bacteroidaceae | | | | | 1 | | | 1 | 1 | 1 | 1 | | | 1 | | 6 |
| Bifidobacteriaceae | | 1 | | | | | | | | | | | | | | 1 |
| Campylobacteraceae | 1 | 1 | | 2 | | | | | 1 | | | 1 | | | | 6 |
| Clostridiaceae | | | 2 | 1 | | 1 | 1 | | | | | | | | | 5 |
| Clostridiales Incertae Sedis XI | | | | 1 | | | | | | | | 1 | | | | 2 |
| Corynebacteriaceae | | | 1 | | | | | | | | | | | | | 1 |
| Desulfovibrionaceae | | | | | | 1 | | 1 | 1 | | | | | | 1 | 4 |
| Enterococcaceae | 1 | | 1 | | | | | | | | 1 | | | | | 3 |
| Erysipelotrichaceae | | | | | | | | 1 | 1 | 1 | | 1 | 1 | | | 5 |
| Eubacteriaceae | | 1 | | | | | | 1 | | | 1 | | | | 1 | 4 |
| Fusobacteriaceae | | 1 | 1 | 1 | | 1 | | | | 1 | | 1 | | 1 | 1 | 8 |
| Helicobacteraceae | 1 | | 2 | 1 | | | 1 | | | | | | | 1 | | 6 |
| Lachnospiraceae | 2 | 1 | 4 | 3 | 2 | 3 | 4 | 2 | 3 | 1 | 4 | 5 | 4 | 3 | 5 | 46 |
| Lactobacillaceae | 3 | | | | 3 | | 1 | 1 | 1 | | 1 | 2 | | 3 | | 15 |
| Leuconostocaceae | 1 | | | | | | 1 | | | | | | | | | 2 |
| Methanomassiliicoccaceae | | | | | | | 1 | | | | | | 1 | | | 2 |
| Micrococcaceae | | | 2 | 1 | | | | | | | | | | | | 3 |
| Moraxellaceae | | | | | 1 | 1 | | | | 1 | | 1 | | | | 4 |
| Oligosphaeraceae | | | | | 1 | | | | | 1 | | 1 | 1 | 1 | | 5 |
| Pasteurellaceae | 1 | | 1 | | | | | | | | | | | | | 2 |
| Peptococcaceae 1 | | | | 1 | | | 1 | | | | | 1 | | | | 3 |
| Peptostreptococcaceae | | | | 2 | | 1 | 2 | | 2 | | | | 1 | 1 | 1 | 10 |
| Porphyromonadaceae | | 1 | | 2 | 1 | 2 | | 1 | | 1 | 1 | | | | 1 | 10 |
| Prevotellaceae | 3 | 2 | | 1 | 8 | 2 | 3 | 4 | | 3 | 2 | 1 | 4 | 1 | 3 | 37 |
| Ruminococcaceae | 4 | 1 | 4 | 2 | 1 | 5 | 3 | 7 | 4 | 11 | 7 | 7 | 5 | 3 | 8 | 72 |
| Spirochaetaceae | | | | 1 | | | | | | | | | | 2 | | 3 |
| Staphylococcaceae | 1 | | | | | 1 | 1 | 1 | 2 | | 1 | 1 | 1 | | | 9 |
| Streptococcaceae | | | 1 | | 3 | | | | | | 1 | | 1 | 1 | | 7 |
| Sutterellaceae | | | | 1 | | | | | | | | | | | | 1 |
| Synergistaceae | | | | | | | | 1 | 1 | | | | | | | 2 |
| Veillonellaceae | | 1 | | | | | | | | | | | 1 | 1 | | 3 |
| Unassigned | 7 | 15 | 6 | 4 | 4 | 7 | 6 | 4 | 7 | 4 | 5 | 2 | 4 | 6 | 4 | 85 |

*Note*: Results refer to Stage 1 of rectal microbiome sampling.

**TABLE 3** Number of operational taxonomic units, as organized by family, used for each trait.

| | BEL | BF | CDG | HAM | IMF | LD | LOI | M.L | M.a | M.b | SCOL | SFIR | SMAR | SSF | pH | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Acidaminococcaceae | | | | | | | 1 | | | | | | | | | 1 |
| Bacteroidaceae | | | | | | | | | | | | | | | 1 | 1 |
| Clostridiaceae | | | | 1 | | | | | 1 | | | | | | 1 | 3 |
| Clostridiales Incertae Sedis XI | | | | | | | | | | | | | | 1 | | 1 |
| Coriobacteriaceae | | 2 | | | | | | | | | | | 2 | | | 4 |
| Desulfovibrionaceae | | | | | | | | | | | | 1 | | | | 1 |
| Erysipelotrichaceae | | | 1 | 1 | | | | | | | | | | 1 | | 3 |
| Eubacteriaceae | | | | | | 1 | | | | 1 | | | | | | 2 |
| Lachnospiraceae | 8 | 8 | 2 | 9 | 5 | 6 | 5 | 5 | 8 | 11 | 6 | 6 | 7 | 3 | 6 | 95 |
| Lactobacillaceae | | | | | | | | 1 | 6 | | 2 | | | 2 | | 11 |
| Methanobacteriaceae | | | | | | 1 | | 1 | | | | | | | | 2 |
| Peptococcaceae 1 | 1 | | 1 | | 1 | | 1 | | | | 1 | | | | 1 | 6 |
| Porphyromonadaceae | 2 | | 1 | | | 1 | | | | | 1 | 3 | 2 | | | 10 |
| Prevotellaceae | 1 | 4 | 1 | | 5 | 2 | 3 | | 4 | 4 | 1 | 3 | 6 | | | 34 |
| Ruminococcaceae | 6 | 4 | 6 | 5 | 2 | 3 | 5 | 4 | 3 | 3 | 5 | 4 | 2 | 4 | 5 | 61 |
| Spirochaetaceae | 1 | | | | 1 | 1 | | 2 | | 1 | 1 | | 2 | | | 9 |
| Streptococcaceae | | | | 1 | | | | | | | | | | | | 1 |
| Subdivision5_genera_incertae_sedis | | | 1 | | | | | | | | | | | | | 1 |
| Succinivibrionaceae | | 1 | 1 | | | | 1 | 1 | | | | | | | | 4 |
| Sutterellaceae | | | | | | | | | | | | | | | 1 | 1 |
| Veillonellaceae | | | | | | 1 | | | | | | | | 2 | | 3 |
| Unassigned | 6 | 6 | 11 | 8 | 11 | 9 | 9 | 11 | 3 | 5 | 8 | 8 | 4 | 12 | 10 | 121 |

*Note*: Results refer to Stage 1 of rectal microbiome sampling.

**TABLE 4** Number of operational taxonomic units, as organized by family, used for each trait.

| | BEL | BF | CDG | HAM | IMF | LD | LOI | M.L | M.a | M.b | SCOL | SFIR | SMAR | SSF | pH | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacteroidaceae | | 1 | 2 | | 2 | 1 | | 1 | | | 1 | | 1 | | | 9 |
| Clostridiaceae | 1 | 1 | | | | 1 | 1 | | 1 | 2 | | | | 1 | | 8 |
| ClostridialesIncertae Sedis XI | | | | | | | | | | | | | | 1 | | 1 |
| Coriobacteriaceae | | | | | | | | | | | | 1 | | | | 1 |
| Corynebacteriaceae | | | | | | | | | | | | | | 1 | | 1 |
| Desulfovibrionaceae | | | | | | | | | | | | | 1 | | | 1 |
| Erysipelotrichaceae | | 1 | | | | | | 1 | 1 | | | | 1 | 1 | | 5 |
| Helicobacteraceae | | | | | | | | | | | 1 | | | | | 1 |
| Lachnospiraceae | 8 | 4 | 7 | 7 | 3 | 6 | 5 | 4 | 6 | 8 | 1 | 8 | 6 | 3 | 6 | 82 |
| Methanobacteriaceae | | | | | | 1 | | | | | | | | | | 1 |
| Peptostreptococcaceae | | 1 | | | 1 | 1 | | 1 | 1 | 1 | 2 | | 1 | | | 9 |
| Porphyromonadaceae | 1 | 1 | | | 1 | | | 1 | | | | | | | | 4 |
| Prevotellaceae | 2 | | 1 | | 1 | 1 | 1 | | | 1 | | 3 | | 1 | 2 | 13 |
| Ruminococcaceae | 7 | 5 | 10 | 6 | 2 | 3 | 4 | 6 | 4 | 5 | 7 | 5 | 4 | 4 | 8 | 80 |
| Spirochaetaceae | | | | | 1 | 1 | | | | | | | | | | 2 |
| Streptococcaceae | | 1 | | | | | 1 | 1 | | | | | | 1 | | 4 |
| Veillonellaceae | | | | | | | | | 1 | | | | | | | 1 |
| Unassigned | 6 | 10 | 5 | 12 | 14 | 10 | 13 | 10 | 11 | 8 | 13 | 8 | 11 | 12 | 9 | 152 |

*Note*: Results refer to Stage 1 of rectal microbiome sampling.

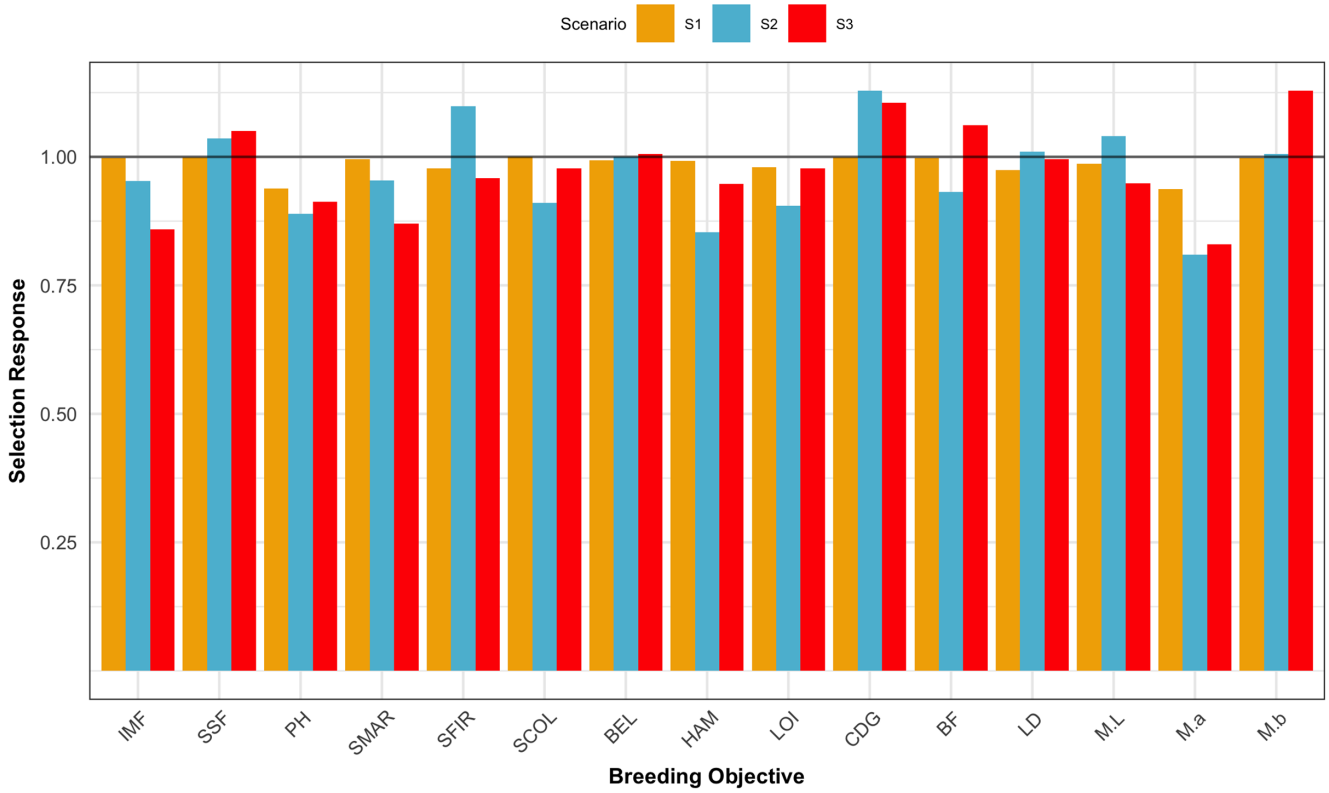


**FIGURE 2** Expected genetic progress achieved using indirect selection based on microbial information relative to direct selection. BEL, Belly weight; BF, Backfat depth; CDG, Carcass average daily gain; HAM, Ham weight; IMF, Intramuscular fat; LD, Loin depth; LOI, Loin weight; M.a, Minolta a*; M.b, Minolta b*; M.L, Minolta L*; PH, PH; S1, Microbiome measured at weaning; S2, Microbiome measured at mid-test; S3, Microbiome measured at off-test; SCOL, Subjective colour; SFIR, Subjective firmness; SMAR, Subjective marbling; SSF, Warner–Bratzler slice shear force.

additional information to the validation set. A REIS above the value of 1 indicates that selection for (favourable) microbiome composition could yield genetic progress larger than direct selection.

According to the variance components estimates, indirect selection for gut microbiome composition seldom outperformed direct selection. Indirect selection did not outperform direct selection for PH, LOI and M.a. Indirect and direct selection were equivalent for IMF, SMAR, SCOL and HAM only when S1 was used. In contrast, direct selection showed better response when using S2 and S3. Indirect selection showed better response for SSF (S2 and S3 only), SFIR (S2 only), CDG (S2 and S3 only), BF (S3 only) and M.b (S3 only). All other traits showed equivalent direct or indirect selection.

## 3.2 | Multiple-trait genomic prediction

Results from the cross-validation are reported in Figures 3 and 4. Prediction accuracy is reported on the *y*-axis (with error bars obtained by bootstrapping), while the different scenarios are reported on the *x*-axis. The plots compare ABLUP and GBLUP for the single-trait model ST and the multiple-trait models that include the MF. The latter are split for including the MF over S1, S2 and S3, and for the number of MF from each stage included (1, 2, 5, 10 or 25).

The different BGT showed variable accuracy even for the ST model. For some traits, no difference was found between ABLUP and GBLUP. These were LD (0.10), HAM (0.10), SSF (0.10), SFIR (0 vs. 0.06), SCOL (0.11 vs. 0.18), M.L (0.09 vs. 0.19), M.a (0.05 vs. 0.20), M.b (0 vs. 0.10). Traits that showed a difference between ABLUP and GBLUP were CDG (0.02 vs. 0.19), BF (−0.04 vs. 0.28), BEL (0.05 vs. 0.20), LOI (0 vs. 0.12), IMF (−0.08 vs. 0.28), PH (−0.08 vs. 0.06), SMAR (−0.03 vs. 0.24). In general, GBLUP performance was equal to or better than ABLUP.

The variability in prediction accuracy among the traits was even larger with the MF-enabled models. Considering an average between the ABLUP and GBLUP models, some traits showed a strong increase in prediction accuracy with the MF-enabled models. These were IMF and SMAR, mostly when including up to 5 MF from S1 or S3; BEL and PH when including MF from S1 and S2; HAM when including few MF from S1 and S2; BF and CDG in almost any case of using microbial features. Prediction

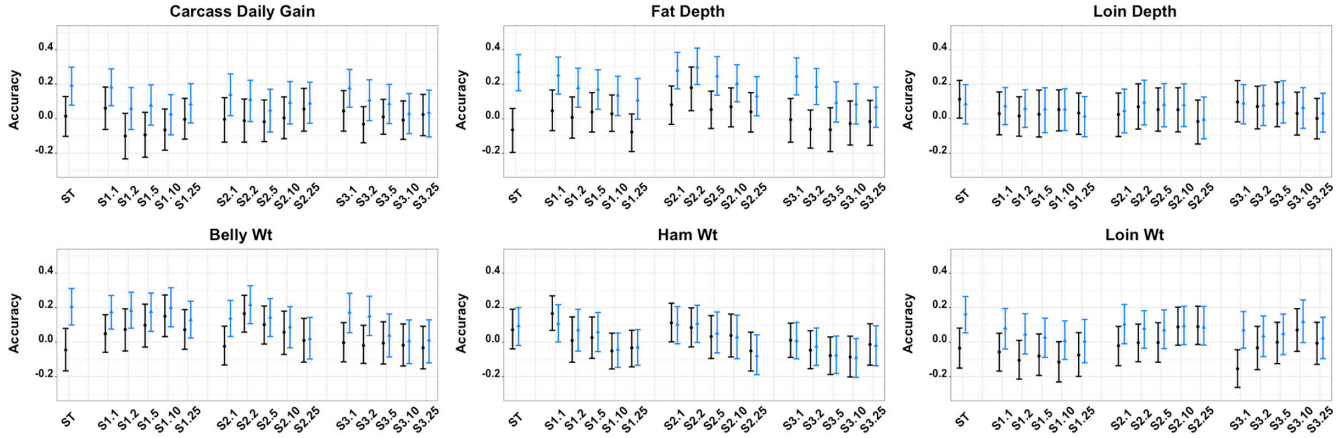


**FIGURE 3** Prediction accuracy (with confidence intervals of bootstrapped values) for nine of the studied traits. Black dots and lines: Pedigree-based BLUP; Blue dots and lines: Genomic-based BLUP; ST: Single-trait model; S1.1-2-5-10-25: Multiple-trait models incorporating the top 1-2-5-10-25 microbial features obtained at stage 1 (S1, weaning); S2.1-2-5-10-25: Multiple-trait models incorporating the top 1-2-5-10-25 microbial features obtained at stage 2 (S2, on-test); S3.1-2-5-10-25: Multiple-trait models incorporating the top 1-2-5-10-25 microbial features obtained at stage 3 (S3, off-test).

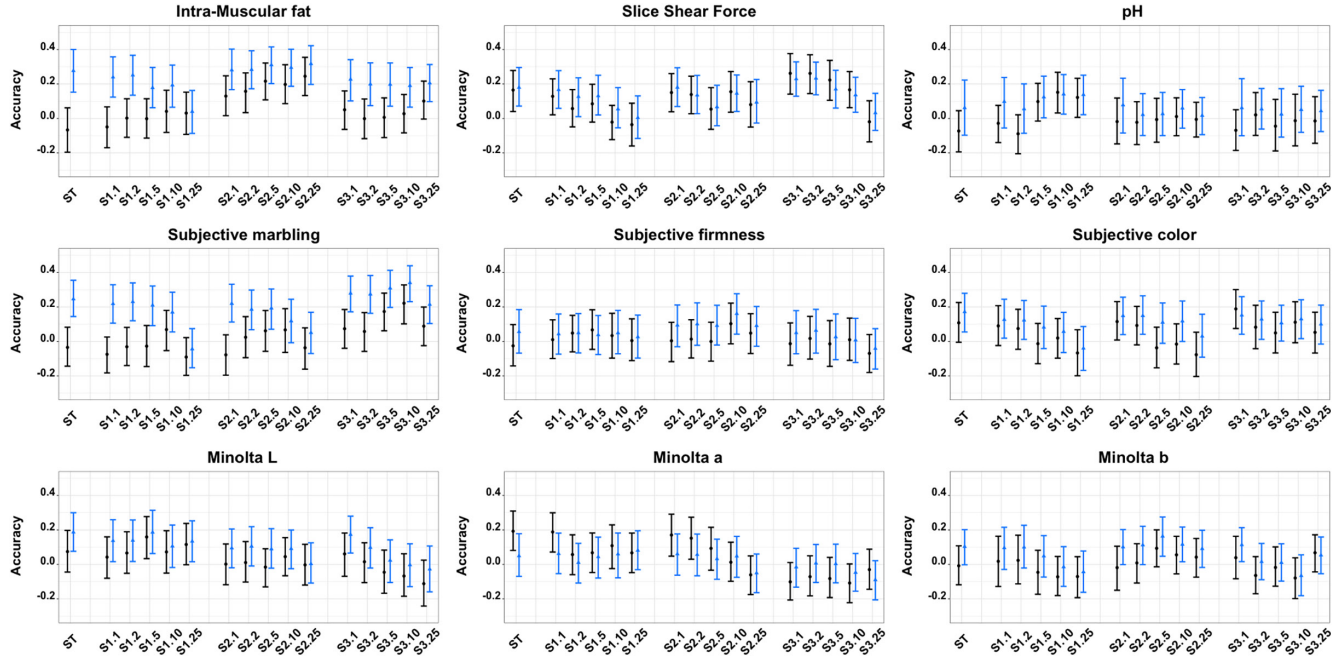


**FIGURE 4** Prediction accuracy (with confidence intervals of bootstrapped values) for six of the studied traits. Black dots and lines: Pedigree-based BLUP; Blue dots and lines: Genomic-based BLUP; ST: Single-trait model; S1.1-2-5-10-25: Multiple-trait models incorporating the top 1-2-5-10-25 microbial features obtained at stage 1 (S1, weaning); S2.1-2-5-10-25: Multiple-trait models incorporating the top 1-2-5-10-25 microbial features obtained at stage 2 (S2, on-test); S3.1-2-5-10-25: Multiple-trait models incorporating the top 1-2-5-10-25 microbial features obtained at stage 3 (S3, off-test).

accuracy for MF-enabled models was lower than ST for several traits when referring to the performance of the same relationship matrix used (from pedigree or genomic markers). These were SSF, SCOL, M.L, M.a and M.b.

Among the traits that showed an increase in prediction accuracy using the MF-enabled models, some showed this increase to be large enough that the MF-enabled ABLUP models would outperform the ST GBLUP models (when this one showed better performance than the ST GBLUP model). This was the case for IMF, SMAR, pH, BF, LOI and BEL.

# 4 | DISCUSSION

## 4.1 | Selection of relevant gut microbiome features

In this study, we used the MF as correlated traits in a multiple-trait model to improve genetic progress and model's predictive ability. This study did not aim to estimate the shared microbial architecture among traits since this was already reported in a previous study based on the

same data (Khanal et al., 2019). The MF that were selected to be included for each trait might not totally reflect the complexity of the association among the traits conditionally on the gut microbial composition. The overlapping of MF across the traits was moderate, and less than 50% of the total MF used were included for more than one trait, while Khanal et al. (2019) showed some degree of shared microbial architecture.

The selection of the 25 most relevant features for all traits considered gave results that are similar to another former study on the same population where the relationship at the phenotypic level was estimated (Maltecca et al., 2021). Results show the same genera being the most relevant for their association at the phenotypic and genetic levels. However, the number of MF that underlie this association seems to be different. In addition, the results from this study appear to be consistent with another study that aimed to estimate genetic correlations between MF and different traits in swine in an unrelated population (Aliakbari et al., 2021).

The genera *Prevotella*, belonging to the family *Prevotellaceae*, and *Blautia*, belonging to the family *Lachnospiraceae*, were the most frequent and consistent in having MF selected in the present study. Few MF were found to be significantly associated in Maltecca et al. (2021), although Aliakbari et al. (2021) estimated a strong genetic correlation between these genera and backfat depth.

Genera *Lactobacillus* (*Lactobacillaceae*) and *Streptococcus* (*Streptococcaceae*) had a strong and consistent relevance in the present study (although diminishing as age increased for the former), which was also found at the genetic level by Aliakbari et al. (2021) and at the phenotypic level by Maltecca et al. (2021). The genera *Coprococcus* (*Lachnospiraceae*) and *Faecalibacterium* (*Ruminococcaceae*) had a small but consistent relevance over time in the present study, with the same contribution at the phenotypic level in Maltecca et al. (2021) and a moderately strong genetic correlation with backfat depth in Aliakbari et al. (2021). The genus *Dorea* (*Lachnospiraceae*) showed a moderate relevance mostly at S1 and S2 was not found to be associated at the phenotypic level but showed a strong genetic correlation with backfat dept in Aliakbari et al. (2021). The genus *Oscillibacter* (*Ruminococcaceae*) showed relevance in this study at S1 and S2, but only a small association at the phenotypic level. *Ruminococcus* (*Ruminococcaceae*) had a relevance that, in this study, increased as the age of the animal progressed, which was also previously found at the phenotypic level but not confirmed by other studies at the genetic level. Similarly, *Roseburia* (*Lachnospiraceae*) had strong relevance at S2 in the present study, previously showing a strong association at the phenotypic level at S3. In the present study, *Campylobacter* (*Campylobacteraceae*) and *Staphylococcus* (*Staphylococcaceae*) were relevant at S1, but no association was found in the other studies.

It should also be considered that the selection of the MF was performed in a way that could be suboptimal. Given the complexity of the microbiota and its compositional nature (Gloor et al., 2017), selecting the features on a one-by-one basis for their association at the genetic level with the traits of interest might neglect the dependency among the features themselves. While this issue was partially addressed by estimating the variance components with the 26-trait model, other methods could probably handle the high-dimensionality of the microbial features more efficiently (Lopez-Cruz et al., 2020; Runcie et al., 2021).

## 4.2 | Relevance of gut microbiota information

The microbiota information could serve as a correlated indicator trait in a selection index.

While the expected genetic progress, based on variance components estimation, showed little advantage in selecting for the MF rather than the trait itself, the cross-validation suggested otherwise. Some traits showed increased prediction accuracy with the MF-enabled models regardless of the use of host genomic information. MF-based selection showed improvements over direct selection for traits such as IMF and SMAR, with MF from S1 being the most informative. Similarly, CDG, BF and LOI showed a less relevant increase, but S1 remained the most informative microbial information source. Most of these traits showed diminishing returns in including a larger number of MF, suggesting that few MF are truly related to the trait of interest on the genetic level.

All the traits mentioned (except for CDG) are expensive to measure, require carcass dissection and, in some cases, require expensive wet-lab analyses. Sampling with a rectal swab and performing 16S rRNA sequencing could be a cost-effective solution to cut the selection cost for these traits. Using microbial composition in the selection index would not replace the need to keep and update a reference population with all the relevant phenotypes available. Still, it could improve the efficiency of the breeding scheme (Martínez-Álvaro et al., 2022).

Likely, the shared genetic architecture between the MF and the other traits is to be studied more in depth to gain the most out of this strategy. Nonetheless, we must remark that the method proposed here is too simplistic and does not fully account for the dimensionality and complexity of the microbial composition. While this is the first attempt to carry out such a study, further research on this topic is

warranted. Also, changes in the bio-informatic part of the pipeline could lead to changes in the results. Rarefaction could be avoided and better strategies for normalization could be adopted. This will have to be addressed in further studies.

To the best of our knowledge, no other study has used MF with the same methodology proposed here. The partial consistency of the present estimates with those reported by Aliakbari et al. (2021) suggests that the shared genetic architecture that is emerging here is validated in unrelated populations. If that were the case, this could support the use of MF in selection indices, both in swine and other livestock species (Haas et al., 2022; Martinez Boggio et al., 2022; Martínez-Álvaro et al., 2022; Saborío-Montero et al., 2020, 2021).

### 4.3 | Gut microbiota sequencing versus host genotyping

In the present study, we also aimed to compare the value of generating data for gut microbial composition versus obtaining host (pig) genomic information using an SNP chip. Since we ran this study on 15 different traits with different genetic and microbial architecture, this study could serve as a ground for comparing the two sources of information.

This comparison was only carried out using cross-validation since the change due to phenotyping would have only been through an increase/decrease in (G)EBV prediction accuracy, which we considered not fully representative of the shared genetic architecture between the MF and the other traits.

In the cross-validation, the performance of GBLUP was generally better than the performance of ABLUP. However, the performance of ST-GBLUP should be compared to the performance of the top-performing MF-enabled ABLUP model. This comparison would allow to compare the values of each technology directly. For some traits, in this sense, some of the MF-enabled ABLUP models seemed to have better prediction (i.e. IMF, SMAR, PH, BF, BEL) than the ST-GBLUP. Results suggest that there could be some value in adding microbial information rather than host genomic information to the model, although this seems limited to some traits and circumstances (e.g. time of microbial characterization). In fact, the use of correlated indicator traits in selection is known to be valuable when the magnitude of the genetic correlations reaches a given value (e.g. 0.5; see Calus & Veerkamp, 2011). Other sources of information have been used in multiple-trait genomic prediction models. These include infrared spectroscopy in dairy cattle, where both Cecchinato et al. (2020) and van den Berg et al. (2020) concluded that the advantage of including the correlated trait depends on the shared genetic architecture between the breeding goal and its predictor.

## 5 | CONCLUSION

The present study shows a simple method to account for gut microbial information in breeding value prediction models. We focused on several expensive-to-measure traits that are economically important to the swine breeding industry, encompassing growth, tissue deposition, carcass composition and meat quality.

This study revealed the relevance of the microbial information compared to the host genomic. Since we studied 15 different traits, some of them highlighted the relevance of one source of information rather than the other. In some cases, gut microbial information improved the model's predictive ability more than host genomic information did. In others, host genomic information was a more valuable source of information. The latter seemed to be the most frequent case.

Further research should focus on accounting for the compositional and high-dimensional nature of the microbiota data. Results from the present study could have penalized the contribution of this kind of data due to its limited use in the models. Methods to account for higher dimensionality of the phenotype(s) are available and will be the subject of the upcoming research.

### FUNDING INFORMATION

This study is part of the project “Re-defining growth efficiency accounting for the interaction between host genome and commensal gut bacteria” funded by The National Pork Board Association and part of the project “From host to guest and back” funded by The Maschhoffs LLC and North Carolina State University.

### CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

### ORCID

*Francesco Tiezzi* https://orcid.org/0000-0002-4358-9236
*Christian Maltecca* https://orcid.org/0000-0002-9996-4680

## REFERENCES

Aliakbari, A., Zemb, O., Billon, Y., Barilly, C., Ahn, I., Riquet, J., & Gilbert, H. (2021). Genetic relationships between feed efficiency and gut microbiome in pig lines selected for residual feed intake. *Journal of Animal Breeding and Genetics*, *138*, 491–507. https://doi.org/10.1111/jbg.12539

Bergamaschi, M., Maltecca, C., Schillebeeckx, C., McNulty, N. P., Schwab, C., Shull, C., & Tiezzi, F. (2020). Heritability and genome-wide association of swine gut microbiome features with growth and fatness parameters. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-66791-3

Bergamaschi, M., Tiezzi, F., Howard, J., Huang, Y. J., Gray, K. A., Schillebeeckx, C., & Maltecca, C. (2020). Gut microbiome composition differences among breeds impact feed efficiency in swine. *Microbiome*, *8*(1), 1–15. https://doi.org/10.1186/s40168-020-00888-9

Bonfatti, V., Vicario, D., Degano, L., Lugo, A., & Carnier, P. (2017). Comparison between direct and indirect methods for exploiting Fourier transform spectral information in estimation of breeding values for fine composition and technological properties of milk. *Journal of Dairy Science*, *100*(3), 2057–2067. https://doi.org/10.3168/jds.2016-11951

Calus, M. P. L., & Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genetics Selection Evolution*, *43*, 26. https://doi.org/10.1186/1297-9686-43-26

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S., & Bennewitz, J. (2017). Host genome influence on gut microbial composition and microbial prediction of complex traits in pigs. *Genetics*, *206*, 1637–1644. https://doi.org/10.1534/genetics.117.200782

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336. https://doi.org/10.1038/nmeth.f.303

Cecchinato, A., Toledo-Alvarado, H., Pegolo, S., Rossoni, A., Santus, E., Maltecca, C., Bittante, G., & Tiezzi, F. (2020). Integration of wet-lab measures, Milk infrared spectra, and genomics to improve difficult-to-measure traits in dairy cattle populations. *Frontiers in Genetics*, *11*, 1131. https://doi.org/10.3389/fgene.2020.563393

Crespo-Piazuelo, D., Migura-Garcia, L., Estellé, J., Criado-Mesas, L., Revilla, M., Castelló, A., & Folch, J. M. (2019). Association between the pig genome and its gut microbiota composition. *Scientific Reports*, *9*, 8791. https://doi.org/10.1038/s41598-019-45066-6

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., & Gordon, J. I. (2013). The long-term stability of the human gut microbiota. *Science*, *341*, 1237439. https://doi.org/10.1126/science.1237439

Fang, S., Xiong, X., Su, Y., Huang, L., & Chen, C. (2017). 16S rRNA gene-based association study identified microbial taxa associated with pork intramuscular fat content in feces and cecum lumen. *BMC Microbiology*, *17*(1), 162. https://doi.org/10.1186/s12866-017-1055-x

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224

Godinho, R. M., Bergsma, R., Silva, F. F., Sevillano, C. A., Knol, E. F., Lopes, M. S., & Guimarães, S. E. (2018). Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. *Journal of Animal Science*, *96*(3), 817–829. https://doi.org/10.1093/jas/skx011

Haas, V., Vollmar, S., Preuß, S., Rodehutscord, M., Camarinha-Silva, A., & Bennewitz, J. (2022). Composition of the ileum microbiota is a mediator between the host genome and phosphorus utilization and other efficiency traits in Japanese quail (*Coturnix japonica*). *Genetics Selection Evolution*, *54*(1), 20. https://doi.org/10.1186/s12711-022-00697-8

Khanal, P., Maltecca, C., Schwab, C., Fix, J., & Tiezzi, F. (2021). Microbiability of meat quality and carcass composition traits in swine. *Journal of Animal Breeding and Genetics*, *138*(2), 223–236. https://doi.org/10.1111/jbg.12504

Khanal, P., Maltecca, C., Schwab, C., Gray, K., & Tiezzi, F. (2019). Genetic parameters of meat quality, carcass composition and growth traits in commercial swine. *Journal of Animal Science*, *97*, 3669–3683. https://doi.org/10.1093/jas/skz247

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature*, *444*, 1022–1023. https://doi.org/10.1038/nature4441022a

Lopez-Cruz, M., Olson, E., Rovere, G., Crossa, J., Dreisigacker, S., Mondal, S., Singh, R., & de Los Campos, G. (2020). Regularized selection indices for breeding value prediction using hyperspectral image data. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-65011-2

Lu, D., Tiezzi, F., Schillebeeckx, C., McNulty, N. P., Schwab, C., Shull, C., & Maltecca, C. (2018). Host contributes to longitudinal diversity of fecal microbiota in swine selected for lean growth. *Microbiome*, *6*, 4. https://doi.org/10.1186/s40168-017-0384-1

Magoc, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957–2963.

Maltecca, C., Dunn, R., He, Y., McNulty, N. P., Schillebeeckx, C., Schwab, C., & Tiezzi, F. (2021). Microbial composition differs between production systems and is associated with growth performance and carcass quality in pigs. *Animal Microbiome*, *3*(1), 1–20. https://doi.org/10.1186/s42523-021-00118-z

Martinez Boggio, G., Meynadier, A., Buitenhuis, A. J., & Marie-Etancelin, C. (2022). Host genetic control on rumen microbiota and its impact on dairy traits in sheep. *Genetics Selection Evolution*, *54*(1), 77. https://doi.org/10.1186/s12711-022-00769-9

Martínez-Álvaro, M., Mattock, J., Auffret, M., Weng, Z., Duthie, C. A., Dewhurst, R. J., Cleveland, M. A., Watson, M., & Roehe, R. (2022). Microbiome-driven breeding strategy potentially improves beef fatty acid profile benefiting human health and reduces methane emissions. *Microbiome*, *10*(1), 166. https://doi.org/10.1186/s40168-022-01352-6

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., & Lee, D. H. (2002). *BLUPF90 and related programs (BGF90)*. In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production: 19–23 August 2002. Montpellier.

Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2015). zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and*

*Intelligent Laboratory Systems*, *143*, 85–96. https://doi.org/10.1016/j.chemolab.2015.02.019

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.

Putz, A. M., Tiezzi, F., Maltecca, C., Gray, K. A., & Knauer, M. T. (2015). Variance component estimates for alternative litter size traits in swine. *Journal of Animal Science*, *93*(11), 5153–5163. https://doi.org/10.2527/jas2015-9416

Quan, J., Cai, G., Ye, J., Yang, M., Ding, R., Wang, X., Zheng, E., Fu, D., Li, S., Zhou, S., Liu, D., Yang, J., & Wu, Z. (2018). A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios. *Scientific Reports*, *8*(1), 4536. https://doi.org/10.1038/s41598-018-22692-0

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramayo-Caldas, Y., Mach, N., Lepage, P., Levenez, F., Denis, C., Lemonnier, G., Leplat, J.-J., Billon, Y., Berri, M., Doré, J., Rogel-Gaillard, C., & Estellé, J. (2016). Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME Journal*, *10*(12), 2973–2977. https://doi.org/10.1038/ismej.2016.77

Ramayo-Caldas, Y., Prenafeta-Boldú, F., Zingaretti, L. M., Gonzalez-Rodriguez, O., Dalmau, A., & Quintanilla, R. (2020). Gut eukaryotic communities in pigs: Diversity, composition and host genetics contribution. *Animal Microbiome*, *2*, 18. https://doi.org/10.1186/s42523-020-00038-4

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., Hyslop, J. J., Waterhouse, A., Freeman, T. C., Watson, M., & Wallace, R. J. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genetics*, *12*(2), e1005846. https://doi.org/10.1371/journal.pgen.1005846

Runcie, D. E., Qu, J., Cheng, H., & Crawford, L. (2021). MegaLMM: Mega-scale linear mixed models for genomic predictions with thousands of traits. *Genome Biology*, *22*(1), 1–25. https://doi.org/10.1186/s13059-021-02416-w

Saborío-Montero, A., Gutiérrez-Rivas, M., García-Rodríguez, A., Atxaerandio, R., Goiri, I., Lopez de Maturana, E., Jiménez-Montero, J. A., Alenda, R., & González-Recio, O. (2020). Structural equation models to disentangle the biological relationship between microbiota and complex traits: Methane production in dairy cattle as a case of study. *Journal of Animal Breeding and Genetics*, *137*(1), 36–48. https://doi.org/10.1111/jbg.12444

Saborío-Montero, A., López-García, A., Gutiérrez-Rivas, M., Atxaerandio, R., Goiri, I., García-Rodriguez, A., Jiménez-Montero, J. A., González, C., Tamames, J., Puente-Sánchez, F., Varona, L., Serrano, M., Ovilo, C., & González-Recio, O. (2021). A dimensional reduction approach to modulate the core ruminal microbiome associated with methane emissions via selective breeding. *Journal of Dairy Science*, *104*, 8135–8151. https://doi.org/10.3168/jds.2020-20005

Schloss, P. D., & Handelsman, J. (2006). Toward a census of bacteria in soil. *PLoS Computational Biology*, *2*, e92. https://doi.org/10.1371/journal.pcbi.0020092

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*, 863–864. https://doi.org/10.1093/bioinformatics/btr026

Schneeberger, M., Barwick, S. A., Crow, G. H., & Hammond, K. (1992). Economic indices using breeding values predicted by BLUP. *Journal of Animal Breeding and Genetics*, *109*(1–6), 180–187. https://doi.org/10.1111/j.1439-0388.1992.tb00395.x

See, G. M., & Knauer, M. T. (2019). Associations with four generations of divergent selection for age at puberty in swine. *Journal of Animal Science*, *97*(6), 2320–2328. https://doi.org/10.1093/jas/skz129

Tiezzi, F., Fix, J., Schwab, C., Shull, C., & Maltecca, C. (2021). Gut microbiome mediates host genomic effects on phenotypes: A case study with fat deposition in pigs. *Computational and Structural Biotechnology Journal*, *19*, 530–544. https://doi.org/10.1016/j.csbj.2020.12.038

Valente, B. D., Rosa, G. J., Gianola, D., Wu, X. L., & Weigel, K. (2013). Is structural equation modeling advantageous for the genetic improvement of multiple traits? *Genetics*, *194*(3), 561–572. https://doi.org/10.1534/genetics.113.151209

van den Berg, I., Ho, P. N., Luke, T. D. W., Haile-Mariam, M., Bolormaa, S., & Pryce, J. E. (2020). The use of milk mid-infrared spectroscopy to improve genomic prediction accuracy of serum biomarkers. *Journal of Dairy Science*, *104*(2), 2008–2017. https://doi.org/10.3168/jds.2020-19468

van den Boogaart, K. G., & Tolosana-Delgado, R. (2008). "Compositions": A unified R package to analyze compositional data. *Computers & Geosciences*, *34*(4), 320–338. https://doi.org/10.1016/j.cageo.2006.11.017

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Weishaar, R., Wellmann, R., Camarinha-Silva, A., Rodehutscord, M., & Bennewitz, J. (2020). Selecting the hologenome to breed for an improved feed efficiency in pigs—A novel selection index. *Journal of Animal Breeding and Genetics*, *137*(1), 14–22. https://doi.org/10.1111/jbg.12447

Wilson, K. B., Overholt, M. F., Hogan, E. K., Schwab, C., Shull, C. M., Ellis, M., Grohmann, N. S., Dilger, A. C., & Boler, D. D. (2016). Predicting pork loin chop yield using carcass and loin characteristics. *Journal of Animal Science*, *94*, 4903–4910. https://doi.org/10.2527/jas.2016-0610

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Tiezzi, F., Schwab, C., Shull, C., & Maltecca, C. (2024). Multiple-trait genomic prediction for swine meat quality traits using gut microbiome features as a correlated trait. *Journal of Animal Breeding and Genetics*, *00*, 1–16. https://doi.org/10.1111/jbg.12887