



## Article

# Estimation of Complex-Trait Prediction Accuracy from the Different Holo-Omics Interaction Models

Qamar Raza Qadri <sup>1</sup>, Qingbo Zhao <sup>1</sup>, Xueshuang Lai <sup>1</sup>, Zhenyang Zhang <sup>2</sup>, Wei Zhao <sup>1</sup>, Yuchun Pan <sup>2,3</sup>  
and Qishan Wang <sup>2,3,4,\*</sup>

<sup>1</sup> School of Agriculture and Biology, Department of Animal Science, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup> Department of Animal Breeding and Reproduction, College of Animal Science, Zhejiang University, Hangzhou 310030, China

<sup>3</sup> Hainan Institute, Zhejiang University, Yongyou Industry Park, Yazhou Bay Sci-Tech City, Sanya 572000, China

<sup>4</sup> Zhejiang Key Laboratory of Dairy Cattle Genetic Improvement and Milk Quality Research, Hangzhou 310030, China

\* Correspondence: wangqishan@zju.edu.cn

**Abstract:** Statistical models play a significant role in designing competent breeding programs related to complex traits. Recently; the holo-omics framework has been productively utilized in trait prediction; but it contains many complexities. Therefore; it is desirable to establish prediction accuracy while combining the host's genome and microbiome data. Several methods can be used to combine the two data in the model and study their effectiveness by estimating the prediction accuracy. We validate our holo-omics interaction models with analysis from two publicly available datasets and compare them with genomic and microbiome prediction models. We illustrate that the holo-omics interactive models achieved the highest prediction accuracy in ten out of eleven traits. In particular; the holo-omics interaction matrix estimated using the Hadamard product displayed the highest accuracy in nine out of eleven traits, with the direct holo-omics model and microbiome model showing the highest prediction accuracy in the remaining two traits. We conclude that comparing prediction accuracy in different traits using real data showed important intuitions into the holo-omics architecture of complex traits.

**Keywords:** holo-omics; model selection; complex trait; prediction accuracy; breeding program; random effect



**Citation:** Qadri, Q.R.; Zhao, Q.; Lai, X.; Zhang, Z.; Zhao, W.; Pan, Y.; Wang, Q. Estimation of Complex-Trait Prediction Accuracy from the Different Holo-Omics Interaction Models. *Genes* **2022**, *13*, 1580. <https://doi.org/10.3390/genes13091580>

Academic Editors: Ruihua Huang, Pinghua Li and Liming Hou

Received: 3 August 2022

Accepted: 30 August 2022

Published: 2 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Genomic prediction is used in various fields and has become an essential tool for the animal breeding program [1]. Due to high throughput sequencing (HTS) technologies, the analysis of genetic data from host and microbiome, altogether termed holo-omics, has led to a growing interest in selecting traits in livestock and studying complex phenotypes. Over the past few years, many methodologies have been designed using the hologenome domains to study nutrient acquisition [2], immune response [3], and body development [4]. It has helped to understand the interconnection and coregulation systems that shape host phenotypes and critical functional pathways [5]. The genome-wide SNP data is widely used to construct an SNP-based matrix termed the genomic relationship matrix (GRM) and predict the accuracy of genomic selection [6].

Similarly, the sequencing data from 16S rRNA has been used to construct a microbial composition-based matrix known as the microbial relationship matrix (MRM), and it is based on relative abundance and an operational taxonomic unit (OTU) [7]. The MRM and cluster analysis of large-scale microbiome data has allowed for research and can predict complex traits, mainly related to feed efficiency and methane emission in animals [8–10]. Additionally, microbiome diversity has been linked to various human traits with a moderate prediction accuracy [11] and a higher prediction accuracy than genomic prediction in

animals [12]. The GRM and MRM have been used collectively in the holo-omics framework for selection programs in livestock [13]. A significant drawback with the holo-omics framework is that it requires complex statistical models to compute the genetic and microbiome effects, resulting in excessive computational processing and timeframe. The complexity of the model makes it challenging to capture the genetic and microbiome effects jointly. Linear mixed models (LMMs) are key tools to fit SNP and OTUs as random effects to explore complex traits' genetic architecture [14]. One similar study used the holo-omics dataset in a standard linear model to predict complex traits [15].

Inspired by the complicated interaction between the hologenome affecting the complex traits prediction, this paper evaluates the accuracy of three holo-omics interaction models based on (i) covariance between random effects genome-based restricted maximum likelihood (CORE-GREML); (ii) Hadamard product; and (iii) direct genomic and microbiome effect. The above interaction models were evaluated using real animal genotype and microbiome datasets against the noninteractive models that directly use only genomic and microbiome's solo effects.

## 2. Methods and Methodology

### 2.1. Real Data Analysis

We have used two publicly available datasets to estimate the prediction accuracy of the complex traits. The first dataset was from the Ruminomics – 1000 cows' study [16], large enough to perform prediction accuracy, whereas the second dataset was from the gut microbial composition study, which consisted of 207 pigs [12]. From Ruminomics – 1000 cows' data, we only selected the high-density array genotypes of Holstein cows from UK and Italian farms ( $n = 795$ ). The cows were group-housed in loose housing barns with postpartum between 10 and 40 weeks, had an average live weight of 680 kg (Italy) and 678 kg (UK), received the standard diet for at least 14 days, and had no health issues present in them. In total, 120,321 SNPs were retained based on a minimum allele frequency of 0.01. Prokaryotes' 16S rRNA short-read sequencing data were downloaded from Short Reads Archive (SRA) under project accession number PRJNA517480. The sequences were analyzed using QIIME2 to generate relative abundance and OTU tables [17]. The total abundance rate at 20% retained 734 OTUs. In addition, we analyzed traits from cattle data that reflect a significant interplay between genome and microbiome, such as milk yield (milk, protein, fat, lactose, and fat-corrected milk (FCM)) and methane emission ( $\text{CH}_4$  g/d,  $\text{CH}_4$  DMI, and  $\text{CH}_4$  ECM). Similarly, from the pig data, we analyzed feed conversion (FC), feed intake (FI), and daily gain (DG), which are related to the gut microbial composition. The pigs belonged to the German Piétrain breed (housed at one experimental farm), with an initial weight of 30 kg and a final weight of 105 kg. The average age of the animals was 188 ( $\pm 14$ ) days during the slaughter time. The details of the datasets are shown in Table 1.

**Table 1.** Details of real animal datasets used to study prediction accuracy.

| Name                            | Subject | Sample Size | No. of SNPs | No. of OTUs | Fixed Effects  | Traits  |
|---------------------------------|---------|-------------|-------------|-------------|--|---|
| Ruminomics—1000 cows study      | Cattle  | 795 *       | 120,321     | 734         | Animal farm  | Milk<br>Fat<br>Protein<br>Lactose<br>FCM<br>$\text{CH}_4$ g/d<br>$\text{CH}_4$ DMI<br>$\text{CH}_4$ ECM |
| Gut microbial composition study | Pig     | 207         | 51,970      | 1870        | Slaughter weight, weight and age at the test station | Daily Gain<br>Feed Intake<br>Feed Conversion  |

\* From Dataset 1, we filtered only the Holstein dairy cows and used 795 subjects for the analysis.

## 2.2. Evaluation of the Models

A linear mixed modeling approach was used to perform statistical analysis using the BGLR R package [18]. We have used the Bayesian model (BayesB) to fit the effects in the model as it allows markers to have different effects and variance, along with better measures of fit and simple trait architecture. In addition, we used a Bayesian reproducing kernel Hilbert space (RKHS) based approach to fit the genomic and microbiome data as random effects in the interaction models, as RKHS will have better data-driven performance due to the large data size [19].

## 2.3. Noninteractive Models

In Models 1a and 1b, we only considered the number of casual SNPs influencing the complex traits to test the genomic effect. We called these noninteractive models because neither genome nor microbiome effects are fit to predict complex traits (Figure 1). The fixed effects in the models may vary because the datasets we used were from two different studies. The models are as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{g} + \mathbf{e} \quad (1a)$$

where  $\mathbf{y}$  is the vector of complex traits (from Ruminomics – 1000 cows' study, i.e., milk yield and methane emission-related parameters),  $\boldsymbol{\beta}$  is a vector of fixed effect (i.e., animal farm),  $\boldsymbol{\gamma}$  is the animal's dietary effect fitted by using BayesB,  $\mathbf{X}$  and  $\mathbf{Z}$  are the corresponding design matrix,  $\mathbf{g}$  is a vector with random animal genomic effect fitted by using RKHS, and  $\mathbf{e}$  is the residual term with residual variance  $\sigma_e^2$ . The distribution of the animal genome effect is  $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$ , with  $\mathbf{G}$  being the GRM of an individual's SNPs and  $\sigma_g^2$  being the additive genetic variance. The GRM was estimated by using the VanRaden method [20]:

$$\mathbf{G} = \frac{(\mathbf{Z} - 2\mathbf{Q})(\mathbf{Z} - 2\mathbf{Q})^T}{\sum_m 2p_m(1 - p_m)}$$

where  $\mathbf{Z}$  is the gene content matrix with entries 0,1 or 2 for each SNP and each animal, and matrix  $\mathbf{Q}$  contains the frequency  $p_m$  of each SNP  $m$ .

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{SD}\mathbf{SD} + \mathbf{Z}_{pen}\mathbf{pen} + \mathbf{g} + \mathbf{e} \quad (1b)$$

where  $\mathbf{y}$  is the vector of complex traits (from the gut microbial composition dataset, i.e., FC, FI, and DG),  $\boldsymbol{\beta}$  is a vector of fixed covariables (i.e., age and weight measured at the test station and the slaughter weight),  $\mathbf{SD}$  and  $\mathbf{pen}$  are the vectors with slaughter day and pen effect fitted by using BayesB,  $\mathbf{X}$ ,  $\mathbf{Z}_{SD}$ , and  $\mathbf{Z}_{pen}$  are the corresponding design matrix, and  $\mathbf{g}$  and  $\mathbf{e}$  are the same as Model 1a.

In Models 2a and 2b, we only considered the microbiome effect, i.e., the number of OTUs affecting the phenotype. The OTUs in 20% of the samples were used to construct the OTU table and derive the MRM. The models are as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{m} + \mathbf{e} \quad (2a)$$

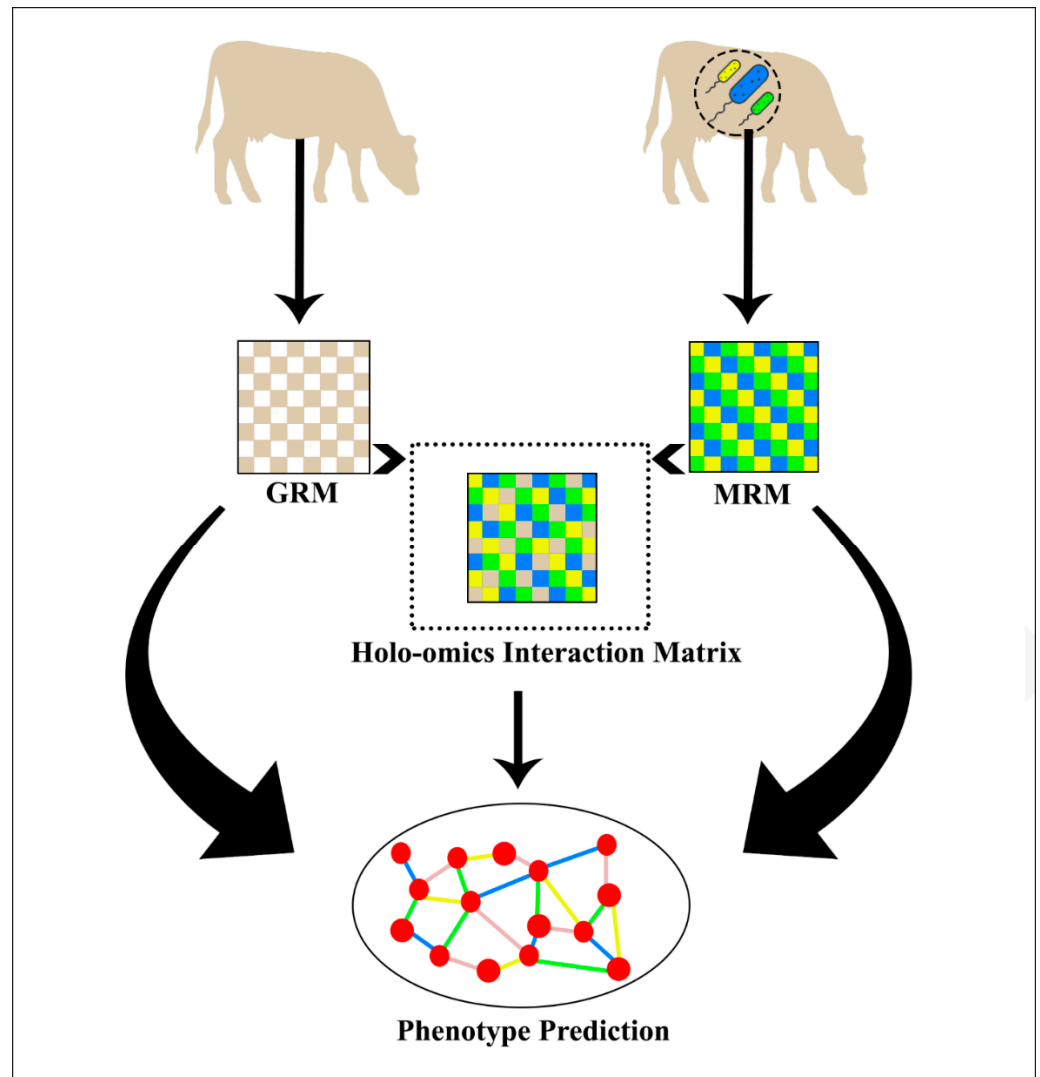
where the model parameters are as described in Model 1a except  $\mathbf{m}$ , which is the random effect of the microbiome fitted by the RKHS method following the multinomial distribution  $\mathbf{m} \sim N(0, \mathbf{M}\sigma_m^2)$ , with  $\mathbf{M}$  being the MRM constructed from OTUs and  $\sigma_m^2$  being the microbial variance. The  $\mathbf{M}$  matrix was estimated as:

$$\mathbf{M} = \frac{\mathbf{R}\mathbf{R}'}{n}$$

where  $\mathbf{R}$  is the matrix with log-transformed prokaryote's relative abundance from all animals and  $n$  is the number of OTUs within animals. The matrix  $\mathbf{R}$  was scaled and recombined into a single matrix before the calculation of  $\mathbf{M}$ .

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_{SD}\mathbf{SD} + \mathbf{Z}_{pen}\mathbf{pen} + \mathbf{m} + \mathbf{e} \quad (2b)$$

where the model parameters are the same as described in **1b**, and  $\mathbf{m}$  is the microbiome's random effect, same as Model **2a**.



**Figure 1.** Graphical representation of the models evaluated. The GRM is estimated from an animal's genome data and the MRM is estimated from the host's metagenome. The holo-omics interaction matrix is estimated by combining the GRM and MRM in several possible ways. Collectively, all the matrices can be used for phenotype prediction and to study their effects.

#### 2.4. Holo-Omics Interactive Models

The three holo-omics interactive models consisted of three different interaction frameworks. In the first interaction, the genomic and microbiome effects were coincidentally fitted into Models **3a** and **3b** to estimate the prediction accuracy. These models assume the direct interaction between the GRM and MRM to predict the complex traits, hence they are called holo-omics direct prediction models. The models are as follows:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{g} + \mathbf{m} + \mathbf{e} \quad (3a)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{SD}\mathbf{SD} + \mathbf{Z}_{pen}\mathbf{pen} + \mathbf{g} + \mathbf{m} + \mathbf{e} \quad (3b)$$

where the model parameters are the same as described in **1a** and **1b**, with  $\mathbf{g}$  being the animal genomic effect as described in **1a**, and  $\mathbf{m}$  being the microbiome's random effect, same as in model **2a**.

The second interaction consists of CORE-GREML [21] that estimates the covariance between two random effects, i.e., genomic and microbiome, by fitting the Cholesky decomposition of the GRM and MRM in an LMM. Therefore, these models are called holo-omics indirect prediction models. The models are as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{g} + \mathbf{m} + \mathbf{g}_c\mathbf{m}_c + \mathbf{e} \quad (4a)$$

where the model parameters are the same as Model **3a** except  $\mathbf{g}_c\mathbf{m}_c$ , which is the holo-omics interaction matrix constructed as follows:

$$\mathbf{g}_c\mathbf{m}_c = \sqrt{\mathbf{M}}.\sqrt{\mathbf{G}}' + (\sqrt{\mathbf{M}}.\sqrt{\mathbf{G}}) \quad (4b)$$

where  $\sqrt{\mathbf{M}}$  is the Cholesky decomposition of the  $\mathbf{M}$  matrix, which is described in **2a**, and  $\sqrt{\mathbf{G}}$  is the Cholesky decomposition of the  $\mathbf{G}$  matrix, which is explained in **1a**, and  $\sigma_{\mathbf{g}_c\mathbf{m}_c}$  is the covariance between the genomic and microbiome effect. We assume no correlation between the residuals and any other random effects.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{SD}\mathbf{SD} + \mathbf{Z}_{pen}\mathbf{pen} + \mathbf{g} + \mathbf{m} + \mathbf{g}_c\mathbf{m}_c + \mathbf{e} \quad (4c)$$

where the model parameters are the same as described in **3b**, with  $\mathbf{g}$  being the genomic effect,  $\mathbf{m}$  being the microbiome effect and  $\mathbf{g}_c\mathbf{m}_c$  is the holo-omics interaction matrix estimated using the CORE-GREML described in **4b**.

The third interaction consists of Hadamard product between the random effects, i.e., GRM and MRM [22]. The Hadamard product matrix is then fitted in the LMM to estimate the prediction accuracy of complex traits and is also called holo-omics indirect prediction models.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{g} + \mathbf{m} + \mathbf{g}_h\mathbf{m}_h + \mathbf{e} \quad (5a)$$

where the model parameters are the same as described in **3a** except  $\mathbf{g}_h\mathbf{m}_h$ , which is the holo-omics interaction matrix, also known as the covariance matrix, which captures the covariance between genome and microbiome. It is estimated as follows:

$$\mathbf{g}_h := (\mathbf{G})_{ij} (\mathbf{M})_{ij} \quad (5b)$$

where  $\mathbf{G}$  is the GRM defined in **1a**, and  $\mathbf{M}$  is the MRM defined in **2a**. The symbol  $\circ$  is defined as the Hadamard product for the two matrices,  $i = 1, 2, 3 \dots, n$  indicates the row, and  $j = 1, 2, 3 \dots, n$  indicates the column. The symbol  $:=$  means that the left-hand side is defined by the right-hand side, and  $\mathbf{G}$  and  $\mathbf{M}$  are multiplied entry-wise. The element-by-element product of  $\mathbf{G}$  and  $\mathbf{M}$  is:

$$(\mathbf{G} \circ \mathbf{M})_{ij} = \begin{bmatrix} G_{1,1}M_{1,1} & G_{1,2}M_{1,2} & \cdots & G_{1,n}M_{1,n} \\ G_{2,1}M_{2,1} & G_{2,2}M_{2,2} & \cdots & G_{2,n}M_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ G_{i,1}M_{i,1} & G_{i,2}M_{i,2} & \cdots & G_{i,j}M_{i,j} \end{bmatrix}$$

The  $\mathbf{g}_h\mathbf{m}_h$  is a square matrix and  $\sigma_{\mathbf{g}_h\mathbf{m}_h}$  is the covariance between the genome and microbiome.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{SD}\mathbf{SD} + \mathbf{Z}_{pen}\mathbf{pen} + \mathbf{g} + \mathbf{m} + \mathbf{g}_h\mathbf{m}_h + \mathbf{e} \quad (5c)$$

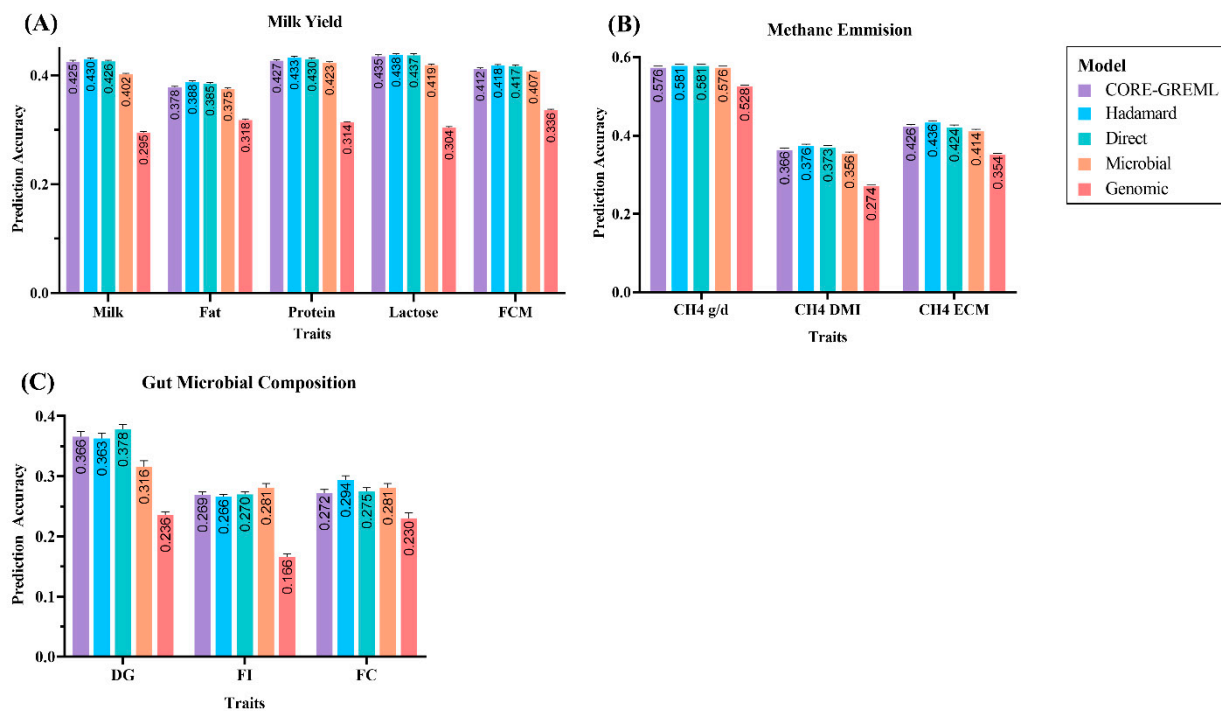
where the model parameters are the same as described in **3b**, with  $\mathbf{g}$  being the genomic effect,  $\mathbf{m}$  being the microbiome effect and  $\mathbf{g}_h\mathbf{m}_h$  is the holo-omics interaction matrix estimated using the Hadamard product described in **5b**.

### 2.5. Prediction Accuracy Evaluation

We used a foldid vector ( $n = 10$ ) to cross-validate the model's prediction accuracy [23]. The vector randomly divided our original sample into 10 sub-samples, and each sub-sample was used as the validation set, and the rest of the folds were treated as the training set. We repeated this process ten times to calculate the square of the correlation coefficient ( $r^2$ ) and used the average calculation as the criteria for prediction accuracy. Finally, one-way ANOVA with multiple comparisons was used to find the significance ( $p$ -value) of estimated prediction accuracy between models. We have used the significance value of  $\alpha = 0.05$  to compare our calculated  $p$ -values and reject the null hypothesis, i.e., mean predicted values are equal. A  $p$ -value less than the  $\alpha$ -value suggests that our calculated prediction accuracy is statistically significant.

### 3. Results

We evaluated the various holo-omics models and tested how the prediction accuracy changed concerning different genomic and microbiome interactions (Figure 2). We not only tested the individual GRM and MRM but also included the sophisticated holo-omics interaction matrix based on the CORE-GREML and Hadamard product. They have been proven to help estimate and detect covariance between random effects using the RKHS method. We calculated the square of the correlation coefficient ( $r^2$ ) by comparing prediction accuracy in different traits and showed important intuitions into the holo-omics architecture of complex traits.



**Figure 2.** (A) Prediction accuracy ( $r^2$ ) of different models in milk yield traits ( $p < 0.0001$ ). (B) Prediction accuracy ( $r^2$ ) of different models in methane emission traits ( $p < 0.0001$ ). (C) Prediction accuracy ( $r^2$ ) of different models in gut composition traits ( $p < 0.0001$ ).

#### 3.1. Prediction Accuracy of Complex Traits from the Ruminomics—1000 Cow's Study

The results of noninteractive and holo-omics interactive prediction accuracy using data from the Ruminomics—1000 cows' study are shown in Table 2. The holo-omics interactive prediction considering the Hadamard interaction framework (rh) showed excellent accuracy in milk yield and methane emission traits. The prediction accuracy of  $0.430 \pm 0.002$  for milk ( $p < 0.0001$ ),  $0.388 \pm 0.002$  for fat ( $p < 0.0001$ ),  $0.433 \pm 0.002$  for protein ( $p < 0.0001$ ),

0.438 ± 0.002 for lactose ( $p < 0.0001$ ), and 0.418 ± 0.002 for FCM ( $p < 0.0001$ ) was estimated from the model. In methane emission-related phenotypes, the prediction accuracy of 0.581 ± 0.001 for CH<sub>4</sub> g/d ( $p < 0.0001$ ), 0.376 ± 0.002 for CH<sub>4</sub> DMI ( $p < 0.0001$ ), and 0.436 ± 0.002 for CH<sub>4</sub> ECM ( $p < 0.0001$ ) was calculated. The holo-omics direct interaction model (rd) also had a prediction accuracy of 0.581 for ± 0.002 CH<sub>4</sub> g/d ( $p < 0.0001$ ). The genomic prediction model estimated the lowest accuracy among all the models. The summary of one-way ANOVA with multiple comparisons and test statistics is provided in Supplementary File S1.

**Table 2.** Accuracy of holo-omics CORE-GREML ( $r_c$ ), holo-omics Hadamard ( $r_h$ ), holo-omics direct ( $r_d$ ), microbiome ( $r_m$ ), and genomic prediction ( $r_g$ ) of milk yield and methane emission traits. The best estimated prediction accuracy values are highlighted in bold text.

| Trait               | Holo-omics Indirect Prediction |                      | Holo-omics Direct Prediction | Microbial Prediction | Genomic Prediction |
|---------------------|--------------------------------|----------------------|------------------------------|----------------------|--------------------|
|                     | $r_c$                          | $r_h$                | $r_d$                        | $r_m$                | $r_g$              |
| Milk                | 0.425 ± 0.003                  | <b>0.430</b> ± 0.002 | 0.426 ± 0.002                | 0.402 ± 0.002        | 0.295 ± 0.002      |
| Fat                 | 0.378 ± 0.002                  | <b>0.388</b> ± 0.002 | 0.385 ± 0.002                | 0.375 ± 0.002        | 0.318 ± 0.002      |
| Protein             | 0.427 ± 0.002                  | <b>0.433</b> ± 0.002 | 0.430 ± 0.002                | 0.423 ± 0.002        | 0.314 ± 0.001      |
| Lactose             | 0.435 ± 0.003                  | <b>0.438</b> ± 0.002 | 0.437 ± 0.003                | 0.419 ± 0.002        | 0.304 ± 0.002      |
| FCM                 | 0.412 ± 0.002                  | <b>0.418</b> ± 0.002 | 0.417 ± 0.002                | 0.407 ± 0.001        | 0.336 ± 0.001      |
| CH <sub>4</sub> g/d | 0.576 ± 0.001                  | <b>0.581</b> ± 0.001 | <b>0.581</b> ± 0.002         | 0.576 ± 0.001        | 0.528 ± 0.002      |
| CH <sub>4</sub> DMI | 0.366 ± 0.002                  | <b>0.376</b> ± 0.002 | 0.373 ± 0.002                | 0.356 ± 0.002        | 0.274 ± 0.001      |
| CH <sub>4</sub> ECM | 0.426 ± 0.002                  | <b>0.436</b> ± 0.002 | 0.424 ± 0.003                | 0.414 ± 0.002        | 0.354 ± 0.001      |

### 3.2. Prediction Accuracy of Complex Traits from the Gut Microbial Composition Study

The prediction accuracy was peculiar across all models while analyzing the gut microbial composition study data (Table 3). The prediction accuracy of 0.294 ± 0.007 ( $p < 0.0001$ ) was estimated in FC using the holo-omics Hadamard interaction framework. The accuracy of holo-omics direct prediction (0.378 ± 0.008,  $p < 0.0001$ ) and microbial prediction (0.281 ± 0.007,  $p < 0.0001$ ) was highest in DG and FI, respectively. Similar to the previous analysis, the genomic prediction accuracy was lower than the other models.

**Table 3.** Accuracy of holo-omics CORE-GREML ( $r_c$ ), holo-omics Hadamard ( $r_h$ ), holo-omics direct ( $r_d$ ), microbiome ( $r_m$ ), and genomic prediction ( $r_g$ ) of gut microbial compositional traits. The best estimated prediction accuracy values are highlighted in bold text.

| Trait | Holo-Omics Indirect Prediction |                      | Holo-Omics Direct Prediction | Microbial Prediction | Genomic Prediction |
|-------|--------------------------------|----------------------|------------------------------|----------------------|--------------------|
|       | $r_c$                          | $r_h$                | $r_d$                        | $r_m$                | $r_g$              |
| DG    | 0.366 ± 0.009                  | 0.363 ± 0.009        | <b>0.378</b> ± 0.008         | 0.316 ± 0.01         | 0.236 ± 0.005      |
| FI    | 0.269 ± 0.005                  | 0.266 ± 0.004        | 0.270 ± 0.004                | <b>0.281</b> ± 0.007 | 0.166 ± 0.005      |
| FC    | 0.272 ± 0.006                  | <b>0.294</b> ± 0.007 | 0.275 ± 0.006                | 0.281 ± 0.007        | 0.230 ± 0.009      |

## 4. Discussion

This study shows a simple approach for modeling holo-omics interactions in LMM to evaluate the prediction accuracy of complex traits in animals. Evaluating models for prediction accuracy before the actual experiment and formulating precise breeding programs using hologenomes in animals is essential. Furthermore, since breeding programs are complicated and expensive, it is necessary to optimize the breeding parameters, which include the size of the population, selection of animals, subject's age, choosing the phenotypes, and density of genotyping concerning the animal's age, along with 16s rRNA sequencing. Therefore, statistical models can be effectively used to design and optimize

breeding programs with respect to the holo-omics domain. Furthermore, it allows for predicting expected outcomes by experimenting with many alternatives, such as holo-omics interaction frameworks, and proves to be computationally less demanding. One approach is to establish the accuracy of models that can successfully predict complex traits in a breeding program. Additionally, it provides insight into essential factors, such as host genome, epigenome, transcriptome, metagenome, meta-transcriptome, proteomes, metabolome, metaproteome, and meta-metabolome.

The prediction accuracy models evaluated here account for the contribution of the host genome and microbiome to successfully predict complex traits concerning milk yield, methane emissions, and gut microbial composition. Since multiple interactions of holo-omics are involved in shaping these traits, it is essential to show to what extent the models can contribute to a breeding program. Ultimately, this will enable optimization with hologenome data, including the animal's genotype and phenotype data, for complex traits. In addition, it also helps to understand the role of holo-omics at multiple stages of the selection program.

The models for prediction of the accuracy of holo-omics interactions in animal data were developed on three concepts: (i) estimating the holo-omics interaction matrix by fitting the Cholesky decomposition of the GRM and MRM, termed as CORE-GREML, (ii) estimating the holo-omics interaction matrix by fitting the Hadamard product of the GRM and MRM, and (iii) direct holo-omics interaction of GRM and MRM. Methods to model these three concepts were developed, and the prediction accuracy was calculated using the real animal genotype and microbiome data from two different studies. However, further validation of these three concepts can be possible, provided that the dataset is larger than the current data used in this study to estimate the prediction accuracy of the complex traits.

#### 4.1. Modeling the Non-Interactive Interaction

Genomic prediction accuracy has gained attention due to its advantage in formulating breeding programs. Estimating accuracy from genomic prediction models largely depends on the reference population's size, the heritability of the trait, effective population size, marker density, etc. [24]. Several studies have experimented with these parameters to test which models gain better prediction accuracy and can be further used in breeding programs, considering complex traits. In contrast, microbial flora has also gained importance in shaping complex traits [9,25,26], and estimating prediction accuracy by using metagenomic profiles has shown its precision. Camarinha-Silva et al. [12] studied the gut microbial community's role in predicting complex traits and showed that the microbial prediction generated a higher accuracy than the genomic prediction. Similarly, in our study, the prediction accuracy also agrees with the previous estimations, i.e., the microbial prediction accuracy ( $r_m$ ) was higher than the genomic prediction accuracy ( $r_g$ ) in all the eleven complex traits (Tables 2 and 3). The gain in prediction accuracy shows the beneficial role of the microbiome in designing breeding programs for complex traits. The term microbiome-association index has been successfully used to explain the fraction of the microbial community involved in analyzing complex traits and reliably estimating the prediction accuracy [27]. In one such study, metagenomic profiles have been successfully used to predict the accuracy of phenotypes in humans and cattle [7]. The considerable prediction accuracy of tested phenotypes, i.e., 0.423 in inflammatory bowel disease (IBD), 0.422 in body mass index (BMI), and 0.553 in methane emission, highlighted the microbiome's significance and importance in outlining the complex traits.

Machine learning techniques have also shown metagenomics' effectiveness in random forest and neural network models. The study concluded the better performance of the random forest model in predicting dissolved organic carbon from plant litter decomposition by using microbial community profiles [28]. Although most of the studies estimated the increase in the accuracy of microbial prediction, still one main question needs to be answered whether the microbial prediction accuracy remains the same at different stages of the host's life since variation in microbe flora is common [29]. Another challenge includes



the difference in the microbial community observed due to the host's ecosystem [30,31]. The above challenges require further research to conclude the reliability of microbial prediction for complex traits, which is outside the scope of our current study.

#### 4.2. Modeling the Interactive Interaction

Since the popularity of the concept of hologenome, the interaction between the symbionts and the host has become limelight and widely studied by the scientific community [32,33]. The holo-omics interactive models used in our study have outperformed noninteractive models in predicting high accuracy in 100% complex traits related to milk yield and methane emission (Table 2). The holo-omics interactive model also predicted higher accuracy in gut microbial compositional traits in two of three complex traits (Table 3). Difford et al. [34] studied the holo-omics interactions in dairy cows and their specific roles in association with methane emissions. In another study, Weishaar et al. [35] further explored the domain of hologenome and used it to study its effect on animal feed efficiency traits. Nyholm et al. [36] investigated the composite interactions between the host's genome and their microbiome by combining the multiomics data. This study showed that integrating holo-omics data will help develop methods that will play a significant role in applied and basic biological research. The diverse prediction accuracy across all models while analyzing the gut microbial composition study data (Table 3) also indicates the importance of integrating different levels of holo-omics data.

In our holo-omics model analysis, the most successful model included the Hadamard interaction framework, in which the element-by-element product of GRM and MRM was estimated to construct a holo-omics interaction matrix. The matrix was further used in the LMM to predict the accuracy of complex traits (Figure 2). Previously, the Hadamard product was efficiently used to study the mechanisms of multimodel learning [37]. Merrick et al. [38] also studied genotype-by-environment (GE) scenarios for genomic prediction accuracy using the Hadamard product and estimated the positive correlation between environments using the variance-covariance matrix of the main effects. In contrast, Pérez-Enciso et al. [15] did not find any improvements in prediction with Hadamard products while using different scenarios of host and microbiome interactions. However, their research showed that the holo-omics prediction accuracy is ~50% better than the microbiome or genomic prediction through simulation and real-time data analysis.

Moreover, we have not estimated prediction accuracy using the Kronecker product framework, which could also be another potential model for testing the prediction accuracy of complex traits. However, several studies suggest that the prediction models using the Hadamard product have continually outperformed the Kronecker product model [38] and consist of a similar covariance structure [39]. Therefore, further studies will be required to conclude the performance of the Kronecker product in the holo-omics framework.

## 5. Conclusions

According to Stewart Ian [40], "Life is a partnership between genes and mathematics." Similarly, this paper describes several approaches for estimating the holo-omics interaction matrix and utilizing them in LMM for complex trait prediction. The Hadamard product interaction is the most successful holo-omics interaction with respect to complex trait prediction accuracy. The mechanisms that shape the host's genetics remain unknown. The host's genome and microbiome generally highlight the critical relationship between them and complex traits. This paper supports the host-microbe relationship and provides strategic results for implementing the holo-omics framework in breeding programs to increase complex trait prediction.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13091580/s1>, Supplementary File S1 The summary of one-way ANOVA with multiple comparisons and test statistics.

**Author Contributions:** Formal analysis, Q.R.Q.; Methodology, Q.Z. and Z.Z.; Project administration, Y.P. and Q.W.; Resources, X.L.; Supervision, Y.P. and Q.W.; Visualization: W.Z.; Writing—original draft, Q.R.Q.; Writing—review and editing, Q.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (Grant number: 2021C02068-1), the National Key Research and Development Program of China (2021YFD1200802), and the National Natural Science Foundation of China (Grant number: U21A20249).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. The datasets can be found at <https://www.science.org/doi/10.1126/sciadv.aav8391> (Ruminomics—1000 cow's study) and <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5500156/> (gut microbial composition study).

**Acknowledgments:** We thank for the support of this work by the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding, the National Key Research and Development Program of China, and the National Natural Science Foundation of China.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Georges, M.; Charlier, C.; Hayes, B. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* **2019**, *20*, 135–156. [[CrossRef](#)] [[PubMed](#)]
2. Hacquard, S.; Garrido-Oter, R.; González, A.; Spaepen, S.; Ackermann, G.; Lebeis, S.; McHardy, A.C.; Dangl, J.L.; Knight, R.; Ley, R. Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* **2015**, *17*, 603–616. [[CrossRef](#)] [[PubMed](#)]
3. Woodhams, D.C.; Bletz, M.C.; Becker, C.G.; Bender, H.A.; Buitrago-Rosas, D.; Diebboll, H.; Huynh, R.; Kearns, P.J.; Kueneman, J.; Kurosawa, E. Host-associated microbiomes are predicted by immune system complexity and climate. *Genome Biol.* **2020**, *21*, 23. [[CrossRef](#)] [[PubMed](#)]
4. Gilbert, J.A.; Blaser, M.J.; Caporaso, J.G.; Jansson, J.K.; Lynch, S.V.; Knight, R. Current understanding of the human microbiome. *Nat. Med.* **2018**, *24*, 392–400. [[CrossRef](#)]
5. Xiao, L.; Estellé, J.; Kiilerich, P.; Ramayo-Caldas, Y.; Xia, Z.; Feng, Q.; Liang, S.; Pedersen, A.Ø.; Kjeldsen, N.J.; Liu, C. A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* **2016**, *1*, 16161. [[CrossRef](#)]
6. Goddard, M.E.; Hayes, B.J.; Meuwissen, T.H. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* **2011**, *128*, 409–421. [[CrossRef](#)]
7. Ross, E.M.; Moate, P.J.; Marett, L.C.; Cocks, B.G.; Hayes, B.J. Metagenomic predictions: From microbiome to complex health and environmental phenotypes in humans and cattle. *PLoS ONE* **2013**, *8*, e73056. [[CrossRef](#)]
8. Li, F.; Li, C.; Chen, Y.; Liu, J.; Zhang, C.; Irving, B.; Fitzsimmons, C.; Plastow, G.; Guan, L.L. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome* **2019**, *7*, 92. [[CrossRef](#)]
9. Vollmar, S.; Wellmann, R.; Borda-Molina, D.; Rodehutschord, M.; Camarinha-Silva, A.; Bennewitz, J. The gut microbial architecture of efficiency traits in the domestic poultry model species Japanese quail (*Coturnix japonica*) assessed by mixed linear models. *G3 Genes Genomes Genet.* **2020**, *10*, 2553–2562. [[CrossRef](#)]
10. Khanal, P.; Maltecca, C.; Schwab, C.; Fix, J.; Tiezzi, F. Microbiability of meat quality and carcass composition traits in swine. *J. Anim. Breed. Genet.* **2021**, *138*, 223–236. [[CrossRef](#)]
11. Rothschild, D.; Leviatan, S.; Hanemann, A.; Cohen, Y.; Weissbrod, O.; Segal, E. An atlas of robust microbiome associations with phenotypic traits based on large-scale cohorts from two continents. *PLoS ONE* **2022**, *17*, e0265756. [[CrossRef](#)] [[PubMed](#)]
12. Camarinha-Silva, A.; Maushammer, M.; Wellmann, R.; Vital, M.; Preuss, S.; Bennewitz, J. Host genome influence on gut microbial composition and microbial prediction of complex traits in pigs. *Genetics* **2017**, *206*, 1637–1644. [[CrossRef](#)] [[PubMed](#)]
13. Ramayo-Caldas, Y.; Zingaretti, L.; Popova, M.; Estellé, J.; Bernard, A.; Pons, N.; Bellot, P.; Mach, N.; Rau, A.; Roume, H. Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. *J. Anim. Breed. Genet.* **2020**, *137*, 49–59. [[CrossRef](#)] [[PubMed](#)]
14. Gusev, A.; Lee, S.H.; Trynka, G.; Finucane, H.; Vilhjálmsson, B.J.; Xu, H.; Zang, C.; Ripke, S.; Bulik-Sullivan, B.; Stahl, E. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **2014**, *95*, 535–552. [[CrossRef](#)]
15. Pérez-Enciso, M.; Zingaretti, L.M.; Ramayo-Caldas, Y.; de Los Campos, G. Opportunities and limits of combining microbiome and genome data for complex trait prediction. *Genet. Sel. Evol.* **2021**, *53*, 65. [[CrossRef](#)]

16. Wallace, R.J.; Sasson, G.; Garnsworthy, P.C.; Tapio, I.; Gregson, E.; Bani, P.; Huhtanen, P.; Bayat, A.R.; Strozzi, F.; Biscarini, F. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci. Adv.* **2019**, *5*, eaav8391. [[CrossRef](#)]
17. Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **2019**, *37*, 852–857. [[CrossRef](#)] [[PubMed](#)]
18. Pérez, P.; de Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [[CrossRef](#)]
19. Chakraborty, S.; Ghosh, M.; Mallick, B.K. Bayesian nonlinear regression for large p small n problems. *J. Multivar. Anal.* **2012**, *108*, 28–40. [[CrossRef](#)]
20. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)]
21. Zhou, X.; Im, H.K.; Lee, S.H. CORE GREML for estimating covariance between random effects in linear mixed models for complex trait analyses. *Nat. Commun.* **2020**, *11*, 4208. [[CrossRef](#)] [[PubMed](#)]
22. Vitezica, Z.G.; Legarra, A.; Toro, M.A.; Varona, L. Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Genetics* **2017**, *206*, 1297–1307. [[CrossRef](#)] [[PubMed](#)]
23. Zhao, W.; Lai, X.; Liu, D.; Zhang, Z.; Ma, P.; Wang, Q.; Zhang, Z.; Pan, Y. Applications of Support Vector Machine in Genomic Prediction in Pig and Maize Populations. *Front. Genet.* **2020**, *11*, 598318. [[CrossRef](#)]
24. Goddard, M. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* **2009**, *136*, 245–257. [[CrossRef](#)]
25. Hooper, L.V.; Littman, D.R.; Macpherson, A.J. Interactions between the microbiota and the immune system. *Science* **2012**, *336*, 1268–1273. [[CrossRef](#)] [[PubMed](#)]
26. Kannan, S.; Krishnamoorthy, G.; Palanichamy, P.; Marudhamuthu, M. Role of Gut Microbiome in Neuromodulation. In *Understanding Host-Microbiome Interactions-An Omics Approach*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 105–122.
27. Rothschild, D.; Weissbrod, O.; Barkan, E.; Kurilshikov, A.; Korem, T.; Zeevi, D.; Costea, P.I.; Godneva, A.; Kalka, I.N.; Bar, N. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **2018**, *555*, 210–215. [[CrossRef](#)] [[PubMed](#)]
28. Thompson, J.; Johansen, R.; Dunbar, J.; Munsky, B. Machine learning to predict microbial community functions: An analysis of dissolved organic carbon from litter decomposition. *PLoS ONE* **2019**, *14*, e0215502. [[CrossRef](#)]
29. Zhang, Z.; Li, D.; Refaey, M.M.; Xu, W.; Tang, R.; Li, L. Host age affects the development of southern catfish gut bacterial community divergent from that in the food and rearing water. *Front. Microbiol.* **2018**, *9*, 495. [[CrossRef](#)]
30. Price, D.C.; Brennan, J.R.; Wagner, N.E.; Egizi, A.M. Comparative hologenomics of two *Ixodes scapularis* tick populations in New Jersey. *PeerJ* **2021**, *9*, e12313. [[CrossRef](#)]
31. Díaz-Sánchez, S.; Estrada-Peña, A.; Cabezas-Cruz, A.; de la Fuente, J. Evolutionary insights into the tick hologenome. *Trends Parasitol.* **2019**, *35*, 725–737. [[CrossRef](#)]
32. Rosenberg, E.; Koren, O.; Reshef, L.; Efrony, R.; Zilber-Rosenberg, I. The role of microorganisms in coral health, disease and evolution. *Nat. Rev. Microbiol.* **2007**, *5*, 355–362. [[CrossRef](#)] [[PubMed](#)]
33. Rosenberg, E.; Zilber-Rosenberg, I. Microbes drive evolution of animals and plants: The hologenome concept. *mBio* **2016**, *7*, e01395-15. [[CrossRef](#)] [[PubMed](#)]
34. Difford, G.F.; Plichta, D.R.; Løvendahl, P.; Lassen, J.; Noel, S.J.; Højberg, O.; Wright, A.-D.G.; Zhu, Z.; Kristensen, L.; Nielsen, H.B. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet.* **2018**, *14*, e1007580. [[CrossRef](#)]
35. Weishaar, R.; Wellmann, R.; Camarinha-Silva, A.; Rodehutsord, M.; Bennewitz, J. Selecting the hologenome to breed for an improved feed efficiency in pigs—A novel selection index. *J. Anim. Breed. Genet.* **2020**, *137*, 14–22. [[CrossRef](#)] [[PubMed](#)]
36. Nyholm, L.; Koziol, A.; Marcos, S.; Botnen, A.B.; Aizpurua, O.; Gopalakrishnan, S.; Limborg, M.T.; Gilbert, M.T.P.; Alberdi, A. Holo-omics: Integrated host-microbiota multi-omics for basic and applied biological research. *IScience* **2020**, *23*, 101414. [[CrossRef](#)] [[PubMed](#)]
37. Kim, J.-H.; On, K.-W.; Lim, W.; Kim, J.; Ha, J.-W.; Zhang, B.-T. Hadamard product for low-rank bilinear pooling. *arXiv* **2016**, arXiv:1610.04325.
38. Merrick, L.F.; Herr, A.W.; Sandhu, K.S.; Lozada, D.N.; Carter, A.H. Optimizing plant breeding programs for genomic selection. *Agronomy* **2022**, *12*, 714. [[CrossRef](#)]
39. Martini, J.W.; Crossa, J.; Toledo, F.H.; Cuevas, J. On Hadamard and Kronecker products in covariance structures for genotype × environment interaction. *Plant Genome* **2020**, *13*, e20033. [[CrossRef](#)]
40. Stewart, I. *Life's Other Secret: The New Mathematics of the Living World*; John Wiley: New York, NY, USA, 1998.