

RESEARCH ARTICLE

Open Access



# Including microbiome information in a multi-trait genomic evaluation: a case study on longitudinal growth performance in beef cattle

Marina Martínez-Álvaro<sup>1,2\*</sup> , Jennifer Mattock<sup>2</sup>, Óscar González-Recio<sup>3</sup>, Alejandro Saborío-Montero<sup>4</sup>, Ziqing Weng<sup>5</sup>, Joana Lima<sup>2</sup>, Carol-Anne Duthie<sup>2</sup>, Richard Dewhurst<sup>2</sup>, Matthew A. Cleveland<sup>5</sup>, Mick Watson<sup>2</sup> and Rainer Roehe<sup>2\*</sup>

## Abstract

**Background** Growth rate is an important component of feed conversion efficiency in cattle and varies across the different stages of the finishing period. The metabolic effect of the rumen microbiome is essential for cattle growth, and investigating the genomic and microbial factors that underlie this temporal variation can help maximize feed conversion efficiency at each growth stage.

**Results** By analysing longitudinal body weights during the finishing period and genomic and metagenomic data from 359 beef cattle, our study demonstrates that the influence of the host genome on the functional rumen microbiome contributes to the temporal variation in average daily gain (ADG) in different months (ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub>, ADG<sub>4</sub>). Five hundred and thirty-three additive log-ratio transformed microbial genes (*alr*-MG) had non-zero genomic correlations ( $r_g$ ) with at least one ADG-trait (ranging from |0.21| to |0.42|). Only a few *alr*-MG correlated with more than one ADG-trait, which suggests that a differential host-microbiome determinism underlies ADG at different stages. These *alr*-MG were involved in ribosomal biosynthesis, energy processes, sulphur and aminoacid metabolism and transport, or lipopolysaccharide signalling, among others. We selected two alternative subsets of 32 *alr*-MG that had a non-uniform or a uniform  $r_g$  sign with all the ADG-traits, regardless of the  $r_g$  magnitude, and used them to develop a microbiome-driven breeding strategy based on *alr*-MG only, or combined with ADG-traits, which was aimed at shaping the rumen microbiome towards increased ADG at all finishing stages. Combining *alr*-MG information with ADG records increased prediction accuracy of genomic estimated breeding values (GEBV) by 11 to 22% relative to the direct breeding strategy (using ADG-traits only), whereas using microbiome information, only, achieved lower accuracies (from 7 to 41%). Predicted selection responses varied consistently with accuracies. Restricting *alr*-MG based on their  $r_g$  sign (uniform subset) did not yield a gain in the predicted response compared to the non-uniform subset, which is explained by the absence of *alr*-MG showing non-zero  $r_g$  at least with more than one of the ADG-traits.

\*Correspondence:

Marina Martínez-Álvaro  
mamaral9@upv.es  
Rainer Roehe  
rainer.roehe@sruc.ac.uk

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions** Our work sheds light on the role of the microbial metabolism in the growth trajectory of beef cattle at the genomic level and provides insights into the potential benefits of using microbiome information in future genomic breeding programs to accurately estimate GEBV and increase ADG at each finishing stage in beef cattle.

## Background

Feed is the largest cost associated with commercial beef producers (comprising 60–70% of the variable production costs [1]), therefore any effort to achieve the most efficient ratio between feed intake and growth rate will significantly contribute to the profitability of the industry. In addition, more efficient animals are needed to address the environmental challenge faced by beef production [2, 3]. Faster growth is associated with more efficient animals [4] because they reach slaughter weight at earlier ages, saving total energy (feed) requirements for body maintenance and thus reducing impact on the environment per kg product. However, the pattern of growth rate is not stable during the finishing phase. Specifically, growth rate declines and fat deposition increases with age [5–7], which has a direct impact on their feed conversion efficiency. Understanding the biology and, for breeding purposes, the genetics underlying growth rate and its pattern along the growing life of the animal [8], and adjusting feeding requirements accordingly, are critical to maximizing the feed conversion efficiency during each of the growing phases [9]. To identify these phenotypic and genetic factors, in this study, we treated average daily gains at different growth stages as different traits [10], and analysed them using multivariate models.

Among different factors that affect bovine growth performance, the microbial metabolism in the rumen plays a key role [11, 12], since cattle obtain ~70% of their energy requirements from volatile fatty acids produced by microbial fermentation [13], and 50–80% of their amino acids requirements from microbial proteins [14, 15]. In addition, the ruminal microbiome produces many microbiome-derived metabolites that act as regulatory signals in the gut-brain [16] and gut-liver axis [17–19], which may influence body composition [20], bone density [21], muscle development [22] and feeding behavior [23]. Given the importance of the rumen microbiome on growth, it is likely that variations in microbial pathways are associated with fluctuations in the growth rate of beef cattle. Host genetics partially determines the rumen microbiome [24–28], as first suggested by Weimer et al. [29] based on an analysis of a ruminal microbiome exchange in dairy cattle. Later, several studies reported moderate heritability estimates for some microbial genera and gene abundances in several beef populations [30]. The main hypothesis of this study was that relevant genomic correlations exist between performance and

microbial traits, and thus our objective was to provide evidence for a shared genetic determination of cattle growth trajectory and the functional microbiome. If these genomic correlations are informative enough, it would be possible to estimate the animal's genetic value for growth rate with greater accuracy when combining the animal weight with abundance information of microbial genes in its rumen, or, if weight records are not available, to evaluate animals based only on their microbiome composition. Furthermore, the large amount of genomic variation on microbial abundances might open the opportunity of exploring the microbiome as a breeding tool to obtain simultaneous desired responses to selection on different traits (e.g. growth rates at different stages) by enhancing specific microbial mechanisms that have favourable genomic correlations with all of these traits.

This study used a unique database of 359 animals with weekly body weight records during the finishing period and a comprehensive identification of their functional core microbiome profiles (i.e. abundances of 3632 microbial genes (MG) present in at least 70% of the animals) from rumen samples collected at slaughter. The first objective was to investigate the genomic correlations between the rumen microbiome and the average daily gains of cattle at different stages of the finishing period. The second objective was to investigate whether genomic evaluation for average daily gain traits can be performed using indirect information from MG (i.e. microbiome-driven breeding strategy), and to evaluate the improvement in the accuracy of EBV when adding microbiome information into the genomic evaluations. The third objective was to explore whether using rumen microbiome abundances, which are genomically correlated to average daily gain at different stages in the same direction, will present any advantage on the responses in these traits.

## Methods

### Animals

Data were obtained from 359 beef cattle used in different experiments [31–35] conducted over five years (2011, 2012, 2013, 2014 and 2017). The animals were from different breeds (rotational crosses of Aberdeen Angus and Limousin breeds, Charolais crosses and purebred Luings) and were fed two basal diets consisting of 480:520 and 80:920 forage:concentrate ratios. Additional file 1:

Table S1 shows the distribution of animals and data among experiments, breeds, and diets.

### Collection of weight data and estimation of average daily gains at different stages during the finishing period

Seventeen body weights of each of the 359 beef cattle were recorded weekly for four months when the animals were between  $394 \pm 32$  and  $505 \pm 33$  days of age, always before feeding using a calibrated weight platform. We divided the stages of growth into four longitudinal average daily gains ( $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$ ), representing four consecutive 4-week intervals (Fig. 1a). Each monthly average daily gain was estimated by linear regression using five weight points, as the slope between weight increase and number of past days.

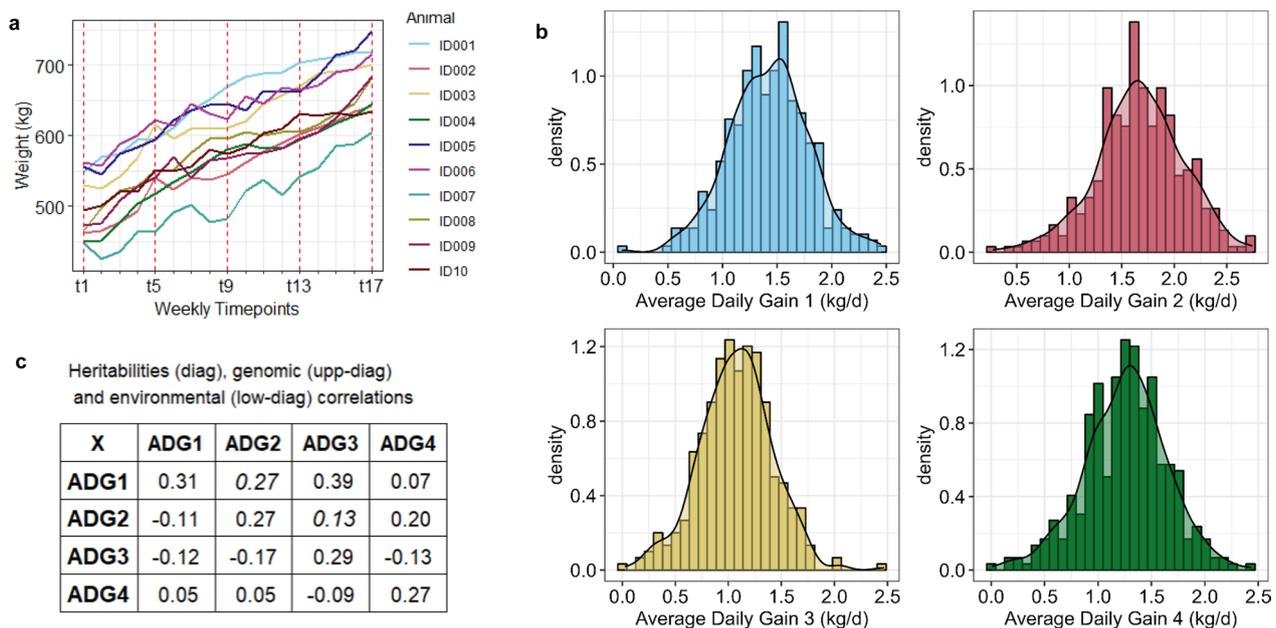
### Collection and sequencing of genomic samples

For host DNA analysis, 6–10 mL of blood from the 359 beef cattle were collected from the jugular or coccygeal vein on live animals or during slaughter in a commercial abattoir. An additional seven blood and 23 semen samples from sires of the cattle were available. Blood was stored in tubes containing 1.8 mg EDTA/mL blood and immediately frozen to  $-20^\circ\text{C}$ . Genomic DNA was isolated from blood samples using the Qiagen QIAamp toolkit and from semen samples using the Qiagen QIAamp DNA Mini Kit, according to the manufacturer's instructions. The DNA concentration and integrity were estimated

using a Nanodrop ND-1000 (NanoDrop Technologies). Genotyping was performed by Neogen Genomics (Ayr, Scotland, UK) using the GeneSeek Genomic Profiler (GGP) Bovine single nucleotide polymorphism (SNP) 50k Chip (GeneSeek, Lincoln, NE). Missing SNP genotypes were imputed using the Beagle 5.2 software [36]. Genotypes were filtered for quality control purposes using PLINK version 1.09b [37]. SNPs were removed from further analysis if they met any of these criteria: unknown chromosomal location according to Illumina's maps [38], SNP call rate lower than 95%, deviation from Hardy–Weinberg proportions ( $\chi^2$  test  $P$ -value  $> 10^{-8}$ ), or minor allele frequency lower than 0.05. Animals showing genotypes with a call rate lower than 90% were also removed. After imputation and filtering, 386 animals and 38,807 SNPs remained for the analyses.

### Collection and sequencing of metagenomic samples and compositional transformation of microbial abundances

For microbial DNA analysis, *post-mortem* digesta samples (approximately 50 mL) were collected from 359 cattle at slaughter, immediately after the rumen was opened to be emptied. All animals were moved from their pens and restricted to feed access between 4 to 5 h prior to slaughter. Five mL of the strained rumen fluid were mixed with 10 mL of phosphate buffered saline (PBS) with glycerol (87%) and stored at  $-20^\circ\text{C}$ .



**Fig. 1** Longitudinal growth rates (ADG) at different 4-week stages during the finishing period. **a** Variation in the growth curves from 10 randomly selected animals in the study. **b** Phenotypic distribution (after correction for systematic effects) of  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$ , with decreasing means (1.57, 1.50, 1.48 and 1.41 kg/day, respectively). **c** Heritabilities, genomic and environmental correlations between  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$

DNA extraction from rumen samples was performed according to the protocol of Yu and Morrison [39] based on repeated bead beating with column filtration, and DNA concentrations and integrity were assessed using the same procedure (Nanodrop ND-1000) as for blood samples. DNA Illumina TruSeq libraries were prepared from microbial genomic DNA and sequenced on an Illumina HiSeq Systems 4000 (samples from 283 animals from experimental years 2011, 2012, 2013, and 2014) [40, 41] or a NovaSeq (samples from 76 animals from experimental year 2017) by Edinburgh Genomics (Edinburgh, Scotland, UK). Paired-end reads ( $2 \times 150$  bp for HiSeq systems 400 and NovaSeq) were generated, ranging from 7.8 to 47.8 GB per sample (between 26 and 159 million paired-end reads). To measure the abundance of known functional MG, the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologue (KO) abundance pipeline KOunt [42] was used (<https://github.com/WatsonLab/KOunt>). Whole metagenome sequencing reads were trimmed for quality using the Fastp tool [43] and assembled using the MEGAHIT assembler [44]. Proteins were predicted using the Prodigal software [45], filtered to remove incomplete proteins, and searched against the KEGG database (<https://www.genome.jp/kegg/ko.html>) [46] (version 2020-10-04) using the KofamScan tool [47]. Hits that exceeded the default KofamScan thresholds were assigned to KO. Proteins that exceeded the threshold for multiple KO or that had no hits were grouped separately. The resulting KO grouping corresponded to a very similar set of sequences. The BWA-MEM algorithm [48] was used to map reads against their assembly and KO abundance was calculated using the BamDeal [49] and BEDTools tools [50]. We identified 7976 KO, which are referred to as MG. To discard non-core microbiome functions, we used only MG that were present in at least 252 of the 359 (70%) samples which resulted in 3632 MG representing 99.55% of the total counts identified in the microbiome dataset. Remaining zeroes were imputed based on a geometrical Bayesian-multiplicative method [51]. To deal with data compositionality, MG abundances were additive log-ratio transformed (*alr*) using the *ribulose-phosphate 3-epimerase* gene (*rpe*; EC:5.1.3.1; KEGG code K01783) as a reference, resulting in 3631 additive log-ratio transformed microbial gene abundances (*alr*-MG). The criteria to select the reference MG was a trade-off between two conditions explained in Greenacre et al. [52]: first, a Procrustes correlation of 0.9974 between the exact log-ratio geometry and the approximate geometry generated by the set of *alr*-MG, which ensures that Euclidean distances between samples are preserved after the transformation and second, a low variance of 0.0379 in

the log relative abundance of *rpe* (coefficient of variation of 5.08%), which further facilitates the *alr* interpretation by reducing it to the numerator part.

The main approach of this study was to explore the genomic associations between longitudinal growth rate and microbiome function, defined as *alr*-MG. Also of interest was the main microbial taxa in our populations associated with the *alr*-MG. To this end, we identified which of the 4941 rumen uncultured genomes (RUG) generated by Stewart et al. [40] had been annotated with our *alr*-MG of interest.

### Estimation of the heritability ( $h^2$ ) of the functional core microbiome and its genomic correlations ( $r_g$ ) with longitudinal average daily gains

Heritabilities of the functional core microbiome were estimated by fitting the 3361 *alr*-MG as observed traits in 3361 univariate genomic models. Each model was as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}. \quad (1)$$

Phenotypic data were assumed to be conditionally normally distributed as:

$$\mathbf{y}|\mathbf{b}, \mathbf{u}, \sigma_e^2 \sim N(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u}, \mathbf{I}\sigma_e^2), \quad (2)$$

where  $\mathbf{y}$  is the vector containing the observed *alr*-MG for each of the 359 animals,  $\mathbf{b}$  is the vector including a combination of diet, breed and experimental year (17 levels) as fixed effects,  $\mathbf{u}$  is the vector of host genomic random effects,  $\mathbf{e}$  is the vector of residuals, and  $\mathbf{X}$  and  $\mathbf{Z}$  are the known incidence matrices relating fixed and host genomic effects, respectively, to the microbial abundance observations. Note, that sequencer effect is nested within year and thus accounted for by including this effect in the model. Random effects had a normal distribution with mean 0 and variances  $\mathbf{G}_{\text{GRM}}\sigma_u^2$  for  $\mathbf{u}$  and  $\mathbf{I}\sigma_e^2$  for  $\mathbf{e}$ , where  $\sigma_u^2$  and  $\sigma_e^2$  are the genomic and residual variances,  $\mathbf{G}_{\text{GRM}}$  is the genomic relationship matrix computed following Method 2 of Van Raden [53] and  $\mathbf{I}$  is an identity matrix of the same order as the number of individuals with data. Bayesian statistics were used [54] to obtain and interpret the results. The analyses were carried out using the BGLR software [55]. The fixed effects were assigned flat priors (i.e., Gaussian prior with a null mean and a very large variance). The random variances were assigned default priors; i.e. scaled-inverse chi square distributions with the prior degree of freedom (df0) equal to 5 and the prior scale parameter  $S0 = \text{var}(Y) \cdot (\text{df0} + \text{number of traits} + 1) \cdot R^2$ ,  $\text{var}(Y)$  being the phenotypic variance of the trait and  $R^2$  being defined as the proportion of variance that one expects, a priori, to be explained by the regression with 0.5 as default value [56]. As features of

the marginal posterior distributions of  $h^2$ , we calculated the median and their highest posterior density interval at 95% probability (HPD<sub>95%</sub>). To test the sensitivity of our estimates from different prior information, we randomly selected 100 *alr*-MG and computed their  $h^2$ , this time using scaled-inverse chi square distributions with different df0 (1, 10 and 20) for genomic variances. We also investigated the distribution of the  $h^2$  for the *alr*-MG under a null hypothesis of no connection between *alr*-MG phenotypes and genotypes; i.e., we re-computed the  $h^2$  of the 3631 *alr*-MG after permuting the data (1000 repetitions) and compared their distribution to the real distribution of  $h^2$ .

The genomic correlations between the 3631 *alr*-MG and each of the average daily gains (ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub> and ADG<sub>4</sub>) were estimated by fitting 4 times 3631 bivariate animal models with the same effects as for Model (1). In this case, the phenotypic data were assumed to be conditionally normally distributed as:

$$\mathbf{y}|\mathbf{b}, \mathbf{u}, \mathbf{R} \sim N(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u}, \mathbf{R}), \quad (3)$$

where  $\mathbf{y}$  is a vector containing the observed average daily gain and *alr*-MG for each of the 359 animals,  $\mathbf{b}$ ,  $\mathbf{u}$ ,  $\mathbf{X}$  and  $\mathbf{Z}$  are as previously described but extended to a bivariate approach. Random effects had a multivariate normal distribution with a mean 0 and variances  $\mathbf{G} \otimes \mathbf{G}_{GRM}$  for  $\mathbf{u}$  and  $\mathbf{R} \otimes \mathbf{I}$  for residuals, where  $\mathbf{G}$  and  $\mathbf{R}$  are the  $2 \times 2$  genomic and residual (co)variance matrices, respectively, and  $\mathbf{I}$  is an identity matrix of the same order as the number of individuals with data. Models were solved using the BGLR software. As before, the fixed effects were assigned flat priors. The (co)variances for the residual and the genomic effects were assigned priors with default hyper parameters set in the unstructured option offered by the software. This is, an inverse Wishart distribution with five and three prior df0 for the residual and the genomic effects, and a prior scale  $2 \times 2$  matrix equal to  $\mathbf{S0} = \mathbf{var}(\mathbf{Y}) * (\text{df0} + 1 + \text{number of traits}) * R^2$ ,  $\mathbf{var}(\mathbf{Y})$  being the  $2 \times 2$  phenotypic variance–covariance matrix of the traits and a default value [56] of 0.5 for  $R^2$ . As features of the marginal posterior distributions of  $r_g$ , we calculated the median and their HPD<sub>95%</sub>. We also computed the probability of the  $r_g$  being higher or lower than 0 when the median is positive or negative, respectively ( $P_0$ ). Non-zero genomic correlations were defined as the estimates with  $P_0 \geq 0.85$ . To test the sensitivity of our estimates from different prior information, we re-computed these  $r_g$ , this time, using inverse Wishart distributions with different df0 (2, 10 and 20) for genomic effects.

In both the univariate and bivariate analyses, the marginal posterior distributions of the estimated parameters were based on Markov Monte Carlo chains consisting of 500,000 iterations, with a burn-in period of 100,000, and

only one of every 50 samples was saved for inferences. Convergence was tested with the R package *coda* [57] by checking the Z criterion of Geweke. Monte Carlo sampling errors were computed using time-series procedures and checked for being at least 10 times lower than the standard deviation of the posterior marginal distribution [54].

#### Heritabilities and $r_g$ between longitudinal average daily gains at different stages of the finishing period

To estimate the  $h^2$  of ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub> or ADG<sub>4</sub> and their  $r_g$ , we fitted six bivariate genomic models including pair-wise combinations of the four longitudinal growth traits as observations. The bivariate models included the same effects as described for Model (1). Phenotypic data was assumed to be conditionally normally distributed as described for Model (3). The analyses were carried out using the BGLR software [55], with the same priors as the default ones defined for Model (3). The marginal posterior distributions of the  $h^2$  were computed averaging the three  $h^2$  marginal posterior distributions obtained from three bivariate models involving each average daily gain trait. To describe the  $h^2$  and  $r_g$  estimates, we calculated the median, HPD<sub>95%</sub> and  $P_0$  (only for  $r_g$ ) of their final marginal posterior distributions.

#### Selection of microbiome variables to maximize accuracy of estimated genomic breeding values of longitudinal average daily gains

For breeding purposes, we considered only the *alr*-MG that had a genetic correlation  $r_g$  with any of the average daily gains achieving  $P_0 \geq 0.85$ . From those, we discarded the *alr*-MG for which the numerators were MG with average relative abundances across all animals  $\geq 0.001\%$ . The microbial gene *rpe* used as denominator had a mean relative abundance of 0.03%. We filtered for a minimum abundance of 0.001% since microbial genes that show low counts in the population are more likely to be unidentified in the rumen of some animals and these generated counts may be subject to more technical variation [58, 59]. The microbiome-driven breeding strategy was based on fitting a multi-trait best linear unbiased prediction (BLUP) model, which includes the goal traits and the selected *alr*-MG as observed traits [60]. Because the number of coefficients to be estimated and the computational cost increase rapidly with the number of traits included in the model, our aim was to select a reduced subset of 32 *alr*-MG among all the candidates, which maximize the accuracy of the EBV of the average daily gain traits. We selected two alternative subsets of 32 *alr*-MG, referred to as non-uniform and uniform  $r_g$  subsets. The first subset (non-uniform  $r_g$ ) was selected based on a stepwise Akaike information criteria (AIC) forward

regression method using the R package MASS [61], where the genomic estimated breeding values (GEBV) of average daily gain traits were the dependent variables, and the GEBV of the *alr*-MG were tested as predictors. Genomic EBV for all traits were obtained from the corresponding univariate analysis described above. First, the best *alr*-MG to predict  $ADG_1$  was selected based on the largest AIC reduction, in a model already including the remaining average daily gain traits  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  as predictors. Next, the best *alr*-MG to predict  $ADG_2$  was added to the model based on the same criterion, in a model already including the remaining ADG traits ( $ADG_1$ ,  $ADG_3$  and  $ADG_4$ ) plus the previously selected *alr*-MG. The one-by-one addition of variables to the model continued for  $ADG_3$  and  $ADG_4$ , and then started again for  $ADG_1$ . We stopped the process after selecting 32 *alr*-MG in the final model (8 per average daily gain trait). The second subset (uniform  $r_g$ ) was selected by the same procedure, but restricting the selection of *alr*-MG to those that had an  $r_g$  with  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  with the same sign, regardless of their magnitude. The reason for this strategy was to test whether the exclusive use of microbial functions with favourable associations with overall growth traits would provide an advantage in terms of the accuracy of the GEBV of ADG.

#### Multitrait BLUP model to evaluate the expected benefits of including *alr*-MG in the genomic evaluation of average daily gain traits

The GEBV of  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  were estimated based on three strategies:

- (1) Using only  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  information (direct strategy).
- (2) Using only information of the subset of 32 *alr*-MG (microbiome-driven breeding strategy), considering the two non-uniform and uniform  $r_g$  alternative subsets.
- (3) Using information from  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  and the subset of 32 *alr*-MG (combined strategy), considering the two non-uniform and uniform  $r_g$  alternative subsets.

In strategy (1), GEBV of average daily gains were estimated by fitting a tetra-variate genomic (G)BLUP analysis that included observed  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$ , and  $ADG_4$  traits, providing the previously computed coefficients of the  $4 \times 4$  genomic and residual (co)variance matrices as fixed. In strategies (2) and (3), GEBV of average daily gains were estimated by fitting a 36-trait multivariate analysis, including the 32 *alr*-MG and missing [strategy (2)] or observed values [strategy (3)] for average

daily gain traits, and introducing the coefficients of the corresponding  $36 \times 36$  genomic and residual (co)variance matrices as fixed. In a previous step, each coefficient of these matrices was estimated by fitting a bivariate model (496 models in total) with the corresponding pairwise combination between *alr*-MG as observed traits. Models were identical to those used to compute genomic parameters between *alr*-MG and average daily gain traits. Once the  $36 \times 36$  residual and genomic variance–covariance matrices were obtained, the latter needed bending in order to be positive definite (tolerance for minimum eigenvalues =  $1 \times 10^{-3}$ ). The difference between original and bent matrices was never larger than the posterior standard error of the corresponding parameters, and none of the  $h^2$  of the trait after bending varied by more than 5% from their original values.

The GEBV estimates were based on Monte Carlo Markov chains, which consisted of 100,000 iterations with a burn-in period of 20,000. To reduce autocorrelation, only 1 of every 100 samples was saved for inference. The GEBV accuracies of  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  under each strategy were computed as follows:

$$\text{Accuracy}_i = \sqrt{1 - \frac{\text{sd}_i^2}{g_{RM_{ii}} * \sigma_u^2}}, \quad (4)$$

where  $\text{sd}_i$  is the standard deviation of the posterior marginal distribution of the GEBV for animal  $i$ ,  $g_{RM_{ii}}$  is the diagonal element of the genomic relationship matrix for animal  $i$  and  $\sigma_u^2$  is the genomic variance of the trait. For the three strategies, prediction of the selection response in  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  was estimated by ranking individuals based on their GEBV for each of the four traits [62], assuming equal economic weights for all the traits, and a selection intensity of 1.755 or proportion of selection of 10%. Prediction of the response to selection for each trait was estimated as the marginal posterior distribution of the difference between the mean estimated GEBV of all animals and the mean of the selected animals, and its median and standard deviation were calculated.

## Results

### Genomic determination of average daily gains at different stages during the finishing period

Average daily gain was largest in the first month ( $ADG_1$  average was 1.57 kg/day), and then decreased slightly over time ( $ADG_2$ ,  $ADG_3$  and  $ADG_4$  averaged 1.50, 1.48 and 1.41 kg/day), with all traits showing a similar adjusted phenotypic variation across animals (coefficient of variation ranging from 25 to 32%, after correction for diet, breed, and year effects). Around 30% of this phenotypic variation was due to genomic variation across

animals in all growth stages, as the  $h^2$  of  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  were 0.31 (HPD<sub>95%</sub> of [0.13, 0.48]), 0.27 [0.12, 0.42], 0.29 [0.13, 0.45] and 0.27 [0.11, 0.43], respectively. In general, a common positive genomic association underlaid the average daily gain over the four considered months of the finishing period, although the magnitude of these genomic correlations ( $r_g$ ) were moderate or low and ranged from 0.39 [0.03, 0.69] between  $ADG_1$  and  $ADG_3$  (with a probability of  $r_g$  being different from 0 ( $P_0$ )=0.97) to 0.07 [- 0.34, 0.47] between  $ADG_1$  and  $ADG_4$  ( $P_0$ =0.63) (see Fig. 1c). An exception was the  $r_g$  between  $ADG_4$  and  $ADG_3$ , which was negative with a probability  $P_0$  of 0.72, but the estimate had a low median of -0.13 and large HPD<sub>95%</sub> of [- 0.50, 0.28] due to the small size of the dataset. These results suggest that different host genes influence the growth rate at different growth stages [10]. The environmental correlations between longitudinal traits were low (ranging from - 0.17 to 0.05) and inconsistent with  $r_g$ , indicating that the random environmental factors were not correlated during the growth of cattle.

#### Heritabilities of functional core microbiome and $r_g$ between average daily gains at different stages and the functional core microbiome

The functional composition of the rumen microbiome was influenced by the host genetics (see Additional file 2: Table S2). The  $h^2$  estimates of the 3631 *alr*-MG ranged from 0.19 [0.10, 0.29] to 0.44 [0.22, 0.68]. However, the estimated variance components were not entirely independent from the prior information used in the Bayesian analyses (see Additional file 2: Table S2, Additional file 3: Figure S1 and Additional file 4: Figure S2) due to the limited size of the dataset. Next, we investigated whether the 3361 *alr*-MG were genomically correlated with any of the longitudinal average daily gains ( $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$ ) and therefore could be used to indirectly estimate their GEBV. Of the 3631 *alr*-MG, 533 had non-zero ( $P_0 \geq 0.85$ )  $r_g$  with at least one of the longitudinal traits. When the overlapping MG were also counted, there were 583 non-zero  $r_g$  estimates: 236 with  $ADG_1$ , 227 with  $ADG_2$ , 44 with  $ADG_3$ , and 76 with  $ADG_4$ , with median values ranging from |0.21| to |0.42| (see Additional file 5: Table S3). The magnitudes of the  $r_g$  estimates (medians) were not completely insensitive to the use of different prior information, particularly when using highly informative priors (see Additional file 5: Table S3 and Additional file 6: Figure S3).

Among the 533 *alr*-MG with non-zero  $r_g$  ( $P_0 \geq 0.85$ ) with any of the average daily gains, the largest overlap occurred between  $ADG_1$  and  $ADG_3$  ( $n=8$ ) followed by that between  $ADG_1$  and  $ADG_2$ , ( $n=32$ ). These results suggest that part of the host genetic determinism that

underlies average daily gains at different stages occurs through differences in host genes that influence the functional core microbiome in the rumen. The functionalities of *alr*-MG with non-zero  $r_g$  with average daily gain at earlier growth stages (either with  $ADG_1$ ,  $ADG_2$  or both), included specific microbial energy-related processes; for example, several subunits of the *energy-converting hydrogenase A* (*ehaA*, *ehaC*, *ehaH*, *ehaK*) and *B* (*ehbC*, *ehbE*, *ehbI*, *ehbJ*, *ehbO*, *ehbQ*, *ehbP*) gene groups presented positive  $r_g$  (medians from 0.22 to 0.30,  $P_0 \geq 0.85$ ), while different subunits of the *F-type H+transporting ATPase* (*ATPF0A*, *ATPF0B*, *ATPF0C*, *ATPF1E*, *ATPF1G*) gene group presented negative  $r_g$  (medians from - 0.26 to - 0.37,  $P_0 \geq 0.88$ ). Earlier average daily gains were also correlated with a large number of *alr*-MG ( $n=31$ ) involved in ribosome biogenesis. Among these, 13 had non-zero positive  $r_g$  from 0.22 to 0.34 only with  $ADG_1$  (e.g. *NMD3*, *SDO1*, *RP-L18Ae*, and *RP-L10e*) and 13 had non-zero negative  $r_g$  from - 0.24 to - 0.34 only with  $ADG_2$  (e.g. *RP-L4*, *RP-L5*, *RP-L6*, and *RP-L7*). Moreover, average daily gains at early stages ( $ADG_1$  and  $ADG_2$ ) had strong and positive  $r_g$  with 11 *alr*-MG involved in methane metabolism including *methyl-coenzyme M* (*mcrG*), *formylmethanofuran* (*fmt*, *fwdD*), *formate dehydrogenase* (*fdhB*) or *coenzyme F<sub>420</sub> hydrogenase* (*frhB*) (medians from 0.22 to 0.30,  $P_0 \geq 0.85$ ).

Among the microbial mechanisms that were linked (non-zero  $r_g$ ) to ADG at later stages of the finishing period (either with  $ADG_3$ , or  $ADG_4$ ), the metabolism of sulphur was largely represented, and included three *alr*-MG with negative  $r_g$  with  $ADG_4$  (*asrC*, *asrB* and *hydB*, all with an  $r_g$  median of - 0.24,  $P_0 \geq 0.86$ ) and three *alr*-MG with positive  $r_g$  with  $ADG_3$  (*moeB*, *MOCS2B* and *dmsB*, with  $r_g$  medians ranging from 0.23 to 0.29,  $P_0 \geq 0.86$ ). Interestingly, different *alr*-MG that are also involved in sulphur metabolism had non-zero positive or negative  $r_g$  with  $ADG_1$  and  $ADG_2$  (*thiS*, *NFU1*, *hydG*, *tusA*). Another microbial mechanism that influences growth rates at later stages was the metabolism of amino and nucleotide sugars (*galT*, *nanA* and *nanE* had negative  $r_g$  with  $ADG_4$  with a median of - 0.24,  $P_0 \geq 0.85$ ), and as before, a few other *alr*-MG in the same pathway correlated positively with  $ADG_1$  and  $ADG_2$  (*legF*, *E1.1.1.374*, and *pgu* with  $r_g$  ranging from 0.26 to 0.34,  $P_0 \geq 0.89$ ).

Although in some cases our results allowed us to associate specific microbial mechanisms to specific growing stages (see above), a common situation was that *alr*-MG that are involved in specific microbial metabolic pathways presented positive or negative non-zero ( $P_0 \geq 0.85$ )  $r_g$  with different longitudinal average daily gains without a clear pattern, which reflects the complexity of the host-microbiome regulation during the four months of growth analysed in this study. This was the case for 27 *alr*-MG

that are involved in the biosynthesis of several amino acids (glycine, serine and threonine; lysine; arginine; cysteine and methionine, phenylalanine tyrosine and tryptophan; or alanine) and 23 *alr*-MG that are involved in the metabolism of carbohydrates (pentose and glucuronate interconversions; fructose and mannose; galactose; starch and sucrose) all with  $r_g$  medians  $\geq |0.22|$ . Another example were the *alr*-MG that are linked to the synthesis of microbial cell wall or membrane components. While some of them that are involved in the biosynthesis of lipopolysaccharides (*lpxK*, *lpxH*, *lptA*, *lptB*, *lptC*, and *ABC-2.LPSE.P*) and peptidoglycans (*spoVD* and *ftsI*) had positive or negative  $r_g$  with  $ADG_1$  and  $ADG_2$  ( $r_g$  between  $|0.22|$  and  $|0.36|$ ,  $P_0 \geq 0.85$ ); others that are linked to the biosynthesis of O-antigens (*wecE*, *wecB*, *pseB* and *pseF*) had positive or negative  $r_g$  to  $ADG_2$  and  $ADG_4$  ( $r_g$  from  $|0.24|$  to  $|0.29|$ ,  $P_0 \geq 0.85$ ).

#### Rumen uncultured genomes (RUG) associated to the 533 *alr*-MG with non-zero $r_g$ with longitudinal growth rates

Only nine of the 533 *alr*-MG numerators were identified in the genome of 4877 RUG out of the 4941 RUG present in the rumen of a subset of 282 cattle from our study, described by Stewart et al. [40] (see Additional file 7: Table S4). The remaining 524 MG were not found in any of the RUG. This could be attributed to the fact that a large part of the data was lost during the binning process and to the removal of low quality RUG; and also to the different methodologies to obtain the MG information [40]. The nine MG were involved in amino acid (*trpA*, *metN*, *ABC.PE.P1* and *ABC.PA.S*) and lipid (*INO1*) transport and metabolism, genetic microbial processes (*polC* and *RP-L6*) or were not characterized (*K07133*). Their corresponding *alr*-MG were genomically correlated ( $P_0 \geq 0.85$ ) with either  $ADG_1$  (*trpA*, *metN*, and *ABC.PE.P1*),  $ADG_2$  (*RP-L6*, *K07133*, and *INO1*),  $ADG_3$  (*nuoJ* and *ABC.PA.S*) or  $ADG_4$  (*polC*). The most ubiquitous MG were *K07133* (found in the genome of 3636 out of the 4877 different RUG), *ABC.PA.S* (in 3042 RUG), *RP-L6* (in 2921 RUG), *trpA* (in 2538 RUG) and *polC* (in 2436 RUG). The KEGG *K07133* was the most represented MG within the genome of 3636 RUG; for example, 378 of these RUG carried between 10 and 64 unique proteins classified as *K07133* in their genome; most of them are identified as *uncultured Bacteroidales bacterium* (67 RUG), *uncultured Bacteroidia bacterium* (62 RUG) or *uncultured Bacteroidetes bacterium* (49 RUG).

Of the 4877 RUG, we identified 820 that carried at least six of the nine MG. Among these, 276 were classified as *uncultured Lachnospiraceae bacterium*; 92 as *Ruminococcaceae bacterium*; 61 as *uncultured Clostridiales bacterium*; 54 as *uncultured Methanobrevibacter sp.*; 40 as *uncultured Erysipelotrichaceae bacterium*; 35 as

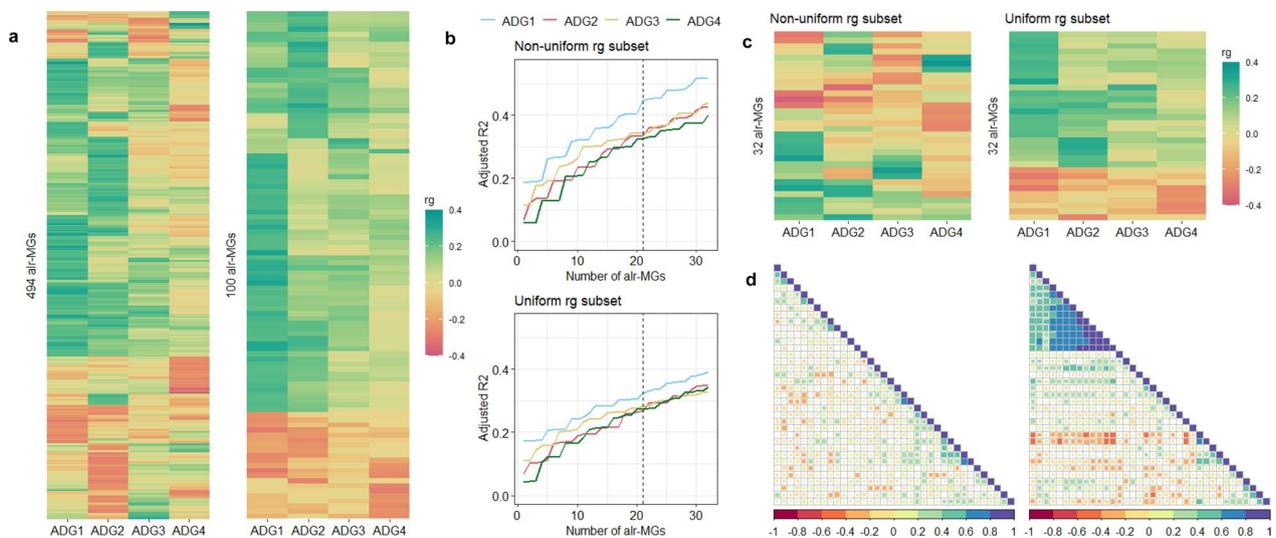
*uncultured Succiniclasicum sp.*, 32 as *uncultured Ruminococcus sp.*; 31 as *Fibrobacter succinogenes* and 26 as *uncultured Prevotellaceae bacterium*.

#### Selection of an optimal subset of microbial gene abundances for genomic evaluations of average daily gains at different stages

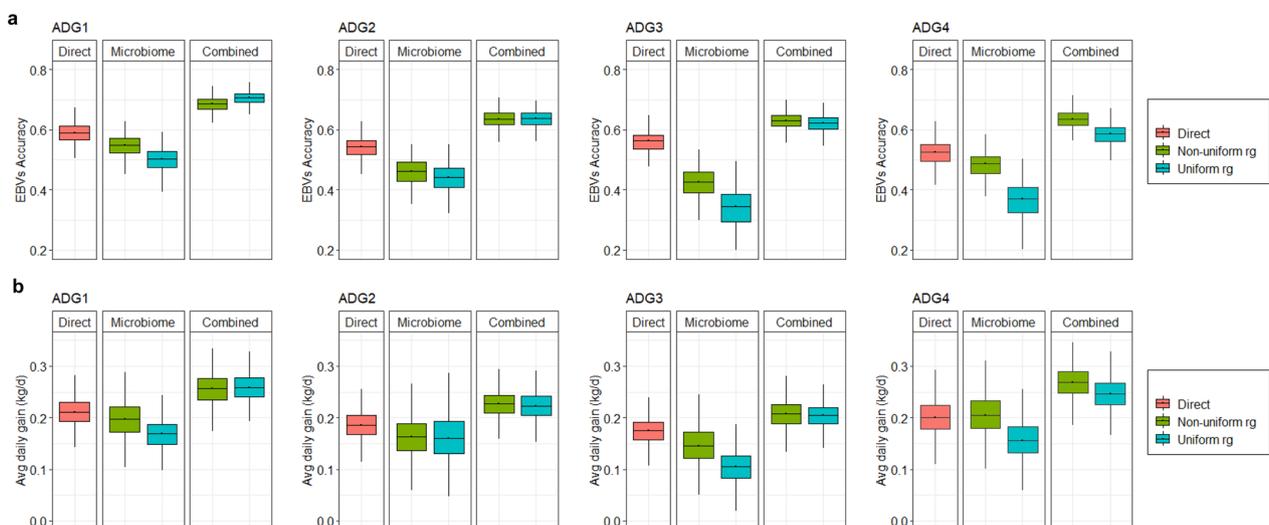
Of the 533 *alr*-MG candidates to be used for average daily gain genomic evaluations, we considered only 494 *alr*-MG with an average relative abundance across all animals  $\geq 0.001\%$  for the selection of the non-uniform  $r_g$  subset (Fig. 2a left). The uniform  $r_g$  subset was selected only from 100 of these 494 *alr*-MG that all had  $r_g$  estimates with all average daily gains with the same sign, regardless of their magnitude (Fig. 2a right). In both cases, the variables were chosen based on a forward-wise variable selection (see Additional file 8: Table S5). The non-uniform  $r_g$  subset of 32 *alr*-MG achieved greater adjusted  $R^2$  in the regression predicting GEBV of average daily gains (0.52, 0.42, 0.44, and 0.40 for  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$ ) than the uniform  $r_g$  alternative (0.40, 0.35, 0.32 and 0.34, respectively, Fig. 2b). Only four *alr*-MG were common in both subsets (*ehbO*, *mshG*, *srlD*, and *traN*). The uniform  $r_g$  subset contained a smaller number of *alr*-MG with a non-zero  $r_g$  with  $ADG_3$  and  $ADG_4$  (5 with  $ADG_4$  and 2 with  $ADG_3$ ) than the non-uniform  $r_g$  alternative (7 with  $ADG_4$ , 5 with  $ADG_3$  and 1 with both) (Fig. 2c). The genomic correlation matrix among the 32 *alr*-MG had a lower degree of redundancy in the first alternative, i.e. the maximum  $r_g$  between *alr*-MG was 0.68 ( $P_0 = 0.87$ ); while in the uniform  $r_g$  subset, the maximum  $r_g$  was 0.88 ( $P_0 = 0.95$ ) (Fig. 2d).

#### Expected benefits in the accuracy of average daily gain GEBV and predicted responses to selection when microbiome traits are included in genomic evaluations

We evaluated the prediction accuracy of GEBV for average daily gains at different stages as well as the prediction of selection responses in three different scenarios using different sources of information: (1) observed  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  and  $ADG_4$  (direct breeding strategy), (2) the functional microbiome information (microbiome-driven breeding strategy), and (3) the combination of observed average daily gains and functional microbiome information (combined breeding strategy). The mean GEBV accuracy in the direct breeding strategy was  $0.58 \pm 0.04$  for  $ADG_1$ ,  $0.54 \pm 0.04$  for  $ADG_2$ ,  $0.56 \pm 0.04$  for  $ADG_3$  and  $0.52 \pm 0.04$  for  $ADG_4$ . Inclusion of microbiome traits in a combined approach increased the prediction accuracies of GEBV for all traits by 11 to 22% (Fig. 3a), depending on the growth trait or subset of *alr*-MG used. Using only



**Fig. 2** Selection of additive log-ratio transformed microbial gene (*alr*-MG) abundances to be included in genomic evaluations. **a** Host-genomic correlations ( $r_g$ ) between average daily gains at different stages of the finishing period (ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub> and ADG<sub>4</sub>) with (see the left side of the panel) 494 *alr*-MG with mean relative abundance of the numerator  $\geq 0.001\%$ , and a genomic correlation ( $r_g$ ) with any of the average daily gain traits (ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub>, ADG<sub>4</sub>) being different from 0 with a probability  $\geq 0.85$ ; or with (see the right of the panel) the 100 *alr*-MG with the additional restriction of having a uniform  $r_g$  sign with all growth traits. **b** Change in adjusted  $R^2$  with an increasing number of *alr*-MG GEBV to 32 as explanatory variables using forward-wise linear regression to predict the GEBV of ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub> and ADG<sub>4</sub> for both alternatives defined in panel **a**. **c** Host-genomic correlations ( $r_g$ ) between average daily gains and the 32 *alr*-MG selected for both alternatives defined in panel **a**. **d** 36-traits genomic correlation matrix including the 32 *alr*-MG of the uniform (left) and non-uniform (right)  $r_g$  subsets and average daily gain traits



**Fig. 3** Prediction accuracies and expected predicted responses to selection on growth traits. **a** Prediction accuracy of genomic breeding values (GEBV) of growth traits (ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub> and ADG<sub>4</sub>) when including exclusively each ADG trait (direct), exclusively 32 microbiome traits that were found informative for each of the ADG traits (Microbiome) or both (Combined). We used two strategies to select the 32 microbiome traits: the most informative *alr*-MG among all those that presented a non-zero genomic correlation with the ADG traits (Non-uniform  $r_g$  subset) or the same criterion but including the restriction their genomic correlation with all ADG traits had a uniform sign (Uniform  $r_g$  subset). **b** Prediction of the selection response expected for each ADG trait when selecting the 10% best animals (i.e., animals with the largest EBV for ADG<sub>1</sub>, ADG<sub>2</sub>, ADG<sub>3</sub> or ADG<sub>4</sub>) under each strategy

microbiome information to estimate the GEBV of growth traits yielded reasonable accuracies ranging from  $0.33 \pm 0.07$  to  $0.55 \pm 0.04$  but never outperformed the direct strategy (7–41% lower accuracies relative to the direct strategy). The subset of *alr*-MG to be used as microbiome information had a stronger influence on GEBV accuracy of average daily gain when the microbiome breeding strategy was applied compared to the combined strategy. Applying the combined strategy, GEBV accuracies were similar using the non-uniform or uniform  $r_g$  subsets at earlier stages ( $0.70 \pm 0.02$  and  $0.68 \pm 0.03$  for  $ADG_1$ ;  $0.64 \pm 0.03$  and  $0.63 \pm 0.03$  for  $ADG_2$ ; and  $0.62 \pm 0.03$  and  $0.63 \pm 0.03$  for  $ADG_3$ ), but using the non-uniform  $r_g$  subset performed better for  $ADG_4$  ( $0.58 \pm 0.04$  vs  $0.63 \pm 0.03$ ). In the microbiome-driven strategy, the accuracies obtained were similar for both *alr*-MG subsets ( $0.44 \pm 0.05$  and  $0.46 \pm 0.05$  for  $ADG_2$ ) but, in the non-uniform strategy, they were higher particularly at later stages ( $0.50 \pm 0.04$  vs  $0.55 \pm 0.04$  in  $ADG_1$ ;  $0.33 \pm 0.07$  vs  $0.42 \pm 0.05$  in  $ADG_3$ ; and  $0.36 \pm 0.06$  vs  $0.48 \pm 0.04$  in  $ADG_4$ ). The lower accuracy for  $ADG_4$  and/or  $ADG_3$  using the uniform  $r_g$  subset is explained by the fact that most of the *alr*-MG that were penalized by the restriction were those with a high  $r_g$  with these two growth traits (Fig. 2a).

The magnitude of the predicted responses to selection that ranked the animals based on their GEBV for  $ADG_1$ ,  $ADG_2$ ,  $ADG_3$  or  $ADG_4$  was consistent with the GEBV prediction accuracies. Assuming a proportion of selection of 10%, the predicted responses to selection in the direct breeding strategy were  $0.20 \pm 0.04$  kg/d for  $ADG_1$ ,  $0.16 \pm 0.04$  kg/d for  $ADG_2$ ,  $0.15 \pm 0.04$  kg/d for  $ADG_3$  and  $0.21 \pm 0.04$  kg/d for  $ADG_4$  (12 to 14% of the trait mean). The expected predicted response to selection was greater when microbiome and growth traits were combined (responses ranged from  $0.20 \pm 0.02$  to  $0.27 \pm 0.03$  kg/d or from 14 to 19% of the trait mean), and there were no differences in terms of the two different microbial information used ( $0.26 \pm 0.03$  vs  $0.26 \pm 0.03$  for  $ADG_1$ ;  $0.23 \pm 0.03$  vs  $0.22 \pm 0.03$  for  $ADG_2$ ;  $0.21 \pm 0.03$  vs  $0.20 \pm 0.02$  for  $ADG_3$ ; and  $0.27 \pm 0.03$  vs  $0.25 \pm 0.03$  for  $ADG_4$ ). However, when only microbiome information was used, a greater predicted response was obtained in later growth stages using the non-uniform compared to using the uniform  $r_g$  subset of *alr*-MG:  $0.15 \pm 0.04$  vs.  $0.10 \pm 0.03$  kg/d in  $ADG_3$ ; and  $0.21 \pm 0.04$  vs  $0.16 \pm 0.04$  kg/d in  $ADG_4$ . Our results demonstrate that the inclusion of microbiome information in breeding evaluations for growth can increase the predictive accuracy of GEBV and response to selection or, if it is used alone, it is an acceptable strategy for conducting genomic evaluations for animals without growth rate records.

## Discussion

Longitudinal variations in average daily gain during cattle growth have important implications for the beef industry. For example, to achieve the highest feed conversion efficiency, nutrient requirements would have to be adjusted accordingly to the growth rate at each stage [1], which is not always practical under farming conditions. A comprehensive knowledge of the biological factors that contribute to these temporal variations offers the opportunity to modulate them towards a more uniform and increased growth pattern. One of these factors is the host's genetic contribution, as indicated by the moderate ( $r_g = 0.31$  and  $0.39$ ) to near zero ( $r_g = 0.07$ ) or even negative ( $r_g = -0.13$ ) genomic correlations between early and late average daily gains during the finishing period. Several studies suggest that candidate genes that affect growth rate at different stages [63–66] are associated with skeletal muscle and adipose tissue development (e.g. *FGF4*, *PLA2G4A*, *ITGA5* or *ANGPTL4*), and some of them are also expressed in the rumen [67]. In our study, we show, for the first time, that the polygenic basis underlying average daily gains at different stages is genetically correlated with the abundance of specific functions of the rumen microbiota. This is likely to be mediated by genetically influenced traits which affect the microbial composition in the rumen, such as rumen empty weight (heavier in less efficient animals) [68], digesta retention time [69], or rumen tissue development and absorptive capacity [70–72]. The majority of the 533 *alr*-MG showed a non-zero  $r_g$  with average daily gain only at one of the stages, with some exceptions for  $ADG_1$  and  $ADG_2$ . Regardless of the magnitude of  $r_g$ , some of these *alr*-MG showed  $r_g$  with a consistent sign for earlier growing stages but changed sign as animals were older ( $ADG_4$ ), or they simply did not follow any specific pattern, contributing to the differential genetic determination underlying longitudinal growth rates. Our study relies on the assumption that the composition of the ruminal functional microbiome sampled at slaughter is reasonably representative of the composition of the rumen microbiome throughout the growth period. This is supported by the longitudinal study of Snelling et al. [73] that was performed with a subsample of 50 animals from our dataset. The ruminal microbiome composition was monitored over 200 days (7 time points) from the early finishing stage until slaughter, and the authors did not detect any significant variability over time, suggesting temporal stability of the core rumen microbiome [73]. In our genetic study, animals were fed two different diets (see Methods), which was included as a fixed effect in the genetic model, assuming that our strategy is applicable under these two different environments (diets). However, it is possible that there is a genetic-by-environment (GxE) interaction on growth

and/or microbial traits, i.e., the genetic effects and/or microbial traits on growth can be different depending on diet. In this case, further analysis is required before the results of this study can be extrapolated to other feeding regimes. To date, the previous study of Roehe et al. [74] showed that there was no interaction between sire type and diet for methane emissions and that the regression coefficients between the abundance of specific microbial genes and methane emissions were similar in two groups of animals that were fed either concentrate or forage-based diets.

Growth rate is a moderately heritable low-cost indicator of feed conversion efficiency [4] because animal weights are relatively easy to record; and therefore body weight gain is considered in almost all beef genetic selection programs [75]. Furthermore, the existence of a  $r_g$  ( $P_0 \geq 0.85$ ) between specific heritable ( $h^2 \geq 0.19$ ) microbial functions and the average daily gains reported, opens up the possibility of using microbial biomarkers as complementary information to improve the accuracy of the GEBV for growth rate (combined breeding strategy). They could even be used as the sole source of information when weights are not recorded (microbiome-driven breeding strategy), given that microbiome information can be recorded for multiple uses, such as predicting feed efficiency, methane emissions or animal health. For breeding purposes, it was necessary to overcome the high-dimensionality of the microbial metabolism that we wanted to use for selection (in our case, the 533 *alr*-MG reduced to 494 *alr*-MG after filtering for average relative abundance  $\leq 0.001\%$ ), which most likely contain highly redundant information. In this study, we proposed to use a forward-wise variable selection method based on GEBV to avoid redundancy by selecting the most informative microbial predictors based on linear regression and maximizing the average daily gain explained variance. It should be noted that one limitation of this methodology is that it does not account for the prediction errors of GEBV. Recently, two alternative approaches for using -omics data for genomic prediction have been proposed, based on a two-step application of linear mixed model equations [76] or on the extension of linear mixed models by multilayer artificial neural networks [77] that do not require selection of -omics variables. These two new approaches suggest increases on breeding value prediction accuracies similar to this study when tested on simulated data (e.g. +12% compared to using genotypic information [76]).

A postulated advantage of a microbiome-driven breeding strategy is that the unfavourable and/or null  $r_g$  between productive traits (e.g., -0.13 between  $ADG_3$  and  $ADG_4$  or 0.07 between  $ADG_1$  and  $ADG_4$ ) may not be necessarily reflected in the overall complexity of

the functional microbiome, and specific *alr*-MG with strong and sign-consistent associations to these traits could be found. The inclusion of these microbial metabolic functions in genomic evaluations could help to overcome these unfavourable correlations, but at the cost of slowing down the response on the production traits. However, we faced two difficulties when testing this hypothesis. First, due to our small dataset, we could not determine whether  $ADG_3$  and  $ADG_4$  were negatively correlated with sufficient evidence. Second, our results showed a lack of microbial functions with at least moderate  $r_g$  with  $ADG_1$  and  $ADG_4$  or  $ADG_3$  and  $ADG_4$ . The use of *alr*-MG that are genomically correlated with average daily gain at different stages in the same direction (regardless of their magnitude) for genomic evaluations did not only provide no advantage over the non-uniform  $r_g$  alternative, but also reduced the accuracy of GEBV and the predicted response obtained at the latter stages. Regardless of the subset of *alr*-MG included in the genomic evaluations, a relevant finding was that combining information from *alr*-MG and average daily gains is the most beneficial scenario to increase their GEBV accuracies (the combined strategy achieved +11–22% greater GEBV accuracy than the direct strategy), and consequently, it achieves the greatest predicted responses to selection (up to 14–19% of the mean for all growth traits). The best performance of a combined breeding strategy compared to a direct or microbiome-driven one was also observed in a study that aimed at mitigating methane emissions [60] of beef cattle. The makeup of the microbiome represents a vast pool of genomic variation which, until now, has been only indirectly exploited when targeting production traits. However, our study demonstrates that targeting the microbiome directly could bring additional progress in genomic selection programmes. One concern with our study is that the results should be taken with caution as the estimation of GEBV is very sensitive to the magnitude of the variance components used, which in our case were estimated with a database of 359 animals, and were partly influenced by the priors. Much larger databases are required to obtain more accurate estimates, but this is currently hampered by the high cost of whole metagenomic deep sequencing and the novelty of using KEGG gene abundance as selection criteria. The feasibility of selection on microbial traits was shown in a selection experiment in pigs based on the abundance of *Prevotella*, *Mitsuokella*, *Treponema* and *Ruminococcus* in faeces, in which responses to selection in their abundances and correlated responses in growth rate and back fat thickness were obtained [78]. In our case, we proposed *alr*-MG as selection information because, according to our previous work, they are more

influenced by the host genome and more informative for phenotypic as well as genomic evaluation of traits than taxonomic composition [60, 74].

The targeted microbial metabolic routes for genomic selection in our microbiome-driven or combined breeding strategy can be biologically interpreted based on the functions of the numerators of *alr*-MG, given that the denominator has a very reduced variation [52]. We found that the *alr*-MG involved in the biosynthesis of ribosomes and DNA replication (*NMD3*, *SDO1*, *RP-L18Ae*, and *RP-L10e*), energy-related processes (*ehaA*, *ehaC*, *ehaH*, *ehaK*, *ehbC*, *ehbE*, *ehbI*, *ehbJ*, *ehbO*, *ehbQ*, and *ehbP*) and methane metabolism (*frhB*, *mcrG*, *fwdD*, and *ptr*) are genomically correlated with average daily gains at early growing periods ( $ADG_1$  and  $ADG_2$ ), mostly positively (see Additional file 5: Table S3). One hypothesis to explain these positive  $r_g$  could rely on the genetic determination for retention time of digesta (RTD) varying among animals. Variations in RTD among cattle could affect average daily gain by altering the abundance and activity of microorganisms in the rumen, the energetic efficiency of their growth, and also the digestibility of the diet [69]. A higher RTD is associated with a greater release of digestion by-products to be directed as nutrients for the host [69], but also with a greater release of substrates for methane production [79]. The objective of this work was exclusively focused on growth rate; however, a selection strategy on the microbiome could be explored to obtain more productive animals that emit less methane. For example, *alr*-MG involved in metabolic pathways that limit the excess of metabolic  $H_2$  substrate (e.g. reductive acetogenesis), consume  $H_2$  in alternative pathways other than methane production (e.g. nitrogen fixation), or inhibit methanogenic organisms (e.g. branched-chain amino acids) proposed in our previous microbiome-driven breeding strategy to reduce methane emissions [60], could be incorporated in the current selection criterion.

In our efforts to uncover the taxonomy of the RUG that are major carriers of the 533 *alr*-MG, in our population [40], we found *uncultured Lachnospiraceae bacterium*, *Ruminococcaceae bacterium uncultured Ruminococcus sp.* and *uncultured Prevotellaceae bacterium*. Interestingly, Li et al. [24] found five SNPs in the host genome being associated with unclassified *Lachnospiraceae* and *Ruminococcus* genera. Some of these SNPs were also associated with feed conversion efficiency in beef cattle [24], and were located within the host gene *RAPH1*, which is involved in the ability of the rumen epithelia to absorb nutrients in cattle [70]. In addition, the abundance of the genus *Ruminococcus* was used as a selection criterion to reduce growth rate in the selection experiment in pigs, while *Prevotella* was used to increase this trait [78].

Other microbial functions with discordant  $r_g$  directions with average daily gain during the finishing period are critical for the synthesis of cell-membrane components including peptidoglycan genes (*spoVD* and *ftsI*), lipopolysaccharide genes (*lpxK*, *lpxH*, *lptA*, *lptB*, *lptC*, and *ABC-2.LPSE.P*) and their O-antigen component of lipopolysaccharides (*wecE*, *wecB*, *pseB* and *pseF*), which are essential in the colonization of Gram-negative bacteria [80] (e.g. members of the *Lachnospiraceae* family). Bacterial lipopolysaccharides are microbiota-derived endotoxins [81, 82] that contribute to host metabolic endotoxemia [83, 84] by boosting intestinal permeability [83, 85–87] and triggering proinflammatory responses in various tissues, including muscle [88], when they bind to  $CD_{14}$ ,  $TLR_4$  [89] muscle receptors. Lipopolysaccharides have been reported to impair insulin action on glucose metabolism in muscle [88], and in extreme cases, elevated circulating levels have been associated with age-related reductions in muscle mass in humans [22]. Our study suggests that the signalling action by lipopolysaccharides, some of which mediated by receptors regulated by host genes (e.g. *TLR\_4*) [90–92], may influence muscle development in growing cattle. Individuals with greater butyrate production in their microbiomes are reported to counteract this effect by reinforcing tight junction assembly, which could prevent translocation of endotoxins and circulating inflammation [93, 94]. This matches with our finding of the *alsD* gene involved in the butyrate metabolism and its positive  $r_g$  with average daily gains at early stages ( $r_g$  was 0.15 ( $P_0=0.75$ ) with  $ADG_1$  and 0.29 ( $P_0=0.91$ ) with  $ADG_2$ ).

It is expected that animals grow faster when they have a greater supply of amino acids from microbial proteins synthesised in the rumen, although it has been suggested that this association is not so straightforward. Several *alr*-MG involved in the biosynthesis of phenylalanine, tyrosine and tryptophan (*aroF*, *trpC* and *trpCF*) and in the metabolism of cysteine and methionine (*mmuM*, *E4.4.1.11* and *yrrT*) or arginine (*alaA* and *E3.5.3.1*) had non-zero  $r_g$  with some growth rates traits and showed an almost entirely inconsistent pattern.

We also found evidence of a genomic association ( $P_0 < 0.85$ ) between *alr*-MG involved in sulphur metabolism and  $ADG_2$ ,  $ADG_3$  or  $ADG_4$ . These included anaerobic sulphite reductase genes (*asrB* and *asrC*), *NFU1* and *dmsB* that are involved in the synthesis of iron-sulphur clusters, and sulphhydrogenase genes (*hydB*, *hydG*) or *thiS*, *tusA*, *moeB* and *MOCS2B* that are involved in the sulphur relay system. In the rumen, the sulphur cycle plays a key role in the synthesis of amino acids (e.g., as a component for their iron/sulphur clusters). Most of the sulphur-containing amino acids (cysteine and methionine) are contained in the microbial proteins available for host

uptake, since just a few ruminal bacteria require them for their own growth [95]. The incorporation of sulphur into amino acids occurs in its sulphite form after assimilatory sulphate reduction [96], but sulphate can alternatively be used as an electron acceptor to generate H<sub>2</sub>S in dissimilatory sulphate reduction [96]. Sulphite reductases are key enzymes regulating assimilatory and dissimilatory sulphate reductions, which may be related to the synthesis of these amino acids and explain the genomic association between these *alr*-MG and longitudinal growth rate.

## Conclusions

This work contributes to filling the current gap in the knowledge of the role of microbial metabolism in the growth pattern of beef cattle. Our results indicate that the host genomics influences the functional rumen microbiome, which is an important biological factor contributing to variation in growth rate during the finishing phase of cattle. For example, host genes that influence microbial gene abundances involved in microbial growth and activity, sulphur and amino acid metabolism and transport, or lipopolysaccharides signalling are influenced by the host genomic background that underlies longitudinal average daily gains. Genomic selection for these specific microbiome functions are proposed as a strategy for breeding for an enhanced growth rate at different stages of the finishing period, although the genomic parameters to be introduced in the models need to be estimated with high accuracy. The potential benefits of using microbiome information to estimate the genetic merit of growth traits can be maximized when combined with weight records that are relatively cost-effective to collect. Our findings contribute to understanding the role of microbial metabolism in the growth trajectory of cattle and provide insights for introducing microbiome information into future genomic breeding programs to improve responses to selection for production traits by permanently shaping the rumen microbiome into a more efficient ecosystem.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-024-00887-6>.

**Additional file 1: Table S1.** Description of the animals used in the study. Experimental design displaying the number of animals within each breed, diet and experiment.

**Additional file 2: Table S2.** Heritabilities of the microbiome. Genomic parameter estimates for the 3631 *alr*-transformed microbial gene abundances and genetic parameters estimates for 100 randomly selected *alr*-transformed abundances of microbial genes when using different degrees of freedom in the chi-squared priors for the random effect variances.

**Additional file 3: Figure S1.** Robustness analysis to test the sensitivity of the heritability estimates from different prior information. (a). Four different scaled-inverse Chi-squared distributions (degrees of freedom (df0) equal to 1, 5, 10 and 20) used as priors for genomic variances in the

Bayesian estimation of variance components of 100 randomly selected *alr*-transformed microbial abundances. As an example, the graph shows the distributions used for *alr*-transformed microbial abundance of K03623. Scale parameter (S0) is equal to 0.48, calculated as  $S0 = \text{var}(y) * (\text{df0} + 1 + \text{number of traits}) * R^2$ , var (y) being the phenotypic variance of the trait (0.148), and R<sup>2</sup> being 0.5. (b). Heritability estimates (means of the marginal posterior distributions) of the 100 *alr*-transformed microbial gene abundances when using different prior information.

**Additional file 4: Figure S2.** Test of the significance of heritabilities by a permutation test. Distribution of 1000 heritability (h<sup>2</sup>) estimates when the data are permuted compared to the h<sup>2</sup> estimates with real phenotypes (in red) for 20 randomly selected *alr*-transformed microbial gene abundances.

**Additional file 5: Table S3.** Results from the bivariate genomic models between growth traits and microbiome traits. Genetic parameters estimates of *alr*-transformed abundances of microbial genes and each average daily gain when using different degrees of freedom in the inverse Wishart distributions for the genomic effects.

**Additional file 6: Figure S3.** Robustness analysis to test the sensitivity of our estimates from different prior information. Estimated 583 genomic correlations (means of the marginal posterior distributions) of the *alr*-transformed microbial gene abundances and any of the longitudinal average daily gains estimated assuming different prior information for genomic effects; this is, an inverse Wishart distribution with different prior degrees of freedom (df): 2, 3 (default), 10 and 20. Scale parameter was var (y) \* (df0 + 1 + number of traits) \* R<sup>2</sup>, var (y) being the phenotypic 2x2 (co) variance matrix between traits, and R<sup>2</sup> being 0.5. Prior distribution for the residual (co)variance was an inverse Wishart distribution with df = 5 and  $S0 = \text{var}(y) * (\text{df0} + 1 + \text{number of traits}) * R^2$ . Only genomic correlations with a probability of being higher or lower than 0 ≥ 0.85 when using default priors are displayed (n=583).

**Additional file 7: Table S4.** Identification of which rumen uncultured genomes carry the microbial genes of interest. Proteins clustered in KEGG orthologous groups (KO) of interest for the microbiome-driven breeding strategy to increase average daily gains at different stages identified in rumen uncultured genomes (RUG).

**Additional file 8: Table S5.** Description of the two subsets of *alr*-MG selected to be included in genomic evaluations. Genomic parameters of the 32 *alr*-transformed microbial gene abundances (*alr*-MG) considered in the microbiome-driven breeding strategy that aims at increasing growth rate at all stages.

## Acknowledgements

The authors are grateful to Laura Nicoll, Claire Broadbent, Lesley Purcell for the excellent technical support in the experimental trials. We thank Bin Zhao for his contribution to the identification and biological description of metagenomic data.

## Author contributions

MM-A, RR and MW conceived and designed the overall study, MM-A, JM, MW and RR designed and executed the bioinformatics and statistical analysis. ZW and MAC provided the genomic information. OG-R, AS-M, JL, C-AD, RJD, and MAC provided essential insight into microbiology, rumen metabolism and animal breeding. MM-A wrote the draft, and subsequently all authors intellectually contributed to the interpretation and presentation of the results and approved the manuscript. All authors read and approved the final manuscript.

## Funding

This project was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC BB/N01720X/1, BB/N016742/1, BB/S006567/1, and BB/S006680/1) and Genus plc. This research is based on data from experiments funded by the Scottish Government, Agriculture and Horticulture Development Board (AHDB) Beef & Lamb, Quality Meat Scotland (QMS), and Department for Environment Food & Rural Affairs (Defra). The funding bodies had no influence in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Marina Martínez-Álvaro thanks the Spanish Ministry of Science and Innovation for a

Ramon y Cajal grant (RYC2021-032618-I) funded by MCIN/AEI/<https://doi.org/10.13039/501100011033> and by European Union NextGenerationEU/PRTR.

#### Availability of data and materials

Metagenomic sequence reads for all rumen samples are available in the European Nucleotide Archive (ENA) under accession projects PRJEB31266, PRJEB21624, PRJEB10338 and PRJEB70912.

#### Declarations

##### Ethics approval and consent to participate

Animal experiments were conducted at the Beef and Sheep Research Centre of Scotland's Rural College (SRUC). The experiment was approved by the Animal Experiment Committee of SRUC and was conducted following the requirements of the UK Animals (Scientific Procedures) Act 1986.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Institute of Animal Science and Technology, Universitat Politècnica de València, 46022 Valencia, Spain. <sup>2</sup>Scotland's Rural College, Easter Bush, Edinburgh EH25 9RG, UK. <sup>3</sup>Instituto Nacional de Investigaciones Agrarias, 28040 Madrid, Spain. <sup>4</sup>Escuela de Zootecnia y Centro de Investigación en Nutrición Animal, Universidad de Costa Rica, San José 11501, Costa Rica. <sup>5</sup>Genus plc, DeForest, WI 53532, USA.

Received: 11 July 2023 Accepted: 22 February 2024

Published online: 15 March 2024

#### References

- Anderson RV, Rasby RJ, Klopstein TJ, Clark RT. An evaluation of production and economic efficiency of two beef systems from calving to slaughter. *J Anim Sci*. 2005;83:694–704.
- Gerber PJ, Steinfield H, Henderson B, Mottet A, Opio C, Dijkman J, et al. Tackling climate change through livestock—a global assessment of emissions and mitigation opportunities. Rome: Food and Agriculture Organization of the United Nations (FAO); 2013.
- Johnson KA, Johnson DE. Methane emissions from cattle. *J Anim Sci*. 1995;73:2483–92.
- Hoque MA, Arthur PF, Hiramoto K, Gilmour AR, Oikawa T. Variance components due to direct genetic, maternal genetic and permanent environmental effect for growth and feed-efficiency traits in young male Japanese Black cattle. *J Anim Breed Genet*. 2007;124:102–7.
- Brown JE, Fitzhugh HA. A comparison of nonlinear models for describing weight-age relationships in cattle. *J Anim Sci*. 1976;42:810–8.
- Fitzhugh HA Jr. Analysis of growth curves and strategies for altering their shape. *J Anim Sci*. 1976;42:1036–51.
- Johnson ZB, Brown CJ, Brown AH. Evaluation of growth patterns of beef cows. Fayetteville: Arkansas Agricultural Experimental Station. 1990
- David I, Huynh Tran VH, Gilbert H. New residual feed intake criterion for longitudinal data. *Genet Sel Evol*. 2021;53:53.
- Shirali M, Doesch-Wilson A, Knap PW, Duthie C, Kanis E, van Arendonk JAM, et al. Nitrogen excretion at different stages of growth and its association with production traits in growing pigs. *J Anim Sci*. 2012;90:1756–65.
- Yin T, König S. Genome-wide associations and detection of potential candidate genes for direct genetic and maternal genetic effects influencing dairy cattle body weight at different ages. *Genet Sel Evol*. 2019;51:4.
- Myer PR, Freetly HC, Wells JE, Smith TPL, Kuehn LA. Analysis of the gut bacterial communities in beef cattle and their association with feed intake, growth, and efficiency. *J Anim Sci*. 2017;95:3215–24.
- Lima J, Auffret MD, Stewart RD, Dewhurst RJ, Duthie CA, Snelling TJ, et al. Identification of rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion efficiency in cattle. *Front Genet*. 2019;10:701.
- Bergman EN. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev*. 1990;70:567–90.
- Loor JJ, Erolimy AA, McCann JC. Dietary impacts on rumen microbiota in beef and dairy production. *Anim Front*. 2016;6:22–9.
- Strom E, Øskov ER. The nutritive value of rumen micro-organisms in ruminants. *Br J Nutr*. 1983;50:463–70.
- Carabotti M, Scirocco A, Maselli MA, Severi C. The gut–brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann Gastroenterol*. 2015;28:203–9.
- Xu T, Tao H, Chang G, Zhang K, Xu L, Shen X. Lipopolysaccharide derived from the rumen down-regulates stearoyl-CoA desaturase 1 expression and alters fatty acid composition in the liver of dairy cows fed a high-concentrate diet. *BMC Vet Res*. 2015;11:52.
- Krajmalnik-Brown R, Ilhan Z-EE, Kang D-WW, DiBaise JK. Effects of gut microbes on nutrient absorption and energy regulation. *Nutr Clin Pract*. 2012;27:201–14.
- Vajro P, Paoletta G, Fasano A. Microbiota and gut-liver axis: their influences on obesity and obesity-related liver disease. *J Pediatr Gastroenterol Nutr*. 2013;56:461–8.
- Krause TR, Lourenco JM, Welch CB, Rothrock MJ, Callaway TR, Pringle TD. The relationship between the rumen microbiome and carcass merit in Angus steers. *J Anim Sci*. 2020;98:skaa287.
- Sjögren K, Engdahl C, Henning P, Lerner UH, Tremaroli V, Lagerquist MK, et al. The gut microbiota regulates bone mass in mice. *J Bone Miner Res*. 2012;27:1357–67.
- Grosicki GJ, Fielding RA, Lustgarten MS. Gut microbiota contribute to age-related changes in skeletal muscle size, composition, and function: biological basis for a gut–muscle axis. *Calcif Tissue Int*. 2018;102:433–42.
- Sommer F, Bäckhed F. The gut microbiota—masters of host development and physiology. *Nat Rev Microbiol*. 2013;11:227–38.
- Li F, Li C, Chen Y, Liu J, Zhang C, Irving B, et al. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome*. 2019;7:92.
- Zhang Q, Difford G, Sahana G, Løvendahl P, Lassen J, Lund MS, et al. Bayesian modeling reveals host genetics associated with rumen microbiota jointly influence methane emission in dairy cows. *ISME J*. 2020;14:2019–33.
- Bergamaschi M, Maltecca C, Schillebeeckx C, McNulty NP, Schwab C, Shull C, et al. Heritability and genome-wide association of swine gut microbiome features with growth and fatness parameters. *Sci Rep*. 2020;10:10134.
- Chen C, Huang X, Fang S, Yang H, He M, Zhao Y, et al. Contribution of host genetics to the variation of microbial composition of cecum lumen and feces in pigs. *Front Microbiol*. 2018;9:2626.
- Tabrett A, Horton MW. The influence of host genetics on the microbiome. *F1000 Research*. 2020;9:84.
- Weimer PJ, Stevenson DM, Mantovani HC, Man SLC. Host specificity of the ruminal bacterial community in the dairy cow following near-total exchange of ruminal contents. *J Dairy Sci*. 2010;93:5902–12.
- Martínez-Álvaro M, Mattock J, Auffret M, Weng Z, Duthie C-A, Dewhurst R, et al. Microbiome-driven breeding strategy potentially improves beef fatty acid profile benefiting human health and reduces methane emissions. *Microbiome*. 2022;10:166.
- Duthie C-A, Rooke JA, Troy S, Hyslop JJ, Ross DW, Waterhouse A, et al. Impact of adding nitrate or increasing the lipid content of two contrasting diets on blood methaemoglobin and performance of two breeds of finishing beef steers. *Animal*. 2016;10:786–95.
- Duthie CA, Troy SM, Hyslop JJ, Ross DW, Roehe R, Rooke JA. The effect of dietary addition of nitrate or increase in lipid concentrations, alone or in combination, on performance and methane emissions of beef cattle. *Animal*. 2018;12:280–7.
- Duthie CA, Haskell M, Hyslop JJ, Waterhouse A, Wallace RJ, Roehe R, et al. The impact of divergent breed types and diets on methane emissions, rumen characteristics and performance of finishing beef cattle. *Animal*. 2017;11:1762–71.
- Rooke JA, Wallace RJ, Duthie CA, McKain N, De Souza SM, Hyslop JJ, et al. Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. *Br J Nutr*. 2014;112:398–407.

35. Somarriba M, Lonis W, Roehe R, Macrae A, Dewhurst RJ, Duthie, Carol-Anne, et al. The effects of a composite chronic stress treatment on fear responses and attention bias in beef cattle. In: Proceedings of the 53rd Congress of the International Society for Applied Ethology (ISAE): 5–9 August 2019; Bergen. 2019.
36. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet.* 2018;103:338–48.
37. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
38. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One.* 2009;4:e5350.
39. Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques.* 2004;36:808–12.
40. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol.* 2019;37:953–61.
41. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat Commun.* 2018;9:870.
42. Mattock J, Martínez-álvarez M, Cleveland MA, Roehe R, Watson M. KOunt—a reproducible KEGG orthologue abundance workflow. *Bioinformatics.* 2023;39:btad483.
43. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90.
44. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics.* 2015;31:1674–6.
45. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119.
46. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
47. Aramaki T, Blanc-Mathieu R, Endo H, Okubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics.* 2020;36:2251–2.
48. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *BioArxiv.* 2013; <http://arxiv.org/abs/1303.3997>
49. BamDeal\_LICENSE at master BGI-shenzhen\_BamDeal GitHub. <https://github.com/BGI-shenzhen/BamDeal/tree/master>. Accessed 03 Mar 2023.
50. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
51. Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat Modelling.* 2015;15:134–58.
52. Greenacre M, Martínez-Álvarez M, Blasco A. Compositional data analysis of microbiome and any-omics datasets: a revalidation of the additive logratio transformation. *Front Microbiol.* 2021;12: 727398.
53. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414–23.
54. Blasco A. Bayesian data analysis for animal scientists: the basics. Cham: Springer; 2017.
55. Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics.* 2014;198:483–95.
56. Pérez-Rodríguez P, de los Campos G. Multitrait Bayesian shrinkage and variable selection models with the BGLR-R package. *Genetics.* 2022;222:112.
57. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News.* 2006;6:7–11.
58. Fernandes DA, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-seq. *PLoS One.* 2013;8:e67019.
59. Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, Crowley TM. A field guide for the compositional analysis of any omics data. *Gigascience.* 2019;8:giz107.
60. Martínez-Álvarez M, Auffret M, Duthie C, Dewhurst R, Cleveland M, Watson M, et al. Bovine host genome acts on rumen microbiome function linked to methane emissions. *Commun Biol.* 2022;5:350.
61. Venables W, Ripley B. *Modern applied statistics with S.* 4th ed. New York: Springer; 2002.
62. Schneeberger M, Barwick SA, Crow GH, Hammond K. Economic indices using breeding values predicted by BLUP. *J Anim Breed Genet.* 1992;109:180–7.
63. Lusk JL. Association of single nucleotide polymorphisms in the leptin gene with body weight and backfat growth curve parameters for beef cattle. *J Anim Sci.* 2007;85:1865–72.
64. Nkrumah JD, Li C, Yu J, Hansen C, Keisler DH, Moore SS. Polymorphisms in the bovine leptin promoter associated with serum leptin concentration, growth, feed intake, feeding behavior, and measures of carcass merit. *J Anim Sci.* 2005;83:20–8.
65. Duan X, An B, Du L, Chang T, Liang M, Yang BG, et al. Genome-wide association analysis of growth curve parameters in chinese simmental beef cattle. *Animals (Basel).* 2021;11:192.
66. Du L, Duan X, An B, Chang T, Liang M, Xu L, et al. Genome-Wide Association study based on random regression model reveals candidate genes associated with longitudinal data in Chinese simmental beef cattle. *Animals (Basel).* 2021;11:2524.
67. Mamedova LK, Robbins K, Johnson BJ, Bradford BJ. Tissue expression of angiopoietin-like protein 4 in cattle. *J Anim Sci.* 2010;88:124–30.
68. Fitzsimons C, Kenny DA, McGee M. Visceral organ weights, digestion and carcass characteristics of beef bulls differing in residual feed intake offered a high concentrate diet. *Animal.* 2014;8:949–59.
69. Hegarty RS. Genotype differences and their impact on digestive tract function of ruminants: a review. *Aust J Exp Agric.* 2004;44:459–67.
70. Kong RSG, Liang G, Chen Y, Stothard P, Guan LL. Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. *BMC Genomics.* 2016;17:592.
71. Meyer AM, Hess BW, Paisley SI, Du M, Caton JS. Small intestinal growth measures are correlated with feed efficiency in market weight cattle, despite minimal effects of maternal nutrition during early to midgestation. *J Anim Sci.* 2014;92:3855–67.
72. Montanholi Y, Fontoura A, Swanson K, Coomber B, Yamashiro S, Miller S. Small intestine histomorphometry of beef cattle with divergent feed efficiency. *Acta Vet Scand.* 2013;55:9.
73. Snelling TJ, Auffret MD, Duthie C-A, Stewart RD, Watson M, Dewhurst RJ, et al. Temporal stability of the rumen microbiota in beef cattle, and response to diet and supplements. *Anim Microbiome.* 2019;1:16.
74. Roehe R, Dewhurst RJ, Duthie CA, Rooke JA, McKain N, Ross DW, et al. Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 2016;12: e1005846.
75. Blasco A. *Mejora genética animal.* Madrid: Editorial Síntesis; 2021.
76. Christensen OF, Börner V, Varona L, Legarra A. Genetic evaluation including intermediate omics features. *Genetics.* 2021;219:iyab130.
77. Zhao T, Zeng J, Cheng H. Extend mixed models to multilayer neural networks for genomic prediction including intermediate omics data. *Genetics.* 2022;221:iyac034.
78. Rogel-Gaillard C, Estellé J, Borey M, Larzul C. Interplay between host genetics and the dynamics of the gut microbiota in pigs. In: Proceedings of the 72nd Annual Meeting of European Association of Animal Production: 30 August–4 September 2021; Davos. 2021.
79. Huhtanen P, Ramin M, Cabezas-García EH. Effects of ruminal digesta retention time on methane emissions: a modelling approach. *Anim Prod Sci.* 2016;56:501–6.
80. Reeves PR. Role of O-antigen variation in the immune response. *Trends Microbiol.* 1995;3:381–6.
81. Amar J, Chabo C, Waget A, Klopp P, Vachoux C, Bermúdez-Humarán LG, et al. Intestinal mucosal adherence and translocation of commensal bacteria at the early onset of type 2 diabetes: molecular mechanisms and probiotic treatment. *EMBO Mol Med.* 2011;3:559–72.
82. Zhao L, Hu P, Zhou Y, Purohit J, Hwang D. NOD1 activation induces proinflammatory gene expression and insulin resistance in 3T3-L1 adipocytes. *Am J Physiol Endocrinol Metab.* 2011;301:E587–98.

83. Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, et al. Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes*. 2007;56:1761–72.
84. Hotamisligil GS. Inflammation and metabolic disorders. *Nature*. 2006;444:860–7.
85. Brun P, Castagliuolo I, Di Leo V, Buda A, Pinzani M, Palù G, et al. Increased intestinal permeability in obese mice: new evidence in the pathogenesis of nonalcoholic steatohepatitis. *Am J Physiol Gastrointest Liver Physiol*. 2007;292:G518–25.
86. Cani PD, Bibiloni R, Knauf C, Waget A, Neyrinck AM, Delzenne NM, et al. Changes in gut microbiota control metabolic diet-induced obesity and diabetes in mice. *Diabetes*. 2008;57:1470–81.
87. Cani P, Delzenne N. The role of the gut microbiota in energy metabolism and metabolic disease. *Curr Pharm Des*. 2009;15:1546–58.
88. Liang H, Hussey SE, Sanchez-Avila A, Tantiwong P, Musi N. Effect of lipopolysaccharide on inflammation and insulin action in human muscle. *PLoS One*. 2013;8:e63983.
89. Dehghan M, Mente A, Zhang X, Swaminathan S, Li W, Mohan V, et al. Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study. *Lancet*. 2017;390:2050–62.
90. Doddapattar P, Gandhi C, Prakash P, Dhanesha N, Grumbach IM, Dailey ME, et al. Fibronectin splicing variants containing extra domain A promote atherosclerosis in mice through toll-like receptor 4. *Arterioscler Thromb Vasc Biol*. 2015;35:2391–400.
91. Michelsen KS, Wong MH, Shah PK, Zhang W, Yano J, Doherty TM, et al. Lack of toll-like receptor 4 or myeloid differentiation factor 88 reduces atherosclerosis and alters plaque phenotype in mice deficient in apolipoprotein E. *Proc Natl Acad Sci USA*. 2004;101:10679–84.
92. Knuefermann P, Nemoto S, Misra A, Nozaki N, Defreitas G, Goyert SM, et al. CD14-deficient mice are protected against lipopolysaccharide-induced cardiac inflammation and left ventricular dysfunction. *Circulation*. 2002;106:2608–15.
93. Peng L, Li ZR, Green RS, Holzman IR, Lin J. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. *J Nutr*. 2009;139:1619–25.
94. Cox MA, Jackson J, Stanton M, Rojas-Triana A, Bober L, Lavery M, et al. Short-chain fatty acids act as antiinflammatory mediators by regulating prostaglandin E2 and cytokines. *World J Gastroenterol*. 2009;15:5549–57.
95. Pittman KA, Bryant MP. Peptides and other nitrogen sources for growth of bacteroides. *J Bacteriol*. 1964;88:401–10.
96. Kandyliis K. The role of sulphur in ruminant nutrition. A review. *Livest Prod Sci*. 1984;11:611–24.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.