



HAL
open science

Predicting feed efficiency traits in growing lambs from their ruminal microbiota

Quentin Le Graverand, Christel Marie-Etancelin, Annabelle Meynadier,
Jean-Louis Weisbecker, D Marcon, Flavie Tortereau

► **To cite this version:**

Quentin Le Graverand, Christel Marie-Etancelin, Annabelle Meynadier, Jean-Louis Weisbecker, D Marcon, et al.. Predicting feed efficiency traits in growing lambs from their ruminal microbiota. *Animal*, 2023, 17 (6), 10.1016/j.animal.2023.100824 . hal-04173640

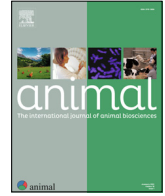
HAL Id: hal-04173640

<https://hal.science/hal-04173640v1>

Submitted on 29 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Predicting feed efficiency traits in growing lambs from their ruminal microbiota



Q. Le Graverand^{a,*}, C. Marie-Etancelin^a, A. Meynadier^a, J.-L. Weisbecker^a, D. Marcon^b, F. Tortereau^a

^aGenPhySE, Université de Toulouse, INRAE, ENVT, 24 Chemin de Borde-Rouge-Auzeville CS 52627, F-31326 Castanet-Tolosan, France

^bINRAE, Unité Expérimentale P3R, Domaine de la Sapinière, F-18390 Osmoy, France

ARTICLE INFO

Article history:

Received 13 December 2022

Revised 7 April 2023

Accepted 14 April 2023

Available online 26 April 2023

Keywords:

Meat sheep

Microbiome

Multivariate analyses

Residual feed intake

Rumen

ABSTRACT

Selecting feed-efficient sheep could improve the sustainability of this livestock production. However, most sheep breeding companies cannot afford to record feed intake to select feed-efficient animals. Past studies underlined the potential of omics data, including microbiota metabarcoding data, as proxies for feed efficiency. The study involved 277 Romane lambs from two lines divergently selected for residual feed intake (**RFI**). There were two objectives: check the consequences of selecting for feed efficiency over the rumen microbiota, and assess the predictive ability of the rumen microbiota for host traits. The study assessed two contrasting diets (concentrate diet and mixed diet) and two microbial groups (prokaryotes and eukaryotes). Discriminant analyses did not highlight any significant effect of sheep selection for residual feed intake on the rumen microbiota composition. Indeed, prokaryotic and eukaryotic microbiota compositions poorly discriminated the RFI lines, with averaged balanced error rates ranging from 45% to 55%. Correlations between host traits (feed efficiency and production traits) and their predictions from microbiota data varied between -0.07 and 0.56 , depending on the trait, diet and sequencing. Feed intake was the most accurately predicted trait. However, predictions from fixed effects and BW were more accurate than or as accurate as predictions from the microbiota. Environmental effects can greatly affect the variability of microbiota compositions. Considering batch and environmental effects should be paramount when the predictive ability of the microbiota is assessed. This study argues why metabarcoding the rumen microbiota is not the best way to predict meat sheep production traits: fixed effects and BW were more cost-effective proxies and they led to similar or better predictive accuracies than microbiota metabarcoding (16S and 18S sequencing).

© 2023 The Author(s). Published by Elsevier B.V. on behalf of The Animal Consortium. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Implications

Selecting sheep for feed efficiency is often impossible for small companies because phenotyping of individual feed intake is expensive: search for predictors of efficiency is therefore crucial. In this study, the predictive ability of rumen microbiota for feed intake and in a lesser extend for efficiency was proven using 277 growing Romane lambs under two diets. But metabarcoding the rumen microbiota is not the best way to predict feed efficiency traits: predictions from BW and systematic effects were as much or more accurate. Selection for feed efficiency did not seem to modify the microbiota abundancies.

Introduction

Feed efficiency is a key factor in the livestock industry, which encompasses many environmental, economic and societal stakes. Improving ruminant feed efficiency can decrease feed inputs, help limit feed-food competition and mitigate environmental impacts from sheep breeding. Additionally, lowering feed intake and feeding costs can improve farmers' profitability (Cruz et al., 2010; Lima et al., 2017). Thus, selecting feed-efficient animals is one way to improve sheep industry sustainability.

Residual Feed Intake (**RFI**) is a feed efficiency criterion proposed by Koch et al. (1963). This criterion is commonly used in genetics as RFI is a heritable trait: the pooled RFI heritability over five studies in small ruminants has been estimated at 0.32 ± 0.15 (Mucha et al., 2022). But whatever the feed efficiency criteria, it requires knowing the amount of ingested feed at the individual level, which remains too expensive for most sheep breeding companies.

* Corresponding author.

E-mail address: quentin.le-graverand@inrae.fr (Q. Le Graverand).

Identifying proxies could be a solution to circumvent the expensive feed intake recording in order to widen the selection of feed-efficient sheep. In cattle or sheep, ruminal or blood metabolites are associated with feed efficiency (Artegoitia et al., 2017; Goldansaz et al., 2020; Touitou et al., 2022). Previous studies also relied on metabarcoding data: both 18S (Zhang et al., 2020; Clemmons et al., 2021) and 16S sequencing (Ellison et al., 2019; Marie-Etancelin et al., 2021) allowed to pinpoint microbial clusters as biomarkers for ruminant feed efficiency.

While looking for sheep feed efficiency biomarkers with microbiota data (16S or 18S), most studies included few animals (<50). These were often selected from a larger population, according to their feed efficiency phenotype. A gap of knowledge remains around the accuracy of feed efficiency predictions from microbiota data when a large number of lambs are used, rather than a small subset having extreme phenotypes. Feed efficiency prediction accuracies may also vary with the diet (concentrate or mixed, for instance). Furthermore, whether a generalisable microbial signature can be identified when sheep are raised across different years is unclear. A study over several years is paramount to assess the generalisability of predictors: environmental effects, such as diet, farming practices, or climate conditions can affect the microbiota composition (Henderson et al., 2015; Belanche et al., 2019; Li et al., 2019; Anderson et al., 2021; Marie-Etancelin et al., 2021; Wei et al., 2021).

The present study focused on male lambs from two Romane lines divergently selected on RFI under a concentrate diet. In this population, the heritability of RFI was found to be moderate ($h^2 = 0.45 \pm 0.08$, Tortereau et al., 2020).

We hypothesised that selecting for feed efficiency might shape the rumen microbiota since it contributes to lamb's digestion and that the rumen microbiota can be considered as a proxy for feed efficiency. Therefore, our objectives were to determine whether selecting RFI modified the rumen microbiota of lambs and whether rumen microbiota can be considered as a reasonable proxy of feed efficiency. These investigations were conducted under two contrasting diets.

Material and methods

Animals

The present study was conducted from 2018 to 2020 on Romane male lambs, at the INRAE Experimental Unit P3R (UE 332 agreement D18-174-01; Osmoy, France, <https://doi.org/10.15454/1.5483259352597417E12>).

Divergent lines

The studied animals were part of a larger design of divergent selection on RFI. Since 2009, feed efficiency was phenotyped in Romane male lambs under a 100% concentrate diet. In 2015, a divergent selection experiment started: two divergent lines were selected for an increased RFI (least feed-efficient line, **RFI+**) or a decreased RFI (most efficient line, **RFI-**) as described in Tortereau et al. (2020). Briefly, animals were divergently selected for RFI under a 100% concentrate diet. For the genetic evaluation, the RFIs of 1900 male and female lambs phenotyped since 2009 were calculated according to Tortereau et al. (2020). Estimated breeding values (**EBVs**) were then computed with PEST software (Groeneveld et al., 1990), considering a heritability of 0.45 (Tortereau et al., 2020) and a pedigree of 6 419 sheep. The mixed model accounted for the pen, suckling method (maternal or artificial), litter size, year of phenotyping and sex as fixed effects.

Every year since 2015, ten to fourteen rams with the lowest and highest RFI breeding values under a concentrate diet were selected

for breeding to produce the next generation of selection. Mating was planned within divergent lines: rams with extreme RFI values were selected and mated with ewes chosen to minimise inbreeding. Two years were necessary to complete one generation of selection.

Husbandry and phenotyping protocol

The present study focused only on a subset of Romane male lambs from the divergent lines described above. Part of the second generation was phenotyped in 2018 (103 lambs), while the third generation was phenotyped in 2019 (101 lambs) and 2020 (73 lambs). Part of the lambs phenotyped in 2018 sired lambs studied in 2020. On average, 7.7 male lambs shared the same sire in the study.

The experimental design is illustrated in Fig. 1. Animals were reared indoors with wood chip litter and fed successively with two different diets: the first one with concentrates exclusively and the second one with a mixed diet.

Phenotyping under a concentrate diet – Every year, lambs born in the experimental unit were gathered in the experimental barn at approximately 10 weeks of age. Then, lambs were adapted to a 100% concentrate *ad libitum* diet and to its distribution by automatic feeders (for feed nutritional values, see Supplementary Table S1-A).

Feed intake was recorded using automatic concentrate feeders during six weeks, between 17 and 23 weeks of age on average (Fig. 1). Depending on the year, lambs were grouped into four to six pens of approximately 20 lambs with homogeneous BWs to prevent fights. One concentrate feeder was available per pen. From 2018 to 2020, a total of 277 lambs were phenotyped under a concentrate diet.

To account for annual variations in feed compositions, net energy intake was computed from feed intake and feed energy densities. Energy densities of feed were estimated with the INRA 2007 system (Baumont et al., 2007) in megajoules of net energy. The average daily energy intake of concentrate (**ADEI_C**) was calculated as the average of daily energy intakes over the six-week period. At the end of the recording period, back ultrasound measurements were carried out to assess the *longissimus dorsi* muscle depth (**MD_C**) and back fat thickness (**BFT_C**). Starting and final BWs were recorded and used to compute the average daily gain (**ADG_C**) over six weeks. Two feed efficiency traits were computed: the feed conversion ratio (**FCR_C**) as the ratio of ADEI_C over ADG_C and, REI_C as the residuals of the linear regression of ADEI_C over characterised energy sinks (Eq. (1)):

$$\text{ADEI}_C = \mu_C + \beta_{1,C} \text{ADG}_C + \beta_{2,C} \text{final BW}_C^{0.75} + \beta_{3,C} \text{MD}_C + \beta_{4,C} \text{BFT}_C + \text{REI}_C \quad (1)$$

where μ_C is the mean ADEI_C. $\beta_{1,C}$ to $\beta_{4,C}$ respectively stand for the four following covariate effects: ADG_C, final metabolic weight (final BW_C^{0.75}), MD_C and BFT_C. Finally, REI_C is the residual energy intake expressed as megajoules of ingested net energy per day.

Phenotyping under a mixed diet – After the first six-week trial, animals were adapted for 6 weeks to a mixed diet delivered by automatic feeders. In 2018, no restricted concentrate feeder was available; thus, forage feeders delivered an *ad libitum* total mixed ration (33% concentrates, 67% hay). In 2019 and 2020, automatic concentrate feeders ensured restricted access to concentrates (up to 700 g/day), and forage feeders delivered *ad libitum* hay separately (see Supplementary Table S1-B for feed nutritional values).

Due to facility limitations (maximum of 35–40 lambs simultaneously), only lambs having extreme RFI EBVs under a concentrate diet were then phenotyped under a mixed diet. In 2018 and 2019, animals were split during two different periods per year: during summer (period 1, from 29 to 35 weeks of age) or during fall (pe-

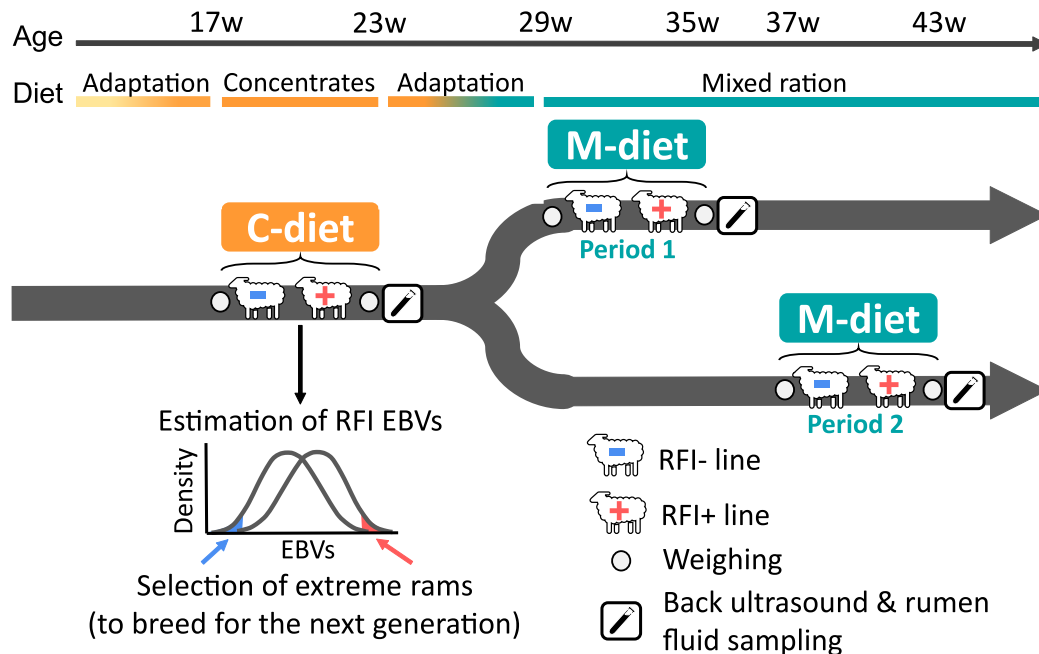


Fig. 1. Experimental design of lamb phenotyping. Abbreviations: C-diet: 100% concentrate diet; EBV: estimated breeding value; M-diet: mixed diet including 2/3 of forages and 1/3 of concentrates; RFI: residual feed intake; RFI-: efficient line with a low residual feed intake; RFI+: inefficient line with a high residual feed intake; w: weeks. Animals were successively phenotyped under the concentrate and mixed diets. Under a mixed diet, two groups of animals were phenotyped at two different periods each year. EBVs of RFI were computed under the concentrate diet. Rams having low RFI EBVs under a concentrate diet were bred to obtain the RFI- line, while rams with high RFI EBVs were selected to obtain the RFI+ line. Average ages are expressed in weeks.

riod 2, from 37 to 43 weeks of age) (Fig. 1). In 2020, only one group was phenotyped during fall (period 2). Within one period, animals were allocated by BW into two pens with 16 animals per pen, on average. Each pen was equipped with two forage feeders and one concentrate feeder in 2019 and 2020. Over the three years, 166 lambs were phenotyped under a mixed diet.

$ADEI_M$, ADG_M and REI_M were estimated under the mixed diet as it was done under the concentrate diet before. However, no FCR_M was calculated under the mixed diet because some individuals had a null or negative ADG_M over this 6-week period.

Data cleaning and overview of host phenotypes

Under one diet, outliers were identified based on feed efficiency (REI, FCR), energy intake (ADEI), growth rate (ADG) and body composition (MD, BFT) traits. The animal was removed from the diet data subset when one of the phenotypes was outside the range $[\mu - 3 SD; \mu + 3 SD]$, where μ is the mean phenotype. Eight animals were filtered out under a concentrate diet (out of 277 lambs), and three were filtered out under a mixed diet (out of 166 animals).

Linear regressions were fitted with the `lm` function included in R software (version 4.1.2) (R Core Team, 2021). All traits recorded under a concentrate diet ($Trait_C$) were regressed as follows (Eq. (2)):

$$Trait_C = \mu_{Trait_C} + \beta_{1,C} \text{ line} + \beta_{2,C} \text{ age} + \beta_{3,C} \text{ suckling} + \beta_{4,C} \text{ year} + \beta_{5,C} \text{ pen|year} + \varepsilon_C \quad (2)$$

where μ_{Trait_C} denotes the trait mean. $\beta_{1,C}$ to $\beta_{5,C}$ respectively stand for the effects of the line, age, suckling, year and pen nested in the year. Finally, ε_C denotes the model residuals.

Traits recorded under a mixed diet were regressed similarly (Eq. (3)):

$$Trait_M = \mu_{Trait_M} + \beta_{1,M} \text{ line} + \beta_{2,M} \text{ age} + \beta_{3,M} \text{ year} + \beta_{4,M} \text{ period|year} + \beta_{5,M} \text{ pen|period|year} + \varepsilon_M \quad (3)$$

where μ_{Trait_M} denotes the trait mean. $\beta_{1,M}$ to $\beta_{5,M}$ respectively stand for the effects of the line, age, year, period nested in the year, and pen nested in the period and the year. ε_M denotes the model residuals. Then, regression solutions (Eqs. (2) and (3)) were used to compute least square means per line and host trait with the `emmeans` R package (version 1.7.3) (Lenth, 2022). Comparison of traits between RFI lines were carried out with the Tukey test. To account for multiple testing, P -values were corrected per diet with Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) (seven traits under a concentrate diet, six under a mixed diet).

16S and 18S ribosomal RNA gene sequencing

Every year, rumen fluid samples were collected at the end of each feed intake recording trial. Sampling was carried out by trained staff, with a medical gastric tube coupled to a vacuum pump. Rumenal samples were immediately frozen in liquid nitrogen.

DNA extraction and sequencing

DNA extraction, amplification and sequencing of microbial fluid samples were carried out in two different batches: one batch with 2018 and 2019 samples and another with 2020 samples. Within each batch, a bead-beating step was carried out with a FastPrep device (MP Biomedicals, Illkirch, France). Then, DNA was extracted with the QIAamp DNA Stool Mini Kit (Qiagen Ltd, West Sussex, UK) from 85 μ L of ruminal fluid.

The V4-V5 region of the 16S ribosomal RNA gene was amplified with the forward 515F (5'-CTTCCCTACACGACGCTCTTCCGATCTGT GYACGCMGCCGCGTA-3') and reverse 928R primers (5'-GGAGTT CAGACGTGTGCTCTTCCGATCTCCCCGYCAATTCMTTTRAGT-3') (Wang and Qian, 2009) for 30 PCR cycles. To barcode samples, an index of six base pairs was added to 928R primers during a second amplification (12 cycles) with forward (5'-GTGYCAGCMGCC-3') and reverse primers (5'-CCCCGYCAATT-3'), plus adapters. Overlap-

ping paired-end reads of 250 base pairs were produced and aligned to obtain full-length reads with Illumina MiSeq technology (Illumina, San Diego, CA, USA).

The V4 region of the 18S ribosomal RNA gene was amplified with the forward 566F (5'-CTTCCCTACAGCAGCTCTCCGATCTCAGCAGCCGCGTAATCC-3') and reverse 1200R primers (5'-GGA GTTCAGACGTGTGCTCTCCGATCTCCCGTGTGAGTCAAATTAAGC-3') (Hadziavdic et al., 2014) for 30 PCR cycles. To barcode samples, an index of six base pairs was added to 1200R primers during a second amplification (12 cycles) with forward (5'-CAGCAGCCGCGGTAATCC-3') and reverse primers (5'-CCCGTGTGAGTCAAATTAAGC-3'). Overlapping and non-overlapping paired-end reads of 250 base pairs were produced.

After multiplexing and amplifications, 16S and 18S reads were purified and loaded on an Illumina MiSeq cartridge (Illumina, San Diego, CA, USA) to be sequenced at the Genomic and Transcriptomic Platform (INRAE, Toulouse, France).

Sequence bioinformatic processing

After sequencing, 16S and 18S reads were processed separately with FROGS tools (version 4.0.1) (Escudié et al., 2018). However, within both sets of sequences, reads from different diets and sequencing batches were treated together (step 1, Fig. 2). Read

processing was performed with the following pipeline: (i) demultiplexing; (ii) reconstruction of 18S amplicon sequences only, (iii) quality control of amplicons based on the presence of primers, ambiguous bases and size (>380 and <500 base pairs for 16S amplicons; >200 and <490 base pairs for 18S amplicons); (iv) clustering into operational taxonomic units (OTUs) with the Swarm algorithm using a difference of 1 between sequences in each aggregation step (Mahé et al., 2014); (v) chimaera removal; (vi) prefiltering, by removing OTUs aggregating less than 0.005% of all sequences; and (vii) taxonomic affiliation with BLAST+ (Camacho et al., 2009) and the Silva 132 16S reference database for bacteria and archaea, or the Silva 138.1 18S database for fungi and protozoa (Quast et al., 2013).

Sequencing data cleaning and overview

For data cleaning and subsequent analyses (steps 2–6, Fig. 2), four distinct compositional datasets were considered: one per diet (concentrate or mixed diet) and amplified gene (16S or 18S). Then, sequencing data were filtered at the sample and OTU levels. Samples with low sequencing depths were associated with low OTU richness and small variations between the number of observed OTUs (Supplementary Fig. S1). Thus, samples with a depth smaller than 7 500 reads were discarded (step 2, Fig. 2). OTU filtering was

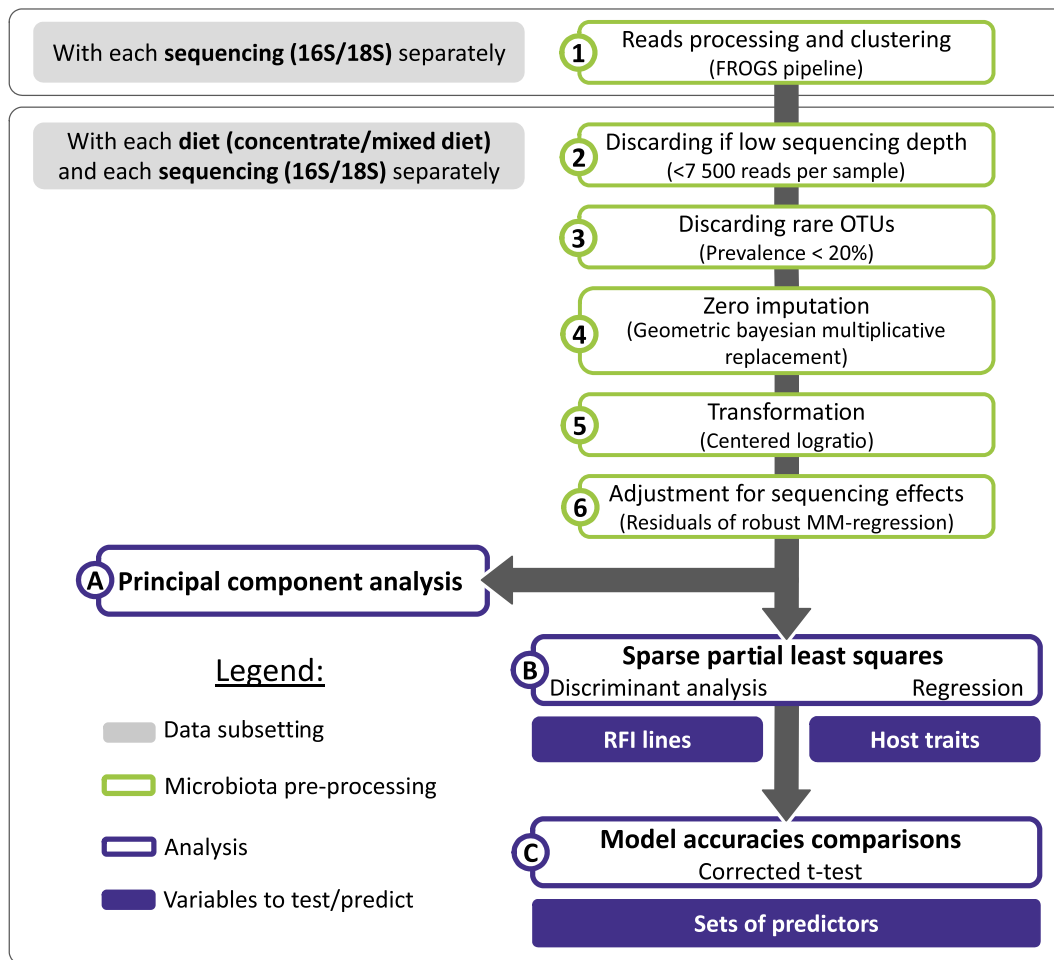


Fig. 2. Workflow from lamb microbiota data handling, to statistical learning and inference. Abbreviations: OTU: operational taxonomic unit; PCA: principal component analysis; RFI: residual feed intake; sPLSDA: sparse partial least squares discriminant analysis; sPLSR: sparse partial least squares regression. Microbiota preprocessing steps are numbered from 1 to 6. Microbiota compositions were studied at the OTU level. Main variation factors of compositions were identified with PCAs. Two kinds of predictive models were used: sPLSDA and sPLSR. Accuracies were compared with corrected *t*-tests. RFI lines were selected for a decreased or an increased estimated breeding value of RFI. Host traits included feed efficiency, feed intake, growth and body composition traits.

performed by removing OTUs with a prevalence lower than 20% under the considered diet (step 3, Fig. 2). The numbers of sequencing reads and OTUs before and after data cleaning are reported in [Supplementary Table S2](#).

To give an overview of microbiota compositions, the sequence of OTU seeds and descriptive statistics are available in [Supplementary Table S3](#) (16S data) and [Supplementary Table S4](#) (18S data). Phyla mean relative abundances were also computed after regrouping OTUs by phyla with the phyloseq R package (McMurdie and Holmes, 2013). Then, at the OTU level, unsupervised learning was used to identify the main variation patterns driving microorganisms' abundances (analysis A, Fig. 2). Prior to principal component analyses (PCAs), several preprocessing steps were carried out to account for compositionality and sequencing effects. Null abundances of OTUs were imputed with the Geometric Bayesian Multiplicative replacement procedure (step 4, Fig. 2) (Martín-Fernández et al., 2015). Then, a Centered LogRatio (CLR) transformation was carried out with OTU abundances (step 5, Fig. 2). CLR coordinate values were adjusted univariately for sequencing effects with a robust MM regression (Maechler et al., 2017) (Eq. (4), step 6, Fig. 2). Robust MM regression was preferred as it is less sensitive to outlying values than least squares linear regressions.

$$CLR(X)_i = \beta_{0,i} + \beta_{1,i}seq_{depth} + \beta_{2,i}seq_{batch} + \beta_{3,i}seq_{batch|plate} + \varepsilon_i \quad (4)$$

where $CLR(X)_i$ stands for the CLR values of the i^{th} OTU. Then, $\beta_{0,i}$ stands for the intercept. For the i^{th} OTU, $\beta_{1,i}$ to $\beta_{3,i}$ represent the effects of sequencing depth, sequencing batches ($n = 2$) and sequencing plates nested in batches ($n = 5$), respectively. Finally, ε_i stands for residuals. Residuals were used as adjusted CLR values to define the principal components (PCs) with the mixOmics R package (version 6.18.1) (Rohart et al., 2017).

Predictions with sparse partial least squares discriminant analyses and regressions

Sparse Partial Least Squares Discriminant Analyses (sPLSDA) or Regressions (sPLSR) were performed to assess how well the RFI lines and host traits could be predicted with microbiota data by using multivariate analyses (analysis B, Fig. 2).

Preprocessing

sPLSDA and sPLSR were carried out with the mixOmics R package (version 6.18.1) (Rohart et al., 2017) on the CLR values adjusted for sequencing effects (steps 4–6, Fig. 2).

Sparse partial least squares discriminant analysis

sPLSDA was carried out to predict the RFI lines to check whether ruminal microbiota can predict efficient vs. inefficient animals.

With sPLSDA, two hyperparameters were tuned through a grid search and cross-validations: the number of components and the number of selected variables per component. The retention of 1 to 20 components was tested, with a step of 1. We tested selecting 2 to 9 variables per component, with a step of 1 and then 10 to the maximum number of variables with a step of 10. To assess hyperparameters, 5-fold cross-validations were repeated 50 times. A joint factor denoting the RFI line and contemporaneous group was used to stratify the cross-validation sets thanks to the createMultiFolds function (Kuhn, 2008): *i.e.* every line, pen and year are represented in all sets and in similar proportions. The tuning criterion was the Balanced Error Rate (BER), calculated as the average of error rates over the RFI lines (Eq. (5)):

$$BER = \frac{1}{2} \times \left(\frac{a}{a+b} + \frac{c}{c+d} \right) \quad (5)$$

with a, b, c and d being the false predicted RFI-, the true predicted RFI+, the false predicted RFI+ and the true predicted RFI-, respectively.

The following rule of thumb was used to select the final hyperparameters: a more complex model (*i.e.*, with more components and/or more selected variables) was retained if the averaged BER decreased by one standard error (Kuhn and Johnson, 2013). The average BERs of the selected models are reported.

Sparse partial least squares regression

sPLSR was used to predict continuous lamb traits: feed efficiency (REI, FCR), energy intake (ADEI), growth rate (ADG) and body composition traits (MD, BFT). The goal was to check how accurately feed efficiency and its components could be predicted under each diet.

With sPLSR, the same two hyperparameters as sPLSDA had to be tuned. Therefore, the same grid search and cross-validation strategies were implemented. The tuning criterion was the RMSE, which is the square root of the mean of squared errors. A more complex model (*i.e.* with more components and/or more variables per component) was retained if the averaged RMSE decreased by one standard error. Different accuracy criteria were calculated and averaged over folds and repetitions: Pearson and Spearman correlations between the true and predicted values of selected models, RMSE, mean absolute error, coefficient of determination, and bias. All metrics are reported in [Supplementary Table S5](#).

Comparing prediction accuracies from different sets of predictors

To predict lamb lines or traits, three sets of predictors were considered: (i) The first set included either 16S- or 18S-adjusted CLR values. (ii) The second set included systematic effects: suckling method (concentrate diet only), phenotyping period (mixed diet only), year, pen, age and final BW. (iii) Finally, the third set of predictors included the systematic effects and CLR values of 16S or 18S OTUs.

Models fitted on different sets of predictors were compared per predicted trait and data subset. Differences between model accuracies were tested with corrected *t*-tests (Bouckaert and Frank, 2004) (Analysis C, Fig. 2). Then, *P*-values were adjusted with Benjamini-Hochberg procedure per predicted trait and data subset.

Results

Data overview

Phenotypic lamb traits

For feed efficiency traits and their components, descriptive statistics are provided in [Table 1](#), along with least square means computed for the two RFI divergent lines.

Under a concentrate diet (17 to 23 weeks of age), lambs ingested an average of 10.98 MJ of net energy per day and grew by 330.91 g/day. It resulted in an average FCR_c of 0.03 MJ/g. Under a mixed diet (29 to 35, or 37 to 43 weeks of age), sheep consumed an average of 8.06 MJ of net energy per day. No FCR_c was computed with a mixed diet because of some negative and null ADG_M values. Regardless of the diet, RFI- animals had significantly lower intakes: with a decrease of -0.95 MJ of net energy/day under a concentrate diet and -0.46 MJ/day under a mixed diet.

Considering the animals under study, the divergence between the two RFI lines was equal to 1.86 genetic SDs of RFI_c least square means. Regardless of the diet, RFI- animals were more feed efficient, with a difference in residual energy intake between the RFI lines equal to 0.69 MJ of net energy/day under a concentrate diet against 0.33 MJ/day under a mixed diet. Under a concentrate diet, the difference in FCR_c was also significant, with RFI+ animals

Table 1
Statistical summary of lamb traits and least square means for residual feed intake lines.

Diet ¹ (n ²)	Trait ³	Population statistics ⁴				Line least square means ⁵		
		Mean	SD	Min	Max	RFI-	RFI+	Adj. P ⁶
C-diet (269)	REI _C , MJ/day	0.000	0.756	-1.701	2.070	-0.353	0.337	<0.001
	FCR _C , MJ/g	0.034	0.006	0.021	0.055	0.032	0.035	<0.001
	ADEI _C , MJ/day	10.98	1.38	7.37	14.14	10.50	11.45	<0.001
	Final BW _C , kg	56.62	6.71	39.74	78.85	55.69	57.49	<0.001
	ADG _C , g/day	331	58	150	501	333	331	0.927
	BFT _C , mm	5.7	0.8	3.9	8.3	5.8	5.8	0.927
	MD _C , mm	28.1	2.2	22.1	35.0	28.0	28.6	0.029
M-diet (163)	REI _M , MJ/day	0.000	0.899	-2.740	2.222	-0.203	0.125	0.013
	ADEI _M , MJ/day	8.06	1.22	4.31	10.50	7.77	8.23	0.008
	Final BW _M , kg	64.49	5.80	47.80	77.90	63.74	64.72	0.196
	ADG _M , g/day	123	67	-58	264	121	128	0.386
	BFT _M , mm	4.5	0.8	2.9	6.9	4.6	4.4	0.196
	MD _M , mm	27.3	2.4	20.8	34.0	27.0	27.5	0.196

Abbreviations: ADEI: average daily energy intake; ADG: average daily gain; BFT: back fat thickness; C-diet: concentrate diet; FCR: feed conversion ratio; M-diet: mixed diet; MD: muscle depth; REI: residual energy intake; RFI-: efficient line with a low residual feed intake; RFI+: inefficient line with a high residual feed intake.

¹ C-diet: sheep were fed a 100% concentrate diet; M-diet: sheep were fed a mixed diet including 2/3 of forage and 1/3 of concentrate.

² Number of animals.

³ Subscripts denote the trait diet: C for the C-diet; M for the M-diet.

⁴ Descriptive statistics based on raw phenotypes, without adjusting for confounding effects.

⁵ Least squares means were computed for lamb lines divergently selected for residual feed intake. Traits were adjusted for the age, suckling method (under the C-diet only), year, pen and phenotyping period (under the M-diet only).

⁶ Tukey's test adjusted *P*-values. The Benjamini-Hochberg procedure was applied per diet.

ingesting 0.31 MJ of net energy/day more than RFI- animals to grow by 100 g/day.

Regardless of the diet, no significant difference was found between the two RFI lines for growth (ADG_C or ADG_M) and most body composition traits (BFT_C, BFT_M, or MD_M). Only MD_C significantly differed between the two lines: RFI- animals had a lower MD by 0.56 mm. Regarding the final BW under a concentrate diet, the RFI- line was significantly lighter (-1.8 kg on average, at approximately 23 weeks of age). Later in life, under the mixed diet, the difference between the BWs of both lines was no longer significant (adjusted *P*-value = 0.196).

Microbiota data

After data cleaning, 582 and 1148 OTUs of 16S sequencing remained under the concentrate and mixed diets, respectively, whereas for 18S sequencing, 124 and 183 OTUs were kept under the two diets, respectively.

OTUs were clustered according to their phylum affiliations, and mean phylum relative abundances are reported in Fig. 3. With 16S sequencing, the most abundant phyla were *Bacteroidetes* under a concentrate diet (Fig. 3A) and *Firmicutes* under a mixed diet (Fig. 3C), while *Euryarchaeota* was the third most abundant phylum under both diets. With 18S sequencing and regardless of the diet (Fig. 3B and D), the *Ciliophora* phylum was largely predominant, with relative abundances ranging between 87 and 96%.

PCA allowed the identification of the main variation patterns of microbiota compositions (Fig. 4). Under a concentrate diet, the variability of the second PC was associated with the year of lamb phenotyping with 16S data (5% of the variance explained, Fig. 4A). Under a mixed diet and regardless of the sequencing, the first PC variability seemed to be mostly tied to the phenotyping period and then the year to a lesser extent (Fig. 4C, D). None of the recorded variables (age at sampling, age at weaning, sampling order, pen, suckling method or RFI line) were associated with the other PCs. No main factor of variability could be identified for 18S data under a concentrate diet (Fig. 4B).

Predictive accuracy of sparse partial least squares models

Predicting the residual feed intake lines

Regardless of the predictors, diet and sequencing, the BERs of sPLSDA were always high (from 0.35 to 0.55) when predicting RFI lines (Table 2).

Systematic effects included fixed effects and final BW. Prediction of the RFI line from systematic effects was never significantly worse than the prediction from microbiota data (*i.e.*, with a higher averaged BER). The addition of the microbiota data to the fixed effect and BW data never decreased BERs significantly.

Predicting host traits

Host traits were predicted with sPLSR by using fixed effects and BW and/or microbiota data. Pearson correlations between actual traits and predictions are reported in Fig. 5, and other metrics are provided in Supplementary Table S5.

Predictions from microbiota data only – Under a concentrate diet, when host feed efficiency was predicted from 16S (Fig. 5A) or 18S (Fig. 5B) data, the average correlations between actual traits and predictions were almost null for REI_C (0.11 for 16S compared to 0.06 for 18S data) or low to moderate for FCR_C (0.35 for 16S, 0.16 for 18S). Under a mixed diet, predicting REI_M led to low average correlations (0.35 with 16S, 0.17 with 18S) (Fig. 5C, D).

Predicting each component of feed efficiency also led to varying average Pearson correlations, depending on the diet and the sequencing dataset. Predicting intakes (ADEI_C, ADEI_M) with 16S or 18S data led to a large range of correlations: from 0.05 with 18S under a concentrate diet to 0.56 with 16S under a mixed diet. Similar results were obtained for growth traits (ADG_C or ADG_M), with correlations ranging from 0.15 with 18S data under a concentrate diet to 0.38 with 18S data under a mixed diet. Finally, average correlations between body composition traits (BFT_C, BFT_M, MD_C and MD_M) and microbiota predictions fluctuated between almost null and moderate values: from -0.07 for BFT_C predicted from 18S data to 0.45 for BFT_M predicted from 16S data.

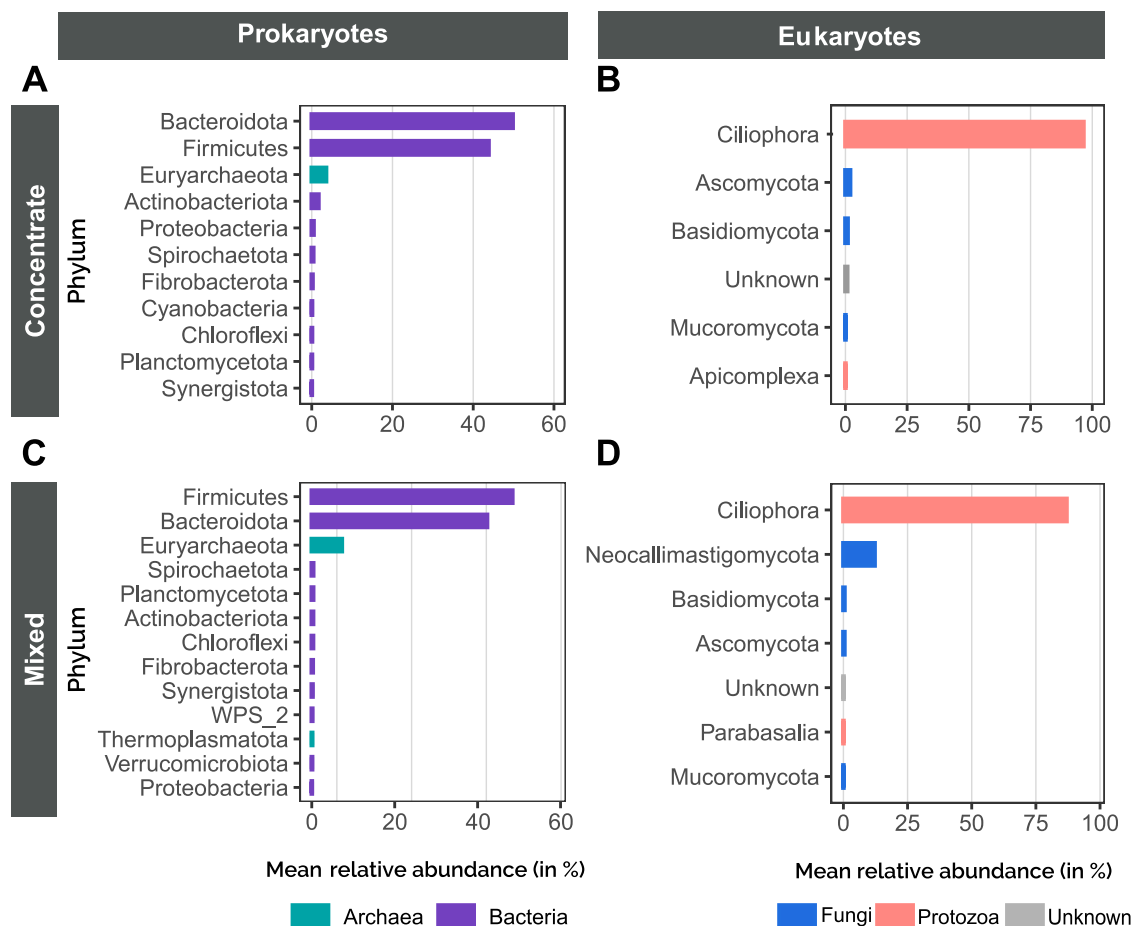


Fig. 3. Mean phylum abundances per sequencing and diet, in lamb ruminal fluids. Relative phylum abundances were computed after data cleaning, per diet and sequencing. (A) 16S sequencing under a concentrate diet; (B) 18S under a concentrate diet; (C) 16S under a mixed diet; (D) 18S under a mixed diet. Taxonomic affiliations were based on the Silva 132 16S and Silva 138.1 18S databases. Taxa affiliated with archaea, bacteria, fungi and protozoa are represented in blue–green, purple, dark blue, and red, respectively.

Predictions from systematic effects – Regardless of diet, trait and sequencing, average correlations for predictions from systematic effects ranged from 0.31 to 0.84. Furthermore, for almost all recorded traits, correlations were significantly higher than predictions derived from 16S or 18S data only. There were only three exceptions: similar accuracies were reached when FCR_C and MD_M were predicted from 16S or systematic effects, same as ADG_M predicted from 18S or systematic effects.

Predictions from microbiota and systematic effects – Finally, combining microbiota data with fixed effects and final BW never significantly improved correlations compared to predictions from fixed effects and BW. Most of the correlations were not significantly different. However, some correlations significantly decreased when 16S data and systematic effects were combined to predict: $ADEI_C$ (by -0.05 units on average), $ADEI_M$ (-0.11) and ADG_C (-0.19). Similar decreases were observed with 18S data when predicting FCR_C (-0.18), $ADEI_C$ (-0.04) and ADG_C (-0.24).

Discussion

DNA extracted from the rumen fluid of fattening lambs was sequenced for the 16S ribosomal RNA gene to detect prokaryotes and the 18S ribosomal RNA gene to study eukaryotic microorganisms. This makes the study one of the few that have quantified both types of microorganisms simultaneously in more than 150 lambs, fed successively with two contrasting diets.

The rumen microbiota was not modified by the selection for feed efficiency

Our results suggest that the ongoing selection for RFI did not modify the rumen microbiota abundances (both prokaryotes or eukaryotes), regardless of the diet.

When we looked at a microbial signature combining several OTUs for predicting the RFI line, by using an sPLSDA multivariate framework, we also observed a low predictive ability ($BER > 0.45$). A relationship between the rumen microbiota and the RFI lines can be doubted, as the microbiota poorly predicts the selected line. Similarly, in dairy sheep, divergent selection on somatic cell count in milk or milk persistency did not modify the ruminal microbiota (averaged BER ranging from 0.50 to 0.71) (Martinez Boggio et al., 2021). Our results were obtained after two and three generations of selection, so it remains to be checked whether longer-term sheep selection for RFI could also significantly alter the rumen microbiota.

In monogastrics, divergent selection of pigs for RFI during 9–10 generations resulted in significant differences in the abundances of 52 faecal bacterial genera between the RFI lines (Aliakbari et al., 2021). With rabbits, discriminant analysis of principal components was able to discriminate a line selected for a decreased RFI during nine generations and the ancestral line based on their caecal bacterial phylotypes (Drouilhet et al., 2016). The discrepancy between these monogastrics results and ours may be due to much more advanced divergence or the microbiota gut localisation that was studied.

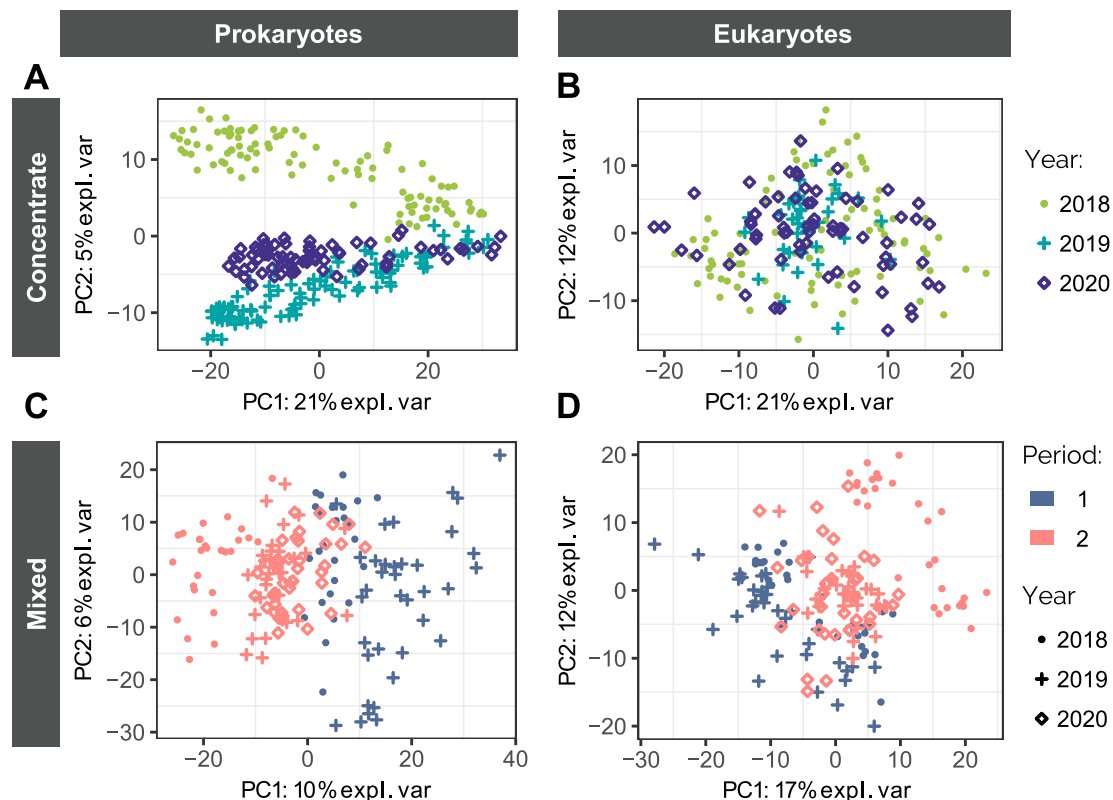


Fig. 4. Principal component analyses of lamb ruminal microbiota, per diet and sequencing. Abbreviations: CLR: centered logratio; expl. var: explained variance; OTU: operational taxonomic unit; PC: principal component. Analyses were carried out per diet and sequencing with OTU CLR values adjusted for sequencing effects. (A) 16S sequencing under a concentrate diet; (B) 18S under a concentrate diet; (C) 16S under a mixed diet; (D) 18S under a mixed diet. Animals phenotyped in 2018, 2019 and 2020 are represented by dots, crosses and diamonds, respectively. Under a mixed diet, animals phenotyped during the first and second periods are represented in dark blue and red, respectively.

Table 2

Balanced error rates while predicting residual feed intake lines of lambs from fixed effects, BW and/or microbiota data.

Diet ¹	Sequencing (n ²)	Set of predictors ³		
		M	S	S + M
C-diet	16S (269)	0.46 ^a (0.06)	0.38 ^a (0.05)	0.46 ^a (0.06)
	18S (205)	0.47 ^b (0.07)	0.35 ^a (0.06)	0.43 ^{ab} (0.06)
M-diet	16S (160)	0.55 ^a (0.08)	0.47 ^a (0.07)	0.53 ^a (0.08)
	18S (161)	0.45 ^a (0.08)	0.49 ^a (0.07)	0.45 ^a (0.08)

Abbreviations: C-diet: concentrate diet; M: microbiota; M-diet: mixed diet; S: systematic effects; S + M: microbiota plus systematic effects.

Predictions were carried out with sparse partial least squares discriminant analyses. Balanced error rates were averaged over 5-fold cross-validations repeated 50 times. SDs are given in brackets.

¹ Diets. C-diet: sheep were fed a 100% concentrate diet; M-diet: sheep were fed a mixed diet (2/3 of forage and 1/3 of concentrate).

² Number of samples.

³ Sets of predictors. M: adjusted centered logratio values of 16S or 18S operational taxonomic units; S: systematic effects, i.e., fixed effects and final BW; S + M: systematic effects and adjusted centered logratio values of 16S or 18S operational taxonomic units.

^{a,b} Balanced error rates with different letters significantly differ (corrected *t*-test *P*-value < 0.05 after adjustment). Comparisons and Benjamini–Hochberg adjustments were made per diet and sequencing.

Predictions of feed efficiency and associated traits are not improved with rumen microbiota data

Assessing the predictive ability of host traits for the rumen microbiota

The predictive ability of host traits from rumen prokaryote or eukaryote microbiota was highly variable according to traits: low prediction accuracies were observed for residual intakes compared to the other studied traits.

Predictions from microbiota compositions only – As a majority of authors did, we fitted models with only microbiota variables to predict feed efficiency traits. Averaged Pearson correlations between sPLSR predictions from microbiota data and actual traits varied with the diet (concentrate or mixed diet), feed efficiency

trait (FCR_C, REI_C or REI_M), and sequencing data used as predictors (16S or 18S). We obtained null to low correlations between actual residual intakes (0.06–0.35) and predictions from microbiota data. This result contradicts that of Ellison et al. (2019), who found a strong correlation between actual sheep RFI and predictions from 16S data (0.71 versus 0.35 in our study for REI_M). Discrepancies between our results might be explained by two main differences. They used a smaller number of animals (16 in their training set, 20 in their testing set), which might not fully reflect the variability of a larger population. Second, they picked animals with extreme and average RFI phenotypes out of a larger population to build the testing set. Selecting discontinuous and contrasting groups might artificially inflate the Pearson correlation (Aggarwal and

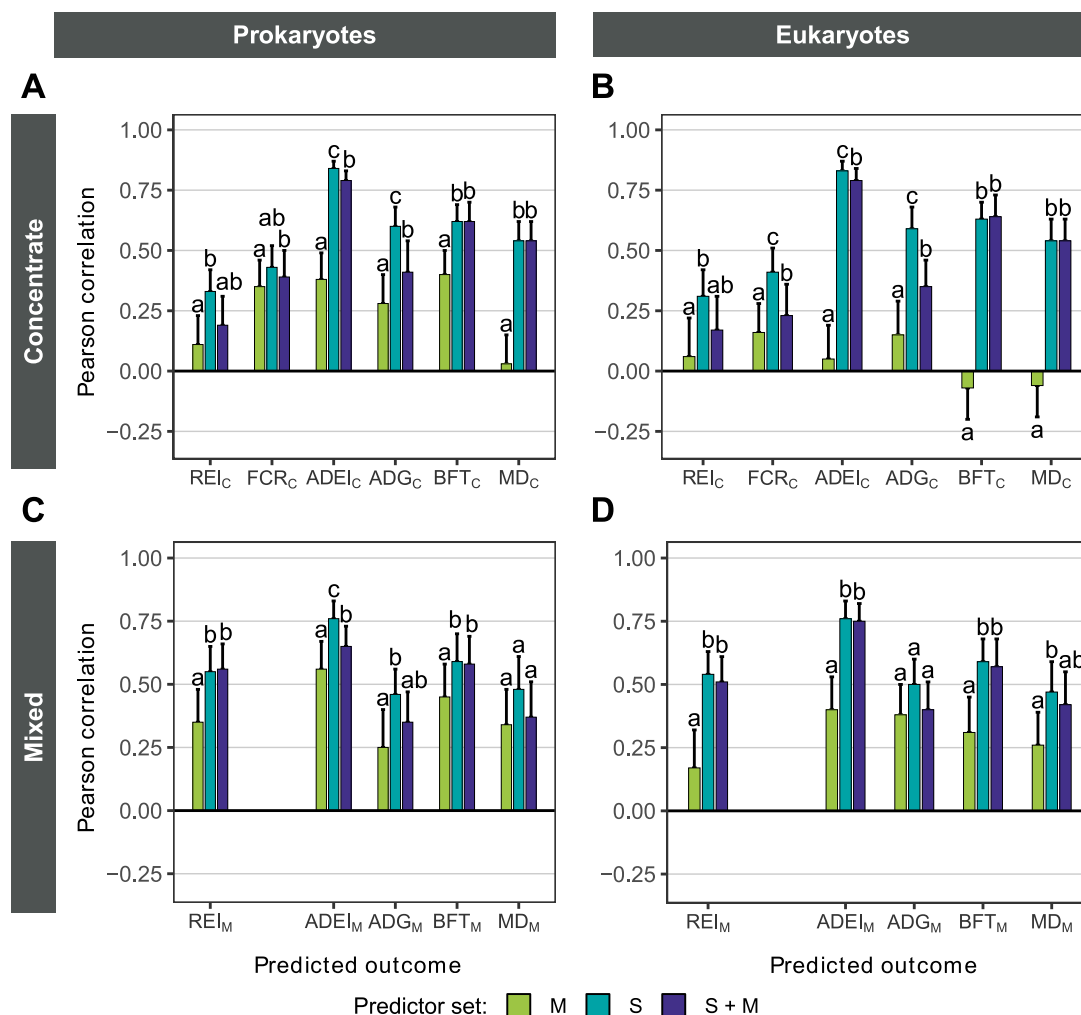


Fig. 5. Average Pearson correlations between lamb feed efficiency, production traits and predictions from systematic effects and/or microbiota data. Abbreviations: ADEI: average daily energy intake; ADG: average daily gain; BFT: back fat thickness; FCR: feed conversion ratio; M: microbiota; MD: muscle depth; REI: residual energy intake; S: systematic effects; S + M: microbiota plus systematic effects. Predictions were carried out with sparse partial least squares regressions. Pearson correlations were averaged over 5-fold cross-validations repeated 50 times. Error bars are equivalent to 1 SD. (A) 16S sequencing under a concentrate diet; (B) 18S under a concentrate diet; (C) 16S under a mixed diet; (D) 18S under a mixed diet. Three sets of predictors were tested, with M: adjusted centred logratio values of 16S or 18S operational taxonomic units (green); S: systematic effects including fixed effects and final BW (blue-green); S + M: systematic effects and adjusted centred logratio values of 16S or 18S operational taxonomic units (purple). Traits were recorded under a concentrate diet (C subscript) or a mixed diet (M subscript). ^{a,b,c}Correlations with different letters significantly differ (corrected *t*-test *P*-value < 0.05 after adjustment). Comparisons and Benjamini–Hochberg adjustments were made per trait, diet and sequencing.

Ranganathan, 2016). Furthermore, it might not reflect the variability of a continuous trait such as RFI. In dairy cows fed a mixed diet (30 animals in the training and 31 in the testing sets), there was a low correlation (0.19) between the ratio of milk production over feed intake and predictions from whole metagenome sequencing data (Delgado et al., 2019). This result is intermediate to our prediction accuracies when FCR was predicted from 16S or 18S data (0.16–0.35) in meat sheep fed a concentrate diet.

When only microbial predictors were used as predictors, the best prediction accuracies were obtained with intake traits (ADEI_C or ADEI_M): correlations ranged between 0.05 and 0.56 on average, with higher correlations for 16S predictors than 18S. Delgado et al. (2019) found an intermediate correlation of 0.39 between actual intakes and predictions of cow DM intake from whole metagenome sequencing data.

As observed before in different contexts, microbiota data alone might be used to predict energy intake, with moderate accuracies, but for feed efficiency traits, accuracies of predictions are very low.

Predictions from a combination of microbiota and systematic effects – Few authors have used a combination of microbiota data

and systematic effects to jointly predict host traits. In rabbits fed feed pellets, the averaged correlation was as low as 0.17 between actual RFI and sPLSR predictions from systematic effects and 16S data (Velasco-Galilea et al., 2021). In sheep-fed concentrates, we obtained slightly lower correlations between REI_C and predictions from analogous predictors (average correlation of 0.11).

The present study also combined systematic effects and 16S predictors to predict not only feed efficiency traits but also their components. With these predictors, ADEI prediction accuracy in sheep fed concentrates was similar to previous findings in monogastrics: our average correlation of 0.79 in sheep is close to that of 0.73 in rabbits (Velasco-Galilea et al., 2021). The same observation holds for ADG prediction accuracy from 16S data and systematic effects: average of 0.41 in our study, against 0.51 in rabbits (Velasco-Galilea et al., 2021), and for BFT prediction accuracy: average correlation of 0.62 in our study against averages that may be over 0.50 in pigs, depending on the sampling time (Maltecca et al., 2019).

Therefore, low to high prediction accuracies can be reached when systematic effects and microbiota are combined as

predictors of host traits. However, it does not demonstrate that predictions benefit from the inclusion of microbiota variables. Prediction models may rely more on systematic effects, such as the live weight, than microbiota variables. Indeed, the present study illustrates that models integrating microbial predictors and systematic effects never significantly outperformed predictions from systematic effects only. Therefore, microbiota predictors did not provide any added value to improve prediction accuracies. Prediction accuracies did not benefit from any synergy between microbial, fixed effects and BW predictors. In addition, systematic effects (*i.e.*, fixed effects and BW here) represented less-expensive predictors than metabarcoding variables. In contrast, Velasco-Galilea et al. (2021) demonstrated that sPLSR models more accurately predicted rabbit RFI when microbiota variables were included than models with systematic effects only. Similarly, Maltecca et al. (2019) showed an improvement in machine learning accuracy by including microbiota data to predict swine growth and carcass traits. This was not the case in our study, regardless of the trait, microbiota sequencing or diet.

The sheep rumen microbiota may not have a negligible predictive ability for feed efficiency or its components. However, the present study suggests that recording systematic effects, such as animal weight and environmental effects, might be more effective than sampling and metabarcoding the rumen fluid. Most studies did not consider the predictive ability of systematic effects to establish a baseline and prove the utility of the microbiota in host trait predictions. We conclude that host trait predictions did not benefit from the inclusion of microbiota predictors because these were compared to cheaper predictors that proved to be more or as effective.

Disentangling ruminal microbiota compositions and the environment is crucial. The current experiment showed that environmental effects such as the year or period of phenotyping shaped the ruminal microbiota composition of lambs (Fig. 4). The influence of environmental effects on the gut microbiota has already been well documented, as shown by geographic location, pen, diet, and farming practice effects (Thompson and Holmes, 2009; Henderson et al., 2015; Siegerstetter et al., 2017; Belanche et al., 2019; Bi et al., 2019; Marie-Etancelin et al., 2021; Wen et al., 2021). For instance, Ellison et al. (2017) found four times more OTUs significantly associated with the diet than OTUs associated with RFI status. Marie-Etancelin et al. (2021) highlighted that the diet impacted 91 bacterial genera out of 114, whereas no significant relationship was found between RFI traits and bacterial genera. Identifying significant associations between host traits and the rumen microbiota might be challenging, partly because of strong environmental effects in the upper gut of ruminating animals.

Therefore, scientists should also assess the predictive potential of systematic effects when using microbiota data for predictions. It would be desirable to include covariates or factors reflecting spatiotemporal changes, as they may affect the microbiota composition (Maltecca et al., 2019) and the predicted phenotype. For instance, animal phenotypes might not be homogeneous across pens due to the allocation of animals, or across years due to experimental variations. Microbiota variations associated with the pen or the year may be spuriously correlated with host phenotypes. Therefore, ignoring systematic effects might have led to overoptimistic claims about the microbiota's potential to predict sheep feed efficiency.

Perspectives for host phenotype predictions

Metabarcoding the rumen microbiota might not be efficient in predicting the production traits presently examined in sheep. However, it might be more useful in predicting greenhouse gas

emissions, given that methane is a direct by-product of fibre fermentation by rumen microorganisms (Hill et al., 2016). Moreover, it was previously estimated that approximately 13% of methane production variations were associated with microbiome variations in dairy cattle (microbiability of 0.13 ± 0.08) (Difford et al., 2018). In sheep, estimates of methane yield microbiability varied with the sequencing method and pipeline, from 0.19 (± 0.07) to 0.41 (± 0.08) (Hess et al., 2020).

To predict sheep feed efficiency, intake, growth or body composition traits, different omics approaches might provide better results. Previously, our RFI lines were better discriminated based on plasma metabolomics than based on rumen metabolomics, but the predictive ability was not assessed (Touitou et al., 2022). This suggests that the rumen in the upper gut would not be the best host component to predict feed efficiency variations. Sequencing the digestive microbiota in a different gut section might be useful. This was already explored with the faecal microbiota to predict pig traits (Maltecca et al., 2019; Aliakbari et al., 2022). In ruminants, we focused on the ruminal microbiota because of its central role in animal nutrition. However, the faecal microbiota, as the end of the entire digestive process, could be relevant. To the best of our knowledge, in ruminants, no peer-reviewed study has used testing sets to assess the predictive accuracies of complex traits from rumen and faecal microbiotas. Nonetheless, Andrade et al. (2022) identified faecal and ruminal microorganisms significantly associated with beef RFI. Furthermore, Monteiro et al. (2022) suggested that the dairy cow faecal microbiota was less correlated with feed intake than the rumen microbiota and that the faecal microbiome was more correlated with production efficiency. They hypothesised that more efficient animals could be associated with a different fermentation profile in the rumen. Then, the host uptake of nutrients might lead to differences in nutrient availability for microorganisms in the lower gut (Monteiro et al., 2022). In addition to the rumen, it might be the accumulation throughout the whole digestive tract of differences in digestion or assimilation that influences the host efficiency. It may also affect the availability of nutrients for microorganisms in the lower gut. This might result in a correlation between host feed efficiency and faecal microbiota. Finally, faecal sampling would be less invasive than rumen fluid sampling.

Supplementary material

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.animal.2023.100824>.

Ethics approval

Animals were reared at the experimental unit P3R (Osmoy, France, experimental approval C18-174-01) in compliance with the French National Regulations for human care and use of animals for research purposes. The protocol was assessed and approved by an ethics committee (APAFIS #12682-2017121509105248 v3) on behalf of the French Ministry for Higher Education, Research and Innovation.

Data and model availability statement

Raw microbiota sequences are publicly available in the NCBI database (accession number: PRJNA880038) with information about sampling conditions. Zootechnical data are not deposited in an official repository, but they are available from the authors upon reasonable request.

Author ORCIDs

Quentin Le Graverand: <https://orcid.org/0000-0002-7995-694X>.

Christel Marie-Etancelin: <https://orcid.org/0000-0001-8950-0420>.

Annabelle Meynadier: <https://orcid.org/0000-0003-0631-6949>.

Jean-Louis Weisbecker: no ORCID.

Didier Marcon: no ORCID.

Flavie Tortereau: <https://orcid.org/0000-0003-1444-4481>.

Author contributions

Quentin Le Graverand: Data curation, Formal analysis, Writing – Original Draft; **Christel Marie-Etancelin:** Conceptualisation; Data curation; Writing – Review and Editing; Supervision; Funding acquisition; **Annabelle Meynadier:** Data curation; Writing – Review and Editing; Funding acquisition; **Jean-Louis Weisbecker:** Data curation, Investigation, Resources; **Didier Marcon:** Investigation, Resources, Supervision; **Flavie Tortereau:** Conceptualisation; Data curation; Writing – Review and Editing; Supervision; Funding acquisition.

Declaration of interest

None.

Acknowledgements

We wish to acknowledge our P3R colleagues (<https://doi.org/10.15454/1.5483259352597417E12>) for caring for animals and their help in collecting samples. We thank M-L Chemit for the estimation of feed dietary values, DNA extraction and amplification. We acknowledge the help of Géraldine Pascal in bioinformatics, to process sequences files with Frogs tools. We are grateful to the Genotoul bioinformatics platform (Toulouse, France) and Sigene group for providing computing and storage resources (Galaxy instance <https://vm-galaxy-prod.toulouse.inra.fr>).

Financial support statement

Authors acknowledge the financial support through the partners of the 2018 Joint Call of the Cofund ERA-Nets FACCE ERA-GAS (Grant No. 696356), ICT-AGRI 2 (Grant No. 618123) and SusAn (Grant No. 696231), for the GrassToGas project. This work also received funding from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER).

References

Aggarwal, R., Ranganathan, P., 2016. Common pitfalls in statistical analysis: The use of correlation techniques. *Perspectives in Clinical Research* 7, 187.

Aliakbari, A., Zemb, O., Billon, Y., Barilly, C., Ahn, I., Riquet, J., Gilbert, H., 2021. Genetic relationships between feed efficiency and gut microbiome in pig lines selected for residual feed intake. *Journal of Animal Breeding and Genetics*, 491–507.

Aliakbari, A., Zemb, O., Cauquil, L., Barilly, C., Billon, Y., Gilbert, H., 2022. Microbiability and microbiome-wide association analyses of feed efficiency and performance traits in pigs. *Genetics Selection Evolution* 54, 1–15.

Anderson, C.J., Koester, L.R., Schmitz-Esser, S., 2021. Rumen Epithelial Communities Share a Core Bacterial Microbiota: A Meta-Analysis of 16S rRNA Gene Illumina MiSeq Sequencing Datasets. *Frontiers in Microbiology* 12, 1–14.

Andrade, B.G.N., Bressani, F.A., Cuadrat, R.R.C., Cardoso, T.F., Malheiros, J.M., de Oliveira, P.S.N., Petrini, J., Mourão, G.B., Coutinho, L.L., Reecy, J.M., Koltes, J.E., Neto, A.Z., de Medeiros, S.R., Berndt, A., Palhares, J.C.P., Afli, H., Regitano, L.C.A., 2022. Stool and Rumen Microbiome Components Associated With Methane

Emission and Feed Efficiency in Nelore Beef Cattle. *Frontiers in Genetics* 13, 1–12.

Artegoitia, V.M., Foote, A.P., Lewis, R.M., Freetly, H.C., 2017. Rumen fluid metabolomics analysis associated with feed efficiency on crossbred steers. *Scientific Reports* 7, 1–14.

Baumont, R., Dulphy, J.P., Sauvart, D., Meschy, F., Aufrere, J., Peyraud, J.-L., 2007. Valeur alimentaire des fourrages et des matières premières: tables et prévision. In *Alimentation des ruminants. Besoins des animaux et valeur des aliments* (ed. Editions Quae). Editions Quae, Versailles, France.

Belanche, A., Kingston-Smith, A.H., Griffith, G.W., Newbold, C.J., 2019. A multi-kingdom study reveals the plasticity of the rumen microbiota in response to a shift from non-grazing to grazing diets in sheep. *Frontiers in Microbiology* 10, 122.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 289–300.

Bi, Y., Cox, M.S., Zhang, F., Suen, G., Zhang, N., Tu, Y., Diao, Q., 2019. Feeding modes shape the acquisition and structure of the initial gut microbiota in newborn lambs. *Environmental Microbiology* 21, 2333–2346.

Bouckaert, R.R., Frank, E., 2004. Evaluating the replicability of significance tests for comparing learning algorithms. In: *Honghua, D., Ramakrishnan, S., Chengqi, Z. (Eds.), Advances in Knowledge Discovery and Data Mining*. Springer, Sydney, Australia, pp. 3–12.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10, 1–9.

Clemmons, B.A., Shin, S.B., Smith, T.P.L., Embree, M.M., Voy, B.H., Schneider, L.G., Donohoe, D.R., McLean, K.J., Myer, P.R., 2021. Rumenal protozoal populations of angus steers differing in feed efficiency. *Animals* 11, 1–10.

Cruz, G.D., Rodríguez-Sánchez, J.A., Oltjen, J.W., Sainz, R.D., 2010. Performance, residual feed intake, digestibility, carcass traits, and profitability of angus-hereford steers housed in individual or group pens. *Journal of Animal Science* 88, 324–329.

Delgado, B., Bach, A., Guasch, I., González, C., Elcoso, G., Pryce, J.E., Gonzalez-Recio, O., 2019. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. *Scientific Reports* 9, 1–13.

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., Guldbandsen, B., Sahana, G., 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLOS Genetics* 14, e1007580.

Drouilhet, L., Achard, C.S., Zemb, O., Molette, C., Gidenne, T., Larzul, C., Ruesche, J., Tircazes, A., Segura, M., Bouchez, T., Theau-Clément, M., Joly, T., Balmisse, E., Garreau, H., Gilbert, H., 2016. Direct and correlated responses to selection in two lines of rabbits selected for feed efficiency under ad libitum and restricted feeding: I. Production traits and gut microbiota characteristics. *Journal of Animal Science* 94, 38–48.

Ellison, M.J., Conant, G.C., Lamberson, W.R., Cockrum, R.R., Austin, K.J., Rule, D.C., Cammack, K.M., 2017. Diet and feed efficiency status affect rumen microbial profiles of sheep. *Small Ruminant Research* 156, 12–19.

Ellison, M.J., Conant, G.C., Lamberson, W.R., Austin, K.J., Van Kirk, E., Cunningham, H. C., Rule, D.C., Cammack, K.M., 2019. Predicting residual feed intake status using rumen microbial profiles in ewe lambs. *Journal of Animal Science* 97, 2878–2888.

Escudé, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., Pascal, G., 2018. FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics* 34, 1287–1294.

Goldansaz, S.A., Markus, S., Berjanskii, M., Rout, M., Guo, A.C., Wang, Z., Plastow, G., Wishart, D.S., 2020. Candidate serum metabolite biomarkers of residual feed intake and carcass merit in sheep. *Journal of Animal Science* 98, 1–15.

Groeneveld, E., Kovac, M., Wang, T., 1990. PEST, a general purpose BLUP package for multivariate prediction and estimation. In: *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production*, pp. 488–491.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., Troedsson, C., 2014. Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS ONE* 9, 2.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P.H., Abecia, L., Angarita, E., Aravena, P., Arenas, G.N., Ariza, C., Attwood, G.T., Avila, J.M., Avila-Stagno, J., Bannink, A., Barahona, R., Batistotti, M., Bertelsen, M.F., Brown-Kav, A., Carvajal, A.M., Cersosimo, L., Chaves, A.V., Church, J., Clipson, N., Cobos-Peralta, M.A., Cookson, A.L., Cravero, S., Carballo, O.C., Crosley, K., Cruz, G., Cucchi, M.C., De La Barra, R., De Menezes, A.B., Detmann, E., Dieho, K., Dijkstra, J., Dos Reis, W. L.S., Dugan, M.E.R., Ebrahimi, S.H., Eythórsdóttir, E., Fon, F.N., Fraga, M., Franco, F., Friedeman, C., Fukuma, N., Gagić, D., Gangnat, I., Grilli, D.J., Guan, L.L., Miri, V. H., Hernandez-Sanabria, E., Gomez, A.X.I., Isah, O.A., Ishaq, S., Jami, E., Jelincic, J., Kantanen, J., Kelly, W.J., Kim, S.H., Klieve, A., Kobayashi, Y., Koike, S., Kopecny, J., Kristensen, T.N., Krizsan, S.J., LaChance, H., Lachman, M., Lamberson, W.R., Lambie, S., Lassen, J., Leahy, S.C., Lee, S.S., Leiber, F., Lewis, E., Lin, B., Lira, R., Lund, P., Macipe, E., Mamuad, L.L., Mantovani, H.C., Marcopiddo, G.A., Márquez, C., Martin, C., Martinez, G., Martinez, M.E., Mayorga, O.L., McAllister, T.A., McSweeney, C., Mestre, L., Minnee, E., Mitsumori, M., Mizrahi, I., Molina, I., Muenger, A., Munoz, C., Murovec, B., Newbold, J., Nserere, V., O'Donovan, M., Okunade, S., O'Neill, B., Ospina, S., Ouwerkerk, D., Parra, D., Pereira, L.G.R., Pinares-Patino, C., Pope, P.B., Poulsen, M., Rodehutscord, M., Rodriguez, T., Saito, K., Sales, F., Sauer, C., Shingfield, K., Shoji, N., Simunek, J., Stojanović-Radić, Z., Stres, B., Sun, X., Swartz, J., Tan, Z.L., Tapio, I., Taxis, T.M., Tomkins, N., Ungerfeld, E., Valizadeh, R., Van Adrichem, P., Van Hamme, J., Van Hoven, W., Waghorn, G., Wallace, R.J., Wang, M., Waters, S.M., Keogh, K., Witzig, M., Wright, A.D.G.,

- Yamano, Yan, T., Yanez-Ruiz, D.R., Yeoman, C.J., Zambrano, R., Zeitz, J., Zhou, M., Zhou, H.W., Zou, C.X., Zunino, P., . Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* 5, 14567.
- Hess, M.K., Rowe, S.J., Van Stijn, T.C., Henry, H.M., Hickey, S.M., Brauning, R., McCulloch, A.F., Hess, A.S., Kirk, M.R., Kumar, S., Pinares-Patiño, C., Kittelmann, S., Wood, G.R., Janssen, P.H., McEwan, J.C., 2020. A restriction enzyme reduced representation sequencing approach for low-cost, high-throughput metagenome profiling. *PLoS ONE* 15, 1–18.
- Hill, J., McSweeney, C., Wright, A.D.G., Bishop-Hurley, G., Kalantar-zadeh, K., 2016. Measuring Methane Production from Ruminants. *Trends in Biotechnology* 34, 26–35.
- Koch, R.M., Swiger, L.A., Chambers, D., Gregory, K.E., 1963. Efficiency of feed use in beef cattle. *Journal of Animal Science* 22, 486–494.
- Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28, 1–26.
- Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer, New-York, NY, USA.
- Lenth, R.V., 2022. The Comprehensive R Archive Network. emmeans: Estimated Marginal Means, aka Least-Squares Means. Retrieved on 24 August 2022, from <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>.
- Li, R., Teng, Z., Lang, C., Zhou, H., Zhong, W., Ban, Z., Yan, X., Yang, H., Farouk, M.H., Lou, Y., 2019. Effect of different forage-to-concentrate ratios on ruminal bacterial structure and real-time methane production in sheep. *PLoS ONE* 14, e0214777.
- Lima, N.L.L., Ribeiro, C.R. de F., de Sá, H.C.M., Leopoldino Júnior, I., Cavalcanti, L.F.L., Santana, R.A.V., Furusho-Garcia, I.F., Pereira, I.G., 2017. Economic analysis, performance, and feed efficiency in feedlot lambs. *Revista Brasileira de Zootecnia* 46, 821–829.
- Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E.L., di Palma, M.A., 2017. The Comprehensive R Archive Network. Robustbase: basic robust statistics R package. Retrieved on 7 December 2021, from <https://cran.r-project.org/web/packages/robustbase/robustbase.pdf>.
- Mahé, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M., 2014. Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ* 2014, 1–13.
- Maltecca, C., Lu, D., Schillebeeckx, C., McNulty, N.P., Schwab, C., Shull, C., Tiezzi, F., 2019. Predicting Growth and Carcass Traits in Swine Using Microbiome Data and Machine Learning Algorithms. *Scientific Reports* 9, 1–15.
- Marie-Etancelin, C., Tortereau, F., Gabinaud, B., Martinez Boggio, G., Le Graverand, Q., Marcon, D., De Almeida, M.L., Pascal, G., Weisbecker, J.L., Meynadier, A., 2021. Apart From the Diet, the Ruminal Microbiota of Lambs Is Modified in Relation to Their Genetic Potential for Feed Efficiency or Feeding Behavior. *Frontiers in Microbiology* 12, 1–15.
- Martinez Boggio, G., Meynadier, A., Daunis-I-Estadella, P., Marie-Etancelin, C., 2021. Compositional analysis of ruminal bacteria from ewes selected for somatic cell score and milk persistency. *PLoS ONE* 16, 1–16.
- Martín-Fernández, J.A., Hron, K., Templ, M., Filzmoser, P., Palarea-Albaladejo, J., 2015. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Statistical Modelling* 15, 134–158.
- McMurdie, P.J., Holmes, S., 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* 8, 4.
- Monteiro, H.F., Zhou, Z., Gomes, M.S., Peixoto, P.M.G., Bonsaglia, E.C.R., Canisso, I.F., Weimer, B.C., Lima, F.S., 2022. Rumen and lower gut microbiomes relationship with feed efficiency and production traits throughout the lactation of Holstein dairy cows. *Scientific Reports* 12, 1–14.
- Mucha, S., Tortereau, F., Doeschl-Wilson, A., Rupp, R., Conington, J., 2022. Animal Board Invited Review: Meta-analysis of genetic parameters for resilience and efficiency traits in goats and sheep. *Animal* 16, 100456.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research* 41, 590–596.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rohart, F., Gautier, B., Singh, A., Lê Cao, K.A., 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology* 13, 1–19.
- Siegerstetter, S.C., Schmitz-Esser, S., Magowan, E., Wetzels, S.U., Zebeli, Q., Lawlor, P. G., O'Connell, N.E., Metzler-Zebeli, B.U., 2017. Intestinal microbiota profiles associated with low and high residual feed intake in chickens across two geographical locations. *PLoS ONE* 12, 1–23.
- Thompson, C.L., Holmes, A.J., 2009. A window of environmental dependence is evident in multiple phylogenetically distinct subgroups in the faecal community of piglets. *FEMS Microbiology Letters* 290, 91–97.
- Tortereau, F., Marie-Etancelin, C., Weisbecker, J.L., Marcon, D., Bouvier, F., Moreno-Romieux, C., François, D., 2020. Genetic parameters for feed efficiency in Romane rams and responses to single-generation selection. *Animal* 14, 681–687.
- Touitou, F., Tortereau, F., Bret, L., Marty-Gasset, N., Marcon, D., Meynadier, A., 2022. Evaluation of the Links between Lamb Feed Efficiency and Rumen and Plasma Metabolomic Data. *Metabolites* 12, 304.
- Velasco-Galilea, M., Piles, M., Ramayo-Caldas, Y., Sánchez, J.P., 2021. The value of gut microbiota to predict feed efficiency and growth of rabbits under different feeding regimes. *Scientific Reports* 11, 1–18.
- Wang, Y., Qian, P.Y., 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4, 10.
- Wei, X., Dong, Z., Cheng, F., Shi, H., Zhou, X., Li, B., Wang, L., Wang, W., Zhang, J., 2021. Seasonal diets supersede host species in shaping the distal gut microbiota of Yaks and Tibetan sheep. *Scientific Reports* 11, 1–11.
- Wen, C., Yan, W., Mai, C., Duan, Z., Zheng, J., Sun, C., Yang, N., 2021. Joint contributions of the gut microbiota and host genetics to feed efficiency in chickens. *Microbiome* 9, 1–23.
- Zhang, Y., Li, F., Chen, Y., Wu, H., Meng, Q., Guan, L.L., 2020. Metatranscriptomic Profiling Reveals the Effect of Breed on Active Rumen Eukaryotic Composition in Beef Cattle With Varied Feed Efficiency. *Frontiers in Microbiology* 11, 1–12.