Journal of Animal Science, 2022, **100**, 1–14
https://doi.org/10.1093/jas/skac231
Advance access publication 1 July 2022
Microbiology and Microbiome

Exploring methods to summarize gut microbiota composition for microbiability estimation and phenotypic prediction in swine

Yuqing He[†,1], Francesco Tiezzi[†,‡], Jicai Jiang[†], Jeremy Howard[‖], Yijian Huang[‖], Kent Gray[‖], Jung-Woo Choi[$], Christian Maltecca[†]

[†]Department of Animal Science, North Carolina State University, Raleigh, NC 27607, USA
[‡]Department of Agriculture, Food, Environment and Forestry, University of Florence, Firenze 50144, Italy
[‖]Smithfield Premium Genetics, Rose Hill, NC 28458, USA
[$]College of Animal Life Sciences, Division of Animal Resource Science 1 Gangwondaehak-gil, Chuncheon-si, Gangwon-do, 24341, Republic of Korea

[1]Corresponding author: yhe22@ncsu.edu

Abstract

The microbial composition resemblance among individuals in a group can be summarized in a square covariance matrix and fitted in linear models. We investigated eight approaches to create the matrix that quantified the resemblance between animals based on the gut microbiota composition. We aimed to compare the performance of different methods in estimating trait microbiability and predicting growth and body composition traits in three pig breeds. This study included 651 purebred boars from either breed: Duroc (n = 205), Landrace (n = 226), and Large White (n = 220). Growth and body composition traits, including body weight (**BW**), ultrasound backfat thickness (**BF**), ultrasound loin depth (**LD**), and ultrasound intramuscular fat (**IMF**) content, were measured on live animals at the market weight (156 ± 2.5 d of age). Rectal swabs were taken from each animal at 158 ± 4 d of age and subjected to 16S rRNA gene sequencing. Eight methods were used to create the microbial similarity matrices, including 4 kernel functions (Linear Kernel, **LK**; Polynomial Kernel, **PK**; Gaussian Kernel, **GK**; Arc-cosine Kernel with one hidden layer, **AK1**), 2 dissimilarity methods (Bray-Curtis, **BC**; Jaccard, **JA**), and 2 ordination methods (Metric Multidimensional Scaling, **MDS**; Detrended Correspondence analysis, **DCA**). Based on the matrix used, microbiability estimates ranged from 0.07 to 0.21 and 0.12 to 0.53 for Duroc, 0.03 to 0.21 and 0.05 to 0.44 for Landrace, and 0.02 to 0.24 and 0.05 to 0.52 for Large White pigs averaged over traits in the model with sire, pen, and microbiome, and model with the only microbiome, respectively. The GK, JA, BC, and AK1 obtained greater microbiability estimates than the remaining methods across traits and breeds. Predictions were made within each breed group using four-fold cross-validation based on the relatedness of sires in each breed group. The prediction accuracy ranged from 0.03 to 0.18 for BW, 0.08 to 0.31 for BF, 0.21 to 0.48 for LD, and 0.04 to 0.16 for IMF when averaged across breeds. The BC, MDS, LK, and JA achieved better accuracy than other methods in most predictions. Overall, the PK and DCA exhibited the worst performance compared to other microbiability estimation and prediction methods. The current study shows how alternative approaches summarized the resemblance of gut microbiota composition among animals and contributed this information to variance component estimation and phenotypic prediction in swine.

Lay Summary

Gut microbiota has received significant research attention in farm animals because of its close relationship with host performance. We chose eight approaches to create a square covariance matrix that characterizes the relationship among animals based on their gut microbiota composition. Then, we fitted this information with linear models to evaluate the proportion of phenotypic variance explained by gut microbiota composition and predict host growth and body composition traits in three pig breeds. We found that different matrices had varying performance in predicting host phenotypes, but the results highly depended on the trait and breed considered in the prediction. Our findings highlight possible alternative approaches to incorporate gut microbiome data in regression models and emphasize the value of gut microbiome data in better understanding complex traits in pigs with diverse genetic backgrounds.

Key words: Gut microbiota, microbial similarity matrix, microbiability, prediction, swine

Abbreviations: AK1, arc-cosine kernel with one hidden layer; ASV, amplicon sequence variant; BC, Bray-Curtis; BF, Backfat thickness; BW, body weight; DCA, detrended correspondence analysis; DIC, deviance information criterion; DR, Duroc; GK, gaussian kernel; IMF, intramuscular fat; JA, Jaccard; LD, loin depth; LK, linear kernel; LLPM, log-likelihood evaluated at posterior mean; LR, Landrace; LW, Large White; MDS, metric multidimensional scaling; pD, effective number of parameters; PK, polynomial kernel; PMLL, posterior mean of the log-likelihood

Introduction

Trillions of diverse microorganisms create a vast and complex ecosystem in the gastrointestinal tracts of humans and animals (Savage, 1977; Cryan et al., 2019). The term "microbiota" is commonly used to refer to the entire structure of this ecological community in microbial research (Turnbaugh et al., 2007). Benefitting from continuous reductions in the cost of next-generation sequencing, it is now possible to document

Received February 17, 2022 Accepted June 28, 2022.

the gut microbiota on a broad scale or in a longitudinal pattern. Several studies have tracked the gut microbiota compositions through the lifetime of pigs, from birth to slaughter (Wang et al., 2019; Li et al., 2020). Furthermore, the gut microbiota composition has been associated with variations in host phenotypes, such as meat quality and carcass traits (Khanal et al., 2021), feed efficiency (Camarinha-Silva et al., 2017), nutrient digestibility (Verschuren et al., 2020), and feeding behavior (He et al., 2022). This proportion of phenotypic variance accounted for by the microbiota composition was termed "microbiability" (Difford et al., 2016).

The OTUs (operational taxonomic units) or ASVs (amplicon sequence variants) tables are the most common data formats used in the analysis of current microbiome research (Xia et al., 2018; Prodan et al., 2020). The microbiome information across animals in a group can be quantified in a square covariance matrix **M** and fitted in the mixed linear model using methodologies developed for genomic prediction (Ross et al., 2013). Currently, most studies created the microbial relationship matrix for microbiability estimation and phenotype prediction using the method inspired by Ross and colleagues (Ross et al., 2013; Camarinha-Silva et al., 2017; Difford et al., 2018; Khanal et al., 2020, 2021; Verschuren et al., 2020). This approach begins with an $n \times p$ matrix $\boldsymbol{X}$ (n is the number of samples and p is the number of microbes) containing the standardized and log-transformed counts and then creates an $n \times n$ microbial relationship matrix $\boldsymbol{M}$ through $(1/p)\mathrm{XX}^{\mathrm{T}}$ (Ross et al., 2013).

However, microbiome data have a complex structure with several unique characteristics, such as high dimensions, sparseness, and great dispersity, which increase the difficulty for analysis (Xia et al., 2018; Coenen et al., 2020). Other methods, such as dissimilarity and ordination metrics, can be considered alternative approaches to measure the differences in microbiota composition across samples to inform the regression models. In recent research on livestock animals, Maltecca et al. (Maltecca et al., 2019) investigated the role of the Jensen-Shannon distance matrix in the model to predict host phenotypes. Saborío-Montero and colleagues (Saborío-Montero et al., 2021) created different dissimilarity and ordination matrices on ruminal microbiota and assessed their contributions in estimating microbiability for methane emission in cattle.

To our knowledge, the use of various microbial relationship matrices to investigate the association between the microbial community and host phenotypes has not been studied in swine. However, an appropriate method for modeling large-scale microbiome data is critical to better understand the role of gut microbiota in pork production. Thus, in this study, we selected eight approaches from the general kernel, dissimilarity, and ordination methods that are extensively used in microbial and genomic research to create the matrices quantifying the relationship among animals based on their gut microbiota composition. We aimed to assess how effective the matrices are at estimating microbiability and predicting host growth and body composition traits in three pig breeds.

Materials and Methods

Data

The data used in this study were collected from animals kept in a commercial routine operated by Smithfield Premium Genetics (SPG; Rose Hill, NC). Thus, animal use approval was not required for this study. Our previous work provided a general description of animals and their rearing conditions (He et al., 2022). Briefly, data were collected on 651 boars from either 3 breeds: Duroc (**DR**; n = 205), Landrace (**LR**; n = 226), and Large White (**LW**; n = 220). Growth and body composition traits, including body weight (**BW**), ultrasound backfat thickness (**BF**), ultrasound loin depth (**LD**), and ultrasound intramuscular fat (**IMF**) content, were measured on live animals that had reached the market weight (around 120 kg) at an average age of 156 ± 2.5 d. Ultrasound pictures were captured over the last 3 ribs using an Aloka 500 Ultrasound machine (Corometrics Medical Systems, Wallingford, CT). Descriptive statistics of traits for each breed are depicted in Fig. 1. This study used the rectal swabs taken from each pig at 158 ± 4 d of age for microbiome.

16S rRNA gene sequencing and microbiome data

Rectal swabs were subjected to 16S rRNA gene sequencing, and the procedures were detailed in our previous studies (Lu et al., 2018; He et al., 2022). The sequencing was performed in the DNA Sequencing Innovation Lab at the Center for Genome Sciences and Systems Biology at Washington University (St. Louis, MO). The bioinformatics pipeline for processing the sequencing data was described in previous publications (Lu et al., 2018; He et al., 2022). An amplicon sequence variant (**ASV**) feature table along with taxonomic information was constructed. The ASVs were removed if the overall counts were less than 1,000 across all samples and the prevalence rate was less than 0.05. A total of 824 ASVs passed the quality control and were used in subsequent analyses.

Statistical analysis

Microbial relationship/similarity matrices

This study used eight methods to construct the microbial similarity matrix $\boldsymbol{M}$ capturing the resemblance among animals based on their gut microbiota abundance. First, we chose the 4 general kernel functions described by Montesinos-López and colleagues (Montesinos-López et al., 2021): Linear Kernel (**LK**), Polynomial Kernel (**PK**), Gaussian Kernel (**GK**), and Arc-cosine Kernel with one hidden layer (**AK1**) to create the **M** matrix in R environment (R Core Team, 2021). These methods have been effectively used in mixed linear models to describe the genomic relationship and guide the prediction (Schaid, 2010; Jiang and Reif, 2015; Cuevas et al., 2019; Souza et al., 2019). In this study, we treated the LK as the baseline method because it is the most common method used in current microbiome research (Ross et al., 2013; Camarinha-Silva et al., 2017; Difford et al., 2018; Khanal et al., 2020, 2021; Verschuren et al., 2020). Before applying the kernel functions, the ASV counts were log-transformed, then centered and scaled (Montesinos-López et al., 2021).

Second, we chose two dissimilarity approaches: Bray-Curtis (**BC**) and Jaccard (**JA**), to measure the differences in the gut microbiota composition between pairs of animals. The BC dissimilarity is often applied on microbial count data, considering the difference in the abundance of individual microbes and the total abundance of the samples (Greenacre and Primicerio, 2014). The JA dissimilarity, which is based on the presence/absence of microbes, particularly considers rare and low-abundance microbes in the measurement

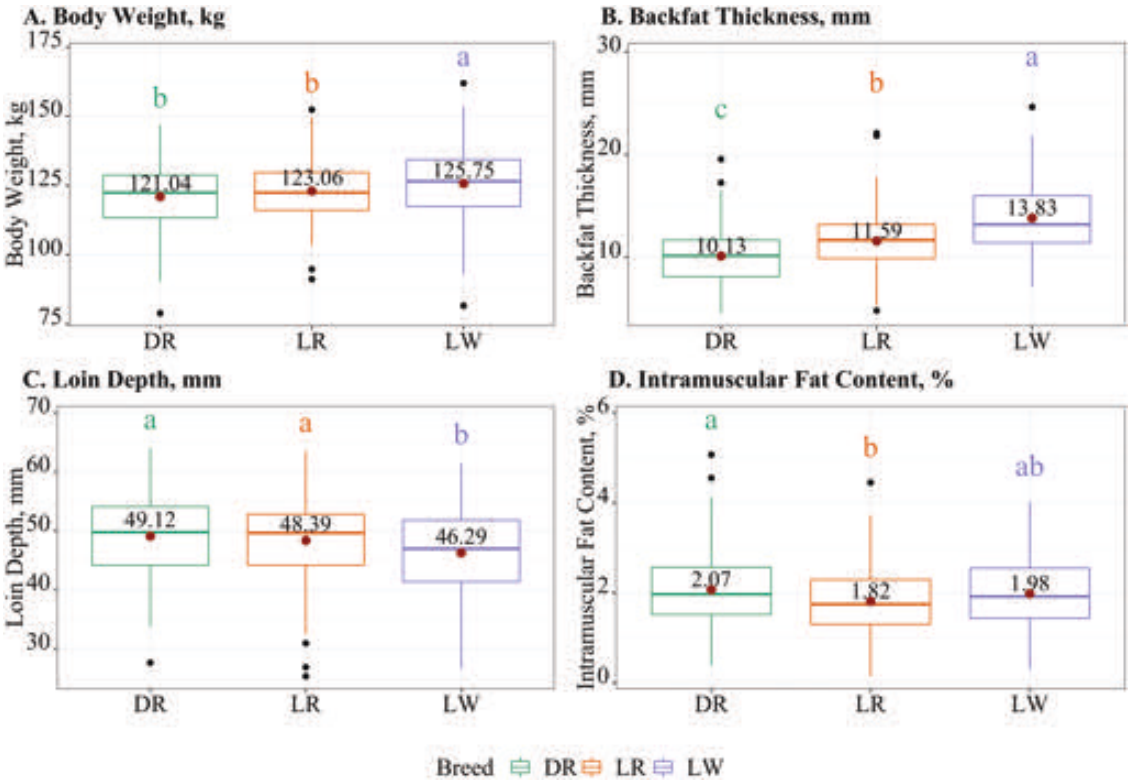

Figure 1. Descriptive statistics of growth and body composition traits by breed. Dots in the middle are the mean with values labeled. Different letters a, b, and c denote $P < 0.05$ for each trait among breeds. Colors represent each breed group. DR: Duroc; LR: Landrace; LW: Large White.

(Greenacre and Primicerio, 2014). The dissimilarity matrices were created using the *vegdist* function of *vegan* package (Oksanen et al., 2020) on the log-transformed ASV count data in R environment (R Core Team, 2021). The resulting matrices were square, with zero on the diagonal and off-diagonal values ranging from zero to one. The ***M*** matrix was then obtained as one minus the dissimilarity matrix.

Finally, 2 distance-based ordination methods: standard Metric Multidimensional Scaling (**MDS**; also known as Principal Coordinates Analysis) and Detrended Correspondence Analysis (**DCA**) were used to build the ***M*** matrices from the BC dissimilarity matrix. The standard MDS calculates the coordinates of samples based on their microbiome composition in multidimensional space to determine their similarity (Zhang and Takane, 2010). This method is mainly used to visualize clusters of samples in a low-dimensional space depending on the resemblance/difference of their microbiota composition (Nasidze et al., 2009; Larsen and Dai, 2015). The DCA is another distance-based ordination technique often used to detect microbial resemblance between samples while also addressing the distortion problem that other linear ordination methods suffer from (Hill and Gauch, 1980; Podani and Miklós 2002). The procedures for constructing the ***M*** matrix using MDS and DCA methods were similar to those published by Saborío-Montero et al. (2021). Ordinates for MDS and projections for DCA were obtained using the *ordinate* function of the *phyloseq* package (McMurdie and Holmes, 2013) in R environment (R Core Team, 2021). Table 1 shows the functions and brief descriptions of the eight approaches to create the ***M*** matrix.

Estimation of variance components and microbiability

We used the following 3 models to estimate the proportion of phenotypic variance accounted for by the gut microbiota composition, rearing environment, and host factor within each breed group.

1. Model_S_P: $y_{ijk} = \mu + sire_i + pen_j + e_{ijk}$
2. Model_S_P_M: $y_{ijkl} = \mu + sire_i + pen_j + m_k + e_{ijkl}$
3. Model_M: $y_{kl} = \mu + m_k + e_{kl}$

In these models, y was the phenotypic trait measured, μ was the overall intercept, $sire_i$ was the random effects of ith sire, pen_j is the random effect of jth pen, m_k contains the random effect of gut microbiota with $m \sim N(0, M\sigma^2_m)$, where ***M*** is the microbial relationship matrix created by one of the eight methods described in the previous section and σ^2_m is the microbial variance, e is the residual. The effects of sire, pen, and residual were assumed to be distributed as $N(0, I\sigma^2_s)$, $N(0, I\sigma^2_p)$, and $N(0, I\sigma^2_e)$, respectively; where ***I*** was an identity matrix, and σ^2_s, σ^2_p, and σ^2_e were the variances for the respective effects.

We fitted a standard microbial best linear unbiased prediction (MBLUP) model implemented of *BGLR* package (Pérez and De Los Campos, 2014) in R environment (R Core Team, 2021). The analysis was carried out using a Markov chain Monte Carlo (MCMC) algorithm with 150,000 iterations, 50,000 iterations discarded as burn-in, and a thinning interval of 10 iterations. Model convergence was checked by examining the trace plots for each parameter posterior distribution and using *geweke.diag* function with the default setting implemented of *CODA* package (Plummer et al., 2006)

Table 1. Summary of methods and the associated dissimilarity matrix to build microbial similarity matrix

Group	Method[1]	Input Data	Function[2]	Construction of *M* matrix
General Kernel	LK	Centred and scaled log(count + 1)	$LK = XX^T(\frac{1}{p})$	LK
	PK		$PK = (XX^T(\frac{1}{p}))^3$	PK
	GK		$GK = e^{-\frac{1}{p}[X^TX-2X^TX+X^TX]}$	GK
	AK1		$\theta = \cos^{-1}(\frac{X^TX}{\|X\|\|X\|})$ $AK1 = \frac{1}{\pi}\|X\|\|X\|[\sin(\theta)+(\pi-\theta)\cos(\theta)]$	AK1
Dissimilarity	BC	log(count + 1)	$BC = \frac{\sum\|a_i - b_i\|}{\sum(a_i + b_i)}$	1 – BC
	JA		$JA = \frac{y+z}{x+y+z}$	1 – JA
Ordination	MDS	log(count + 1)	$BC = \frac{\sum\|a_i - b_i\|}{\sum(a_i + b_i)}$ X = Vectors	$XX^T(\frac{1}{p})$
	DCA		$BC = \frac{\sum\|a_i - b_i\|}{\sum(a_i + b_i)}$ X = Projections	$XX^T(\frac{1}{p})$

[1]LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis.
[2]For general kernel matrices, the matrix ***X*** contained log-transformed, centered, scaled ASV abundance, with a dimension of $n \times p$, where n is the number of samples and p is the number of ASVs. For dissimilarity matrices, a_i and b_i were the log-transformed counts of ASV_i in samples a and b. The x is the number of species shared by two samples, y and z are the number of species unique in each pairwise sample. In the Chi-square distance, c_i is the total counts of ASV_i across samples.

in R environment (R Core Team, 2021). The proportion of phenotypic variance explained by the microbial composition was calculated as the σ^2_m over the total phenotypic variance.

$$m^2 = \frac{\sigma^2_m}{\sigma^2_m + (\sigma^2_s + \sigma^2_p) + \sigma^2_e}$$

Microbial prediction and cross-validation

To better evaluate the role of gut microbiome information summarized in the covariance matrix for predicting host phenotypes, we performed the prediction using the Model_M with one of eight microbial similarity matrices within each breed group. A four-fold cross-validation strategy was used to split the data into training (~75% of individuals) and validation (~25% of individuals) sets based on the relatedness of sires in each breed group. In each round, the training set contained progenies of 75% of sires, while the validation set had the progenies of the remaining sires. Pearson's correlation between predicted and measured phenotypes in the validation set was computed as the prediction accuracy. Additionally, mean squared errors (MSE) and associated standard deviations were estimated and reported.

Post-analysis on microbiability estimates and prediction accuracy

To comprehensively compare the microbiability estimates and predictive performance among models using different microbial matrices, ANOVA was performed using the PROC GLM in SAS (v9.4, SAS Institute, Carry, NC, USA). The following regression model was fitted:

$$y_{ijkl} = \mu + B_i + T_j + M_k + BT_{ij} + e_{ijkl}$$

where y_{ijkl} was the microbiability estimate or predictive accuracy of each breed/trait/matrix combination, B_i was the fixed effects of breed (3 levels: DR, LR, and LW), T_j was the fixed effects of trait (4 levels: BW, BF, LD, and IMF), M_k was the fixed effects of matrix (8 levels: LK, PK, GK, AK1, BC, JA, MDS, and DCA), BT_{ij} was the interaction effects between breed and trait, and e_{ijkl} was the residual with a distribution assumption $N(0, I\sigma^2_e)$. The least squares mean (**LSM**) was compared using Tukey's multiple comparisons test in the LSMEANS statement of the PROC GLM. An adjusted *P*-value equal or less than to 0.05 was considered significant.

Moreover, we reported the Deviance Information Criterion (**DIC**), estimated effective number of parameters (**pD**), log-likelihood evaluated at posterior mean (**LLPM**), and posterior mean of the log-likelihood (**PMLL**) to compare Bayesian models for microbiability estimation and phenotypic prediction using different microbial similarity matrices.

Results

Data summary

Descriptive statistics of growth and body composition traits of pigs are depicted by breed in Fig.1. Mean values of each trait were compared among the three breeds using one-way ANOVA with Tukey's multiple comparisons test. An adjusted *P*-value of equal or less than 0.05 was considered significant. Differences in all traits among breeds were observed (Fig. 1). The greatest IMF was observed in DR (2.07%), followed by LW (1.98%) and then LR (1.82%) pigs. The LW had greater BW and BF but lower LD than DR and LR pigs ($P < 0.05$).

The microbial composition at the family level and alpha diversity measured by the Shannon, Simpson, and Inverse Simpson indices for three breeds were reported in a previous study in our group (Bergamaschi et al., 2020b). Pearson's correlation coefficients (r) between off-diagonal elements (beta-diversity) of the microbial similarity matrices in each breed group are depicted in Fig. 2. In general, the correlations between matrices had a similar pattern across breeds. Among the matrices created by general kernel methods, the LK and AK1 showed higher correlations than other pairs, ranging from 0.63 to 0.79 in all three breeds. The strongest correlations

(r = 0.99 to 1.00) were found between two dissimilarity matrices (JA and BC), while small correlations (r = 0.09 to 0.31) were noticed between the two ordination matrices (MDS and DCA). Strong positive correlations (r = 0.60 to 0.97) were also found between LK and MDS, BC and GK, JA and GK, and MDS and AK1 matrices. The beta-diversity captured by JA and BC matrices was distinct from that by DCA, AK1, and PK with negative to small positive associations in three breeds.

Estimation of variance components and microbiability

Using all three models, we estimated the variance components and computed the proportion of host phenotypic variances due to pen, sire, or gut microbiota composition. The pen factor explained a minor proportion (7% to 25%) of variances for BW, BF, and IMF, but it accounted for 41% to 48% of variances for LD across three breeds in the Model_S_P_M (Fig. S1). Similarly, the sire factor accounted for 7% to 22% of variances across traits and breeds. The proportions of phenotypic variances due to pen and sire were similar between the Model_S_P and Model_S_P_M (Fig.3). Because the microbiome was included in the model, the proportion of phenotypic variances due to residuals was lower in the Model_S_P_M than in Model_S_P. Variations in the microbiability estimates were observed among matrices used, traits, and breed groups. The microbiability estimated by Model_S_P_M ranged from 0.06 to 0.22, 0.03 to 0.20, and 0.03 to 0.32 for BW, from 0.11 to 0.29, 0.02 to 0.24, and

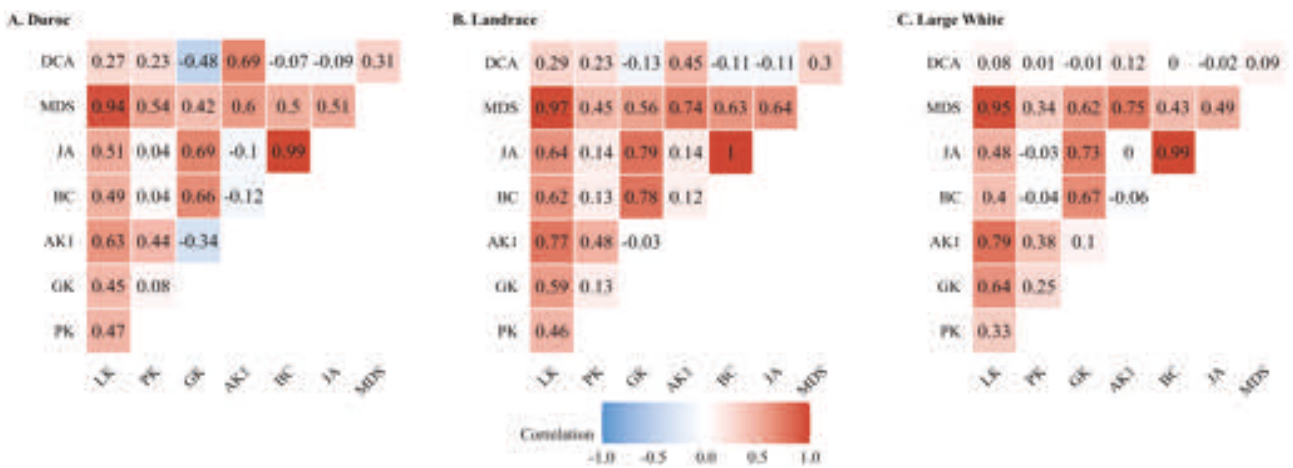

Figure 2. Pearson's correlation coefficient between the off-diagonal elements of microbial similarity matrices by breed. Legend: Colors represent negative and positive correlations. LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis.

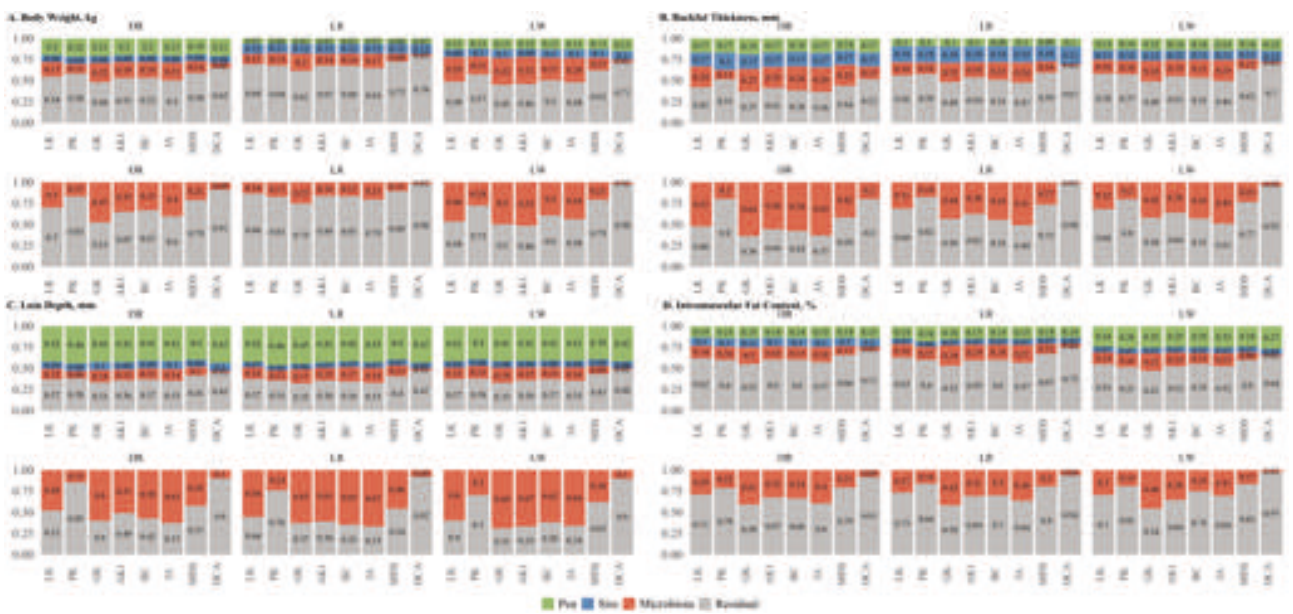

Figure 3. Proportion of total phenotypic variance explained by pen, sire, microbiota, or residual in the Model_S_P_M and Model_M for each trait by breed. Legend: Plots for Model_S_P_M are in the odd row, while for Model_M are in the even row. Colors represent the proportion of phenotypic variance explained by pen, sire, microbiota composition, or residual. The x axis represents 8 matrices. The proportion value is indicated on the y axis. DR: Duroc; LR: Landrace; LW: Large White; LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis.

0.02 to 0.25 for BF, from 0.02 to 0.14, 0.03 to 0.18, and 0.02 to 0.18 for LD, and from 0.05 to 0.20, 0.04 to 0.24, and 0.02 to 0.21 for IMF, in DR, LR, and LW pigs, respectively. Overall, the microbiability estimates obtained by Model_M were higher than those estimated by Model_S_P_M. The Model_M estimated the microbiability ranging from 0.09 to 0.47, 0.02 to 0.25, and 0.02 to 0.52 for BW, from 0.20 to 0.64, 0.02 to 0.51, and 0.05 to 0.49 for BF, from 0.10 to 0.63, 0.08 to 0.67, and 0.10 to 0.69 for LD, and from 0.09 to 0.41, 0.06 to 0.42, and 0.03 to 0.46 for IMF, in DR, LR, and LW pigs, respectively. Interestingly, the gut microbiota in the Model_M absorbed the majority of phenotypic variances that were explained by pen, sire, and microbiota together in the Model_S_P_M.

We numerically ranked the microbiability estimates for the given trait in each breed from highest to lowest values obtained using different matrices in the model, and we summarized the frequency and proportion of the rank by matrix and model in Table 2. In both models, the GK, JA, BC, and AK1 tended to obtain greater microbiability values than other matrices, while the DCA had the lowest microbiability in most cases. In addition, we performed ANOVA to statistically compare the microbiability estimates among the levels of matrix, breed, or trait. Figure 4 shows the LSM of microbiability with a 95% confidence interval and the significance of contrasts. In agreement with the results in Table 2, the microbiability estimates obtained by GK, JA, BC, and AK1 were significantly higher ($P < 0.05$) than those obtained by the remaining matrices in both models averaged over breeds and traits. The microbiability differed between breeds depending on the trait. The LW pigs had higher microbiability for BW compared to DR and LR pigs ($P < 0.0001$) in the Model_S_P_M and LR pigs ($P < 0.0001$) in the Model_M. The microbiability estimated in DR pigs was greater for BF than that in other breeds ($P < 0.0001$) in both models. When comparing across traits within each breed, the microbiability

Table 2. Frequency and proportion of the ranking for the microbiability estimates using different matrices in each model[1]

Matrix[2]	Model	Frequency (proportion, %)							
		I	II	III	IV	V	VI	VII	VIII
LK	Model_S_P_M			1 (8.3)		7 (58.3)	4 (33.3)		
PK				2 (16.7)	2 (16.7)		4 (33.3)	3 (25.0)	1 (8.3)
GK		**7 (58.3)**	**3 (25.0)**	2 (16.7)					
AK1		**1 (8.3)**	**1 (8.3)**	4 (33.3)	5 (41.7)	1 (8.3)			
BC			**2 (16.7)**	3 (25.0)	3 (25.0)	3 (25.0)	1 (8.3)		
JA		**4 (33.3)**	**6 (50.0)**		2 (16.7)				
MDS						1 (8.3)	3 (25.0)	8 (66.7)	
DCA								1 (8.3)	11 (91.7)
LK	Model_M			1 (8.3)	1 (8.3)	9 (75.0)	1 (8.3)		
PK				1 (8.3)			3 (25.0)	7 (58.3)	1 (8.3)
GK		**7 (58.3)**	**2 (16.7)**	3 (25.0)					
AK1		**1 (8.3)**	**2 (16.7)**	2 (16.7)	6 (50.0)	1 (8.3)			
BC			**3 (25.0)**	3 (25.0)	4 (33.3)	2 (16.7)			
JA		**4 (33.3)**	**5 (41.7)**	3 (16.7)	1 (8.3)				
MDS							8 (66.7)	4 (33.3)	
DCA								1 (8.3)	11 (91.7)

[1]The microbiability estimates were numerically ranked from the highest to lowest values for each trait within each breed group. The sum of the frequency for each line is 12 due to four traits in 3 breeds. The Roman numbers represent the rank. The Arabic numbers represent the frequency and proportion for each rank. The top 4 best matrices are highlighted in bold.
[2]LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis.

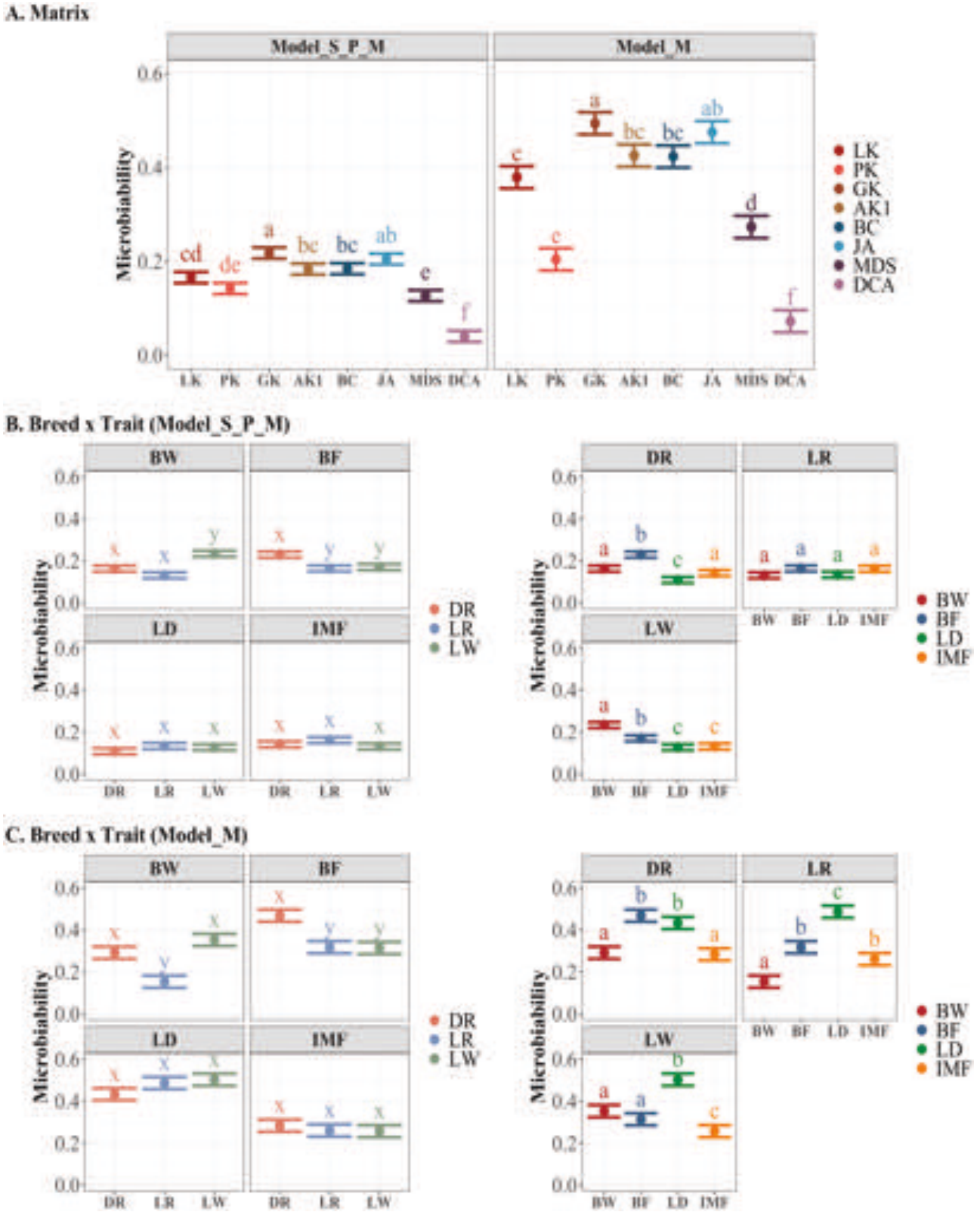

Figure 4. Least squares means of microbiability with 95% confidence interval and contrasts among levels of factors interested. Different letters denote $P < 0.05$ for each level of the factor interested. (A) LSM of microbiability and contrasts among eight matrices used in the Model_S_P_M and Model_M. LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis. (B) LSM of microbiability and contrasts among three breeds for Model_S_P_M and Model_M. DR: Duroc; LR: Landrace; LW: Large White. C. LSM of microbiability and contrasts among four traits for Model_S_P_M and Model_M. B. BW: body weight; BF: backfat thickness; LD: loin depth; IMF: intramuscular fat.

estimates for BF in DR pigs and BW and BF in LW pigs were higher ($P < 0.05$) than that for the rest traits in the Model_S_P_M, whereas they were higher ($P < 0.0001$) for BF and LD in DR pigs and LD in LR and LW pigs than that for the rest traits in the Model_M. The disparity in the contrasts of microbiability among traits within each breed between the two models may be due to the microbiota in Model_M accounting for the majority of the phenotypic variations explained by pen and sire in the Model S_P_M, as described in the preceding paragraph.

Table S1 summarizes several criteria for model fit using various matrices for variance component estimation. The GK, AK1, JA, and BC had a better fit in both models, with smaller DIC and higher LLPM, PMLL, and pD values than the remaining matrices. The Model_S_P_M had smaller DIC values compared to the Model_M because more variables were fitted in the model. Interestingly, the Model_S_P_M had smaller pD values than the Model_M when using GK, AK1, JA, or BC matrix in the model.

Evaluation of predictive performance

Figure 5 illustrates the prediction accuracy for each trait by breed using the Model_M with one of the eight matrices. Each boxplot summarizes four accuracy values obtained by the four-fold cross-validation. For the given trait in each breed group, there were differences in prediction accuracy between matrices used. The PK and DCA had lower accuracy than the other matrices in most predictions, except for predicting the BW in LR pigs and IMF in DR and LR pigs. The Model_M was more accurate in predicting BW for DR and LW than LR pigs and BF for DR pigs than the rest breeds, while had similar accuracies in prediction LD and IMF among the three breeds. Interestingly, a greater standard deviation of prediction accuracy across four-folds was observed in DR pigs for predicting LD and IMF compared to the other breeds.

Table 3 summarizes the mean and standard deviation of MSE over four folds in each breed group. Matrices with greater prediction accuracy also had lower MSE estimates. In addition, the MSE estimates for the given trait differed among the three breeds. In contrast to the prediction accuracy, the Model_M had smaller MSE in LR pigs compared to DR and LW pigs, regardless of the matrix used to predict BW. Furthermore, the DR pigs had smaller MSE when predicting BF and LD but larger when predicting IMF, compared to the other two breeds. We also numerically ranked the prediction accuracy from greatest to lowest values and MSE in the opposite direction. Table 4 shows the frequency and proportion distribution of the rank across the eight matrices used in the predictive model. The matrices ranked similarly in terms of accuracy and MSE. The BC, MDS, LK, and AK1 matrices performed relatively better than the other matrices in the prediction.

To evaluate the factors in the experimental design, we performed a post-analysis on the prediction accuracy. Figure 6 depicts the LSM of prediction accuracy with a 95% confidence interval and the significance of contrasts among levels of interest factors. Overall, the LSM of prediction accuracy ranged from 0.12 to 0.28 across matrices used averaged over traits and breeds (Fig. 6A). The PK and DCA matrices in the model got significantly lower accuracy than the other matrices in the prediction ($P < 0.05$). No significant difference in the accuracy was observed among the models using LK, GK, AK1, BC, JA, or MDS matrix ($P > 0.05$). Variations in prediction accuracy among breeds highly depended on the trait predicted. The gut microbiota performed better in predicting BW for DR and LW pigs ($P < 0.0001$) and BF for DR pigs (P

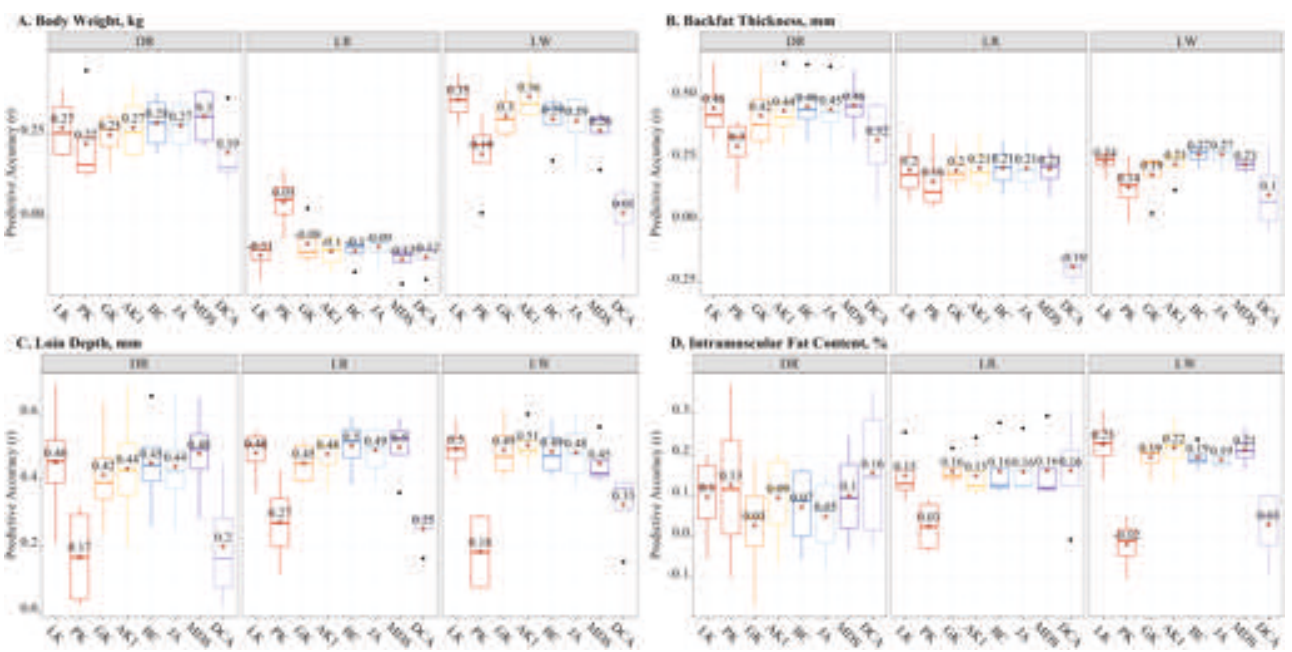

Figure 5. Prediction accuracy obtained by Model_M with one of 8 matrices for each trait by breed. Legend: Each boxplot was created from the accuracy values of 4-fold cross-validation in the prediction. Star sign with the label is the mean of prediction accuracy over four values from the cross-validation for each matrix/trait/breed combination. Colors represent different matrices used in the model. DR: Duroc; LR: Landrace; LW: Large White; LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis.

Table 3. Mean-squared error of prediction using the Model_M with 8 microbial matrices[1]

Trait[2]	Breed[3]	Matrix[4]							
		LK	PK	GK	AK1	BC	JA	MDS	DCA
BW, kg	DR	125.32 (34.99)	131.24 (37.79)	125.85 (38.01)	125.01 (35.02)	124.20 (35.58)	124.82 (36.34)	**123.7 (35.03)**	128.54 (33.44)
	LR	112.90 (26.71)	**107.68 (25.32)**	110.55 (24.61)	112.13 (25.76)	112.59 (26.59)	111.67 (26.23)	114.92 (27.90)	113.66 (28.61)
	LW	**133.01 (45.47)**	145.95 (41.39)	138.98 (43.43)	133.53 (44.48)	138.36 (45.98)	139.22 (45.78)	140.45 (47.36)	148.23 (45.91)
BF, mm	DR	5.21 (0.77)	6.57 (1.59)	5.61 (1.13)	5.32 (0.87)	5.18 (0.91)	5.31 (0.95)	**5.07 (0.93)**	5.80 (1.25)
	LR	8.34 (1.08)	8.73 (1.39)	8.42 (1.33)	8.32 (1.51)	8.24 (1.07)	8.29 (1.16)	**8.23 (1.04)**	9.23 (1.43)
	LW	10.33 (3.44)	11.11 (3.73)	10.66 (3.43)	10.39 (3.39)	**10.20 (3.39)**	10.21 (3.41)	10.39 (3.45)	11.10 (3.92)
LD, mm	DR	33.42 (10.83)	42.45 (17.41)	36.43 (14.17)	34.53 (11.75)	33.57 (11.27)	34.49 (12.23)	**31.73 (9.83)**	41.17 (14.42)
	LR	34.45 (6.18)	43.21 (10.29)	37.63 (8.66)	34.95 (6.65)	33.77 (5.78)	34.97 (6.93)	**33.27 (4.82)**	42.87 (10.67)
	LW	37.92 (14.70)	47.54 (14.14)	40.27 (15.49)	**37.90 (14.64)**	39.03 (15.27)	39.68 (15.91)	39.91 (14.39)	45.90 (19.23)
IMF, %	DR	**0.79 (0.35)**	0.80 (0.37)	0.81 (0.37)	**0.79 (0.35)**	0.81 (0.36)	0.81 (0.36)	0.80 (0.34)	0.80 (0.35)
	LR	0.57 (0.11)	0.58 (0.13)	**0.56 (0.12)**	0.57 (0.11)	0.57 (0.11)	0.57 (0.12)	0.57 (0.11)	0.57 (0.09)
	LW	**0.57 (0.11)**	0.61 (0.13)	0.59 (0.13)	0.58 (0.11)	0.58 (0.12)	0.58 (0.13)	0.58 (0.12)	0.62 (0.14)

[1]Data are presented as the mean (SD) over 4 folds of the cross-validation for each trait/breed/matrix combination. Least MSE mean across the 8 matrices used is in bold for each combination.
[2]BW: body weight (kg); BF: backfat thickness (mm); LD: loin depth (mm); IMF: intramuscular fat content (%).
[3]DR: Duroc; LR: Landrace; LW: Large White.
[4]LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis.

< 0.0001), with no differences among the three breeds when predicting LD and IMF (Fig. 6B). Within each breed group, the gut microbiota better informed the model to predict BF (r = 0.42; P < 0.0001) in DR pigs and LD in all three breeds (DR: r = 0.38; LR: r = 0.43; LW: r = 0.43; P < 0.05) compared with the remaining traits (Fig. 6C). Table S2 summarizes several parameters for model comparisons. The Model_M with the GK, JA, BC, or AK1 matrix had a lower DIC and higher LLPM, PMLL, and pD than the other matrices. The GK matrix in the model achieved the best fit, with the lowest DIC and the highest pD.

Discussion

Various metrics have been investigated for their ability to measure the similarity or difference in microbiota composition among individuals (Barwell et al., 2015; Lin et al., 2015; Yang et al., 2021). To the extent of our knowledge, this is the first study to evaluate the use of different methods measuring the variations in gut microbiota among animals in estimating microbiability and predicting host phenotypes in swine.

We observed variations in the beta-diversity measured by different approaches through calculating the correlation between the off-diagonal elements of pairwise matrices. We found correlations close to unity between the JA and BC matrices in all breeds. Saborío-Montero et al. reported similar results based on the ruminal microbiota in dairy cattle (Saborío-Montero et al., 2021). These two metrices have become popular approaches to comprehensively analyze microbiome data (Li et al., 2016; Knowles et al., 2019; Koh et al., 2019). The Bray-Curtis dissimilarity is a common abundance-based metric that quantifies variation in microbiota composition among samples by taking into account both individual taxa abundance and total abundance of a sample (Bray and Curtis, 1957; Greenacre and Primicerio, 2014; Barwell et al., 2015). On the other hand, the classical Jaccard dissimilarity measures the dissimilarity of microbiota composition among samples based on the presence-absence of species in pairwise samples and is thus more sensitive to the presence of rare species (Jaccard, 1912; Greenacre and Primicerio, 2014). In our study, the beta-diversity among animals captured by the JA or BC matrices was different from that measured by the PK, AK1, or DCA method. This is because components of the microbiome data are evaluated differently when calculating the beta-diversity between pairs of samples using different approaches (Tuomisto, 2010a, 2010b).

Using three models in each breed group, we estimated the proportion of host phenotypic variances accounted by pen, sire, or gut microbiota. Regardless of the matrix used, adding gut microbiota to the model explained more phenotypic variances for all traits than the baseline model with only pen and sire. Moreover, the total proportions of phenotypic variances absorbed by the effects were comparable between the model with pen, sire, and microbiota

Table 4. Frequency and proportion of the ranking for the prediction accuracy and MSE among different matrices used in the Model_M[1]

Prediction performance	Matrix[2]	Frequency (proportion, %)							
		I	II	III	IV	V	VI	VII	VIII
Accuracy	**LK**	**1 (8.3)**	**3 (25.0)**	2 (16.7)	2 (16.7)	2 (16.7)	2 (16.7)		
	PK	1 (8.3)	1 (8.3)					5 (41.7)	5 (41.7)
	GK		1 (8.3)	2 (16.7)	2 (16.7)		6 (50.0)		1 (8.3)
	AK1	**2 (16.7)**	**2 (16.7)**		2 (16.7)	5 (41.7)		1 (8.3)	
	BC	**2 (16.7)**	**3 (25.0)**	2 (16.7)	3 (25.0)	1 (8.3)	1 (8.3)		
	JA	1 (8.3)		4 (33.3)	2 (16.7)	3 (25.0)	1 (8.3)	1 (8.3)	
	MDS	**4 (33.3)**	**1 (8.3)**	2 (16.7)	1 (8.3)	1 (8.3)	2 (16.7)		1 (8.3)
	DCA	1 (8.3)	1 (8.3)					5 (41.7)	5 (41.7)
MSE	**LK**	**3 (25.0)**	**2 (16.7)**	3 (25.0)		2 (16.7)	2 (16.7)		
	PK	1 (8.3)				1 (8.3)		3 (25.0)	7 (58.3)
	GK	1 (8.3)	1 (8.3)		1 (8.3)		8 (66.7)		1 (8.3)
	AK1	**1 (8.3)**	**3 (25.0)**		5 (41.7)	3 (25.0)			
	BC	**1 (8.3)**	**4 (33.3)**	4 (33.3)	1 (8.3)	1 (8.3)	1 (8.3)		
	JA		2 (16.7)	4 (33.3)	3 (25.0)	2 (16.7)		1 (8.3)	
	MDS	**5 (41.7)**		1 (8.3)	2 (16.7)	2 (16.7)	1 (8.3)		1 (8.3)
	DCA			1 (8.3)				8 (66.7)	3 (25.0)

[1]Prediction accuracies were numerically ranked from the highest to lowest values, while the MSE estimates were numerically ranked from the lowest to the highest values, for each trait within each breed group. The sum of the frequency for each line is 12 due to 4 traits in 3 breeds. The Roman numbers represent the rank. The Arabic numbers represent the frequency and proportion for each rank. The top 4 best matrices are highlighted in bold.
[2]LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis.

and the model with only microbiota. Numerous studies in humans have shown that the gut microbiota compositions of individuals who share the same environment or are genetically related are similar (Spor et al., 2011). In pigs, the gut microbiota community is under direct control by the host's genetics, with small to moderate heritabilities (Lu et al., 2018; Bergamaschi et al., 2020a). Maltecca et al. (2021) further confirmed this by finding that sires significantly influenced their offspring's gut microbiota composition. In terms of environmental influence, Hildebrand and colleagues (2013) found that the microenvironment accounted for approximately 30% of the variance in mouse gut microbiota and that mice kept in the same cage exhibited similar gut microbiota. Our findings show that the gut microbiota may capture some variance in the environment and host genetics and could be a valuable source of information for understanding and improving growth and body composition traits in swine, particularly when such systematic information is unavailable. Estimates of microbiability were higher for GK and JA, along with better model fit compared to other methods in both Model_S_P_M and Model_M. Interestingly, the Model_S_P_M with GK, AK1, JA, and BC matrices had lower DIC values than Model_M with the same matrices without increasing model complexity as measured by pD. The DIC, which accounts for both model fit and complexity, is a criterion for model selection in Bayesian statistics (Spiegelhalter et al., 2002). However, due to the growing debate over the use of DIC in model selection (Plummer, 2008), more research is required to validate our findings.

To evaluate the predictive performance of different microbial similarity matrices, we used the microbiome as the only predictor in the Model_M with cross-validation based on sire relatedness. The results were highly dependent on the trait and breed used in the prediction. In most cases, the models with LK, GK, AK1, BC, JA, or MDS matrix performed better than those with PK and DCA matrix. In the current study, we used log-transformed count data as input to generate

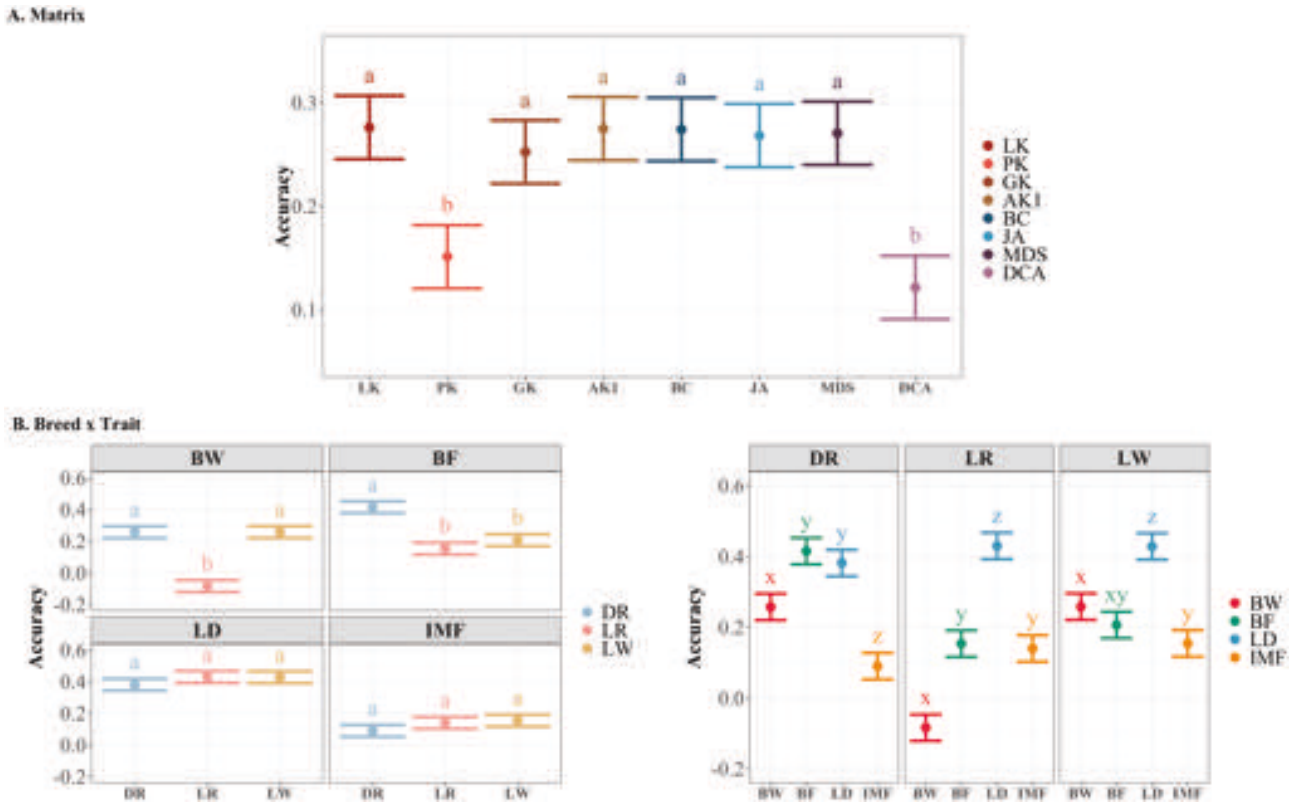

Figure 6. Least squares means of prediction accuracy with 95% confidence interval and contrasts among levels of factors interested in the Model_M. Legend: Different letters denote $P < 0.05$ for each level of the factor interested. (A) LSM of prediction accuracy and contrasts among eight matrices. LK: linear kernel; PK: polynomial kernel; GK: gaussian kernel; AK1: arc-cosine kernel with one hidden layer; BC: Bray-Curtis; JA: Jaccard; MDS: metric multidimensional scaling; DCA: detrended correspondence analysis. (B) LSM of prediction accuracy and contrasts for each trait among three breeds. DR: Duroc; LR: Landrace; LW: Large White. (C) LSM of prediction accuracy and contrasts for each breed among 4 traits. (B) BW: body weight; BF: backfat thickness; LD: loin depth; IMF: intramuscular fat.

various microbial similarity matrices for prediction. Our findings on using LK and PK matrices for phenotype prediction agree with those of Carrieri and colleagues (Carrieri et al., 2016). They looked at the effects of different normalization methods on microbiome count data for predicting host phenotypes from various kernel matrices, including linear and polynomial kernels (Carrieri et al., 2016). Across all datasets studied, the LK outperformed the PK in terms of predictive performance when the input count data were log-transformed (Carrieri et al., 2016). In contrast, the PK could outperform the LK when other data normalization methods were used (Carrieri et al., 2016). We believe that if an appropriate data normalization method for each microbial similarity matrix could be identified, the results would be more robust, but this needs to be investigated further. According to our post-analysis, there was no significant difference in prediction accuracy between models with the LK, GK, AK1, BC, JA, or MDS matrix. The GK, AK1, BC, and JA matrix, on the other hand, outperformed the baseline model with the LK matrix in terms of model fit in the prediction.

Our results showed that the microbiability and predictive ability of gut microbiota composition varied across three breeds for body weight and backfat thickness. Differences in microbiota composition between breeds may contribute to this variation for the given trait. According to our previous study using a related dataset (Bergamaschi et al., 2020b), pigs of different breeds exhibited distinct patterns in the gut microbiota composition in terms of overall diversity and individual taxa. The gut microbiota composition of Duroc pigs was found to be different from that of Landrace and Large White/Yorkshire pigs among the three breeds studied (Alain Pajarillo et al., 2014; Bergamaschi et al., 2020b). As typical terminal sires used in pork production, Duroc pigs have been extensively selected for lean or fat growth (Lonergan et al., 2001). Based on the results of the post-analysis, we found that the microbiota composition explained a greater amount of the phenotypic variance and achieved a higher prediction accuracy for backfat thickness in Duroc pigs than in other breeds. These findings suggest that the gut microbiota composition may carry more valuable information to understand fat deposition in Duroc pigs than in other populations and thus, could be used to guide the selection for related traits. Future research will focus on partitioning microbiability on individual microbes at various taxonomic levels to identify specific taxa that hold primary responsibility for explaining phenotypic variance.

Conclusions

We investigated alternative approaches for quantifying resemblance between animals based on their gut microbiota composition and assessed the use of this information in estimating variance components and predicting host growth and body composition traits in three western breeds. We demonstrated that eight approaches differently characterized the microbial similarity between animals. We observed

variation in the microbiability obtained by different matrices for the given trait. Furthermore, we demonstrated that the gut microbiota could be a valuable source of information for understanding and improving complex traits in swine, particularly when environmental and family information is limited. The prediction accuracy varied across models with different matrices, traits, and breeds. Except for the matrices built using the polynomial kernel and detrended correspondence analysis methods, the remaining matrices achieved comparable prediction accuracy for the given trait in each breed. Models with the matrix constructed by the gaussian kernel, arc-cosine kernel with one hidden layer, Bray-Curtis, and Jaccard methods had a better fit compared those that used matrices built with other methods. The current study shows alternative methods for summarizing microbiota composition between animals and incorporating this information into prediction models, emphasizing the importance of considering genetic background in swine microbiome research.

Conflicts of interest

The authors declare that they have no actual or potential conflicts of interest.

Funding

This study is a part of the project "The role of feeding behavior on gut microbiome composition and its implications for lean growth selection in pigs" funded by The North Carolina Pork Council 2017-1929 grant.

Author's contributions

YH performed the statistical analyses, as well as interpreted the results, and drafted the manuscript. CM and FT designed the experiment and helped draft the manuscript. JH, YH, and KG supervised and performed phenotypic data collection and editing. JJ and JWC contributed to interpretation of the results. All authors contributed to the paper proofreading and approved the final manuscript.

Acknowledgements

The authors acknowledge Smithfield Premium Genetics (Rose Hill, NC, USA) for providing the data used in this study. We also extend our gratitude to Nathan P. McNulty and Constantino Schillebeeckx from Matatu Inc. for their help in sequencing analyses and to Jessica Hoisington-Lopez and MariaLynn Crosby from the DNA Sequencing Innovation Lab at the Center for Genome Sciences and Systems Biology at Washington University in St. Louis for their help in sequencing.

LITERATURE CITED

Alain Pajarillo, E. B., J. Pyo Chae, M. P. Balolong, H. Bum Kim, K. -S. Seo, and D. -K. Kang. 2014. Pyrosequencing-based analysis of fecal microbial communities in three purebred pig lines §. *J. Microbiol.* 52:646–651. doi:10.1007/s12275-014-4270-2.

Barwell, L. J., N. J. B. Isaac, and W. E. Kunin. 2015. Measuring β-diversity with species abundance data. *J. Anim. Ecol.* 84:1112–1122. doi:10.1111/1365-2656.12362 . https://onlinelibrary.wiley.com/doi/full/10.1111/1365-2656.12362.

Bergamaschi, M., C. Maltecca, C. Schillebeeckx, N. P. McNulty, C. Schwab, C. Shull, J. Fix, and F. Tiezzi. 2020a. Heritability and genome-wide association of swine gut microbiome features with growth and fatness parameters. *Sci. Rep.* 10:1–12. doi:10.1038/s41598-020-66791-3 .

Bergamaschi, M., F. Tiezzi, J. Howard, Y. J. Huang, K. A. Gray, C. Schillebeeckx, N. P. McNulty, and C. Maltecca. 2020b. Gut microbiome composition differences among breeds impact feed efficiency in swine. *Microbiome* 8:110. doi:10.1186/s40168-020-00888-9. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-020-00888-9.

Bray, J. R., and J. T. Curtis. 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27:325–349. doi:10.2307/1942268.

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

Carrieri, A. P., N. Haiminen, and L. Parida. 2016. Host phenotype prediction from differentially abundant microbes using RoDEO. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 10477 LNBI:27–41. doi:10.1007/978-3-319-67834-4_3. Available from: https://link.springer.com/chapter/10.1007/978-3-319-67834-4_3.

Coenen, A. R., S. K. Hu, E. Luo, D. Muratore, and J. S. Weitz. 2020. A primer for microbiome time-series analysis. *Front. Genet.* 0:310. doi:10.3389/FGENE.2020.00310.

Cryan, J. F., K. J. O'riordan, C. S. M. Cowan, K. V. Sandhu, T. F. S. Bastiaanssen, M. Boehme, M. G. Codagnone, S. Cussotto, C. Fulling, A. V. Golubeva, et al. 2019. The microbiota-gut-brain axis. *Physiol. Rev.* 99:1877–2013. doi:10.1152/physrev.00018.2018. Available from: https://pubmed.ncbi.nlm.nih.gov/31460832/.

Cuevas, J., O. Montesinos-López, P. Juliana, C. Guzmán, P. Pérez-Rodríguez, J. González-Bucio, J. Burgueño, A. Montesinos-López, and J. Crossa. 2019. Deep kernel for genomic and near infrared predictions in multi-environment breeding trials. *G3 Genes|Genomes|Genetics* 9:2913–2924. doi:10.1534/g3.119.400493.

Difford, G., J. Lassen, and P. Løvendahl. 2016. Genes and microbes, the next step in dairy cattle breeding. In: Book of Abstracts of the 67th Annual Meeting of the European Federation of Animal Science. p. 285.

Difford, G. F., D. R. Plichta, P. Løvendahl, J. Lassen, S. J. Noel, O. Højberg, A. G. Wright, Z. Zhu, L. Kristensen, H. B. Nielsen, et al. 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. T. Leeb, editor. *PLoS Genet.* 14:e1007580. doi:10.1371/journal.pgen.1007580. Available from: https://dx.plos.org/10.1371/journal.pgen.1007580.

Greenacre, M., and R. Primicerio. 2014. *Multivariate analysis of ecological data*. Fundacion BBVA.

He, Y., F. Tiezzi, J. Howard, Y. Huang, K. Gray, and C. Maltecca. 2022. Exploring the role of gut microbiota in host feeding behavior among breeds in swine. *BMC Microbiol.* 22:1. doi:10.1186/s12866-021-02409-6.

Hildebrand, F., T. L. A. Nguyen, B. Brinkman, R. G. Yunta, B. Cauwe, P. Vandenabeele, A. Liston, and J. Raes. 2013. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol.* 14:R4. doi:10.1186/gb-2013-14-1-r4.

Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis: an improved ordination technique. *Classif. Ordination* 2:47–58. doi:10.1007/978-94-009-9197-2_7. Available from: https://link.springer.com/chapter/10.1007/978-94-009-9197-2_7.

Jaccard, P. 1912. The distribution of the flora in the alpine zone.1. *New Phytol.* 11:37–50. doi:10.1111/j.1469-8137.1912.tb05611.x. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.1912.tb05611.x.

Jiang, Y., and J. C. Reif. 2015. Modeling epistasis in genomic selection. *Genetics* 201:759–768. doi:10.1534/genetics.115.177907.

Khanal, P., C. Maltecca, C. Schwab, J. Fix, M. Bergamaschi, and F. Tiezzi. 2020. Modeling host-microbiome interactions for the prediction of meat quality and carcass composition traits in swine. *Genet. Sel. Evol.* 52:41. doi:10.1186/s12711-020-00561-7.

Khanal, P., C. Maltecca, C. Schwab, J. Fix, and F. Tiezzi. 2021. Microbiability of meat quality and carcass composition traits in swine. *J. Anim. Breed. Genet.* 138:223–236. doi:10.1111/jbg.12504. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/jbg.12504.

Knowles, S. C. L., R. M. Eccles, and L. Baltrūnaitė. 2019. Species identity dominates over environment in shaping the microbiota of small mammals. *Ecol. Lett.* 22:826–837. doi:10.1111/ele.13240. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/ele.13240.

Koh, H., Y. Li, X. Zhan, J. Chen, and N. Zhao. 2019. A distance-based kernel association test based on the generalized linear mixed model for correlated microbiome studies. *Front. Genet.* 10:458. doi:10.3389/fgene.2019.00458.

Larsen, P. E., and Y. Dai. 2015. Metabolome of human gut microbiome is predictive of host dysbiosis. *GigaScience* 4:s13742-015-0084–3. doi:10.1186/s13742-015-0084-3.

Li, H., T. Li, D. A. E. Beasley, P. Heděnec, Z. Xiao, S. Zhang, J. Li, Q. Lin, and X. Li. 2016. Diet diversity is associated with beta but not alpha diversity of pika gut microbiota. *Front. Microbiol.* 7:1169. doi:10.3389/fmicb.2016.01169.

Li, Y., W. Xiaofan, W. Xiu-qi, J. Wang, and J. Zhao. 2020. Life-long dynamics of the swine gut microbiome and their implications in probiotics development and food safety. *Gut Microbes* 11:1824–1832. doi:10.1080/19490976.2020.1773748.

Lin, S. W., N. D. Freedman, J. Shi, M. H. Gail, E. Vogtmann, G. Yu, V. Klepac-Ceraj, B. J. Paster, B. A. Dye, G. Q. Wang, et al. 2015. Beta-diversity metrics of the upper digestive tract microbiome are associated with body mass index. *Obesity* 23:862–869. doi:10.1002/OBY.21020. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/oby.21020.

Lonergan, S. M., E. Huff-Lonergan, L. J. Rowe, D. L. Kuhlers, and S. B. Jungst. 2001. Selection for lean growth efficiency in Duroc pigs influences pork quality. *J. Anim. Sci.* 79:2075–2085. doi:10.2527/2001.7982075x.

Lu, D., F. Tiezzi, C. Schillebeeckx, N. P. McNulty, C. Schwab, C. Shull, and C. Maltecca. 2018. Host contributes to longitudinal diversity of fecal microbiota in swine selected for lean growth. *Microbiome* 6:4. doi:10.1186/s40168-017-0384-1. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-017-0384-1.

Maltecca, C., D. Lu, C. Schillebeeckx, N. P. McNulty, C. Schwab, C. Shull, and F. Tiezzi. 2019. Predicting growth and carcass traits in swine using microbiome data and machine learning algorithms. *Sci. Rep.* 9:6574. doi:10.1038/s41598-019-43031-x.

Maltecca, C., R. Dunn, Y. He, N. P. McNulty, C. Schillebeeckx, C. Schwab, C. Shull, J. Fix, and F. Tiezzi. 2021. Microbial composition differs between production systems and is associated with growth performance and carcass quality in pigs. *Anim. Microbiome* 3:57. doi:10.1186/s42523-021-00118-z.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi:10.1371/JOURNAL.PONE.0061217. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0061217.

Montesinos-López, A., O. A. Montesinos-Lóópez, J. C. Montesinos-López, C. A. Flores-Cortes, R. de la Rosa, and J. Crossa. 2021. A guide for kernel generalized regression methods for genomic-enabled prediction. *Heredity (Edinb)* 126:577–596. doi:10.1038/s41437-021-00412-1.

Nasidze, I., J. Li, D. Quinque, K. Tang, and M. Stoneking. 2009. Global diversity in the human salivary microbiome. *Genome Res.* 19:636–643. doi:10.1101/gr.084616.108. Available from: https://genome.cshlp.org/content/19/4/636.full.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, et al. 2020. vegan: Community Ecology Package. R package version 2.5–7. Available from: https://CRAN.R-project.org/package=vegan.

Pérez, P., and G. De Los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495. doi:10.1534/genetics.114.164442. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4196607/?report=abstract.

Plummer, M. 2008. Penalized loss functions for Bayesian model comparison. *Biostatistics* 9:523–539. doi:10.1093/BIOSTATISTICS/KXM049.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. Available from: http://cran.r-project.org/doc/Rnews/Rnews_2006-1.pdf#page=7

Podani, J., and I. Miklós. 2002. Resemblance coefficients and the horseshoe effect in principal coordinates analysis. *Ecology* 83:3331–3343. doi:10.1890/0012-9658(2002)083[3331:RCATHE]2.0.CO;2.

Prodan, A., V. Tremaroli, H. Brolin, A. H. Zwinderman, M. Nieuwdorp, and E. Levin. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One* 15:e0227434. doi:10.1371/JOURNAL.PONE.0227434. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0227434.

Ross, E. M., P. J. Moate, L. C. Marett, B. G. Cocks, and B. J. Hayes. 2013. Metagenomic predictions: from microbiome to complex health and environmental phenotypes in humans and cattle. *PLoS One* 8:e73056. doi:10.1371/JOURNAL.PONE.0073056. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073056.

Saborío-Montero, A., M. Gutiérrez-Rivas, A. López-García, A. García-Rodríguez, R. Atxaerandio, I. Goiri, J. A. Jiménez-Montero, and O. González-Recio. 2021. Holobiont effect accounts for more methane emission variance than the additive and microbiome effects on dairy cattle. *Livest. Sci* 250:104538. doi:10.1016/J.LIVSCI.2021.104538.

Savage, D. C. 1977. Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* 31:107–133. doi:10.1146/annurev.mi.31.100177.000543.

Schaid, D. J. 2010. Genomic similarity and kernel methods II: methods for genomic information. *Hum. Hered.* 70:132–140. doi:10.1159/000312643. Available from: https://www.karger.com/Article/FullText/312643.

Souza, L. M., F. R. Francisco, P. S. Gonçalves, E. J. Scaloppi Junior, V. Le Guen, R. Fritsche-Neto, and A. P. Souza. 2019. Genomic selection in rubber tree breeding: a comparison of models and methods for managing G×E interactions. *Front. Plant Sci.* 10:1353. doi:10.3389/fpls.2019.01353. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2019.01353.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B (Statistical Methodology)* 64:583–639. doi:10.1111/1467-9868.00353. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9868.00353.

Spor, A., O. Koren, and R. Ley. 2011. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.* 9:279–290. doi:10.1038/nrmicro2540.

Team, R. C. 2021. *R: a language and environment for statistical computing*. Available from: https://www.r-project.org/

Tuomisto, H. 2010a. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography (Cop.).* 33:2–22. doi:10.1111/J.1600-0587.2009.05880.X. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0587.2009.05880.x.

Tuomisto, H. 2010b. A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. . *Ecography (Cop.)* 33:23–45. doi:10.1111/J.1600-0587.2009.06148.X. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0587.2009.06148.x.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. 2007. The human microbiome project. *Nature* 449:804–810. doi:10.1038/nature06244.

Verschuren, L. M. G., D. Schokker, R. Bergsma, A. J. M. Jansman, F. Molist, and M. P. L. Calus. 2020. Prediction of nutrient digestibility in grower-finisher pigs based on faecal microbiota composition. *J. Anim. Breed. Genet*. 137:23–35. doi:10.1111/JBG.12433. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/jbg.12433.

Wang, X., T. Tsai, F. Deng, X. Wei, J. Chai, J. Knapp, J. Apple, C. V. Maxwell, J. A. Lee, Y. Li, et al. 2019. Longitudinal investigation of the swine gut microbiome from birth to market reveals stage and growth performance associated bacteria. *Microbiome* 7:109. doi:10.1186/s40168-019-0721-7. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0721-7.

Xia, Y., J. Sun, and D.-G. Chen. 2018. *What are microbiome data? In: Statistical Analysis of Microbiome Data with R. ICSA Book Series in Statistics*. Singapore: Springer; p. 29–41. doi:10.1007/978-981-13-1534-3_2.

Yang, Z., F. Xu, H. Li, and Y. He. 2021. Beyond samples: a metric revealing more connections of gut microbiota between individuals. *Comput. Struct. Biotechnol. J*. 19:3930. doi:10.1016/J.CSBJ.2021.07.009. Available from: https://www.ncbi.nlm.nih.gov//pmc/articles/PMC8319210.

Zhang, Z., and Y. Takane. 2010. Multidimensional Scaling. In: Peterson, P., E. Baker, and B. McGaw, editors. *International Encyclopedia of Education*. 3rd ed. Elsevier; p. 304–311. doi:10.1016/B978-0-08-044894-7.01348-8.