Sepulveda, B., Vander Jagt, C., Pryce, J., Chamberlain, A., Wang, J., Retegan, R., Marett, L., Williams, S., Jacobs, J., Xiang, R. (2026) Prediction of enteric methane emissions from faecal metagenome of dairy cattle, Proceedings 13th World Congress on Genetics Applied to Livestock Production 24323: https://doi.org/10.31274/wcgalp.24323 © 2026 *Sepulveda, et al.* CC BY-NC-ND

Prediction of enteric methane emissions from faecal metagenome of dairy cattle

B.J. Sepulveda[1*], C.J. Vander Jagt[1], J.E. Pryce[12], A.J. Chamberlain[12], J. Wang[1], R.M. Retegan[1], L.C. Marett[23], S.R.O. Williams[3], J.L. Jacobs[23], and R. Xiang[12]

[1]Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Bundoora, Australia; [2]La Trobe University, Bundoora, Australia; [3]Agriculture Victoria Research, Ellinbank, Australia; [*]Corresponding author: boris.sepulveda@agriculture.vic.gov.au

Introduction

Enteric methane emissions (EME) from ruminants contribute to global warming. EME is produced primarily from the activity of methanogenic microorganisms in the rumen. The ruminal metagenome can potentially be used as an indicator of EME (Gonzalez-Recio et al., 2023). However, collecting sufficient ruminal samples for ruminal metagenome profiling is challenging in terms of animal welfare and the logistics of restraining cows during sampling. As an alternative to the ruminal metagenome, the faecal metagenome could be evaluated as a predictor of EME. Sampling faeces is less labour-intensive and invasive than ruminal sampling, making it an attractive alternative for sampling large populations. Additionally, the faecal metagenome could possibly infer the ruminal metagenome which is important for predicting EME. More research is needed to determine whether faecal metagenomic profiles can predict EME and how they relate to ruminal metagenomic profiles. Here we apply shotgun long-read metagenomic sequencing to samples from 46 Australian dairy cows and use machine-learning models to assess how well faecal metagenomic features predict EME traits. We compare this performance against models based on ruminal metagenomic features and against models constructed from non-metagenomic predictors, most of which are commonly available in dairy production systems. We also investigate the association between ruminal and faecal metagenomes.

Materials & Methods

Methane production (MeP; g/d), energy-corrected milk (ECM), and methane intensity (MeI = MeP/ECM) were recorded in 46 dairy cows. Dry matter intake (DMI), and methane yield (MeY = MeP/DMI), were predicted in 26 of these animals. All animals were fed similar diets across six experiments. A total of 87 ruminal fluid and 120 faecal samples were collected (some animals had more than one sample). Ruminal fluid was collected via oesophageal probe, filtered through cheesecloth, and frozen at −80°C. Faecal samples were collected from fresh faecal pats and frozen at −80°C. Microbial DNA was extracted using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research). Sequencing libraries were prepared with the Native Barcoding Kit 96 V14 (Oxford Nanopore Technologies, ONT) and sequenced on a PromethION 2 (ONT). Basecalling was performed with the FAST model of Dorado (versions 7.3.11–7.11.2) and reads ≥250 bp were retained. Samples were retained if they had N50 ≥ 2,000 and ≥1 Gb of Q10 reads. Taxonomic and functional annotation of reads was performed using SqueezeMeta (versions 1.6.5 and 1.7.2) (Tamames and Puente-Sánchez, 2019) to obtain four types of metagenome features: superkingdom, genus, KEGG orthology groups (KOs), and Clusters of Orthologous Groups (COGs). Reads assigned to plants or host taxa were negligible (<0.02%). Unclassified reads and those assigned to non-microbial species were ignored. A core of 7,962 metagenomic features present in rumen and faeces for all animals were retained (prevalence = 100%).
Absolute abundance matrices were constructed separately for each feature type (superkingdom, genus, KOs, and COGs). Each matrix had dimensions of animals × features and contained the

number of reads assigned to each feature for each animal. Relative abundance matrices were then derived by expressing each feature's abundance as a proportion of the total abundance of the same type of features within the same animal. Feature abundances in each relative abundance matrices were normalized within animals using the isometric log-ratio (ILR) transformation (Aitchison, 1986) and then scaled and centred across animals using the scale function in R (R Core Team, 2026) to obtain the matrix **X**. The first five principal components (PCs) of each **X** from the rumen were computed using the prcomp function from the R package stats (R Core Team, 2026). These first five PCs accounted for 100% of the variance at the superkingdom level, 66% at the genus level, 51% for KOs, and 52% for COGs. The coefficient of determination (R^2) between the PCs and MeP was calculated. The contribution of the original variables to each ILR-based PC was quantified by extracting the coefficients of each Centered-Log Ratio on the PC axis (clr_loadings), and the percentage of variance in each PC explained by each variable's squared loading relative to the total loading (contribution_pct).
Random Forest regression models were implemented using the randomForest function from the R package randomForest (Liaw and Wiener, 2002). We evaluated four predictor sets for each EME trait: ruminal metagenome features (**Set 1**), faecal metagenome features (**Set 2**), non-metagenomic predictors (**Set 3**), and the non-metagenomic without the "experiment" predictor (**Set 4**). **Set 3** included experiment (six levels), age at calving, days in milk, ECM (excluded for MeI), and body weight. The "experiment" factor was excluded from **Set 4** because it reflects study-specific conditions not available for commercial prediction. We also tested the predictability of the ruminal PCs using **Sets 2 to 4**. All models were fitted using 10-fold cross-validation, and predictive performance was quantified using R^2 between observed and predicted values in the validation folds.

Results

For MeP, the ruminal metagenome (Set 1; $R^2 = 0.52$) and the full non-metagenomic predictor set (Set 3; $R^2 = 0.54$) showed the strongest predictive performance, followed by the faecal metagenome (Set 2; $R^2 = 0.43$; Figure 1).

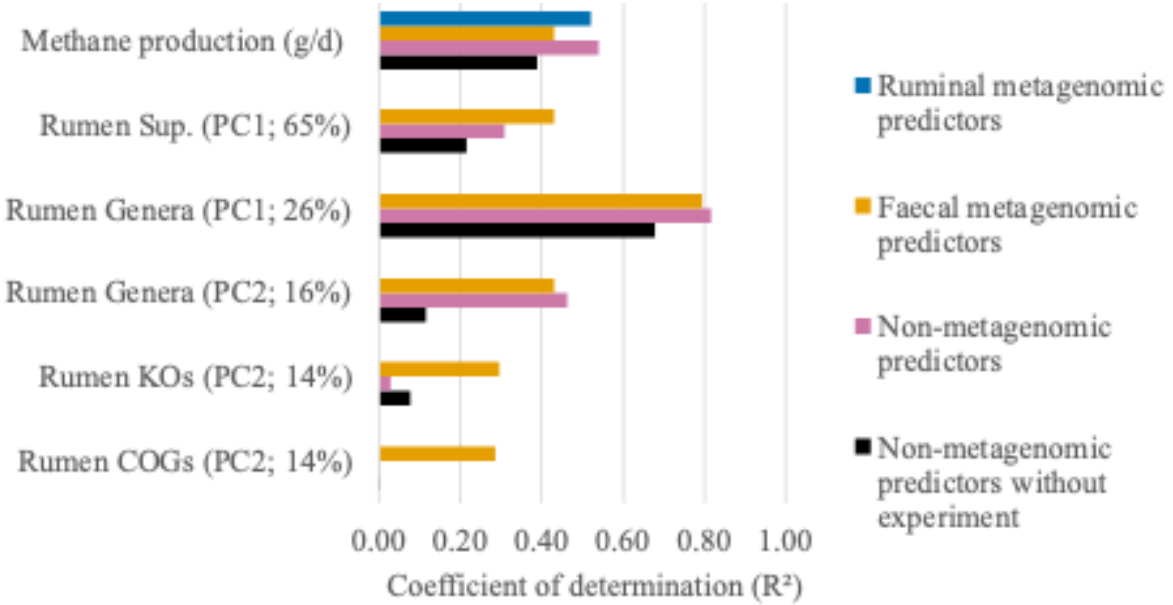

Figure 1. Predictive performance of Random Forest models for methane production and for principal components (PCs) derived from ruminal metagenomic matrices. Four predictor sets were evaluated: ruminal metagenomics (Set 1), faecal metagenomics (Set 2), non-metagenomics (Set 3), and non-metagenomic without the "experiment" predictor (Set 4). For each PC, the label indicates the PC number and the percentage of variance it explains. KOs: KEGG Orthology groups; COGs: Clusters of Orthologous Groups; Sup.: Superkingdom.

When the "experiment" predictor was removed from the non-metagenomic predictor set (Set 4), predictive performance decreased to $R^2 = 0.39$, indicating that a substantial portion of the accuracy of the full non-metagenomic model was attributable to "experiment". For MeY, only Sets 3 and 4 achieved meaningful predictive accuracy (R^2 up to 0.27), whereas Sets 1 and 2 performed poorly ($R^2 < 0.10$). For MeI, Sets 3 and 4 remained the most informative (R^2 up to 0.53), with Set 1 contributing modest predictive power ($R^2 = 0.12$) and Set 2 showing minimal predictive signal ($R^2 < 0.05$).

Set 2 predicted ruminal superkingdom PC1 (65% variance), genus PC1 and PC2 (26% and 16%), and PC2 of both KO and COG matrices (14% each), with R^2 values from 0.29–0.79. Set 3 performed worse than Set 2 for ruminal superkingdom PCs and similarly for ruminal genus PCs, while Set 4 showed lower predictability across all ruminal PCs. Neither Set 3 nor Set 4 predicted the functional PCs (KO-PC2 and COG-PC2) from the rumen, whereas Set 2 did. At the superkingdom level, ruminal PC1 was driven by opposing clr_loadings from Bacteria and Eukaryota, together explaining ~75% of the contribution_pct. At the genera level, methanogenic Archaea (e.g., *Methanobrevibacter*, *Methanosphaera*) contributed to both ruminal PC1 and PC2, with summed contribution_pct of ~10–15%. In ruminal KO and COG PCs each feature contributed <1%. The ruminal PCs showed R^2 between 0.14 and 0.19 with MeP, except for genera PC2 ($R^2 = 33\%$).

Discussion

This study evaluated the prediction of EME traits in Australian dairy cows using Random Forest models fitted to shotgun long-read metagenomes and non-metagenomic covariates. For MeP, the ruminal metagenome (Set 1) achieved moderate predictive accuracy ($R^2 = 0.52$; Figure 1), supporting the value of rumen microbial profiles for methane prediction. The full non-metagenomic predictor set (Set 3) performed similarly ($R^2 = 0.54$), indicating that routinely recorded covariates can explain a meaningful proportion of MeP variation within this study. These results are consistent with previous studies in different animals of the same population using long-read rumen metagenomics and machine-learning approaches, which reported R^2 values of 0.38 to 0.51 for EME traits (Hsieh et al., 2024).

The faecal metagenome (Set 2) also showed moderate predictive performance for MeP ($R^2 = 0.43$; Figure 1). In contrast, earlier short-read studies reported weak linkage between ruminal and faecal metagenomes (Ross et al., 2012), possibly reflecting the lower taxonomic and functional resolution of short-read sequencing compared with long-read approaches (Nature Methods, 2023). The reason faecal profiles can predict MeP is unclear, but one explanation is collinearity among predictors, as shared variation may exist between the ruminal and faecal metagenomes and the non-metagenomic predictors. Supporting this idea, several PCs derived from the ruminal metagenome were predictable from the faecal metagenome and non-metagenomic predictors, with R^2 values up to 0.79 and 0.82, respectively.

The predictive reliability for MeP obtained with the ruminal metagenome (Set 1; $R^2 = 0.52$; Figure 1) and the faecal metagenome (Set 2; $R^2 = 0.43$) fell within the range defined by the non-metagenomic models when the "experiment" predictor was included (Set 3; $R^2 = 0.54$) or excluded (Set 4; $R^2 = 0.39$). This pattern suggests that metagenomic profiles may encode biological and environmental signals that overlap with the information captured by the experiment. Ruminal superkingdom PC1 reflected a strong Bacteria–Eukaryota contrast (~75% of contribution_pct), consistent with previous studies, where this axis was also associated with EME (Saborío-Montero et al., 2021; Sepulveda et al., 2025). At the genus level, methanogens contributed to both ruminal PC1 and PC2 (~10–15% contribution_pct in each PC), and genera PC2 showed the strongest MeP association ($R^2 = 0.33$). Prediction of this methane-associated ruminal genera PC2 dropped when the experiment was removed as a predictor (from Set 3: $R^2 = 0.46$ to Set 4: $R^2 = 0.12$) yet remained moderate when using the

faecal metagenome (Set 2: $R^2 = 0.43$). Additionally, the faecal metagenome also predicted PC2 from the rumen KOs and COGs ($R^2 = 0.29$). These results suggest that the faecal metagenome may be associated with rumen microbial features linked to enteric methane production.
Although laboratory and computational costs are similar for ruminal and faecal metagenomics, faecal sampling is easier, less invasive, and avoids animal-welfare concerns. Further validation of our results is required and clarification of the predictability of MeY and MeI is needed, particularly for MeY, which was calculated for only 26 animals. In addition, future research should estimate the heritability of faecal metagenome features and their genetic correlations with EME traits and with ruminal metagenomes, as such information is essential for assessing the potential integration of faecal metagenomic data into breeding programs.

Conclusions

Faecal metagenomes showed moderate to high predictive reliability for methane production ($R^2 = 0.43$) and for rumen metagenome features associated with these emissions, suggesting they could possibly serve as indicators of rumen metagenome profiles linked to methane. These results may reflect the improved resolution of long-read sequencing. Larger, more diverse datasets and genetic parameters are needed to confirm these results and evaluate the potential integration of faecal metagenomic information into methane-reduction breeding strategies.

Acknowledgements

The authors are grateful for the support of DairyBio (Melbourne, Australia), funded by Dairy Australia, the Gardiner Foundation, and Agriculture Victoria.

References

Aitchison J. (1986). The statistical analysis of compositional data. Chapman and Hall London, London.
Gonzalez-Recio O., Scrobota N., López-Paredes J., Saborío-Montero A., Fernández A., López de Maturana E., Villanueva B., Goiri I., Atxaerandio R., and García-Rodríguez A. (2023) animal 17:100780. 10.1016/j.animal.2023.100780
Hsieh W.Y., Ho P.N., Sepulveda B.J., Chamberlain A.J., Marett L.C., Williams S.R.O., Jacobs J.L., Pryce J.E., Goddard M.E., Bolormaa S., and Xiang R. (2024) Proc. ADSS 2024, Christchurch, New Zealand.
Liaw A., and Wiener M. (2002) R News.
Nature Methods. (2023) Nat Methods 20(1):1-1. 10.1038/s41592-022-01759-x
R Core Team. (2026). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Ross E.M., Moate P.J., Bath C.R., Davidson S.E., Sawbridge T.I., Guthridge K.M., Cocks B.G., and Hayes B.J. (2012) BMC Genet. 13(1):53. 10.1186/1471-2156-13-53
Saborío-Montero A., López-García A., Gutiérrez-Rivas M., Atxaerandio R., Goiri I., García-Rodriguez A., Jiménez-Montero J.A., González C., Tamames J., Puente-Sánchez F., Varona L., Serrano M., Ovilo C., and González-Recio O. (2021) J Dairy Sci 104(7):8135-8151. 10.3168/jds.2020-20005
Sepulveda B.J., González-Recio O., Chamberlain A.J., Khansefid M., Cocks B.G., Wang J., Prowse-Wilkins C.P., Marett L.C., Williams S.R.O., Jacobs J.L., García-Rodríguez A., Jiménez-Montero J.A., and Pryce J.E. (2025) J Dairy Sci 108(8):8619-8636. doi.org/10.3168/jds.2024-25436
Tamames J., and Puente-Sánchez F. (2019) Front Microbiol 9(3349)10.3389/fmicb.2018.03349