UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Ciencia Animal

Modelos de predicción de emisiones de metano en vacuno de carne con datos de microbiota y métodos de aprendizaje automático

Trabajo Fin de Máster

Máster Universitario en Mejora Genética Animal y Biotecnología de la Reproducción

AUTOR/A: Saez Torillo, Santiago Nicolas

Tutor/a: Martínez Álvaro, Marina

CURSO ACADÉMICO: 2023/2024

MÁSTER INTERUNIVERSITARIO EN MEJORA GENÉTICA ANIMAL Y BIOTECNOLOGÍA DE LA REPRODUCCIÓN

Modelos de predicción de emisiones de metano en vacuno de carne con datos de microbiota y métodos de aprendizaje automático

Trabajo de Fin de Máster

Santiago N. Saez Torillo

Tutora:

Dra. Marina Martínez Álvaro

Junio de 2024

Universitat Politècnica de València

Departamento de Ciencia Animal

Agradecimientos

Quisiera expresar mi más sincero agradecimiento a todas aquellas personas que han sido fundamentales para la realización de este Trabajo de Fin de Máster. Sin su apoyo y guía, este logro no habría sido posible.

En primer lugar, quiero agradecer a mi tutora, Marina, por su invaluable orientación, paciencia y apoyo constante a lo largo de todo el proceso. Sus sugerencias, conocimientos y tiempo dedicado han sido esenciales para la culminación de este trabajo. Su compromiso con mi desarrollo académico y personal ha sido una fuente de inspiración y motivación.

A mi familia, les debo todo mi agradecimiento y amor. A mis padres, Silvia y Gustavo por su incansable apoyo, y por creer siempre en mí. Gracias por ser mi pilar de fortaleza y por alentarme a seguir adelante, incluso en los momentos más difíciles. A mis hermanos, por su comprensión y ánimo constante, y por estar siempre a mi lado. A mi sobrino Hilario, por ser el motivo para seguir adelante y esforzarme cada día más. A mi abuela Tati por siempre darme ese empujón que necesito cuando estoy mal y saber como calmar mis nervios. LOS AMO.

También quiero agradecer a Agos, Cris, Pedro, Ilyas, Toñi, Pilar, Noelia y Agustín por toda la ayuda brindada durante el máster, son un equipo genial y estoy muy agradecido de ahora formar parte de él.

Al Centro Agronómico Mediterráneo de Zaragoza (IAMZ–CIHEAM) por financiar mis estudios durante estos dos años de Máster.

Finalmente, quisiera agradecer a mis amigos Jose, Diego y Manu quienes han compartido conmigo esta travesía académica. Y a Benja quien me ha acompañado en la recta final de este máster. Su compañía, apoyo y colaboración han sido vitales para mantener la motivación y alcanzar esta meta.

Índice

Índice de tablas

Índice de figuras

Abreviaturas

- **ALR:** Transformación *log-ratio* aditiva
- **alr-MG**: Genes microbianos transformados por ALR
- **AUC**: Área bajo la curva ROC (Característica Operativa del Receptor)
- **CH_4 :** Metano
- **CLR:** Transformación *log-ratio* centrada
- **clr-MT**: Géneros microbianos transformadas por CLR
- **CO_2:** Dióxido de carbono
- **DMI:** Ingesta de materia seca
- **GHG:** Gases de efecto invernadero
- **ML:** *Machine learning* (aprendizaje automático)
- **MP:** Producción de metano (g CH_4/día)
- **MY:** Rendimiento de metano (g CH_4/ kg DMI)
- **N_2O:** Óxido nitroso
- **OMM**: Organización Meteorológica Mundial
- **PCA:** Análisis de Componentes Principales
- **PLS:** Regresión de Mínimos Cuadrados Parciales
- **R^2:** Coeficiente de determinación
- **RF:** *Random Forest*
- **RM:** Metano residual (g CH_4/día)
- **SF_6:** Hexafluoruro de azufre
- **SV:** Selección de variables
- **XGB:** *eXtreme Gradient Boosting*

Resumen

El metano (CH_4) es uno de los principales gases de efecto invernadero y tiene 28-34 veces mayor potencial de capturar calor en la atmosfera que el CO_2. La fermentación entérica del ganado contribuye significativamente a estas emisiones, representando un 32% de las emisiones globales de CH_4 antropogénico. La mejora genética es una estrategia atractiva para disminuir estas emisiones ya que es permanente y acumulativa, pero necesita de la recopilación de gran cantidad de datos. Sin embargo, la medición directa del CH_4 puede llegar a ser muy costosa o difícil de llevar a cabo a gran escala. Por lo tanto, el desarrollo de métodos indirectos para estimar el CH_4 utilizando *proxies* fáciles de medir como predictores es una opción atractiva. Los datos metagenómicos podrían servir como indicador para la predicción indirecta de las emisiones de CH_4 dada su estrecha relación.

El presente estudio propone modelos predictivos de tres fenotipos de CH_4: Producción de CH_4 (MP, g CH_4/día), Rendimiento de CH_4 (MY, g CH_4/kg de ingesta de materia seca diaria) y CH_4 Residual (RM, g CH_4/día) utilizando datos de microbiota (3,631 abundancias de genes y 1,136 abundancias de géneros, transformadas con *log-ratio* aditivo y centrado) ruminal de 287 animales.

Los animales eran de diferentes razas, fueron alimentados con dos dietas diferentes, y pertenecieron a diferentes experimentos. Todos estos efectos se testaron y corrigieron en los fenotipos de CH_4 y en las variables microbianas antes de ajustar los algoritmos. Las emisiones de CH_4 se midieron en cámaras de respiración y el microbioma se midió tomando muestras ruminales al sacrificio y haciendo secuenciación del metagenoma completo.

Se ajustaron los algoritmos los algoritmos Regresión de Mínimos Cuadrados Parciales (PLS), *eXtreme Gradient Boosting* (XGB) y *Random Forest* (RF) para predecir el CH_4 de manera continua, para clasificación en 5 categorías (*score* de emisiones, de bajos a altos) y en animales extremos (25% de animales que menos emiten y 25% de los que más). Todos los modelos se ajustaron utilizando un 70% de los datos como entrenamiento y dejando el 30% restante como validación externa. Además, se estimó un rango de la precisión en validación externa partiendo 100 veces la base de datos en un conjunto de entrenamiento y otro de validación de manera aleatoria. Para tratar de evitar el sobreajuste, además de ajustar los algoritmos con todas las variables microbianas, se realizó dos estrategias más para evitar el sobreajuste: un análisis de componentes principales (PCA) y usar los componentes principales como variables predictivas; y selección de variables.

Los resultados de predicción continua sufrieron sobreajuste para XGB y RF, pero no para PLS. En el conjunto de validación, el valor de predicción en validación externa (Q^2) promedio máximo fue de 0,09±0,05. Con selección de variables se logró conseguir un valor de 0,18±0,08. Sin embargo, la estrategia de compresión por PCA empeoró los resultados. Para la clasificación en 5 categorías se alcanzó un área bajo la curva en validación externa (AUCv) de 0.68 ± 0.03, sin haber mejorado por selección de variables ni tampoco por PCA. Para clasificar animales extremos con la base de datos completa se alcanzó una AUC_V de 0.63±0.07, mejorando hasta 0.75 ±0.06 con selección de variables.

En resumen, la composición microbiana tomada al sacrificio es un predictor bastante pobre de las emisiones de CH_4, tanto si se expresan como MP, MY o RM. Esto puede ser debido al tamaño muestra reducido, o a la limitación de la metagenómica para distinguir entre microbios activos y no activos. Sin embargo, la información microbiana podría usarse para clasificar a los animales en función de sus emisiones de CH_4, o clasificar animales extremos, con una precisión limitada. Esto podría ser útil como un primer filtro para discernir que animales vale la pena medir en cámaras de respiración en un programa de mejora de disminución de CH_4.

Abstract

Methane (CH_4) is one of the main greenhouse gases and has 28-34 times greater potential to capture heat in the atmosphere than CO_2. Enteric fermentation of livestock contributes significantly to these emissions, accounting for 32% of global anthropogenic CH_4 emissions. Genetic improvement is an attractive strategy to reduce these emissions since it is permanent and cumulative, but it requires the collection of a large amount of data. However, direct measurement of CH_4 can be very expensive or difficult to carry out on a large scale. Therefore, developing indirect methods to estimate CH_4 using easy-to-measure proxies as predictors is an attractive option. Metagenomic data could serve as an indicator for indirect prediction of CH_4 emissions given their close relationship.

The present study proposes predictive models of three CH_4 phenotypes: CH_4 Production (MP, g CH_4/day), CH_4 Yield (MY, g CH_4/kg daily dry matter intake) and Residual CH_4 (RM, g CH_4/ day) using ruminal microbiota data (3,631 gene abundances and 1,136 genus abundances, transformed with additive and centered log-ratio) from 287 animals.

The animals were of different breeds, were fed two different diets, and belonged to different experiments. All these effects were tested and corrected in the CH_4 phenotypes and in the microbial variables before adjusting the algorithms. CH_4 emissions were measured in respiration chambers and the microbiome was measured by taking rumen samples at slaughter and sequencing the entire metagenome.

The Partial Least Squares Regression (PLS), eXtreme Gradient Boosting (XGB) and Random Forest (RF) algorithms were adjusted to predict CH_4 continuously, for classification into 5 categories (emissions score, from low to high). and in extreme animals (25% of animals that emit the least and 25% of those that emit the most). All models were fitted using 70% of the data as training and leaving the remaining 30% as external validation. In addition, a range of precision in external validation was estimated by splitting the database 100 times into a training set and a validation set randomly. To try to avoid overfitting, in addition to adjusting the algorithms with all microbial variables, two more strategies were carried out to avoid overfitting: a principal component analysis (PCA) and using the principal components as predictive variables; and variable selection.

The continuous prediction results suffered from overfitting for XGB and RF, but not for PLS. In the validation set, the maximum average prediction value in external validation (Q^2) was 0.09±0.05. With variable selection, a value of 0.18±0.08 was achieved. However, the PCA compression strategy worsened the results. For the classification into 5 categories, an area under the curve in external validation (AUC_v) of 0.68 ± 0.03 was achieved, without having improved by variable selection or by PCA. To classify extreme animals with the complete database, an AUC_V of 0.63 ± 0.07 was achieved, improving to 0.75 ± 0.06 with variable selection.

In summary, the microbial composition taken at slaughter is a rather poor predictor of CH_4 emissions, whether expressed as MP, MY or RM. This may be due to the small sample size, or the limitation of metagenomics to distinguish between active and non-active microbes. However, microbial information could be used to classify animals based on their CH_4 emissions, or classify extreme animals, with limited accuracy. This could be useful as a first filter to discern which animals are worth measuring in respiration chambers in a CH_4 depletion breeding program.

1. INTRODUCCIÓN

1.1. Panorama actual de emisiones de gases de efecto invernadero

Las emisiones de gases de efecto invernadero (GHG) se han convertido en una preocupación internacional urgente. Un informe de la Organización Meteorológica Mundial (OMM) muestra que las medias globales de las concentraciones atmosféricas de dióxido de carbono (CO_2), metano (CH_4) y óxido nitroso (N_2O), tres de los principales gases de efecto invernadero, alcanzaron nuevos máximos en 2022: 417,9 ± 0,2 ppm para el CO_2, 1923 ± 2 ppb (partes por billón) para el CH_4 y 335,8 ± 0,1 ppb para el N_2O, lo que supone un incremento, respectivamente, del 150 %, 264 % y 124 % en comparación con el año 1750, es decir, con la era pre-industrial (World Meteorological Organization, 2023). Si hacemos foco en la concentración de CH_4 atmosférico, en 2023 ha aumentado en 10.9 ppb, con respecto a las mediciones del año anterior (NOAA, 2024).

La agricultura y sus residuos representan aproximadamente el 50-60% de las emisiones globales de CH_4 antropogénico. La fermentación entérica del ganado contribuye aproximadamente en un 32% de estas emisiones (IPCC, 2023). Con la proyección de que la población mundial superará los 9 mil millones de personas para el año 2050, y un incremento previsto en el consumo de carne de vacuno del 153% en comparación con 2010 (FAO, 2011), se anticipa un aumento sustancial en las emisiones de GHG derivadas de la producción de proteína animal. Sin embargo, la huella ambiental de estos productos dependerá de la especie y el tipo de proteína que produzca. En la Figura 1 se muestran las diferencias en cantidad y tipo de GHG emitidos entre especies rumiantes (bovinos, búfalos, cabras y ovejas) y monogástricas (cerdos y pollos). Mientras que, en los sistemas de cría de rumiantes, el CH_4 entérico representa una gran proporción de las emisiones totales, en los sistemas de cría de monogástricos, el CH_4, CO_2 y/o N_2O derivados de la producción de piensos, el cambio en el uso del suelo y la gestión del estiércol son los principales contribuyentes. En general, la producción de carne es uno de los principales productores de GHG, con dos terceras partes de los 6.2 Gigatoneladas de equivalentes de CO_2 generados por la producción de alimentos de origen animal en 2015 (FAO, 2023).

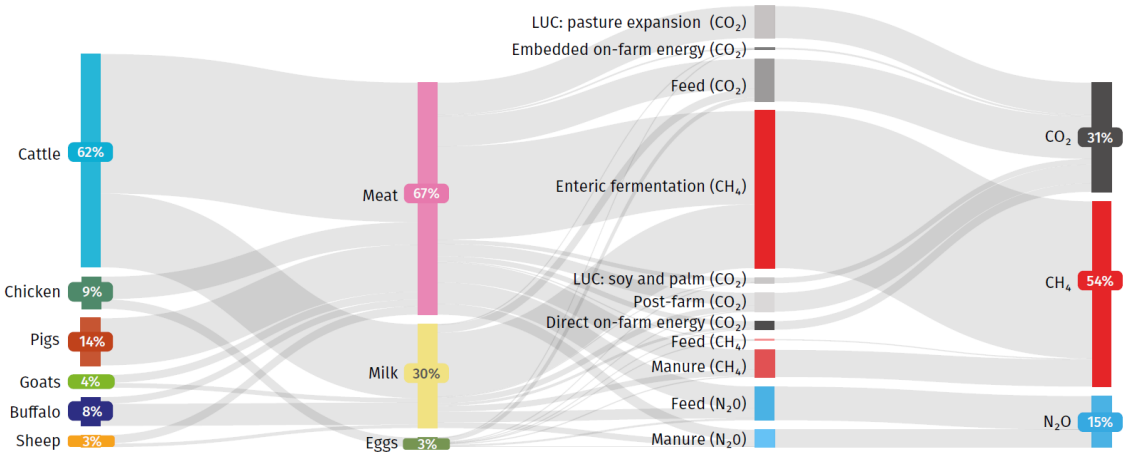


Figura 1. Diagrama de Sankey de fuentes de emisión de gases de efecto invernadero por especies, productos, fuentes de emisiones y gases. Porcentajes basados en un total de 6.2 Gigatoneladas de equivalentes de CO_2. Datos obtenidos en GLEAM3 disponible en https://www.fao.org/gleam/dashboard/en/ Tomado de FAO 2023: Pathways towards lower emissions: A global assessment of the greenhouse gas emissions and mitigation options from livestock agrifood systems (FAO, 2023).

Para mitigar el calentamiento global y mantener el aumento de la temperatura por debajo de 2 °C, se ratificó en 2015 el Acuerdo de París, un tratado internacional jurídicamente vinculante bajo la Convención Marco de las Naciones Unidas sobre el Cambio Climático aprobado por 196 países (ONU, 2015). Para lograr este objetivo, la estrategia más efectiva es disminuir las emisiones de CH_4 debido a su alto potencial de calentamiento (entre 28-34 veces más capacidad de atrapar el calor en la atmosfera que el CO_2 (IPCC, 2014)) y su corto tiempo de vida en la atmósfera, por lo que su disminución repercutiría al calentamiento global a corto plazo. Para llegar a cumplir los objetivos de este acuerdo, es necesario reducir las emisiones de CH_4 provenientes de los rumiantes entre un 11% y un 30% para 2030, y entre un 24% y un 47% para 2050, en comparación con los niveles de 2010 (Arndt et al., 2022). La industria ganadera tiene un gran interés en desarrollar estrategias para mejorar la eficiencia de producción de carne y minimizar las emisiones de CH_4. No solo por disminuir su huella de carbono y cumplir los objetivos del Acuerdo de París; sino también para maximizar el aprovechamiento del pienso ya que el animal pierde entre el 2 y el 12% de la energía total ingerida al producir CH_4 (de Haas et al., 2011; K. A. Johnson & Johnson', 1995; Lassey et al., 1997).

1.2. Fenotipos de CH_4

Las emisiones de CH_4 pueden expresarse de diferentes maneras, constituyendo fenotipos diferentes, con sus ventajas y desafíos particulares. A continuación, se resumen los fenotipos de CH_4 más utilizados a nivel global, basados en los trabajos de Manzanilla-Pech et al., 2016 y de Haas et al., 2017:

- Producción de CH_4 (MP): gramos por día o litros por día de CH_4 producido por el animal. Tiene la desventaja de que está altamente correlacionado con la ingesta media diaria de materia seca (DMI) y con su nivel de producción (por ejemplo, en vacuno de carne la correlación con peso vivo es de 0.58±0.03, Donoghue et al., 2013).
- Rendimiento de CH_4 (MY): gramos o litros de CH_4 por kg de DMI. El problema de este fenotipo es que al ser un ratio entre dos caracteres, tiene ciertas dificultades a la hora de seleccionar (por ejemplo, si uno de los caracteres es más variable que el otro, tiende a ser más seleccionado).
- Intensidad de CH_4 (MI): gramos o litros de CH_4 por kg de carne producida (o en el caso de vacuno de leche, por litro de leche). Este es un carácter de interés para el consumidor ya que está relacionado con el producto que llega a su mesa. Este fenotipo también tiene la misma limitación que en el caso anterior, al tratarse de un carácter ratio, su incorporación en la selección puede ser difícil.
- CH_4 residual (RM): es la diferencia entre la producción de CH_4 observada y la esperada. Su estimación se calcula teniendo en cuenta el peso del animal y su consumo de alimento. La ventaja de este fenotipo es que presenta correlaciones favorables con caracteres de eficiencia (Herd et al., 2016; Manzanilla-Pech et al., 2021, 2022). Sin embargo, tiene como inconveniente que es una estima de un residuo, las cuales no son independientes entre sí.

1.3. Síntesis de CH_4 en el rumen

Los carbohidratos incorporados en la dieta son la principal fuente de energía para los rumiantes. En el rumen, los polisacáridos (mayormente celulosa, hemicelulosa y almidón) son hidrolizados a glucosa y a otros monosacáridos que luego serán metabolizados en ácidos grasos volátiles (VFA) (propionato, butirato y acetato) y CO_2 (Figura 2). Durante este proceso de metabolización se genera hidrogeno metabólico ([H*]) que reduce los cofactores NAD en el proceso de fermentación. Para que el proceso continúe, los cofactores deben ser oxidados. Esto ocurre de

dos maneras. Por un lado, estos cofactores reducidos pueden ser utilizados en la generación de VFA (butirato y propionato). Por otro lado, [H*] puede generar hidrogeno molecular (H_2) gracias a la actividad de la enzima hidrogenasa. El mismo debe ser disipado de alguna forma para evitar aumentar la presión luminal en el rumen, que puede reducir la eficiencia de la fermentación y generar problemas de salud al animal. Este proceso de disipación del H_2 puede ocurrir de diferentes formas: por acetogénesis reductiva (lo cual conlleva generación de acetato, aprovechable por el animal), reducción de sulfuro, nitrógeno o fumarato o por metanogénesis (Janssen & Kirs, 2008). La metanogénesis hidrogenotrófica es la manera principal en que el H_2 es utilizado en la fermentación ruminal (Janssen, 2010). Otro tipo de metanogénesis es la metilotrófica, donde ciertas arqueas metanógenas pueden generar CH_4 por desmetilación de compuesto orgánicos (Poulsen et al., 2013). Esta forma de generar CH_4 ocurre gracias a la acción de gran variedad de enzimas desmetilazas y metiltransferasas (Neill et al., 1978). Otra manera en que se puede generar CH_4, es a través de la vía acetoclástica en donde el acetato es transformado en CH_4 y CO_2, aunque esta vía es minoritaria en el rumen (García et al., 2000).

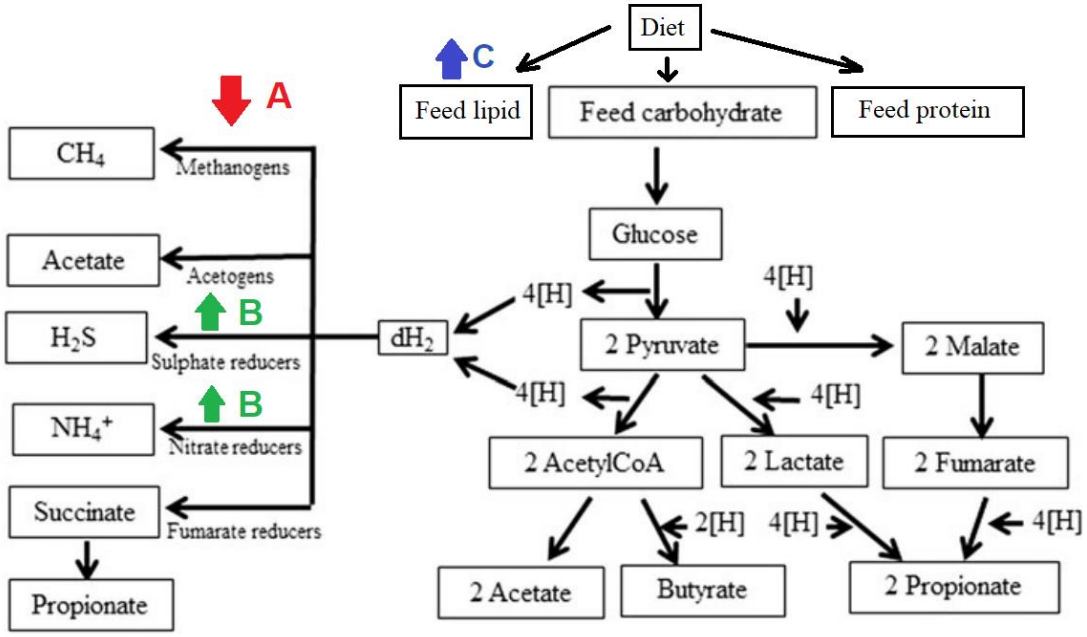


Figura 2. Esquema de metabolización de carbohidratos en el rumen de las vacas, rutas de aprovechamiento del hidrógeno (H_2) y estrategias de mitigación del metano (CH_4). Adaptado de Beauchemin et al., 2020. A) El uso de inhibidores de la metanogénesis; B) El uso de sulfatos y nitratos promueven vías de consumo de H_2 alternativas a la metanogénesis; C) El aumento de la ingesta de lípidos en la dieta.

1.4. Estrategias de mitigación de CH_4 en ganadería

Existen muchas estrategias para mitigar el CH_4 entérico producido por los rumiantes. A modo de resumen podemos nombrar:

1.4.1. Estrategias nutricionales

La manipulación de la dieta, siguiendo diferentes estrategias, ha demostrado ser una manera efectiva de mitigación de las emisiones de CH_4 (Beauchemin et al., 2009; Hristov et al., 2013; Knapp et al., 2014).

Una de las estrategias consiste en incorporar en las dietas aditivos con acción inhibitoria de la vía metanogénica (Figura 2, A). Un inhibidor que está cobrando mucha fuerza es el 3-nitrooxypropanol (3-NOP) conocido comercialmente como Bovaer®. El mismo, ha logrado

reducciones en las emisiones de CH_4 de entre un 22-35% en ganado, sin detrimento de su productividad ni bienestar (Hristov et al., 2015; Vyas et al., 2018). Sin embargo, estudios como el de McGinn et al., (2019) sugieren que, con el tiempo, los microorganismos podrían adaptarse a estos inhibidores, por lo que es un área de investigación en la que se debe profundizar. Otra estrategia para inhibir la vía metanogénica ha sido la incorporación de aceptores de electrones inorgánicos como nitrato y sulfato (Figura 2, B) en las dietas para favorecer las vías de generación de sulfuro de hidrogeno (H_2S) y amoniaco (NH_4^+) (Ungerfeld et al., 2007). Estudios de suplementación de nitrato sugieren que un exceso de nitrato sería perjudicial para los animales debido al riesgo de envenenamiento (Lee et al., 2017).

Un método que tambien se ha estudiado es la suplementación con lípidos (Figura 2, C), ya que los ácidos grasos de cadena larga resultan tóxicos para las arqueas metanógenas (Silva et al., 2016). Además, para una dieta isoenergética, se pueden agregar más lípidos y disminuir los carbohidratos, esto provoca una reducción de entre el 1% y 5% en CH_4 (g/día) por cada 10 g de grasa que se agrega a cada kg de materia seca en la dieta (Grainger & Beauchemin, 2011; Patra, 2013).

1.4.2. Inmunización y control biológico

Otro tipo de estrategias para reducir las emisiones de CH_4 consiste en modular o defaunar grupos específicos de microbios relacionados con la metanogénesis, por ejemplo, a través de la vacunación contra microorganismos metanogénicos. Actualmente el laboratorio AgResearch (Hamilton, New Zealand) está desarrollando una vacuna contra proteínas de superficie de microrganismos metanógenos ruminales (https://www.nzagrc.org.nz/vaccine.html). Hay que ser muy cautos con este tipo de estrategias ya que existe una gran diversidad de microorganismos metanogénicos (Wright et al., 2007) y el nicho que dejen los metanógenos que se eliminen con la vacuna podría ser ocupado por otras especies productoras de CH_4 (Williams et al., 2009), haciendo que este método no sea efectivo.

Con el uso de probióticos es posible la estimulación de poblaciones de microbios ruminales capaces de disminuir las emisiones de CH_4. Por ejemplo, se ha probado inocular microorganismo acetógenos que favorecen la acetogénesis reductiva. Sin embargo, los resultados han sido insatisfactorios ya que las bacterias acetogénicas no pueden competir por el H_2 con las arqueas metanogénicas en el rumen, aunque se inoculen en grandes cantidades (López et al., 1999; Nollet et al., 1998).

Por otro lado, se ha estudiado la defaunación de protozoos como alternativa, ya que los protozoos actúan como hospedadores de arqueas metanógenas, facilitando un ambiente propicio y sustratos para la producción de CH_4. Sin embargo, se ha visto que si bien se producen efectos beneficiosos con esta estrategia (aumentan las concentraciones propionato y emite menos CH_4), tienden a perderse con el tiempo (Z. Li et al., 2018).

1.4.3. Mejora genética

La mejora genética es una estrategia atractiva para lograr animales neutros en carbono, ya que la respuesta genética es permanente y acumulativa en el tiempo, aunque para llevarse a cabo se requiere una gran cantidad de datos (Pickering et al., 2015a). Negussie et al., (2017) y Wall et al., (2010) proponen que se puede reducir las emisiones de CH_4 por selección genética a través de dos estrategias: (1) Selección por emisiones de CH_4 medidas directamente o indirectamente, predichas a través de *proxies*. (2) Selección por animales más eficientes en su producción de leche o carne (ej: eficiencia alimentaria, longevidad) (Basarab et al., 2013; de Haas et al., 2011; Hegarty et al., 2007). La primera opción es tentadora ya que los distintos fenotipos de emisiones de CH_4 presentan heredabilidades moderadas de 0.22 ± 0.11 para MP, 0.19 ± 0.10 para MY y 0.23 ± 0.10 para MI, lo que los convierte en buenos objetivos de selección. La segunda opción

es efectiva según la heredabilidad del carácter de eficiencia y su correlación genética con el fenotipo de CH_4. En este sentido, MP tiene una correlación genética con el RFI de 0.76 ± 0.09, y una heredabilidad de 0.28, lo que indicaría que una selección para disminuir este carácter de eficiencia también disminuiría MP (Manzanilla-Pech et al., 2022).

A pesar de todo el interés global en disminuir los GHG en el mundo, Canadá es el único país que ha incluido las emisiones de CH_4 en sus índices de selección de vacuno como objetivo de mejora (Braian Van Doormaal, 2023). En las evaluaciones genéticas se incluye un carácter denominado Eficiencia de Metano (ME) el cual es una predicción del CH_4 emitido en función del espectro infrarrojo de la leche que produce el animal (Sweett, 2023). En Europa aún no se ha incluido el CH_4 como un objetivo de mejora en ningún índice (de Haas et al., 2017), ni tampoco existen políticas que incentiven la inclusión de estos caracteres en los objetivos de mejora de los índices de selección, necesario para predecir su peso económico. En España, González-Recio et al. (2020) han estimado el peso económico de MP en vacuno de leche, paso necesario para incluir este carácter en los índices de selección (González-Recio et al., 2020). Ellos han asignado al CH_4 peso económico de entre los € [−1.21, -0.32] por kg de CH_4 producidos, según diferentes futuros escenarios de impuestos y perdida energética.

En Nueva Zelanda, existen en la actualidad líneas divergentes de ovejas seleccionadas por MY que presentan una diferencia de emisiones media entre líneas del 12% (Rowe et al., 2019). La selección directa por bajo CH_4 muestra que ha disminuido el tamaño del rumen de los animales y ha modificado su composición microbiana. Además, ha mostrado respuestas correlacionadas favorables en otros caracteres productivos como producción de lana y composición de ácidos grasos de la carne. Este último resultado sería muy beneficioso si se replicara en el ganado bovino, tal y como sugiere Martínez-Álvaro et al. 2022 aunque esto aún está por confirmarse (Martínez-Álvaro et al., 2022).

1.5. Medición directa de las emisiones de CH_4

Se han desarrollado diferentes alternativas para la medición directa e individualizada del CH_4 que emite el ganado. La elección de la técnica dependerá de muchos factores tales como el costo, el nivel de precisión adecuado, el tipo de explotación, la escala y el diseño de los experimentos a realizar (Bhatta & Enishi, 2007).

1.5.1. Cámaras de respiración

Las cámaras miden la concentración de CH_4 recolectando el aliento exhalado del animal mientras el mismo permanece en la cámara, generalmente de 1 a 3 días (Figura 3) (Pickering et al., 2015b; Storm et al., 2012a). Hay dos tipos de cámaras, las de circuito abierto y las de circuito cerrado. En el circuito abierto, se hace pasar un flujo de aire de composición conocida por la cámara y se mide su cambio de composición cuando sale; mientras que en el circuito cerrado el aire se recircula en la cámara (regulando el nivel de oxígeno) y a través de un sistema de monitorización se mide el CH_4 producido por el animal (Turner & Thornton, 1966). La medición con este tipo de cámaras se considera el estándar de referencia debido a que se puede controlar el ambiente y se puede medir la fiabilidad y estabilidad de estos instrumentos (K. A. Johnson & Johnson', 1995). Sin embargo, el principal inconveniente de esta técnica es su alto coste. Además, al retener al animal en una cámara cerrada, se corre el riesgo de afectar su comportamiento y disminuir el DMI (Ellis et al., 2007). Como el DMI está directamente relacionado con las emisiones de CH_4, su disminución no solo afectaría las emisiones totales sino también las estimaciones derivadas, como la pérdida de energía bruta (Ellis et al., 2007). Para evitar este efecto, estudios han demostrado que si los animales se adaptan a estar en las cámaras ~24h antes de las mediciones, su DMI no se ve afectado. (Hellwing et al., 2014). Otros

intentos de reducir este efecto consisten en situar las cámaras en establos con más animales o hacer cámaras con paredes transparentes de forma que el animal pueda verse con los otros (Figura 3B) (Storm et al., 2012b). Aun así, se ha cuestionado que los resultados obtenidos en cámaras no se puedan aplicar a animales en libertad, por ejemplo, animales criados de manera extensiva (K. Johnson et al., 1994; O'kelly & Spiers, 1992).

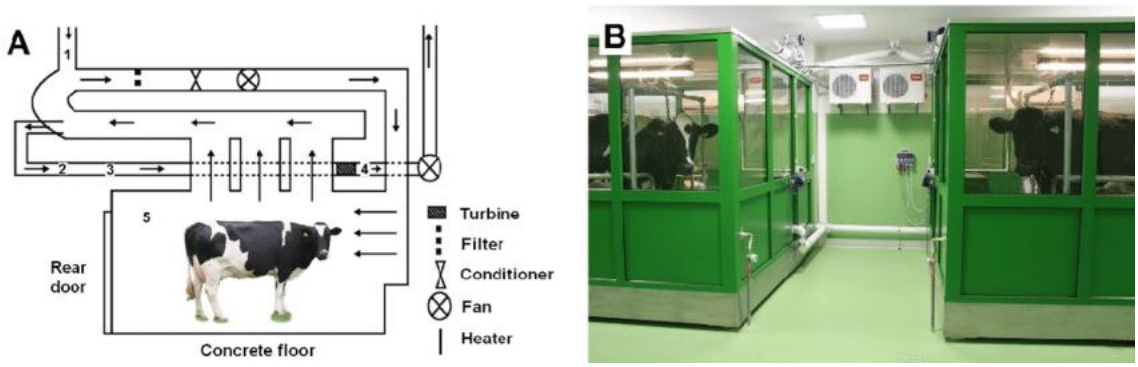


Figura 3. A) Esquema de la cámara de respiración de circuito abierto que muestra los flujos de aire (adaptado de Grainger et al., (2007). B) Instalaciones de investigación de la Universidad de Ciencias de la Vida de Poznan, Polonia (Fuente: http://globalresearchalliance.org/country/ Polonia/ Consultado: 1 de junio de 2024).

1.5.2. *Green Feed*

El sistema *Green Feed* consiste en la integración de un medidor de emisiones de CH_4 en los comederos de los animales. El aparato consta de un extractor que aspira aire sobre la cabeza del animal, para hacerlo pasar por la nariz y la boca y luego recoger el aire a través de un tubo de escape (Figura 4A). Los animales son atraídos a la estación con una pequeña cantidad de alimento, lo que permite medir las emisiones de CH_4 repetidamente y en diferentes momentos del día (Figura 4B). La desventaja es que las mediciones dependen de las visitas voluntarias de los animales al alimentador (Dressler et al., 2024). Esta técnica presenta una correlación de alrededor de 0.10 con las cámaras de respiración lo que la hace una técnica muy poco precisa (Hammond et al., 2015).

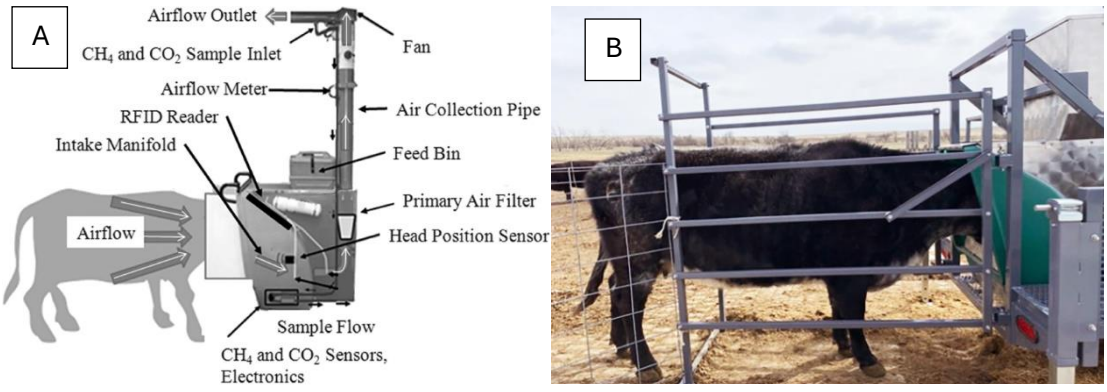


Figura 4. A) Esquema del sistema *Green Feed (*C-Lock Inc., Rapid City, SD). RFID = identificación por radiofrecuencia (Huhtanen et al., 2015). B) Una vaca en la estación de campo Cottonwood de la Universidad de Dakota, USA usando el *GreenFeeder* (Hector Menendez et al., 2023).

1.5.3. Técnica de hexafluoruro de azufre (SF_6):

Esta es una técnica invasiva en la que se introduce en el rumen del animal, de manera no quirúrgica, un tubo de permeación de un gas trazador: el SF_6. La idea básica es que la emisión de CH_4 se puede medir si se conoce la tasa de emisión de un gas trazador desde el rumen (Berndt et al., 2014) (Figura 5A). La proporción de CH_4 a SF_6 en la respiración se mide durante 24 horas y se repite durante un período de cinco a ocho días. La principal ventaja de esta técnica es que es la única que permite la medición de animales en libertad, pastando en el campo (Figura 5B). Sin embrago, requiere una calibración cuidadosa del método. Se debe medir precisamente la tasa de liberación del SF_6 ya que afectará las estimaciones de emisiones si no se determina correctamente (K. A. Johnson et al., 2007). Se han determinado correlaciones de 0.84 con las cámaras de respiración (Deighton et al., 2013) por lo que resultan una alternativa interesante y más aún si se necesita medir el CH_4 a animales criados en extensivo.

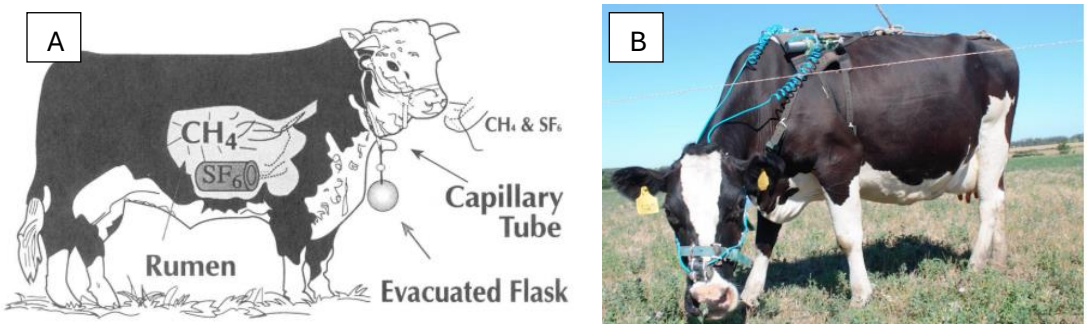


Figura 5. A) Esquema del funcionamiento del sistema de medición con SF_6 tomado de Johnson et al., 1994. B) Vaca utilizando el sistema de medición de metano con SF_6. Fotografía del libro: Protocolo para determinación de emisiones de metano en rumiantes, INIA Uruguay (Ciganda et al., 2022)

En resumen, la medición directa del CH_4 con cámaras de respiración se considera el estándar de referencia debido a su precisión, pero no es una alternativa viable en la producción a gran escala debido a que es económicamente muy costoso y lleva mucho tiempo para llevarse a cabo (Hammond et al., 2016). Otros métodos directos, como los trazadores de SF_6 o los sistemas *Green-feed*, requieren una calibración cuidadosa o no son muy precisos y dependen de las visitas voluntarias de los animales al alimentador (Dressler et al., 2024). Por lo tanto, el desarrollo de métodos indirectos para estimar el CH_4 basados en modelos matemáticos que utilizan *proxies* fáciles de medir como predictores es una opción atractiva.

1.6. Medición indirecta de las emisiones de CH_4: ecuaciones de predicción

El método de estimación de las emisiones de CH_4 más extendido a nivel productivo son las ecuaciones Tier 2, desarrolladas por el IPCC para tratar de crear un inventario que cuantifique las emisiones entéricas de la ganadería a nivel global (Hatfield et al., 2006). Estos modelos se basan en medidas de peso e ingesta media diaria y factores de emisión estándar que reflejan las condiciones promedio de producción y manejo en diferentes regiones del mundo (Hatfield et al., 2006). A pesar de su utilidad, los modelos Tier 2 tienen una capacidad limitada de capturar variaciones individuales en las emisiones de CH_4, que pueden ser importantes debido a diferencias el manejo del ganado, genética y las condiciones ambientales (Hristov et al., 2013).

Además de las ecuaciones Tier 2, se han desarrollado métodos individualizados para estimar las emisiones de CH_4 utilizando *proxies* como DMI (Ellis et al., 2007), peso vivo del animal (Yan et al., 2009) o eficiencia alimentaria. Por ejemplo, en 2007 Ellis et al. lograron un R^2 de 0.44 utilizando como predictor el DMI en ganado de carne sin utilizar validación externa (Ellis et al., 2007). Por otra parte, Yan et al. en 2009 lograron un R^2: 0.26 simplemente con el peso vivo del animal (Yan et al., 2009). Otro *proxy* que se utiliza en ganadería lechera son los ácidos grasos en

leche, los mismos están directamente relacionados con la digestión en el rumen, lo que los convierte en buenos candidatos como variables predictivas (Sweett, 2023). Utilizando ecuaciones ajustadas con el perfil de ácidos grasos se han encontrado resultados variables de predicción, con R^2 que van desde 0.47 hasta 0.80 utilizando distintos métodos, aunque no se reportan los rendimientos de predicción en validación externa (Rico et al., 2016; van Lingen et al., 2014)

1.7. Estimación de las emisiones de CH_4 a través de la composición microbiana del rumen

El rumen alberga un ambiente anaeróbico donde cohabitan una comunidad microbiana compleja y diversa que incluye bacterias, archaeas, protozoos, fungis y virus. Estas comunidades microbianas colaboran en la descomposición de la celulosa y otros polisacáridos presentes en la dieta del ganado, produciendo VFA que son absorbidos y utilizados como fuente de energía por el animal (Morgavi et al., 2013). Como se ha explicado anteriormente, este proceso produce grandes cantidades de H_2 como subproducto. Las arqueas metanógenas lo utilizan para producir su propia energía y en este proceso liberan CH_4. Debido a la estrecha relación que existe entre los microorganismos ruminales y las emisiones de CH_4, las abundancias microbianas (géneros y genes) podrían ser un *proxie* interesante para predecir estas emisiones. De hecho, múltiples estudios han encontrado correlaciones positivas entre microorganismos metanógenos y emisiones de CH_4 (Aguinaga Casañas et al., 2015; Arndt et al., 2015; Martínez-Álvaro et al., 2020; Miller et al., 2023; Wallace et al., 2015); incluso, se han reportado trabajos que utilizan el perfil metagenómico completo para predecir MP con un R^2 0.21 en validación externa utilizando el algoritmo M-BLUP (Ross et al., 2013). Una limitación de las predicciones de ciertos fenotipos con datos de metagenómica es que el microbioma del huésped puede ser bastante variable; es decir, puede cambiar según el momento en que se toma la muestra, por cambios en la dieta, las instalaciones u otros factores ambientales (Negussie et al., 2017). Sin embargo, Lima et al., (2022) ha reportado que la composición microbiana del rumen tomada en varios puntos temporales, tiene una capacidad de predecir las emisiones de CH_4 medidas en cámaras similar a lo largo del tiempo, de entre 60-80%, con un modelo de Regresión de Mínimos Cuadrados Parciales (PLS) y sin validación externa (Lima et al., 2022).

1.8. Potencial de los algoritmos de inteligencia artificial para predicción de fenotipos

En el campo de la ganadería, cada vez hay más información disponible con múltiples variables (por ej. ómicas, imágenes de video) que hace imposible su manejo mediante métodos estadísticos lineales clásicos. Aunque existen técnicas multivariantes como Regresión de Mínimos Cuadrados Parciales (PLS) para regresión o PLS discriminante (PLS-DA) para clasificación, estos algoritmos están diseñados para capturar relaciones lineales entre predictores y la variable respuesta. Por el contrario, los algoritmos basados en árboles de regresión/clasificación (como por ejemplo *Random Forest* (RF), *eXtreme Gradient Boosting* (XGB)), pueden capturar relaciones lineales y no lineales entre las variables predictoras y la variable respuesta. La principal ventaja de este tipo de algoritmos es su alta adaptabilidad pudiéndose aplicar a una amplia gama de tipos de datos y dominios problemáticos, que no podría ser resueltos por estadística clásica, incluidos aquellos con patrones complejos y no lineales (Hastie et al., 2001). Además, son capaces de analizar grandes bases de datos de manera más eficiente en tiempo y espacio computacional.

RF es un algoritmo que construye múltiples árboles de decisión durante el entrenamiento y produce la media de las predicciones individuales de todos los árboles (Breiman, 2001). XGB, es un algoritmo de *boosting* que construye modelos secuenciales donde cada nuevo modelo

intenta corregir los errores de los modelos anteriores, resultando en un modelo final muy preciso para muchas aplicaciones (Chen & Guestrin, 2016). El principal problema de los algoritmos de ML es el sobreajuste, sobre todo en el caso donde el número de variables predictoras y, por tanto, coeficientes a estimar, es mucho mayor que el número de datos. Los algoritmos de ML entrenados con datos insuficientes o de baja calidad pueden conducir a modelos que se ajustan demasiado bien a los datos de entrenamiento, capturando ruido y otros patrones irrelevantes, pero que no generalizan bien a otros conjuntos de datos (B. Zhang et al., 2023). Para salvar este problema, se debe realizar una cuidadosa optimización de los hiperparámetros del modelo, generalmente a través de procedimientos de validación cruzada, para encontrar un equilibrio entre poco sesgo (error en el ajuste) y poca varianza (comportamiento similar en diferentes bases de datos). Una vez optimizado, se puede validar la capacidad de predicción del modelo a través de un conjunto externo de datos, y comprobar si son capaces de generalizar.

Cuando se cuentan con muchas más variables que observaciones, disminuir la dimensionalidad de las variables predictoras puede prevenir el sobreajuste. Una opción es proyectarlas en un subespacio de menores dimensiones que capture su máxima variabilidad (Análisis de componentes principales, PCA) (Vinutha et al., 2023), y usar estas variables latentes como predictoras en un modelo de predicción/clasificación (Bhagat & Kumar, 2023). Otra alternativa es realizar selección de variables predictoras (SV), ya que la inclusión de variables irrelevantes puede introducir ruido en el modelo, reduciendo su capacidad para capturar relaciones informativas en los datos (Guyon & De, 2003). La selección de variables no solo optimiza el rendimiento de los algoritmos, sino que también permite centrarse en estas variables para su interpretación biológica.

Existen múltiples estudios donde se han utilizado modelos de ML para predecir CH_4. Por ejemplo, un estudio reciente Zhang et al. (2023) utilizó géneros de bacterias ruminales para predecir la producción de CH_4 en ovejas a través de RF (Zhang et al., 2023) con un error cuadrático medio de predicción de entre 3 y 2.85. En vacuno de leche, Wallace et al. 2019 predijo el MP utilizando un *core* de microorganismos y algoritmos de ML consiguiendo modelos con R^2 de 0.05-0.4 dependiendo de la granja de los que provenían los datos, a través de una validación cruzada tipo *leave-one-out* (Wallace et al., 2019).

La hipótesis de este trabajo es que aplicación de técnicas avanzadas de ML podría mejorar la precisión de las estimaciones de emisiones de CH_4 con datos de microbiota del rumen frente a modelos lineales, facilitando el desarrollo de estrategias más efectivas para obtener estimas de CH_4 a gran escala. Sin embargo, es necesario validar estos algoritmos con datos externos para evaluar si verdaderamente es una herramienta útil para la ganadería.

2. OBJETIVOS

2.1. Objetivo General

El objetivo de este trabajo es desarrollar y validar algoritmos *machine learning* de predicción/clasificación para emisiones de CH_4 en vacuno de carne usando datos de composición microbiana del rumen. Para ello, se comparará el rendimiento en datos externos de algoritmos multivariantes que capturan relaciones lineales (PLS y PLS-DA) y algoritmos de que capturan relaciones lineales y no lineales (RF, XGB).

2.2. Objetivos Específicos

2.2.1. Evaluar la capacidad de la composición microbiana para predecir de manera continua diferentes fenotipos de CH_4 (MY, MP y RM) utilizando diferentes algoritmos.

2.2.2. Evaluar la capacidad de la composición microbiana para clasificar los animales según sus emisiones de CH_4 en un sistema de puntuación (*score*) del 1-5 (1: Bajos emisores, 2: Emisores bajos-medios, 3:

Emisores medios, 4: Emisores medio-altos y 5: Altos emisores) o en emisiones de CH_4 extremas (2 categorías, Altos y Bajos).

2.2.3. Evaluar si usar técnicas de reducción de dimensiones del microbioma como PCA y/o selección de variables mejora el rendimiento de los algoritmos.

2.2.4. Interpretar biológicamente los resultados obtenidos de la selección de variables para ver su relación con el CH_4.

3. MATERIALES Y MÉTODOS

3.1. Datos

3.1.1. Animales

Los datos se obtuvieron de 287 novillos de 659±54 kg de peso y 528±38 días de edad. Los mismos eran de diferentes razas (cruce rotacional de razas Aberdeen Angus (n=76) y Limousin (n=72), cruces Charolais (n=68) y raza pura Luing (n=67)). Estos animales fueron alimentados con dos dietas diferentes una forrajera y otra concentrada, con proporciones de forraje/concentrado de 480/520 (n=182) y 80/920 (n=105), respectivamente. Los animales procedían de diferentes experimentos (Duthie et al., 2016, 2017, 2018; Rooke et al., 2014) realizados durante 5 años (2011, 2012, 2013, 2014 y 2017) en la misma granja y bajo las mismas condiciones de hospedaje. Cada uno de estos experimentos fueron aprobado por el Comité de Experimentación Animal de SRUC y se llevó a cabo siguiendo los requisitos de la Ley de Animales (Procedimientos Científicos) del Reino Unido de 1986.

3.1.2. Medición de las emisiones de CH_4

Se midieron las emisiones de CH_4 individualmente de los 287 animales durante 48 h dentro de seis cámaras de respiración de circuito abierto (Figura 6; Rooke et al., 2014). Una semana antes de las mediciones los animales se alojaron en cámaras de entrenamiento exactamente iguales a las de medición y además se adaptaron a las cámaras reales durante 24h. Dentro de cada experimento, los animales fueron asignados a las cámaras de respiración en un diseño aleatorio para raza y dieta. Los animales fueron alimentados una vez al día y se registró el DMI.

Figura 6. Fotografía de las cámaras de respiración de circuito abierto del *Scotland's Rural College, Edinburgh* (SRUC), utilizadas para obtener los datos de emisiones de metano de este trabajo.

Las emisiones de CH_4 obtenidas en la cámara se expresaron como producción de CH_4 diaria (MP, g CH_4/día). De este fenotipo y con los datos de DMI se calculó el rendimiento de CH_4 (MY, g CH_4/kg DMI). Además, se calculó el CH_4 residual (RM, g CH_4/día) (Herd et al., 2014) corrigiendo MP por DMI y el peso de los animales:

$$MP\ (g\ CH4/día) = \beta 0 \mp \beta 1 \cdot Peso \mp \beta 2 \cdot DMI\ +\ e$$

Donde el *Peso* es el tomado al ingreso de la cámara de CH_4 (la media y la sd de los pesos de los animales fue de 659±54 kg), DMI es la ingesta de materia seca diaria durante la estancia en la cámara (11.03± 1.04kg/día) y *e* es el CH_4 residual (RM).

Además de contemplar las emisiones de CH_4 como un fenotipo continuo, también se contempló su clasificación en 5 categorías (CH_4 *score*, 1 - 5) o en 2 categorías extremas (Altos y Bajos emisores). Todas las categorías se obtuvieron con los datos de CH_4 ajustados por el efecto de dieta. En el primer caso, se dividió cada fenotipo de CH_4 en 5 categorías equilibradas en número de animales (Tabla 1). Se comprobó que las diferencias entre una categoría y la siguiente fueron relevantes en todos los casos (probabilidad de relevancia (P_R)=1.00), usando como valor relevante un tercio de la sd del carácter (Tabla 2). Este análisis se resolvió con estadística Bayesiana usando con el paquete de R RabbitR (Martínez-Álvaro et al., 2023).

Tabla 1. Distribución de individuos por categorías en cada fenotipo de metano, con la media y desviación típica (sd) de sus emisiones.

Fenotipo[1]	**CH_4 *Score*[2]**	**Nº individuos**[3]	**Media ± sd**[4]
MY	1	53	9.89 ± 0.96
[g CH_4/kg	2	44	12.1 ± 0.60
DFI]	3	60	13.9 ± 0.52
	4	53	15.7 ± 0.56
	5	61	19.1 ± 2.2
	TOTAL	271	14.3 ± 3.4
MP	1	54	89.2 ± 17
[g CH_4/día]	2	85	130 ± 9.6
	3	86	162 ± 10
	4	36	197 ± 10
	5	19	277 ± 56
	TOTAL	280	151 ± 51
RM	1	32	-59.2 ± 9.8
[g CH_4/día]	2	75	-28.3 ± 9.5
	3	84	0.92 ± 8.9
	4	39	29.7 ±7.9
	5	27	102 ± 60
	TOTAL	257	0 ± 48

[1]Fenotipos de metano: MP= Producción de metano, MY= rendimiento de metano, RM= metano residual.
[2]Categoria de emisión de metano según la cantidad que emitan 1-Bajos emisores, 2-Emisores bajos-medios, 3-Emisores medios, 4-Emisores medio-altos y 5-Altos emisores.
[3]Cantidad de individuos en cada categoría.
[4]Media y desviación típica de las emisiones de metano de esa clasificación (media±sd).

Tabla 2. Diferencias en emisiones de metano entre los grupos de clasificación en 5 categorías (CH_4 *Score*).

Fenotipo[1]	**Diferencia**[2]	**Mediana**[3]	**HPD**[4]	**P_0**[5]	**P_R**[6]
MY	**D_{1-2}**	-2.21	[-2.68 ,-1.73]	1.00	1.00
[g CH_4/kg	**D_{2-3}**	-1.85	[-2.33 ,-1.37]	1.00	1.00
DFI]	**D_{3-4}**	-1.76	[-2.21 ,-1.33]	1.00	1.00
	D_{4-5}	-3.41	[-3.86 ,-2.98]	1.00	1.00
MP	**D_{1-2}**	-40.94	[-47.13 ,-34.53]	1.00	1.00
[g CH_4/día]	**D_{2-3}**	-32.29	[-37.72 ,-26.87]	1.00	1.00
	D_{3-4}	-34.16	[-41.22 ,-26.65]	1.00	1.00
	D_{4-5}	-80.15	[-90.33 ,-69.77]	1.00	1.00

RM	**D_{1-2}**	-31.28	[-40.07 ,-22.69]	1.00	1.00
[g CH_4/día]	**D_{2-3}**	-29.01	[-35.29 ,-22.42]	1.00	1.00
	D_{3-4}	-28.67	[-36.91 ,-21.11]	1.00	1.00
	D_{4-5}	-72.23	[-82.11 ,-61.83]	1.00	1.00

[1]Fenotipos de metano: MP= Producción de metano, MY= rendimiento de metano, RM= metano residual.
[2]Categorias entre las que se está comparando si existen diferencias **D_{i-h}** diferencia entre el grupo i y el h.
[3]Mediana de la distribución marginal posterior (MDP) de la diferencia entre categorías.
[4]Intervalo de Máxima Densidad Posterior con un 95% de confianza de las MDP de las diferencias.
[5]Probabilidad de la diferencia de ser mayor o menor que 0.
[6]Probabilidad de la diferencia de ser mayor o menor de un valor relevante, tomando un tercio de la desviación típica como valor relevante.

En el segundo caso, se tomaron los individuos del cuartil 1 (Q_1) y el cuartil 4 (Q_4) en cada uno de los fenotipos de CH_4, para generar una nueva base de datos de individuos de baja producción de CH_4 y de alta producción de CH_4 respectivamente (Tabla 3). Se comprobó que las diferencias entre una categoría y la siguiente fueron relevantes en todos los casos (PR=1.00, Tabla 4).

Tabla 3. Distribución de individuos por categorías extremas en cada fenotipo de metano, con su media y desviación típica (sd).

Fenotipo[1]	**Categoría[2]**	**Nº individuos[3]**	**Media ± sd[4]**
MY	Bajos	67	10.22 ± 1.06
[g CH_4/kg DFI]	Altos	67	18.86 ± 2.25
MP	Bajos	70	95.34 ± 18.81
[g CH_4/día]	Altos	70	214.19 ± 49.03
RM	Bajos	64	-48.14 ± 13.22
[g CH_4/día]	Altos	64	59.92 ± 52.34

[1]Fenotipos de metano: MY: MP= Producción de metano, MY= rendimiento de metano, RM= metano residual.
[2]Categoria de emisión de metano según la cantidad que emitan: Bajos- pertenecen al 25% de animales que menos emiten; Altos- pertenecen al 25% de animales que más emiten.
[3]Cantiadad de individuos en cada categoría.
[4]Media y desviación típica de las emisiones de metano de esa clasificación (media±sd).

Tabla 4. Diferencias entre Altos y Bajos emisores de metano.

Fenotipo[1]	**Mediana[3]**	**HPD[3]**	**P0[4]**	**PR[6]**
MY	-118.86	[-132.04, -107.34]	1.00	1.00
MP	-107.81	[-120.49, -94.61]	1.00	1.00
RM	-8.65	[-9.24, -8.06]	1.00	1.00

[1]Fenotipos de metano: MP= Producción de metano [g CH_4/día], MY= rendimiento de metano [g CH_4/kg DFI], RM= metano residual [g CH_4/día].
[2]Mediana de la diferencia entre categorías.
[3]Intervalo de Máxima Densidad Posterior con un 95% de confianza de las MDP de las diferencias.
[4]Probabilidad de la diferencia de ser mayor o menor que 0.

[5]Probabilidad de la diferencia de ser mayor o menor de un valor relevante, tomando un tercio de la desviación típica como valor relevante.

3.1.3. Medición de la composición microbiana del rumen mediante secuenciación del metagenoma completo

Inmediatamente después del sacrificio, se recolectaron 5 mL de líquido ruminal, y se mezclaron con 10 mL de PBS y glicerol (87%). La extracción de ADN se realizó siguiendo el protocolo de Yu y Morrison (Yu Z. and Morrison M, 2004). Las bibliotecas de ADN Illumina TruSeq se prepararon a partir de ADN genómico y se secuenciaron en sistemas Illumina HiSeq 4000.

Para la anotación filogenética, las lecturas se alinearon con las bases de datos Hungate 1000 (Seshadri et al., 2018) y RefSeq (Pruitt et al., 2007) utilizando el software Kraken (Wood & Salzberg, 2014). Para la anotación funcional, las lecturas se alinearon con la base de datos de Kyoto Encyclopedia of Genes and Genomes Orthologue utilizando el programa KOunt (Mattock et al., 2023). De los 1178 géneros y 7976 genes microbianos identificados, se seleccionaron 1136 géneros y 3632 genes microbianos presentes en al menos el 70% de los animales. Los genes y géneros que tenían cero *counts* fueron imputados basándose en un método multiplicativo bayesiano geométrico (GBM) (Martín-Fernández et al., 2015), este paso es necesario para poder hacer las transformaciones *log-ratio*.

Las abundancias de genes se transformaron utilizando la transformación *log-ratio* aditiva (ALR) utilizando el gen microbiano *rpe* (ribulose-phosphate 3-epimerase, código KEGG K01783) como referencia o denominador. La transformación se realizó utilizando la siguiente formula:

$$ALR(j \mid ref) = log(X_j/X_{ref}), \qquad j = 1,\ldots,J, \qquad j \neq ref$$

Donde J es el número de variables microbianas, X_{ref} es la variable elegida como referencia y X_j es cada una de las variables diferentes a X_{ref}. Como resultado se obtuvieron 3631 abundancias de genes ALR transformadas (alr-MG). El criterio para seleccionar el gen K01783 de referencia fue una compensación entre dos condiciones explicadas en Greenacre et al., (2021). En primer lugar, una correlación de Procrustes de 0,9974 entre la geometría *log-ratio* y la geometría aproximada generada por el conjunto de alr-MG, lo que garantiza que las distancias euclídeas entre muestras se conserven después de la transformación. En segundo lugar, una varianza baja de 0,0379 en el logaritmo de la abundancia relativa del gen *rpe* (coeficiente de variación del 5,08%), lo que facilita la interpretación de ALR. A nivel taxonómico, no se encontró ningún género microbiano que cumpliera con los criterios para ser usado como referencia en el ALR por lo que se decidió realizar una transformación *log-ratio* centrada (CLR). La misma se calcula con la siguiente formula:

$$CLR(j) = \log\big(X_j/g(X)\big), \qquad j = 1,\ldots,J$$

Donde *g(X)* es la media geométrica de cada individuo (Greenacre, 2018). Como resultado se obtuvieron 1136 abundancias de géneros CLR transformadas (clr-MT).

3.2. Análisis estadístico

3.2.1. Análisis exploratorios y detección de outliers

En primer lugar, se realizaron las tablas de contingencia para observar la distribución de los animales por efectos fijos: experimento, dieta y raza para descartar que hubiera efectos confundidos. También se realizó un análisis exploratorio de cada uno de los fenotipos de CH_4 y se detectaron *outliers* univariantes tomando como criterio 3 sd desde la media.

Para identificar los animales *outliers* en su composición microbiana (clr-MT y alr-MG), así como para detectar posibles patrones entre los animales, se utilizó el PCA. El mismo se testó con dos tipos de escalado de los datos. El primero se decidió realizar con autoescalado (lo que le da a cada variable la misma importancia) y el segundo PCA se realizó con un escalado por bloques. Este se hace ponderando la dispersión de cada variable por el número de variables en su bloque y así darle la misma importancia a los datos que vienen de alr-MG (m_{b1}=3631) y clr-MT (m_{b2}=1136). La ecuación para el cálculo de escalado por bloques es la siguiente:

$$X_{ik\ (escalado)} = \frac{X_{ik}}{\sqrt{m_{bi}} * sd_k}$$

Donde cada observación (X_{ik}) de un alr-MG o clr-MT (K), se divide por la desviación típica de K (sd_k) multiplicada por la raíz cuadrada del número de variables que tiene ese bloque (m_b) (van den Berg et al., 2006). Se analizaron los gráficos de T^2-Hotteling (T^2) y suma de cuadrados residual (SCR) en ambos PCA. El criterio para considerar como *outliers* a algún animal fue que superara dos veces el límite de confianza de 0.99 de la T^2 o de SCR en el análisis PCA.

3.2.2. Análisis de los efectos de dieta, experimento y raza.

Se analizó el efecto de la raza, año de experimentación y dieta sobre los fenotipos de CH_4 con un modelo lineal, ajustándolos como efectos fijos. Por otro lado, para estudiar estos efectos sobre las variables microbianas se realizó un análisis a través de un PERMANOVA con el paquete de R vegan (Oksanen et al., 2024).

3.2.3. Ajuste y validación de los algoritmos

Todos los algoritmos se ajustaron independientemente para cada fenotipo de CH_4, usando la combinación de 4767 variables microbianas (3631 alr-MG y 1136 clr-MT) como predictoras. Tanto las variables microbianas como los fenotipos se utilizaron tras ser corregidos por los efectos correspondientes.

Se dividió la base de datos en un conjunto de entrenamiento (70% de los animales), con el que se ajustaron los modelos, y un conjunto de validación (30% de los animales) con el que se validaron los modelos de manera externa (Figura 7). La optimización de los modelos se realizó usando validación cruzada (CV) sobre el conjunto de entrenamiento. Ésta consistió en subdividir el conjunto de entrenamiento en un subconjunto de CV-calibración (70%) y uno de CV- validación (30%). Con el CV-calibración, se probaron los algoritmos con todas las combinaciones posibles de valores de hiperparámetros y se predijo el CV-validación. Se eligieron aquellos hiperparámetros que maximizaban la R^2 (en predicción, estimada como el cuadrado de la correlación de los valores predichos y los reales) o el Área Bajo la Curva ROC (AUC, en clasificación) del conjunto de entrenamiento (AUC_e). Una vez el valor de los hiperparámetros se optimizó, se ajustó un modelo (modelo optimizado) en el que se evaluó, de manera completamente externa, el poder de predicción/clasificación del modelo (Q^2/AUC_v) con el conjunto de validación externo (Figura 7).

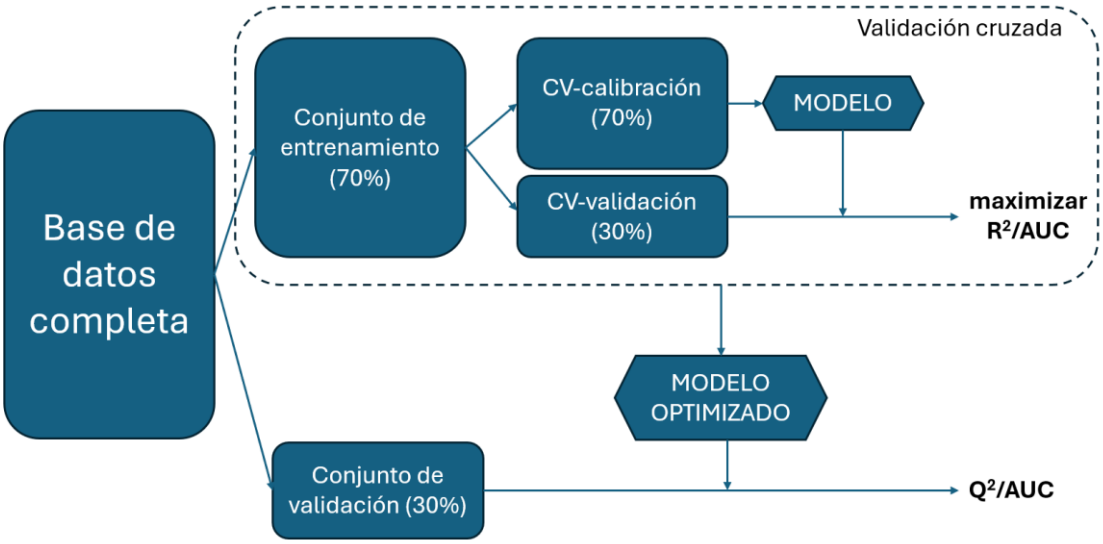


Figura 7. Esquema de doble validación cruzada utilizado en la optimización de los hiperparámetros de los diferentes algoritmos (RF, XGB, PCA-XGB y PCA-RF).

Por último, para calcular un rango del rendimiento del modelo optimizado en validación externa (AUC_v o Q^2), se dividió la base de datos en un conjunto de entrenamiento y un conjunto de validación 100 veces, de manera aleatoria. En cada división, se ajustó el modelo optimizado con el conjunto de entrenamiento, y se utilizó para predecir/clasificar el conjunto de validación. Así, podemos dar una idea de cómo se comportaría el modelo optimizado al enfrentarse a distintas bases de datos y determinar su rango de precisión (Figura 8).

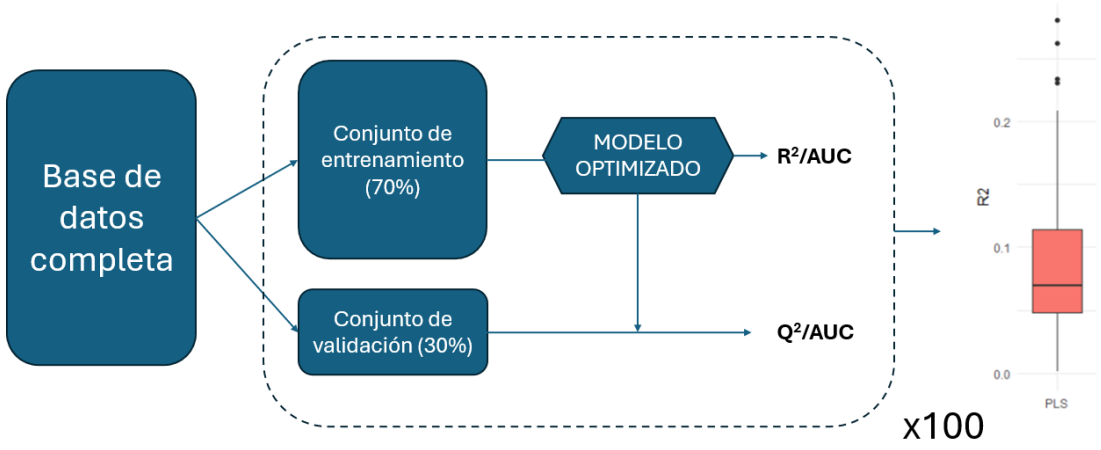


Figura 8. Esquema de estimación de la precisión de las predicciones/clasificaciones con la base de datos y los modelos optimizados de cada uno de los algoritmos.

3.2.4. Algoritmos de predicción y clasificación:

En este trabajo se utilizaron 3 algoritmos diferentes para predicción y clasificación de las emisiones de CH_4. Las características de estos algoritmos se detallan a continuación:

3.2.4.1. Regresión de Mínimos Cuadrados Parciales (PLS):

Este algoritmo siguió un esquema levemente diferente al descrito en la Figura 7. En este caso, primero se realizó un PLS o un PLS-DA con el conjunto de entrenamiento, según fuera

predicción o clasificación, con el paquete *mdatools* (Kucheryavskiy, 2020). El hiperparámetro número de componentes se optimizó con una CV de 5 *folds* y 10 repeticiones. De este primer PLS/PLS-DA con el número de componentes óptimo se seleccionaron las variables más importantes en función de dos criterios: aquellos en los que su influencia en la proyección (VIP) tenían valores mayores a 0.8 y además que tenían coeficientes de regresión cuyo intervalo de confianza Jackknife no incluía al cero. Con las variables seleccionadas, se ajustó otro modelo PLS/PLS-DA, que se utilizó para predecir/clasificar el conjunto de validación.

Finalmente, se dividió la base de datos en un conjunto de entrenamiento y en otro de validación de forma aleatoria 100 veces, se ajustó el modelo PLS con las variables y el número de componentes optimizada en el conjunto de entrenamiento, y se utilizó para predecir/clasificar el conjunto de validación y así obtener los rangos de Q^2 y AUC_v (Figura 8).

3.2.3.2. Random Forest (RF)

Se utilizó el algoritmo el RF a través del paquete de R *randomForest* (Andy Liaw & Matthew Wiener, 2002). Los hiperparámetros optimizados por CV fueron: número de árboles que se van a generar en el modelo (*ntree*), número de variables que se le da a elegir al algoritmo en la formación de cada nodo (*mtry*), número de nodos máximos que puede tener cada árbol (*max_nodes*) y el número mínimo de individuos que debe tener cada nodo (*nodesize*). Para intentar evitar el sobreajuste, y tras una análisis exploratorio, el rango de valores a elegir se limitó a *ntree*: [300, 500, 800]; *mtry:* [p/3, p/6, sqrt(p), p/10, p/20] (donde p es el número de variables con las que se entrena el algoritmo)*; max_nodes:*[5, 10, 15, 25, 35]; *nodesize*: [2, 4, 8, 10, 15, 20].

3.2.3.3. XGBoost (XGB)

Se utilizó el algoritmo XGB del paquete *xgboost* (Chen et al., 2020). Los hiperparámetros optimizados por CV fueron: Número de iteraciones o rondas de *boosting* (*nrounds*), profundidad máxima del árbol de decisión (*max_depth*), la tasa de aprendizaje (*eta*), la mínima pérdida de reducción requerida para hacer una partición adicional en un nodo de árbol (*gamma*) y la proporción de variables a ser muestreadas para cada árbol (*colsample_bytree*). Los rangos de valores en los que se probaron estos hiperparámetros fueron: *nrounds*: [300,500]; *max_depth:* [10, 20, 30]; *eta:* [0.01, 0.1, 0.3]; *gamma:* [0.5, 1, 3, 6, 10]. El valor de *subsample* (proporción de observaciones a ser muestreadas para cada árbol) se fijó en 0.5.

3.2.5. Estrategias de reducción de la dimensionalidad de la microbiota

3.2.5.1. PCA+RF y PCA+XGB

La primera estrategia de reducción de dimensionalidad fue realizar un PCA con los datos de microbiota, y usar las proyecciones de los individuos obtenidos en los componentes principales como variables predictoras para el ajuste de los algoritmos RF (PCA+ RF) y XGB (PCA +XGB) (Gupta et al., 2022). Se utilizaron todas las componentes que se obtenían con el PCA ya que componentes que no explican tanta variabilidad podrían estar mucho más correlacionadas con la respuesta que aquellas que son muy explicativas a nivel de varianza general de los datos en las variables predictoras. Para ser rigurosos con la validación externa, el PCA se ajustó solamente con el conjunto de entrenamiento, y se usó este modelo PCA para obtener las proyecciones en el conjunto de validación (196 componentes). El resto de los pasos se realizó como se ha descrito previamente para RF y XGB.

3.2.5.2. Selección de variables

La segunda estrategia de reducción de dimensionalidad fue ajustar los algoritmos con un conjunto reducido de variables microbianas seleccionadas. Las variables fueron elegidas a través de un consenso entre las metodologías PLS, RF y XGB, es decir, se hizo selección de variables con cada una por separado y luego se tomó las variables que eran elegidas por los 3 métodos. Esta forma de reducir las dimensiones de la base de datos está respaldada por la tesis de máster de Duro-Vizcaino (2024) que será expuesta en paralelo a este trabajo. Ella ha visto que, al utilizar un único algoritmo para seleccionar variables, aparecen un gran número de variables que son falsos positivos (es decir, variables sin un verdadero valor biológico), y una alternativa para evitar esto es utilizar las variables seleccionadas en una combinación de algoritmos.

Los criterios que se utilizaron en cada algoritmo para la selección de variables fueron los siguientes:

- Para el PLS se seleccionaron aquellas variables que tenían un VIP>0.8 y que sus coeficientes de regresión no incluyeran el cero en su intervalo de confianza Jackknife (0.95).
- Para el RF se calculó la importancia de cada variable por permutación, permitiendo seleccionar variables correlacionadas (Boulesteix et al., 2012). Se calcula permutando los valores de cada variable predictora y re-caculando la pérdida de precisión del modelo respecto al mismo con la variable sin permutar (Andy Liaw & Matthew Wiener, 2002). Se seleccionaron aquellas variables que al ser permutadas provocaban una pérdida de precisión mayor a la media.
- Para el XGB se seleccionaron todas aquellas variables que el modelo seleccionó por el criterio de *Gain*. Este criterio consiste en medir la mejora en precisión aportada por una variable en cada nodo donde participa. (Chen & Guestrin, 2016). El algoritmo da *Gain* 0 a todas aquellas variables que no ha utilizado para ningún árbol. Se eligieron las variables con *Gain*>0.

Además, se analizaron tanto las variables que habían sido elegidas en cada fenotipo como así también aquellas que se compartían entre los tres fenotipos, en búsqueda de evidencia biológica de genes o géneros que en bibliografía se los relacionara con las emisiones de CH_4.

3.3. Estimas de emisiones de CH_4 con las ecuaciones Tier 2

Las ecuaciones Tier2 son la forma que actualmente recomienda el IPCC para estimar la producción de CH_4 a nivel global. Estas ecuaciones son más precisas que sus antecesoras Tier 1, ya que incorporan variables como DMI, el tipo de forraje y la eficiencia de conversión de alimento en CH_4. En nuestro trabajo, se estimó la MP de los animales para comparar la precisión de las ecuaciones Tier 2 frente a la predicción de nuestros algoritmos con datos de microbiota.

Para estimar las emisiones de CH_4 se utilizó la siguiente ecuación (Dong et al., 2006):

$$MP_{Tier2}\left(g\frac{CH4}{día}\right) = DMI\left(\frac{kg}{día}\right) x\ EF\ (g\frac{CH4}{kg\ DMI})$$

Donde DMI es la ingesta de materia seca en kg, EF es el factor de emisiones específico que representa la cantidad de CH_4 producido por kilogramo de DMI y depende de la composición de la dieta y su digestibilidad. En este estudio, corresponde usar un EF de 23.3 para una dieta forrajera de 480/520 y de 21 para una dieta basada en concentrado 80/920 (Dong et al., 2006). Con las predicciones y el valor verdadero de MP se calculó el R^2 de las predicciones Tier2 como la correlación entre los valores predichos y MP, elevada al cuadrado.

4. RESULTADOS

4.1. Análisis exploratorio, detección de datos anómalos y efectos fijos

Se encontraron 7 animales con datos de MP muy por encima del límite fijado y con un valor exactamente igual para todos (492 g CH_4 /día). Los datos de estos animales se consideraron errores de anotación y no se consideraron para este estudio. Los animales con composiciones microbiana anómalas se detectaron a través de sus errores cuadráticos de predicción (SPE) y la T^2 de Hotelling (T^2) en el PCA escalado por bloques y el autoescalado. Se establecieron como anómalos aquellos individuos que superaban 2 veces el umbral de 0.99 del límite de confianza de estas métricas (Ferrer, 2007). Ningún animal supero el umbral de T^2 ni de SPE en ninguno de los PCA's (Material Suplementario 1).

Nuestra muestra de animales presentó unas emisiones medias de MP de150.58±50.86g CH_4/día, de MY de 14.36±3.43g CH_4/kg DFI y de RM de 0±48.26g CH_4/día (medias ajustadas por los efectos fijos). En las tablas de contingencia no se detectaron efectos confundidos por lo que se pudo evaluar cada uno de los efectos. La dieta fue relevante en MY y en MP, asumiendo como valor relevante 1/3 de la sd del carácter. La diferencia entre dietas (Forraje – Concentrado) en MY tuvo una mediana de -7.60 g CH_4/kg DMI con un Intervalo de Máxima Densidad Posterior con un 95% de confianza ($HPD_{95\%}$) de [-8.41,-6.73], y en MP de -70.84 [-83.24,-58.65] g CH_4/día (P_R = 1.00 en ambos casos). No se observaron diferencias relevantes para experimento y raza. El RM de estimó con MP ajustado por sus efectos relevantes (dieta), por lo que no estudiaron los efectos en este carácter.

Los efectos de dieta, experimento y raza sobre las 4767 abundancias microbianas (clr-MT y alr-MG) se evaluaron de manera multivariante con un análisis PERMANOVA. La dieta explicó un 9.2% de la varianza total del microbioma (p-valor: 0.001) y el experimento explicó un 5.1% de la varianza (p-valor: 0.001); pero no se pudo detectar un efecto de raza para un nivel de confianza alfa del 5%. El ajuste de los algoritmos se realizó con los fenotipos de CH_4 ajustados por el efecto dieta, y todas las variables microbianas ajustadas por los efectos de dieta y experimento. Las correcciones se relaizaron usando un modelo lineal univariante para cada variable y guardando los residuos. La Figura 9 muestra un PCA del microbioma antes y después de las correcciones.

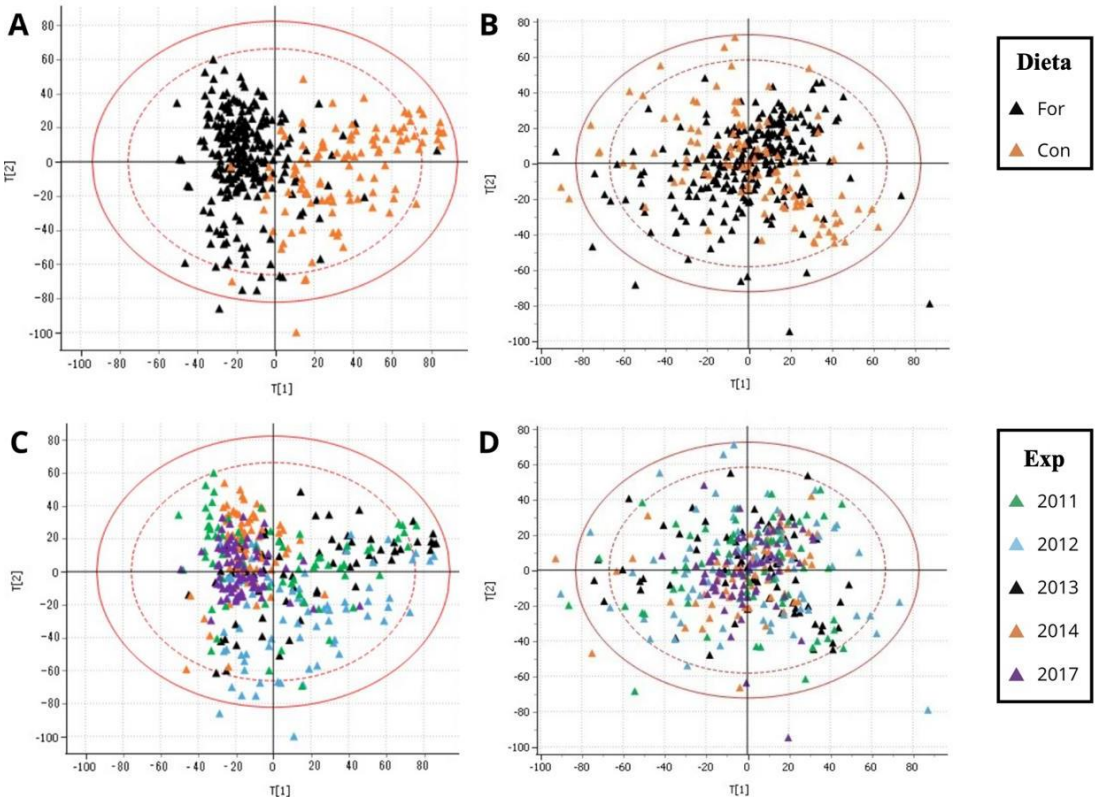


Figura 9. *Scoreplots* de los componentes 1 y 2 resultantes de un PCA del microbioma pintando los individuos por dieta y experiemento (Exp), antes y después de ajustar por dichos efectos. Coloreado por dietas antes (A) y después (B) de corregir;) Coloreado por el experimento antes (C) y después (D) de corregir.

4.2. Algoritmos de predicción de fenotipos de CH_4

La Tabla 5 muestra los resultados del ajuste de los algoritmos de predicción para los tres fenotipos de CH_4 usando 4767 variables microbianas. En todos los casos, los modelos de RF y XGB mostraron un ajuste (R^2) muy alto con los datos de entrenamiento, en un rango de 0.87-0.96, excepto en el caso de XGB en MY cuyo R^2 fue de 0.39±0.07 (media ± sd). Sin embargo, estos modelos sufrieron un claro sobreajuste a pesar de los esfuerzos de optimización por validación cruzada y generalizaron mal al ser usados para predecir datos externos, (Q^2 de entre 0.02-0.05). El algoritmo PLS presentó un comportamiento diferente y no sufrió sobreajuste. En general, su ajuste con datos de entrenamiento fue bajo (R^2 de entre 0.08-0.21). Sin embargo, a la hora de predecir datos externos, fue el que mejores resultados obtuvo, con una Q^2 de 0.086 ± 0.048 en RM.

Para tratar de evitar el sobreajuste y mejorar las predicciones, se probaron dos alternativas de reducción de la dimensionalidad de las variables predictoras.

La primera, fue realizar un PCA del microbioma y usar las proyecciones como variables predictoras de los algoritmos. Para eso se realizó un PCA con las 4767 variables y se conservaron todas las componentes para utilizar como predictoras. Esto es, 70% de 280 animales = 196 componentes, ver Material Suplementario 2. La decisión de conservar todas las componentes, aunque expliquen un bajo porcentaje de variabilidad se fundamenta en que componentes que expliquen poca variabilidad en la base de datos, pueden estar muy correlacionadas con la respuesta. Además, todas estas componentes son ortogonales, así que perder algunas variables podría significar perder toda la información predictiva de los datos.

El uso de las proyecciones de todas las componentes como variables predictoras en XGB y RF (PCA-XGB y PCA-RF) no lograron reducir el sobreajuste, ni si quiera con hiperparámetros muy restrictivos en la complejidad de los modelos. El ajuste en los datos de entrenamiento R^2 fue de entre 0.90-0.96, y su capacidad de predicción en datos externos Q^2 de entre 0.009-0.022 (lo máximo conseguido fue una Q^2 de 0.022 ±0.020 para PCA-RF), cero a efectos prácticos.

Tabla 5: Ajuste de predicción de los algoritmos en los datos de entrenamiento (R^2) y predicción (Q^2). Valores expresados como media ± sd.

		PLS[3]		**RF[4]**		**XGB[5]**	
Carácter[1]	**Estrategia[2]**	**R^2**	**Q^2**	**R^2**	**Q^2**	**R^2**	**Q^2**
MP, g CH_4/día	Completa	0.21±0.20	0.06±0.04	0.96±0.01	0.05±0.04	0.90±0.02	0.05±0.04
	PCA	-	-	0.95±0.01	0.02±0.02	0.95±0.01	0.01±0.02
	SV	0.27±0.04	0.18±0.08	0.95±0.00	0.17±0.08	0.87±0.02	0.15±0.07
MY, g CH_4/kg DMI	Completa	0.08±0.01	0.06±0.04	0.87±0.02	0.03±0.03	0.39±0.07	0.02±0.02
	PCA	-	-	0.90±0.02	0.01±0.01	0.87±0.01	0.01±0.02
	SV	0.15±0.27	0.11±0.05	0.94±0.01	0.15±0.06	0.88±0.01	0.14±0.05
RM, g CH_4/día	Completa	0.21±0.20	0.09±0.05	0.93±0.00	0.04±0.04	0.89±0.02	0.04±0.04
	PCA	-	-	0.95±0.01	0.01±0.01	0.93±0.01	0.01±0.02
	SV	0.24±0.11	0.13±0.06	0.91±0.01	0.13±0.07	0.86±0.02	0.15±0.07

[1] MP= Producción de metano, MY= rendimiento de metano, RM= metano residual

[2]Los modelos se ajustaron utilizando como variables predictoras las 4767 variables microbianas (Completa), las proyecciones de los animales en 196 componentes principales obtenidas del análisis de las 4767 variables (PCA) o un número reducido de variables microbianas seleccionadas: MP: 280, MY: 211 y RM:210.

[3]Regresión de Mínimos Cuadrados Parciales

[4]*Random Forest*

[5]*eXtreme Gradient Boosting*

La segunda alternativa para reducir la dimensión de las variables predictoras fue la selección de variables. Para ello, se decidió seleccionar las variables que coincidían como seleccionadas en los 3 algoritmos PLS, RF y XGB (Figura 10). La selección de variables en PLS se realizó siguiendo el criterio de VIP>0.8 y coeficiente de regresión que no incluyan el cero en su intervalo de confianza Jackknife. La selección de variables en RF se realizó utilizando la importancia por permutación (%IncMSE); y en XGB se seleccionó por el criterio de *Gain* (ver material y métodos).

Las variables seleccionadas por los 3 algoritmos para cada fenotipo de CH_4 se muestran en la Figura 10. Al seleccionar variables con distintos algoritmos, se puede observar que XGB fue el algoritmo más estricto (988 en MP, 946 en MY y 776 RM) y RF el más laxo (2489 en MP, 2449 en MY y 2477 RM). En total, se seleccionaron 230 variables para MP (4.82% del total), 210 RM (4.41%) y 211 para MY (4.43%) como predictoras en los algoritmos. El listado de variables seleccionadas puede consultarse en el material suplementario (Material Suplementario 3). Aunque las variables se seleccionaron acorde a su importancia en los algoritmos de predicción, esta base de datos con variables seleccionadas se utilizó también en los algoritmos de clasificación en 5 categorías y de individuos extremos que se mostrarán más adelante.

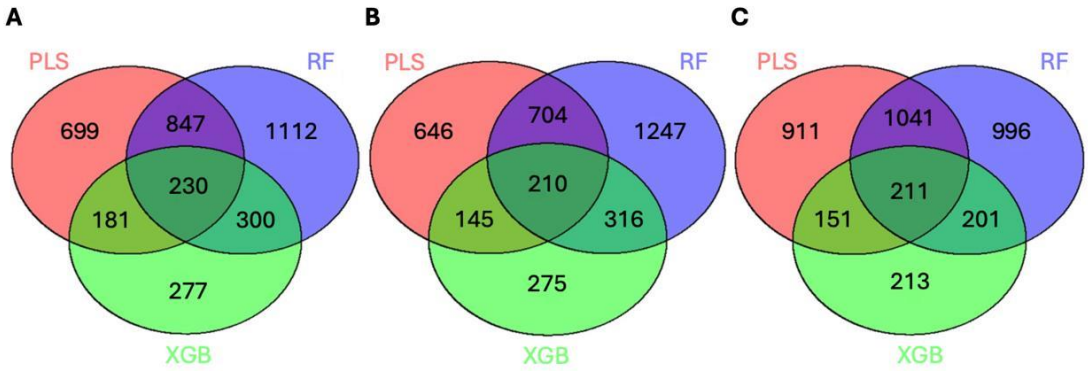


Figura 10. Diagramas de Venn para las variables seleccionadas por cada algoritmo para cada fenotipo de emisión de metano: A)MP; B)RM; C)MY. En verde son las seleccionadas por XGB, en rojo por PLS y en azul por RF.

La selección de variables mejoró el rendimiento de los algoritmos al enfrentarse a bases de datos externas. Los ajustes a los datos de entrenamiento siguen siendo altos para RF y XGB con R^2 que oscilan entre 0.85-0.95 (Tabla 5). Por otro lado, las Q^2 fueron de entre 0.10-0.18. Si bien estos resultados de predicción aun distan mucho de un modelo predictivo aplicable, la selección de variables incrementó la Q^2 en un 208% para PLS, 371% para XGB y 428% para RF, en promedio, con respecto a utilizar todas las variables.

4.3. Algoritmos de clasificación en 5 categorías (*scores 1-5*)

Para ver si el microbioma era capaz de clasificar a los animales en un *score* de 5 categorías de emisiones de CH_4 se decidió dividir a los animales de forma equilibrada en 1) Bajos emisores, 2) Emisores bajos-medios, 3) Emisores medios, 4) Emisores medio-altos y 5) Altos emisores. Esto se hizo con la intención de dar un *score* a los animales según sus emisiones de CH_4.

En la Tabla 6 se pueden ver los resultados del ajuste de los algoritmos de clasificación usando las 4767 variables microbianas.

Tabla 6. Ajuste de los algoritmos de clasificación en emisiones de CH_4 extremas en 5 categorías (1- Bajos emisores, 2- Emisores bajos-medios, 3-Emisores medios, 4-Emisores medio-altos y 5-Altos emisores) en los datos de entrenamiento (Área bajo de curva, AUC_e) y en los datos de validación (AUC_V). Valores expresados como media ± sd.

		PLS-DA[3]		RF[4]		XGB[5]	
Carácter[1]	**Estrategia[2]**	**AUC_e**	**AUC_V**	**AUC_e**	**AUC_V**	**AUC_e**	**AUC_V**
MP, g CH_4/día	Completa	0.66±0.02	0.61±0.04	0.82±0.02	0.57±0.04	0.98±0.02	0.57±0.05
	PCA	-	-	0.66±0.03	0.53±0.04	0.89±0.05	0.56±0.06
	SV	0.67±0.02	0.59±0.04	0.69±0.01	0.59±0.04	0.90±0.03	0.60±0.04
MY, g CH_4/kg DMI	Completa	0.66±0.01	0.60±0.03	0.85±0.02	0.55±0.04	0.74±0.06	0.55±0.04
	PCA	-	-	0.65±0.04	0.51±0.04	0.88±0.05	0.51±0.05
	SV	0.70±0.03	0.61±0.04	0.81±0.02	0.57±0.04	0.90±0.03	0.60±0.04
RM, g CH_4/día	Completa	0.71±0.04	0.68±0.07	0.82±0.02	0.56±0.04	0.77±0.04	0.55±0.04
	PCA	-	-	0.83±0.08	0.53±0.04	0.93±0.02	0.52±0.06
	SV	0.66±0.04	0.58±0.05	0.85±0.02	0.58±0.04	0.60±0.06	0.56±0.06

[1] MP= Producción de metano, MY= rendimiento de metano, RM= metano residual

[2]Los modelos se ajustaron utilizando como variables predictoras las 4767 variables microbianas (Completa), las proyecciones de los animales en 196 componentes principales obtenidas del análisis de las 4767 variables (PCA) o un número reducido de variables microbianas seleccionadas: MP: 280, MY: 211 y RM:210.

[3]Regresión de Mínimos Cuadrados Parciales discriminante
[4]*Random Forest*
[5]*eXtreme Gradient Boosting*

Los resultados no fueron tan sobreajustados como en el caso de la predicción continua. Usando todas las variables, se obtuvieron valores de AUC_e de entre 0.66-0.85, a excepción del XGB en MP que tuvo un AUC_e de 0.98±0.02. En clasificación de muestras externas se obtuvieron AUC_V que oscilaban entre 0.55-0.68, llegando a alcanzar una media de AUC_V de 0.68 ± 0.03 con PLS en MY (Tabla 6).

Al utilizar la estrategia de condensar las 4767 variables de la base de datos por PCA, los resultados de ajuste mostraron AUC_e de entre 0.66-0.82 para PCA+RF y 0.88-0.93 para PCA+XGB. Sin embargo, las clasificaciones de bases de datos externas para ambas combinaciones de algoritmos no distaron mucho del azar, con AUC_V de entre 0.51-0.56 consiguiendo el máximo resultado con PCA-XGB en MP con una AUC_V de 0.56±0.06 (Tabla 6).

En este caso, la selección de variables no mejoró la capacidad de generalización de los modelos respecto la estrategia completa en la mayoría de los casos, incluso en otros hubo un leve detrimento. Obteniendo resultados de ajuste de AUCe entre 0.60-0.89 con los datos de entrenamiento y AUCv entre 0.55-0.64 con los datos de validación, la máxima media de AUCv fue en MY con PLS (Tabla 6).

4.4. Algoritmo de clasificación de animales extremos

Finalmente, testamos la capacidad del microbioma para clasificar animales con fenotipos de emisiones de CH_4 extremas, es decir aquellos animales que pertenecían al 25% de menos emisiones (Q_1, Bajos) y al 25% de más emisiones (Q_4, Altos), para cada fenotipo.

Los resultados con la base de datos completa, mostraron una precisión de ajuste AUCe más baja para PLS-DA (0.63-0.77) que para RF (0.85-0.89) y XGB (0.95-0.96), en todos los fenotips. Sin embargo, a la hora de clasificar el set de datos externos, las AUCv de los 3 algoritmos fueron similares, con valores de entre 0.55-0.63, Los resultados con PCA, no mejoraron las AUC_e ni en los datos de entrenamiento (0.58-0.59 en PCA-RF y 0.88-0.93 en PCA-RF) ni en validación externa (AUCv =0.51-0.58) en ninguno de los dos algoritmos.

Por último, con selección de variables, los algoritmos aumentaron su capacidad de generalización. Se obtuvieron AUCe de ajuste entre 0.83-0.98. Mientras que la clasificación de extremos en validación externa todos los algoritmos lograron superar el AUCv 0.60 (umbral para considerarlos modelos aceptables). Los valores de AUCv estuvieron entre 0.65-0.75, en muestras externas. El máximo valor promedio de AUCv se consiguió por PLS clasificando el fenotipo MY con un AUCv de 0.75± 0.06.

Tabla 7. Ajuste de los algoritmos de clasificación en emisiones de CH_4 extremas (Alto y Bajos emisores) en los datos de entrenamiento (Area bajo de curva, AUC_e) y en los datos de validación (AUC_d). Valores expresados como media ± sd.

		PLS-DA[3]		**RF[4]**		**XGB[5]**	
Carácter[1]	**Estrategia[2]**	**AUC_e**	**AUC_V**	**AUC_e**	**AUC_V**	**AUC_e**	**AUC_V**
MP, g CH_4/día	Completa	0.63±0.04	0.61±0.07	0.88±0.02	0.62±0.07	0.95±0.02	0.63±0.07
	PCA	-	-	0.60±0.06	0.58±0.09	0.89±0.05	0.56±0.06
	SV	0.88±0.02	0.75±0.04	0.88±0.02	0.71±0.07	0.93±0.02	0.68±0.06
MY, g CH_4/kg	Completa	0.67±0.03	0.56±0.06	0.85±0.03	0.58±0.07	0.96±0.02	0.58±0.07
	PCA	-	-	0.58±0.05	0.53±0.07	0.93±0.02	0.52±0.06

DMI	SV	0.85±0.02	0.67±0.07	0.84±0.03	0.65±0.07	0.98±0.01	0.65±0.07
RM, g CH_4/día	Completa	0.77±0.02	0.55±0.06	0.89±0.02	0.60±0.07	0.96±0.02	0.60±0.07
	PCA	-	-	0.58±0.04	0.55±0.07	0.88±0.05	0.51±0.05
	SV	0.83±0.03	0.65±0.07	0.91±0.02	0.70±0.07	0.98±0.01	0.68±0.07

[1] MP= Producción de metano, MY= rendimiento de metano, RM= metano residual

[2]Los modelos se ajustaron utilizando como variables predictoras las 4767 variables microbianas (Completa), las proyecciones de los animales en 196 componentes principales obtenidas del análisis de las 4767 variables (PCA) o un número reducido de variables microbianas seleccionadas: MP: 280, MY: 211 y RM:210.

[3]Regresión de Mínimos Cuadrados Parciales discriminante

[4]*Random Forest*

[5]*eXtreme Gradient Boosting*

4.5. Predicción de MP con las ecuaciones Tier 2

Las predicciones de MP utilizando las ecuaciones Tier 2 y el DMI de los animales de este estudio presentaron una R^2 de 0.29 con respecto a los datos de medición de MP en cámara (Figura 11).

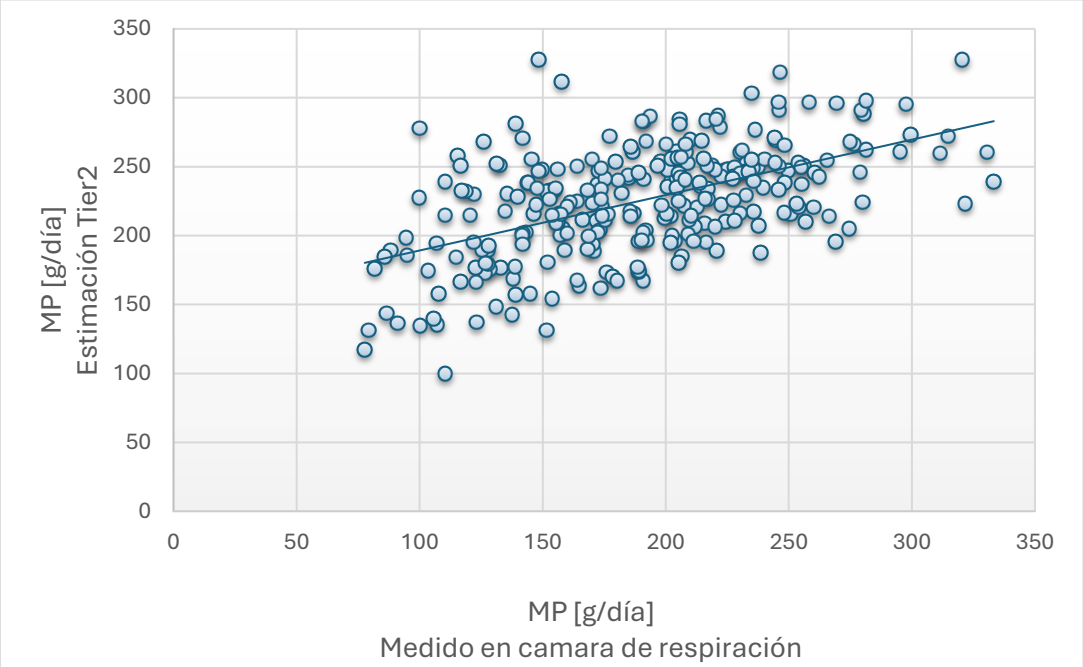


Figura 11. Correlación entre las estimaciones Tier2 y los valores reales de producción de metano medido en cámara de respiración de los animales utilizados en este trabajo.

4.6. Análisis de variables seleccionadas

De las 230 variables seleccionadas para MP, 150 fueron genes y 80 géneros; para las 211 de MY fueron 135 y 76 y para las 210 de RM 129 y 81. Solo hubo 4 genes/géneros que estuvieron presentes en los tres fenotipos: *Candidatus methanoperedens*, *Methanolobus*, *Pirellula* y el gen K03723. Los géneros microbianos que fueron elegidos pertenecían a 26 Phylums diferentes. Entre los que más géneros tenían fueron *Euryarchaeota* y *Proteobacteria*, seguido por *Firmicutes*, *Ascomycota*, *Actinobacteria* y *Ascomycota* y otros Phylums minoritarios. En cuanto a los genes, se encontraron 373 genes diferentes que pertenecían a rutas metabólicas: reparación del ADN, metabolismo energético, metabolismo de aminoácidos, etc. Además de gran cantidad de genes relacionados al metabolismo del metano: *mtmB, mcrA, mcrB, mtrC y mtaC*.

5. DISCUSIÓN

5.1. La microbiota como predictor de las emisiones de metano

Este trabajo muestra que la composición microbiana tomada al sacrificio es un predictor bastante pobre de las emisiones de CH_4, tanto si se expresan como MP, MY o RM. Esto podrían deberse a varios motivos. Por un lado, se podría pensar que la microbiota en un punto (tomada al sacrificio en este caso) no es una buena predictora del CH_4 debido a que el microbioma no es estático, sino que puede variar a lo largo de la vida del animal, y son sus variaciones lo que se puede asociar a las emisiones de CH_4 (Li et al., 2019). Sin embargo, estudios han mostrado que existe una cierta estabilidad temporal en la capacidad predictiva del microbioma sobre MY, en un subgrupo las vacas utilizadas en este trabajo y 7 momentos de medición del microbioma que incluyen la entrada a las cámaras de respiración y al sacrificio (Lima et al., 2022).

Otra limitación podría consistir en que, al hacer metagenómica del fluido ruminal se obtienen los genes y géneros de todo lo que estaba presente en ese fluido, sin importar si el microorganismo estaba viable o no en ese momento (Dungan et al., 2023). Es por esto que los datos metagenómicos por sí solos suelen ser insuficientes para determinar la actividad microbiana o predecir el fenotipo del huésped (New & Brito, 2020). Esto es un problema, ya que puede llevarnos a sacar conclusiones erróneas de asociación de abundancias de microorganismos con fenotipos, o no capturar las asociaciones reales entre el metabolismo microbiano y la síntesis de CH_4. Sin embargo, existen alternativas a la metagenómica que permiten salvar este problema. Una es la prueba de Viabilidad Molecular (MVT), que correlaciona la viabilidad con la capacidad de sintetizar un precursor de rRNA específico de la especie que se detecta mediante la medición de PCR cuantitativa con transcriptasa inversa (RT-qPCR) (Cangelosi & Meschke, 2014). Otra alternativa para detectar estos microorganismos viables y además registrar los genes transcripcionalmente activos de una comunidad microbiana es realizar la metatranscriptómica del rumen (Aguiar-Pulido et al., 2016), ya que solo un microorganismo viable es capaz de transcribir su material genético. Este tipo de análisis ha sido probado en ganado vacuno para discernir si existía un cambio en la transcriptómica del rumen entre grupos de animales divergentes para RFI, encontrando ciertos microrganismos y genes diferencialmente distribuidos en ambos grupos (F. Li & Guan, 2017). Otras ómicas como la proteómica y la metabolómica del contenido ruminal tambien podrían darnos información útil de los microorganismos que realmente están viables en el rumen. La proteómica tiene la limitacion de que para la identificación y asignación taxonómica se tiene una escasa cobertura de las proteínas en los conjuntos de referencia, lo que hace muy difícil su utilización (Denman et al., 2018). Además, tanto para metabolómica como para proteómica, se necesitan técnicas para poder separar las proteínas y metabolitos de los microorganismos de los de la dieta. Además, las muestras de rumen pueden contener compuestos polifenólicos, como taninos y ácidos húmicos, que pueden interferir con el análisis mediante modificaciones que dificultan la identificación de los péptidos (Makkar et al., 1995; Snelling & Wallace, 2017).

Otra posible causa de las limitaciones predictivas del microbioma puede ser que la heterogeneidad de nuestra base de datos (4 razas de animales diferentes, alimentados con dos dietas), no haya podido ser corregida de manera satisfactoria con un modelo lineal. Si bien se utilizaron métodos de corrección para estos efectos, se corrige por la estima de un efecto general, como si estos efectos afectaran a todos los individuos por igual (Blasco, 2021). Por ejemplo, es posible que, bajo diferentes dietas, los microbios involucrados en la metanogénesis sean distintos (Miller et al., 2023). O, dado el efecto genético del hospedador sobre el microbioma (John Wallace et al., 2019), diferentes razas utilicen diferentes rutas metanogénicas para producir CH_4 ruminal.

Para testar la hipótesis de la heterogeneidad entre dietas (fue el efecto más relevante), se realizó una pequeña prueba de predicción de MP dentro de dieta (ya que fue el efecto más relevante), y se comparó con los resultados obtenidos con la base de datos corregida. Para evitar un efecto de tamaño muestral, en los tres casos se utilizaron bases de datos con 90 individuos, muestreados al azar (para la base de datos corregida, se tomaron al azar 45 animales de cada dieta). Los resultados fueron similares en todos los casos, con Q^2=0.04±0.03, 0.03±0.02 y 0.04±0.04 para XGB, 0.06±0.06, 0.11±0.08 y 0.08±0.06 para PLS y 0.2±0.03, 0.05±0.05 y 0.03±0.04 para forraje, concentrado y corregido respectivamente. Esto podría indicar, que el problema de la baja precisión de predicción no viene por la heterogeneidad de los datos debido a la dieta, o si tiene, no tenemos potencia muestral para detectarlo.

Actualmente, los modelos recomendados por el IPCC para estimar las emisiones de CH_4 del ganado se basan en las ecuaciones Tier 2. Estas ecuaciones permiten realizar estimaciones de DMI y el EF, los cuales son utilizados para predecir la MP de los animales. En nuestro estudio, se utilizó el DMI derivado de datos observacionales y el EF obtenido de tablas del IPCC (Dong et al., 2006). Las predicciones resultantes comparadas con los valores verdaderos de MP mostraron un R^2 de 0.29 (Figura 11). La máxima predicción continua de MP lograda basada en el microbioma fue de Q^2 de 0.15, lo que representa un 48% menos en comparación con las predicciones basadas en el modelo Tier 2. Si bien estos resultados no son alentadores para la microbiota, no debemos descartar su uso combinada con otro tipo de información, ya que podría describir fuentes de variación diferentes en las emisiones de CH_4 que otros métodos no son capaces de capturar.

A pesar de las malas predicciones que se obtuvieron de manera continua, fue posible generar una clasificación en 5 categorías de emisiones con un AUC_V máximo de 0.64±0.04, lo que es un modelo aceptable para un *score* con 5 categorías diferentes (Yang & Berdine, 2017). Además, la clasificación de animales extremos mostró una AUC_V de 0.75 ±0.06, un modelo considerado bueno (Yang & Berdine, 2017). En caso de que los datos de secuenciación de microbioma se obtuvieran a bajo coste (en este estudio, el precio ronda en torno a unos 70 €/muestra); este *score* (o clasificación en extremos) podría ser utilizable para escoger de alguna manera qué animales evaluar en cámaras, en caso de no poderlos medir todos, para así optimizar la utilidad de los datos de CH_4 medidos en un programa de mejora. La mejora genética por este carácter no solo podría tener beneficios por la disminución del CH_4, sino que también se ha visto en otras especies que puede traer consigo mejora de otros caracteres productivos (Rowe et al., 2019).

5.2. Modelos de *machine learning* vs. algoritmos lineales

A través de los algoritmos de ML se esperaba poder capturar relaciones no lineales que explicaran mejor la relación del microbioma con el CH_4 con respecto a modelos que no pueden capturar estas relaciones. Sin embargo, el algoritmo PLS mostró mejores resultados que XGB y RF en validación externa, con Q^2 de 0.18. Otros estudios con métodos lineales como el M-BLUP han conseguido R^2 0.21 en validación externa (Ross et al., 2013). Esto podría estar indicando que estos algoritmos no son capaces de capturar las relaciones que presenta la microbiota con el CH_4; bien porque el número de observaciones utilizada sea insuficientes, o porque en la práctica, estas asociaciones no lineales no existan. Quizás aumentando el número de animales, se podrían mejorar los resultados.

Por otro lado, al analizar los resultados de predicción/clasificación con las bases de datos completa, condensada por PCA y con selección de variables, se pudo ver que el uso de PCA como un compresor de la información no parece ser una estrategia adecuada para la predicción de CH_4 utilizando datos de microbioma, ya que sus precisiones siempre son las peores de las 3

estrategias. Aunque PCA puede reducir la dimensionalidad y simplificar el modelo (Gupta et al., 2022), no necesariamente mejora la capacidad de predicción de nuevos datos. La pérdida de información relevante durante la reducción de dimensionalidad podría ser la causa de la disminución capacidad predictiva observada.

La selección de variables en cambio mostró una mejora en la mayoría de los resultados tanto de predicción como de extremos. Las razones por las cuales la selección de variables podría estar mejorando las predicciones son: (1) la reducción del sobreajuste, ya que al tener menos variables se disminuye la complejidad del modelo, mejorando su capacidad de generalización; (2) la disminución del ruido, ya que se eliminan variables irrelevantes para la predicción/clasificación que se está llevando a cabo. Estos resultados van acorde a otras investigaciones con microbiota de otras especies en donde la selección de variables ha mejorado considerablemente las precisiones de las estimaciones (Flemer et al., 2018; Yachida et al., 2019).

5.3. Interpretación de las variables seleccionadas

En la estrategia de selección de variables, se seleccionaron 230 variables para MP, 210 RM y 211 para MY (Material suplementario 3). Se determinó cuáles de estas variables eran comunes a los 3 fenotipos, encontrando 4 genes/géneros comunes (Material suplementario 4). Esto se hizo porque se tenía la hipótesis de que si una variable era seleccionada por los 3 algoritmos en los tres fenotipos podría estar indicando implicaciones biológicas comunes, se exprese como se exprese la emisión de CH_4. Los 4 generos/genes que todos los algoritmos eligieron en todos los fenotipos fueron: *Candidatus methanoperedens, Methanolobus, Pirellula* y el gen *K03723* que codifica para un factor de acoplamiento de reparación de transcripción (*mfd*).

Tanto *Candidatus methanoperedens* como *Methanolobus* son arqueas metanógenas, pero siguen vías diferentes de generación del CH_4. *Candidatus methanoperedens* es un género de arqueas metanotróficas anaerobias que oxidan CH_4 en presencia de nitrato o nitrito (McIlroy et al., 2023). Mientras que *Methanolobus,* es un género de arqueas metanógenas que utilizan compuestos metilados como metanol y metilaminas para producir CH_4 (G. Zhang et al., 2008). Ligado a esta arquea metilotrofica, se identificaron entre las variables seleccionadas varios genes de metil-transferasas tales como: *mtmB, mcrA, mcrB, mtrC y mtaC* (Tsola et al., 2024). Cabe destacar que, entre estos genes, 2 de ellos (*mcrA, mcrB*) son subunidades de la misma proteína (metil-coenzima M reductasa) (Hallam et al., 2003). Esta es una enzima clave en la metanogénesis metilotrofica ya que cataliza el paso final en la producción de CH_4 en arqueas metanogénicas, transformando metil-coenzima M en CH_4. Además, ha descripto que la presencia y abundancia de este gen en el microbiota ruminal es un indicador directo del potencial metanogénico del microbioma (Aryee et al., 2023).

Por otro lado, el gen *mfd* es un factor de acoplamiento de reparación de transcripción que, si bien no está directamente relacionado a la producción de CH_4, podría estarlo de manera indirecta. Algunas hipótesis que tenemos sobre este gen regulador es que podría estar implicado en vías metabólicas relacionadas no descritas previamente en la literatura. O incluso que tenga funciones no caracterizadas que podrían estar relacionadas con la producción de CH_4 (Scott & Oeffinger, 2016).

Por último, *Pirellula,* es un género de bacterias muy abundante en el rumen (0.4±0.2% en nuestros datos) perteneciente a la familia *Planctomycetes* (Daugaliyeva et al., 2022). Tampoco pudimos encontrar literatura que la relacione directamente con el CH_4, por lo que es difícil determinar porque es tan relevante para nuestro análisis.

Por otro lado, entre los géneros microbianos que se seleccionaron por los algoritmos en los distintos fenotipos, destacan 55 arqueas pertenencientes al phylo *Euryarchaeota* (donde se incluyen las arqueas metanogénicas). Algunas de estas son: *Methanocaldococcus, Methanomassiliicoccus, Methanomicrobium, Methanolobus, Methanospirillum, Methermicoccus, Methanotorris, Methanosalsum.* Varias de ellas relacionadas en bibliografía a las emisiones de CH_4 en vacuno (Aryee et al., 2023; Rooke et al., 2014; Tapio et al., 2017). Además, 150 bacterias pertenencientes al Phylo *Proteobacteria*. Las mismas han sido relacionadas en literatura con animales que emiten gran cantidad de CH_4 (Wallace et al., 2015). Además, las proteobacterias producen acetato en el rumen, una vía de aprovechamiento de los cofactores reducidos alternativa a la metanogénesis (Kersters et al., 2006). Por lo que es lógico que estos grupos de bacterias sean relevantes para nuestro análisis.

6. CONCLUSIÓN

- La composición microbiana tomada al sacrificio es un predictor bastante pobre de las emisiones de CH_4, tanto si se expresan como MP, MY o RM, obteniendo un Q^2 promedio máximo de: 0,18±0,08. Sin embargo, podría usarse para clasificar a los animales en función de sus emisiones de CH_4 en 5 esperando precisiones de AUC_V de 0.68 ± 0.03, o clasificar animales extremos con una AUC_V de 0.75 ±0.06. Esto podría ser útil como un primer filtro para discernir que animales vale la pena medir en cámaras de respiración en un programa de mejora de disminución de CH_4.
- Tanto en predicciones continuas como categóricas y extremos, el algoritmo PLS tuvo un mejor rendimiento que los algoritmos de *machine learning*, XGB y RF, que resultaron altamente sensibles al sobreajuste y mostraron una capacidad para generalizar pobre.
- La selección de variables mejora la capacidad predictiva tanto en predicción como clasificación de animales extremos, mientras que la base de datos condensada por PCA tiende a perder información relevante, reduciendo la capacidad de predicción.
- Para todos los fenotipos de CH_4, los modelos seleccionaron como relevantes arqueas metanógenas como *Candidatus methanoperedens* y *Methanolobus* y genes como *mtmB, mcrA, mcrB, mtrC y mtaC*.

7. REFERENCIAS

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., & Narasimhan, G. (2016). Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. *Evolutionary Bioinformatics*, *12s1*, EBO.S36436. https://doi.org/10.4137/EBO.S36436

Aguinaga Casañas, M. A., Rangkasenee, N., Krattenmacher, N., Thaller, G., Metges, C. C., & Kuhla, B. (2015). Methyl-coenzyme M reductase A as an indicator to estimate methane production from dairy cows. *Journal of Dairy Science*, *98*(6), 4074–4083. https://doi.org/10.3168/jds.2015-9310

Andy Liaw, & Matthew Wiener. (2002). *Classification and Regression by randomForest*.

Arndt, C., Hristov, A. N., Price, W. J., McClelland, S. C., Pelaez, A. M., Cueva, S. F., Oh, J., Dijkstra, J., Bannink, A., Bayat, A. R., Crompton, L. A., Eugène, M. A., Enahoro, D., Kebreab, E., Kreuzer, M., McGee, M., Martin, C., Newbold, C. J., Reynolds, C. K., … Yu, Z. (2022). Full adoption of the most effective strategies to mitigate methane emissions by ruminants can help meet the 1.5 °C target by 2030 but not 2050. *Proceedings of the National Academy of Sciences*, *119*(20). https://doi.org/10.1073/pnas.2111294119

Arndt, C., Powell, J. M., Aguerre, M. J., & Wattiaux, M. A. (2015). Performance, digestion, nitrogen balance, and emission of manure ammonia, enteric methane, and carbon dioxide in lactating cows fed diets with varying alfalfa silage-to-corn silage ratios. *Journal of Dairy Science*, *98*(1), 418–430. https://doi.org/10.3168/jds.2014-8298

Aryee, G., Luecke, S. M., Dahlen, C. R., Swanson, K. C., & Amat, S. (2023). Holistic View and Novel Perspective on Ruminal and Extra-Gastrointestinal Methanogens in Cattle. In *Microorganisms* (Vol. 11, Issue 11). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/microorganisms11112746

Basarab, J. A., Beauchemin, K. A., Baron, V. S., Ominski, K. H., Guan, L. L., Miller, S. P., & Crowley, J. J. (2013). Reducing GHG emissions through genetic improvement for feed efficiency: effects on economically important traits and enteric methane production. *Animal*, *7*, 303–315. https://doi.org/10.1017/S1751731113000888

Beauchemin, K. A., McAllister, T. A., & McGinn, S. M. (2009). Dietary mitigation of enteric methane from cattle. *CABI Reviews*, 1–18. https://doi.org/10.1079/PAVSNNR20094035

Beauchemin, K. A., Ungerfeld, E. M., Eckard, R. J., & Wang, M. (2020). Review: Fifty years of research on rumen methanogenesis: Lessons learned and future challenges for mitigation. *Animal*, *14*(S1), S2–S16. https://doi.org/10.1017/S1751731119003100

Berndt, A., Boland, T., Deighton, M., Gere, J., Grainger, C., Hegarty, R., Iwaasa, A., Koolaard, J., & Lassey, K. (2014). Guidelines for use of sulphur hexafluoride (SF6) tracer technique to measure enteric methane emissions from ruminants. *New Zealand Agricultural Greenhouse Gas Research Centre, New Zealand*.

Bhagat, M., & Kumar, D. (2023). Performance evaluation of PCA based reduced features of leaf images extracted by DWT using random Forest and XGBoost classifier. *Multimedia Tools and Applications*, *82*(17), 26225–26254. https://doi.org/10.1007/s11042-023-14370-9

Bhatta, R., & Enishi, O. (2007). Measurement of Methane Production from Ruminants. *Asian-Australasian Journal of Animal Sciences*, *20*(8), 1305–1318. https://doi.org/10.5713/ajas.2007.1305

Blasco, A. (2021). *Mejora Genetica Animal*. EDITORIAL SINTESIS.

Boulesteix, A., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Mining and Knowledge Discovery*, *2*(6), 493–507. https://doi.org/10.1002/widm.1072

Braian Van Doormaal. (2023, January 16). *Canada is Global Leader for Delivering Methane Evaluations*. Https://Lactanet.ca/En/Canada-Leader-Methane-Evaluations/.

Cangelosi, G. A., & Meschke, J. S. (2014). Dead or Alive: Molecular Assessment of Microbial Viability. *Applied and Environmental Microbiology*, *80*(19), 5884–5891. https://doi.org/10.1128/AEM.01763-14

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Zhou, T. (2020). Extreme Gradient Boosting [R Package Xgboost Version 1.2. 0.1]. *In Proc. ACM SIGKDD Int. Conf. Knowl. Discov.*, 13–17.

Ciganda, V., Simón, C., & Mariotta, J. (2022). *Protocolo para Determinación de Emisión de Metano en Rumiantes: "Técnica de Trazador SF6 para Períodos de Medición Prolongados"* (E. L. G. P. T. del L. M. Uruguay. http://www. inia. uy Editado por la Unidad de Comunicación y Transferencia de Tecnología del INIA Avda. Italia 6201, Ed.; 2da Edicion). INIA, Uruguay.

de Haas, Y., Pszczola, M., Soyeurt, H., Wall, E., & Lassen, J. (2017). Invited review: Phenotypes to genetically reduce greenhouse gas emissions in dairying. *Journal of Dairy Science*, *100*(2), 855–870. https://doi.org/10.3168/jds.2016-11246

de Haas, Y., Windig, J. J., Calus, M. P. L., Dijkstra, J., de Haan, M., Bannink, A., & Veerkamp, R. F. (2011). Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. *Journal of Dairy Science*, *94*(12), 6122–6134. https://doi.org/10.3168/jds.2011-4439

Deighton, M. H., O'Loughlin, B. M., Williams, S. R. O., Moate, P. J., Kennedy, E., Boland, T. M., & Eckard, R. J. (2013). Declining sulphur hexafluoride permeability of polytetrafluoroethylene membranes causes overestimation of calculated ruminant methane emissions using the tracer technique. *Animal Feed Science and Technology*, *183*(3–4), 86–95. https://doi.org/10.1016/j.anifeedsci.2013.04.021

Denman, S. E., Morgavi, D. P., & McSweeney, C. S. (2018). Review: The application of omics to rumen microbiota function. *Animal*, *12*, s233–s245. https://doi.org/10.1017/S175173111800229X

Dong, H., Mangino, J., & McAllister, T. A. (2006). CAPITULO 10: EMISIONES RESULTANTES DE LA GESTIÓN DEL GANADO Y DEL ESTIÉRCOL . In *Directrices del IPCC de 2006 para los inventarios nacionales de gases de efecto invernadero* .

Dressler, E. A., Bormann, J. M., Weaber, R. L., & Rolf, M. M. (2024). Use of methane production data for genetic prediction in beef cattle: A review. *Translational Animal Science*, *8*. https://doi.org/10.1093/tas/txae014

Dungan, A. M., Geissler, L., Williams, A. S., Gotze, C. R., Flynn, E. C., Blackall, L. L., & van Oppen, M. J. H. (2023). DNA from non-viable bacteria biases diversity estimates in the corals Acropora loripes and Pocillopora acuta. *Environmental Microbiome*, *18*(1), 86. https://doi.org/10.1186/s40793-023-00541-6

Duthie, C. A., Haskell, M., Hyslop, J. J., Waterhouse, A., Wallace, R. J., Roehe, R., & Rooke, J. A. (2017). The impact of divergent breed types and diets on methane emissions, rumen characteristics and performance of finishing beef cattle. *Animal*, *11*(10), 1762–1771. https://doi.org/10.1017/S1751731117000301

Duthie, C. A., Rooke, J. A., Troy, S., Hyslop, J. J., Ross, D. W., Waterhouse, A., & Roehe, R. (2016). 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*, *10*(5), 786–795. https://doi.org/10.1017/S1751731115002657

Duthie, C. A., Troy, S. M., Hyslop, J. J., Ross, D. W., Roehe, R., & Rooke, J. A. (2018). 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*, *12*(2), 280–287. https://doi.org/10.1017/S175173111700146X

Ellis, J. L., Kebreab, E., Odongo, N. E., McBride, B. W., Okine, E. K., & France, J. (2007). Prediction of methane production from dairy and beef cattle. *Journal of Dairy Science*, *90*(7), 3456–3466. https://doi.org/10.3168/jds.2006-675

FAO. (2023). *Pathways towards lower emissions– A global assessment of the greenhouse gas emissions and mitigation options from livestock agrifood systems*. FAO. https://doi.org/10.4060/cc9029en

Ferrer, A. (2007). Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. *Quality Engineering*, *19*(4), 311–325. https://doi.org/10.1080/08982110701621304

Flemer, B., Warren, R. D., Barrett, M. P., Cisek, K., Das, A., Jeffery, I. B., Hurley, E., O'Riordain, M., Shanahan, F., & O'Toole, P. W. (2018). The oral microbiota in colorectal cancer is distinctive and predictive. *Gut*, *67*(8), 1454–1463. https://doi.org/10.1136/gutjnl-2017-314814

Food and Agriculture Organization of the United Nations (FAO). (2011). *World Livestock 2011 Livestock in food security*. https://www.fao.org/3/i2373e/i2373e.pdf

Garcia, J.-L., Patel, B. K. C., & Ollivier, B. (2000). Taxonomic, Phylogenetic, and Ecological Diversity of Methanogenic Archaea. *Anaerobe*, *6*(4), 205–226. https://doi.org/10.1006/anae.2000.0345

Global Monitoring Laboratory of the National Oceanic and Atmospheric Administration. (n.d.). *NOAA Methane (CH4) measurements*. Https://Gml.Noaa.Gov/Ccgg/Data/Ch4.Html. https://doi.org/https://doi.org/10.25925/20231001

González-Recio, O., López-Paredes, J., Ouatahar, L., Charfeddine, N., Ugarte, E., Alenda, R., & Jiménez-Montero, J. A. (2020). Mitigation of greenhouse gases in dairy cattle via genetic selection: 2. Incorporating methane emissions into the breeding goal. *Journal of Dairy Science*, *103*(8), 7210–7221. https://doi.org/10.3168/jds.2019-17598

Grainger, C., & Beauchemin, K. A. (2011). Can enteric methane emissions from ruminants be lowered without lowering their production? *Animal Feed Science and Technology*, *166–167*, 308–320. https://doi.org/10.1016/j.anifeedsci.2011.04.021

Grainger, C., Clarke, T., McGinn, S. M., Auldist, M. J., Beauchemin, K. A., Hannah, M. C., Waghorn, G. C., Clark, H., & Eckard, R. J. (2007). Methane emissions from dairy cows measured using the sulfur hexafluoride (SF6) tracer and chamber techniques. *Journal of Dairy Science*, *90*(6), 2755–2766. https://doi.org/10.3168/jds.2006-697

Greenacre, M. (2018). *Compositional Data Analysis in Practice*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429455537

Greenacre, M., Martínez-Álvaro, M., & Blasco, A. (2021). Compositional Data Analysis of Microbiome and Any-Omics Datasets: A Validation of the Additive Logratio Transformation. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.727398

Gupta, I., Sharma, V., Kaur, S., & Singh, A. K. (2022). *PCA-RF: An Efficient Parkinson's Disease Prediction Model based on Random Forest Classification*.

Guyon, I., & De, A. M. (2003). An Introduction to Variable and Feature Selection André Elisseeff. In *Journal of Machine Learning Research* (Vol. 3).

Hallam, S. J., Girguis, P. R., Preston, C. M., Richardson, P. M., & DeLong, E. F. (2003). Identification of Methyl Coenzyme M Reductase A (*mcrA*) Genes Associated with Methane-Oxidizing Archaea. *Applied and Environmental Microbiology*, *69*(9), 5483–5491. https://doi.org/10.1128/AEM.69.9.5483-5491.2003

Hammond, K. J., Crompton, L. A., Bannink, A., Dijkstra, J., Yáñez-Ruiz, D. R., O'Kiely, P., Kebreab, E., Eugène, M. A., Yu, Z., Shingfield, K. J., Schwarm, A., Hristov, A. N., & Reynolds, C. K. (2016). Review of current in vivo measurement techniques for quantifying enteric methane emission from ruminants. In *Animal Feed Science and Technology* (Vol. 219, pp. 13–30). Elsevier B.V. https://doi.org/10.1016/j.anifeedsci.2016.05.018

Hammond, K. J., Humphries, D. J., Crompton, L. A., Green, C., & Reynolds, C. K. (2015). Methane emissions from cattle: Estimates from short-term measurements using a GreenFeed system compared with measurements obtained using respiration chambers or sulphur hexafluoride tracer. *Animal Feed Science and Technology*, *203*, 41–52. https://doi.org/10.1016/j.anifeedsci.2015.02.008

Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-21606-5

Hatfield, J. L., Johnson, D. E., Lassey, K. R., Aparecida De Lima, M., Romanovskaya, A., & Bartram, D. (2006). Emisiones resultantes de la gestión del ganado y del estiércol. In *Directrices del IPCC de 2006 para los inventarios nacionales de gases de efecto invernadero* (2da ed., Vol. 4). IPCC.

Hector Menendez, Jameson Brennan, & Krista Ehlert. (2023). *Range Roundup: Precision Technology to Measure Cattle Methane Emissions and Intake on Western S.D. Rangelands*. https://extension.sdstate.edu/range-roundup-precision-technology-measure-cattle-methane-emissions-and-intake-western-sd

Hegarty, R. S., Goopy, J. P., Herd, R. M., & McCorkell, B. (2007). Cattle selected for lower residual feed intake have reduced daily methane production1,2. *Journal of Animal Science*, *85*(6), 1479–1486. https://doi.org/10.2527/jas.2006-236

Hellwing, A. L. F., Weisbjerg, M. R., & Møller, H. B. (2014). Enteric and manure-derived methane emissions and biogas yield of slurry from dairy cows fed grass silage or maize silage with and without supplementation of rapeseed. *Livestock Science*, *165*(1), 189–199. https://doi.org/10.1016/j.livsci.2014.04.011

Herd, R. M., Arthur, P. F., Donoghue, K. A., Bird, S. H., Bird-Gardiner, T., & Hegarty, R. S. (2014). Measures of methane production and their phenotypic relationships with dry matter intake, growth, and body composition traits in beef cattle 1,2. *J. Anim. Sci*, *92*, 5267–5274. https://doi.org/10.2527/jas2014-8273

Herd, R. M., Velazco, J. I., Arthur, P. F., & Hegarty, R. F. (2016). Associations among methane emission traits measured in the feedlot and in respiration chambers in Angus cattle bred to vary in feed efficiency. *Journal of Animal Science*, *94*(11), 4882–4891. https://doi.org/10.2527/jas.2016-0613

Hristov, A. N., Oh, J., Giallongo, F., Frederick, T. W., Harper, M. T., Weeks, H. L., Branco, A. F., Moate, P. J., Deighton, M. H., Williams, S. R. O., Kindermann, M., & Duval, S. (2015). An inhibitor persistently decreased enteric methane emission from dairy cows with no negative effect on milk production. *Proceedings of the National Academy of Sciences*, *112*(34), 10663–10668. https://doi.org/10.1073/pnas.1504124112

Hristov, A. N., Oh, J., Lee, C., Meinen, R., Montes, F., Ott, T., Firkins, J., Rotz, A., Dell, C., Adesogan, A., Yang, W., Tricarico, J., Kebreab, E., Waghorn, G., Dijkstra, J., & Oosting, S. (2013). *MITIGATION OF GREENHOUSE GAS EMISSIONS IN LIVESTOCK PRODUCTION A review of technical options for non-CO 2 emissions* (Pierre J. Gerber, Benjamin Henderson, & Harinder P.S. Makkar, Eds.; Vol. 177).

Huhtanen, P., Cabezas-Garcia, E. H., Utsumi, S., & Zimmerman, S. (2015). Comparison of methods to determine methane emissions from dairy cows in farm conditions. *Journal of Dairy Science*, *98*(5), 3394–3409. https://doi.org/10.3168/jds.2014-9118

Intergovernmental Panel on Climate Change (IPCC). (2023). Global Carbon and Other Biogeochemical Cycles and Feedbacks. In *Climate Change 2021 – The Physical Science Basis* (pp. 673–816). Cambridge University Press. https://doi.org/10.1017/9781009157896.007

IPCC. (2014). *Climate Change 2014: Synthesis Report.* . Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)].

Janssen, P. H. (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Animal Feed Science and Technology*, *160*(1–2), 1–22. https://doi.org/10.1016/j.anifeedsci.2010.07.002

Janssen, P. H., & Kirs, M. (2008). Structure of the Archaeal Community of the Rumen. *Applied and Environmental Microbiology*, *74*(12), 3619–3625. https://doi.org/10.1128/AEM.02812-07

John Wallace, R., Sasson, G., Garnsworthy, P. C., Tapio, I., Gregson, E., Bani, P., Huhtanen, P., Bayat, A. R., Strozzi, F., Biscarini, F., Snelling, T. J., Saunders, N., Potterton, S. L., Craigon, J., Minuti, A., Trevisi, E., Callegari, M. L., Piccioli Cappelli, F., Cabezas-Garcia, E. H., … Mizrahi, I. (2019). A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. In *Sci. Adv* (Vol. 5). https://www.science.org

Johnson, K. A., & Johnson', D. E. (1995). *Methane Emissions from Cattle*. https://academic.oup.com/jas/article-abstract/73/8/2483/4632901

Johnson, K. A., Westberg, H. H., Michal, J. J., & Cossalman, M. W. (2007). The SF6 Tracer Technique: Methane Measurement From Ruminants. In *Measuring Methane Production From Ruminants* (pp. 33–67). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6133-2_3

Johnson, K., Huyler, M., Westberg, H., Lamb, B., & Zimmerman, P. (1994). COMMUNICATIONS Measurement of Methane Emissions from Ruminant Livestock Using a SF6 Tracer Technique. In *Environ. Sci. Technol* (Vol. 28). https://pubs.acs.org/sharingguidelines

Kersters, K., Lisdiyanti, P., Komagata, K., & Swings, J. (2006). The Family Acetobacteraceae: The Genera Acetobacter, Acidomonas, Asaia, Gluconacetobacter, Gluconobacter, and Kozakia. In *The Prokaryotes* (pp. 163–200). Springer New York. https://doi.org/10.1007/0-387-30745-1_9

Knapp, J. R., Laur, G. L., Vadas, P. A., Weiss, W. P., & Tricarico, J. M. (2014). Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *Journal of Dairy Science*, *97*(6), 3231–3261. https://doi.org/10.3168/jds.2013-7234

Kucheryavskiy, S. (2020). mdatools – R package for chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *198*. https://doi.org/10.1016/j.chemolab.2020.103937

Lassey, K. R., Ulyatt, M. J., Martin, R. J., Walker, C. F., & David Shelton, I. (1997). Methane emissions measured directly from grazing livestock in New Zealand. *Atmospheric Environment*, *31*(18), 2905–2914. https://doi.org/10.1016/S1352-2310(97)00123-4

Lee, C., Araujo, R. C., Koenig, K. M., & Beauchemin, K. A. (2017). Effects of encapsulated nitrate on growth performance, nitrate toxicity, and enteric methane emissions in beef steers: Backgrounding phase1,2. *Journal of Animal Science*, *95*(8), 3700–3711. https://doi.org/10.2527/jas.2017.1460

Li, F., & Guan, L. L. (2017). Metatranscriptomic Profiling Reveals Linkages between the Active Rumen Microbiome and Feed Efficiency in Beef Cattle. *Applied and Environmental Microbiology*, *83*(9). https://doi.org/10.1128/AEM.00061-17

Li, Z., Deng, Q., Liu, Y., Yan, T., Li, F., Cao, Y., & Yao, J. (2018). Dynamics of methanogenesis, ruminal fermentation and fiber digestibility in ruminants following elimination of protozoa: a meta-analysis. *Journal of Animal Science and Biotechnology*, *9*(1), 89. https://doi.org/10.1186/s40104-018-0305-6

Lima, J., Martínez-Álvaro, M., Mattock, J., Auffret, M. D., Duthie, C. A., Cleveland, M. A., Dewhurst, R. J., Watson, M., & Roehe, R. (2022). 633. Host-genomically influenced ruminal microbial genes are temporally stable during the finishing phase in beef cattle. *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, 2616–2619. https://doi.org/10.3920/978-90-8686-940-4_633

Lopez, S., McIntosh, F. M., Wallace, R. J., & Newbold, C. J. (1999). Effect of adding acetogenic bacteria on methane production by mixed rumen microorganisms. *Animal Feed Science and Technology*, *78*(1–2), 1–9. https://doi.org/10.1016/S0377-8401(98)00273-9

Makkar, H. P. S., Blümmel, M., & Becker, K. (1995). Formation of complexes between polyvinyl pyrrolidones or polyethylene glycols and tannins, and their implication in gas production and true digestibility in *in vitro* techniques. *British Journal of Nutrition*, *73*(6), 897–913. https://doi.org/10.1079/BJN19950095

Manzanilla-Pech, C. I. V., Løvendahl, P., Mansan Gordo, D., Difford, G. F., Pryce, J. E., Schenkel, F., Wegmann, S., Miglior, F., Chud, T. C., Moate, P. J., Williams, S. R. O., Richardson, C. M., Stothard, P., & Lassen, J. (2021). Breeding for reduced methane emission and feed-efficient Holstein cows: An international response. *Journal of Dairy Science*, *104*(8), 8983–9001. https://doi.org/10.3168/jds.2020-19889

Manzanilla-Pech, C. I. V., Stephansen, R. B., Difford, G. F., Løvendahl, P., & Lassen, J. (2022). Selecting for Feed Efficient Cows Will Help to Reduce Methane Gas Emissions. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.885932

Manzanilla-Pech, C. I. V, De Haas, Y., Hayes, B. J., Veerkamp, R. F., Khansefid, † M, Donoghue, K. A., Arthur, P. F., & Pryce, J. E. (2016). Genomewide association study of methane emissions in Angus beef cattle with validation in dairy cattle 1. *J. Anim. Sci*, *94*, 4151–4166. https://doi.org/10.2527/jas2016-0431

Martínez-Álvaro, M., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., Wallace, R. J., Shih, B., Freeman, T. C., Watson, M., & Roehe, R. (2020). Identification of Complex Rumen Microbiome Interaction Within Diverse Functional

Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00659

Martínez-Alvaro, M., Ibañez-Escriche, N., & Casto-Rebollo, C. (2023). *RabbitR: Innovation in statistical learning: friendly Bayesian inference in R programmming lenguage* . Congresio internacional sobre aprendizaje, innovacion y cooperacion CINAIC 2023.

Martínez-Álvaro, M., Mattock, J., Auffret, M., Weng, Z., Duthie, C. A., Dewhurst, R. J., Cleveland, M. A., Watson, M., & Roehe, R. (2022). Microbiome-driven breeding strategy potentially improves beef fatty acid profile benefiting human health and reduces methane emissions. *Microbiome*, *10*(1). https://doi.org/10.1186/s40168-022-01352-6

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*(2), 134–158. https://doi.org/10.1177/1471082X14535524

Mattock, J., Mart Inez-A ´ Lvaro, M., Cleveland, M. A., Roehe, R., & Watson, M. (2023). KOunt: a reproducible KEGG orthologue abundance workflow. *Bioinformatics*. https://doi.org/https://doi.org/10.1093/bioinformatics/btad483

McGinn, S. M., Flesch, T. K., Beauchemin, K. A., Shreck, A., & Kindermann, M. (2019). Micrometeorological Methods for Measuring Methane Emission Reduction at Beef Cattle Feedlots: Evaluation of 3-Nitrooxypropanol Feed Additive. *Journal of Environmental Quality*, *48*(5), 1454–1461. https://doi.org/10.2134/jeq2018.11.0412

McIlroy, S. J., Leu, A. O., Zhang, X., Newell, R., Woodcroft, B. J., Yuan, Z., Hu, S., & Tyson, G. W. (2023). Anaerobic methanotroph 'Candidatus Methanoperedens nitroreducens' has a pleomorphic life cycle. *Nature Microbiology*, *8*(2), 321–331. https://doi.org/10.1038/s41564-022-01292-9

Miller, G. A., Auffret, M. D., Roehe, R., Nisbet, H., & Martínez-Álvaro, M. (2023). Different microbial genera drive methane emissions in beef cattle fed with two extreme diets. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1102400

Morgavi, D. P., Kelly, W. J., Janssen, P. H., & Attwood, G. T. (2013). Rumen microbial (meta)genomics and its application to ruminant production. *Animal*, *7*(s1), 184–201. https://doi.org/10.1017/S1751731112000419

Negussie, E., de Haas, Y., Dehareng, F., Dewhurst, R. J., Dijkstra, J., Gengler, N., Morgavi, D. P., Soyeurt, H., van Gastelen, S., Yan, T., & Biscarini, F. (2017). Invited review: Large-scale indirect measurements for enteric methane emissions in dairy cattle: A review of proxies and their potential for use in management and breeding decisions. *Journal of Dairy Science*, *100*(4), 2433–2453. https://doi.org/10.3168/jds.2016-12030

Neill, A. R., Grime, D. W., & Dawson, R. M. C. (1978). Conversion of choline methyl groups through trimethylamine into methane in the rumen. *Biochemical Journal*, *170*(3), 529–535. https://doi.org/10.1042/bj1700529

New, F. N., & Brito, I. L. (2020). What Is Metagenomics Teaching Us, and What Is Missed? *Annual Review of Microbiology*, *74*(1), 117–135. https://doi.org/10.1146/annurev-micro-012520-072314

Nollet, L., Mbanzamihigo, L., Demeyer, D., & Verstraete, W. (1998). Effect of the addition of Peptostreptococcus productus ATCC 35244 on reductive acetogenesis in the ruminal ecosystem after inhibition of methanogenesis by cell-free supernatant of Lactobacillus

plantarum 80. *Animal Feed Science and Technology*, *71*(1–2), 49–66. https://doi.org/10.1016/S0377-8401(97)00135-1

O'kelly, J. C., & Spiers, W. G. (1992). Effect of Monensin on Methane and Heat Productions of Steers Fed Lucerne Hay either ad libitum or at the Rate of 250 g/Hour. In *Aust. J. Agric. Res* (Vol. 43).

Oksanen, J., Simpson, G. L., & Blanchet, F. G. (2024). *vegan: Community Ecology Package* (2.6-6.1). CRAN.

ONU. (2015). *Paris Agreement*. https://unfccc.int/sites/default/files/spanish_paris_agreement.pdf

Patra, A. K. (2013). The effect of dietary fats on methane emissions, and its other effects on digestibility, rumen fermentation and lactation performance in cattle: A meta-analysis. *Livestock Science*, *155*(2–3), 244–254. https://doi.org/10.1016/j.livsci.2013.05.023

Pickering, N. K., Oddy, V. H., Basarab, J., Cammack, K., Hayes, B., Hegarty, R. S., Lassen, J., McEwan, J. C., Miller, S., Pinares-Patiño, C. S., & de Haas, Y. (2015a). Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants. *Animal*, *9*(9), 1431–1440. https://doi.org/10.1017/S1751731115000968

Pickering, N. K., Oddy, V. H., Basarab, J., Cammack, K., Hayes, B., Hegarty, R. S., Lassen, J., McEwan, J. C., Miller, S., Pinares-Patiño, C. S., & de Haas, Y. (2015b). Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants. *Animal*, *9*(9), 1431–1440. https://doi.org/10.1017/S1751731115000968

Poulsen, M., Schwab, C., Borg Jensen, B., Engberg, R. M., Spang, A., Canibe, N., Højberg, O., Milinovich, G., Fragner, L., Schleper, C., Weckwerth, W., Lund, P., Schramm, A., & Urich, T. (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nature Communications*, *4*(1), 1428. https://doi.org/10.1038/ncomms2432

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, *35*(SUPPL. 1). https://doi.org/10.1093/nar/gkl842

Rico, D. E., Chouinard, P. Y., Hassanat, F., Benchaar, C., & Gervais, R. (2016). Prediction of enteric methane emissions from Holstein dairy cows fed various forage sources. *Animal*, *10*(2), 203–211. https://doi.org/10.1017/S1751731115001949

Rooke, J. A., Wallace, R. J., Duthie, C. A., McKain, N., De Souza, S. M., Hyslop, J. J., Ross, D. W., Waterhouse, T., & Roehe, R. (2014). Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. *British Journal of Nutrition*, *112*(3), 398–407. https://doi.org/10.1017/S0007114514000932

Ross, E. M., Moate, P. J., Marett, L., Cocks, B. G., & Hayes, B. J. (2013). Investigating the effect of two methane-mitigating diets on the rumen microbiome using massively parallel sequencing. *Journal of Dairy Science*, *96*(9), 6030–6046. https://doi.org/10.3168/jds.2013-6766

Rowe, S. J., Hickey, S. M., Jonker, A., Hess, M. K., Janssen, P., Johnson, T., Bryson, B., Knowler, K., Pinares-Patino, C., Bain, W., Elmes, S., Young, E., Wing, J., Waller, E., Pickering, N., & Mcewan, J. C. (2019). *SELECTION FOR DIVERGENT METHANE YIELD IN NEW ZEALAND SHEEP-A TEN-YEAR PERSPECTIVE*.

Scott, D. D., & Oeffinger, M. (2016). Nucleolin and nucleophosmin: nucleolar proteins with multiple functions in DNA repair. *Biochemistry and Cell Biology*, *94*(5), 419–432. https://doi.org/10.1139/bcb-2016-0068

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Hadjithomas, M., Varghese, N. J., Paez-Espino, D., Perry, R., Henderson, G., Creevey, C. J., Terrapon, N., Lapebie, P., Drula, E., Lombard, V., Rubin, E., … Cerón Cucchi, M. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. In *Nature Biotechnology* (Vol. 36, Issue 4, pp. 359–367). Nature Research. https://doi.org/10.1038/nbt.4110

Silva, S. A., Salvador, A. F., Cavaleiro, A. J., Pereira, M. A., Stams, A. J. M., Alves, M. M., & Sousa, D. Z. (2016). Toxicity of long chain fatty acids towards acetate conversion by *Methanosaeta concilii* and *Methanosarcina mazei*. *Microbial Biotechnology*, *9*(4), 514–518. https://doi.org/10.1111/1751-7915.12365

Snelling, T. J., & Wallace, R. J. (2017). The rumen microbial metaproteome as revealed by SDS-PAGE. *BMC Microbiology*, *17*(1), 9. https://doi.org/10.1186/s12866-016-0917-y

Storm, I. M. L. D., Hellwing, A. L. F., Nielsen, N. I., & Madsen, J. (2012a). Methods for Measuring and Estimating Methane Emission from Ruminants. *Animals*, *2*(2), 160–183. https://doi.org/10.3390/ani2020160

Storm, I. M. L. D., Hellwing, A. L. F., Nielsen, N. I., & Madsen, J. (2012b). Methods for measuring and estimating methane emission from ruminants. In *Animals* (Vol. 2, Issue 2, pp. 160–183). https://doi.org/10.3390/ani2020160

Sweett, H. (2023, March 21). *LACTANET*. Introducing Methane Efficiency.

Tapio, I., Snelling, T. J., Strozzi, F., & Wallace, R. J. (2017). The ruminal microbiome associated with methane emissions from ruminant livestock. In *Journal of Animal Science and Biotechnology* (Vol. 8, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s40104-017-0141-0

Tsola, S. L., Zhu, Y., Chen, Y., Sanders, I. A., Economou, C. K., Brüchert, V., & Eyice, Ö. (2024). Methanolobus use unspecific methyltransferases to produce methane from dimethylsulphide in Baltic Sea sediments. *Microbiome*, *12*(1), 3. https://doi.org/10.1186/s40168-023-01720-w

Turner, H. G., & Thornton, R. F. (1966). *A RESPIRATION CHAMBER FOR CATTLE*.

Ungerfeld, E. M., Rust, S. R., & Burnett, R. (2007). Increases in microbial nitrogen production and efficiency in vitro with three inhibitors of ruminal methanogenesis. *Canadian Journal of Microbiology*, *53*(4), 496–503. https://doi.org/10.1139/W07-008

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, *7*(1), 142. https://doi.org/10.1186/1471-2164-7-142

van Lingen, H. J., Crompton, L. A., Hendriks, W. H., Reynolds, C. K., & Dijkstra, J. (2014). Meta-analysis of relationships between enteric methane yield and milk fatty acid profile in dairy cattle. *Journal of Dairy Science*, *97*(11), 7115–7132. https://doi.org/10.3168/jds.2014-8268

Vinutha, M. R., Chandrika, J., Krishnan, B., & Kokatnoor, S. A. (2023). EPCA—Enhanced Principal Component Analysis for Medical Data Dimensionality Reduction. *SN Computer Science*, *4*(3), 243. https://doi.org/10.1007/s42979-023-01677-5

Vyas, D., Alemu, A. W., McGinn, S. M., Duval, S. M., Kindermann, M., & Beauchemin, K. A. (2018). The combined effects of supplementing monensin and 3-nitrooxypropanol on methane emissions, growth rate, and feed conversion efficiency in beef cattle fed high-forage and high-grain diets1. *Journal of Animal Science*, *96*(7), 2923–2938. https://doi.org/10.1093/jas/sky174

Wall, E., Simm, G., & Moran, D. (2010). Developing breeding schemes to assist mitigation of greenhouse gas emissions. *Animal*, *4*(3), 366–376. https://doi.org/10.1017/S175173110999070X

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C. A., Hyslop, J. J., Ross, D. W., Waterhouse, A., Watson, M., & Roehe, R. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-2032-0

Williams, Y. J., Popovski, S., Rea, S. M., Skillman, L. C., Toovey, A. F., Northwood, K. S., & Wright, A.-D. G. (2009). A Vaccine against Rumen Methanogens Can Alter the Composition of Archaeal Populations. *Applied and Environmental Microbiology*, *75*(7), 1860–1866. https://doi.org/10.1128/AEM.02453-08

Wood, D. E., & Salzberg, S. L. (2014). *Kraken: ultrafast metagenomic sequence classification using exact alignments*. https://doi.org/doi:10.1186/gb-2014-15-3-r46

World Meteorological Organization. (2023). *WMO GREENHOUSE GAS BULLETIN*. https://library.wmo.int/records/item/68532-no-19-15-november-2023

Wright, A.-D. G., Auckland, C. H., & Lynn, D. H. (2007). Molecular Diversity of Methanogens in Feedlot Cattle from Ontario and Prince Edward Island, Canada. *Applied and Environmental Microbiology*, *73*(13), 4206–4210. https://doi.org/10.1128/AEM.00103-07

Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Watanabe, H., Masuda, K., Nishimoto, Y., Kubo, M., Hosoda, F., Rokutan, H., Matsumoto, M., Takamaru, H., Yamada, M., Matsuda, T., Iwasaki, M., Yamaji, T., Yachida, T., … Yamada, T. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*, *25*(6), 968–976. https://doi.org/10.1038/s41591-019-0458-7

Yan, T., Porter, M. G., & Mayne, C. S. (2009). Prediction of methane emission from beef cattle using data measured in indirect open-circuit respiration calorimeters. *Animal*, *3*(10), 1455–1462. https://doi.org/10.1017/S175173110900473X

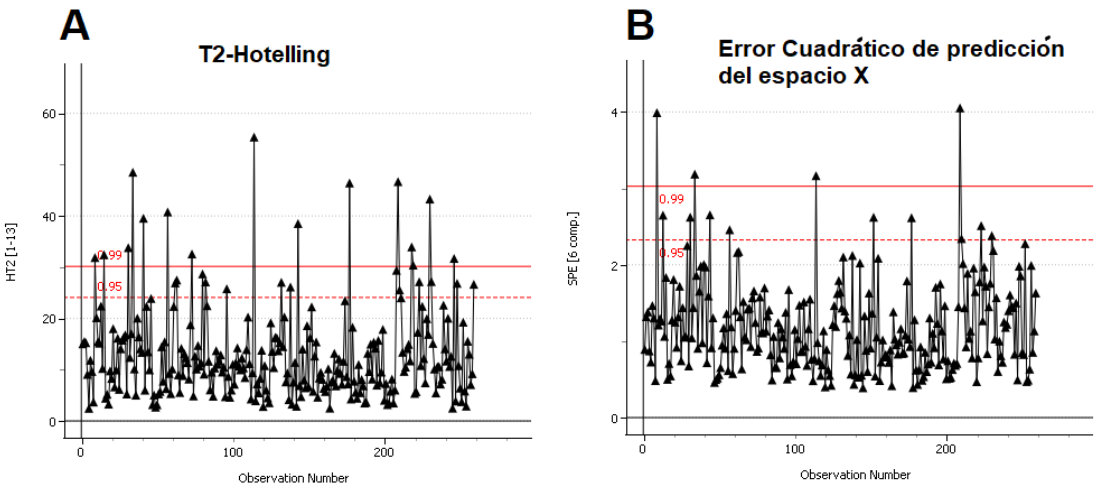
Yang, S., & Berdine, G. (2017). The receiver operating characteristic (ROC) curve. *The Southwest Respiratory and Critical Care Chronicles*, *5*(19), 34. https://doi.org/10.12746/swrccc.v5i19.391

Zhang, B., Lin, S., Moraes, L., Firkins, J., Hristov, A. N., Kebreab, E., Janssen, P. H., Bannink, A., Bayat, A. R., Crompton, L. A., Dijkstra, J., Eugène, M. A., Kreuzer, M., McGee, M., Reynolds, C. K., Schwarm, A., Yáñez-Ruiz, D. R., & Yu, Z. (2023). Methane prediction equations including genera of rumen bacteria as predictor variables improve prediction accuracy. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-48449-y

Zhang, G., Jiang, N., Liu, X., & Dong, X. (2008). Methanogenesis from Methanol at Low Temperatures by a Novel Psychrophilic Methanogen, " *Methanolobus psychrophilus* " sp. nov., Prevalent in Zoige Wetland of the Tibetan Plateau. *Applied and Environmental Microbiology*, *74*(19), 6114–6120. https://doi.org/10.1128/AEM.01146-08

Zhongtang Yu and Mark Morrison. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *BioTechniques* , *36*, 808–812.

8. MATERIAL SUPLEMENTARIO



Material Suplementario 1. Gráfico de errores de cada observación, obtenidos del PCA ajustado con la base de datos de microbioma escalado por bloques. A) T^2 de Hotelling B) Error cuadrático de predicción del espacio de los predictores. En ninguno de los casos se evidencian individuos *outlier*.

N° componente	var_explicada	var_acumulada
1	16.48%	16.48%
2	11.49%	27.97%
3	5.66%	33.63%
4	5.15%	38.78%
5	4.47%	43.26%
6	3.36%	46.61%
7	2.58%	49.19%
8	2.17%	51.36%
9	1.82%	53.18%
10	1.52%	54.70%
11	1.41%	56.11%
12	1.03%	57.14%
13	0.98%	58.12%
14	0.91%	59.03%
15	0.85%	59.88%

16	0.80%	60.68%
17	0.77%	61.44%
18	0.75%	62.19%
19	0.69%	62.89%
20	0.67%	63.56%
21	0.63%	64.19%
22	0.59%	64.78%
23	0.59%	65.37%
24	0.57%	65.94%
25	0.55%	66.49%
26	0.53%	67.02%
27	0.52%	67.54%
28	0.49%	68.04%
29	0.48%	68.52%
30	0.46%	68.98%
31	0.46%	69.44%
32	0.45%	69.89%
33	0.44%	70.33%
34	0.43%	70.76%
35	0.43%	71.19%
36	0.42%	71.61%
37	0.41%	72.02%
38	0.40%	72.42%
39	0.39%	72.81%
40	0.38%	73.19%
41	0.38%	73.57%
42	0.37%	73.94%
43	0.36%	74.30%
44	0.35%	74.65%
45	0.34%	74.99%
46	0.34%	75.33%
47	0.33%	75.66%
48	0.33%	75.99%
49	0.32%	76.31%
50	0.32%	76.63%
51	0.32%	76.95%
52	0.31%	77.26%
53	0.31%	77.57%
54	0.31%	77.88%
55	0.30%	78.18%
56	0.30%	78.48%
57	0.30%	78.78%
58	0.29%	79.07%
59	0.29%	79.36%
60	0.29%	79.64%

61	0.28%	79.92%
62	0.28%	80.19%
63	0.28%	80.47%
64	0.27%	80.74%
65	0.27%	81.01%
66	0.27%	81.27%
67	0.26%	81.54%
68	0.26%	81.80%
69	0.26%	82.05%
70	0.26%	82.31%
71	0.25%	82.56%
72	0.25%	82.81%
73	0.25%	83.06%
74	0.24%	83.30%
75	0.24%	83.55%
76	0.24%	83.79%
77	0.24%	84.02%
78	0.24%	84.26%
79	0.23%	84.49%
80	0.23%	84.72%
81	0.23%	84.95%
82	0.22%	85.17%
83	0.22%	85.39%
84	0.22%	85.60%
85	0.22%	85.82%
86	0.21%	86.03%
87	0.21%	86.24%
88	0.21%	86.45%
89	0.21%	86.66%
90	0.21%	86.86%
91	0.20%	87.07%
92	0.20%	87.27%
93	0.20%	87.47%
94	0.20%	87.67%
95	0.20%	87.87%
96	0.19%	88.06%
97	0.19%	88.25%
98	0.19%	88.44%
99	0.19%	88.63%
100	0.19%	88.82%
101	0.18%	89.00%
102	0.18%	89.18%
103	0.18%	89.36%
104	0.18%	89.54%
105	0.18%	89.72%

106	0.18%	89.90%
107	0.17%	90.07%
108	0.17%	90.24%
109	0.17%	90.41%
110	0.17%	90.58%
111	0.17%	90.75%
112	0.17%	90.92%
113	0.16%	91.08%
114	0.16%	91.24%
115	0.16%	91.41%
116	0.16%	91.57%
117	0.16%	91.73%
118	0.16%	91.88%
119	0.15%	92.04%
120	0.15%	92.19%
121	0.15%	92.34%
122	0.15%	92.49%
123	0.15%	92.64%
124	0.15%	92.79%
125	0.15%	92.93%
126	0.14%	93.08%
127	0.14%	93.22%
128	0.14%	93.36%
129	0.14%	93.50%
130	0.14%	93.64%
131	0.14%	93.78%
132	0.14%	93.92%
133	0.14%	94.05%
134	0.13%	94.19%
135	0.13%	94.32%
136	0.13%	94.45%
137	0.13%	94.58%
138	0.13%	94.71%
139	0.13%	94.84%
140	0.13%	94.96%
141	0.12%	95.09%
142	0.12%	95.21%
143	0.12%	95.33%
144	0.12%	95.45%
145	0.12%	95.57%
146	0.12%	95.69%
147	0.12%	95.80%
148	0.12%	95.92%
149	0.12%	96.03%
150	0.11%	96.15%

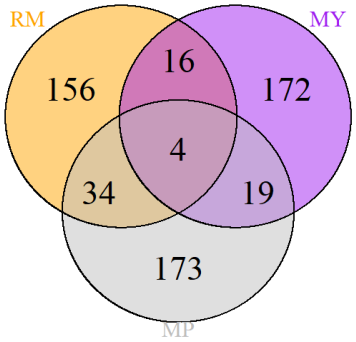
151	0.11%	96.26%
152	0.11%	96.37%
153	0.11%	96.48%
154	0.11%	96.59%
155	0.11%	96.69%
156	0.11%	96.80%
157	0.11%	96.91%
158	0.10%	97.01%
159	0.10%	97.11%
160	0.10%	97.21%
161	0.10%	97.32%
162	0.10%	97.42%
163	0.10%	97.51%
164	0.10%	97.61%
165	0.10%	97.71%
166	0.10%	97.81%
167	0.09%	97.90%
168	0.09%	97.99%
169	0.09%	98.08%
170	0.09%	98.17%
171	0.09%	98.26%
172	0.09%	98.35%
173	0.09%	98.44%
174	0.09%	98.52%
175	0.09%	98.61%
176	0.08%	98.69%
177	0.08%	98.77%
178	0.08%	98.85%
179	0.08%	98.93%
180	0.08%	99.01%
181	0.08%	99.09%
182	0.08%	99.16%
183	0.08%	99.24%
184	0.07%	99.31%
185	0.07%	99.38%
186	0.07%	99.45%
187	0.07%	99.52%
188	0.07%	99.59%
189	0.07%	99.66%
190	0.06%	99.72%
191	0.06%	99.78%
192	0.06%	99.84%
193	0.06%	99.90%
194	0.06%	99.95%
195	0.05%	100.00%

196	0.00%	100.00%

Material suplementario 2. Tabla con la varianza explicada por las componentes principales del PCA realizado sobre el microbioma. Contiene el número de variables, la varianza explicada por cada componente y la varianza acumulada al agregar más componentes.

	MP	**MY**	**RM**	**Media**	**sd**
PLS	1957	2314	1705	1992	306
RF	2489	2449	2477	2472	21
XGB	988	946	776	903	112
Total de VS	3646	4152	3115		
% VS	76,5%	87,1%	65,34%		

Material suplementario 3. Tabla de las variables seleccionadas por cada modelo para los 3 fenotipos por cada modelo. El Total de VS es el total de variables seleccionadas por los 3 odelos en conjunto y el %VS es cuanto representa del total esa cantidad de variables. Y luego la media y la desviación típica (sd) de la cantidad de variables que cada algoritmo ha seleccionado para los 3 fenotipos.



Material suplementario 4. Diagrama de Venn de las variables seleccionadas por todos los algoritmos para cada fenotipo. En naranja tenemos RM, en violeta MY y en gris MP.

Material Suplementario 5. Variables seleccionadas para cada uno de los fenotipos, para los 3 algoritmos: PLS, XGB y RF.

	MP	**RM**	**MY**
1	CandidatusMethanoperedens	Archaeoglobus	Archaeoglobus
2	Halapricum	CandidatusMethanoperedens	CandidatusMethanomethylophilus
3	Haladaptatus	Haloarcula	CandidatusMethanoperedens
4	Halobacterium	Haladaptatus	Halobacterium
5	Haloferax	CandidatusHalobonum	Halobellus
6	Halolamina	Halobaculum	Haloprofundus

7	Methanotorris	Halolamina	Halostagnicola
8	Methanosalsum	Methanobrevibacter	Methanosphaera
9	Pyrococcus	Methanotorris	Methanocaldococcus
10	Methanobacterium	Methanosalsum	Methanomassiliicoccus
11	Methanolobus	Halobiforma	Methanomicrobium
12	Methanosarcina	Halostagnicola	Methanolobus
13	Aciduliprofundum	Natrinema	Methanospirillum
14	Halorhabdus	Methanobacterium	Methermicoccus
15	Halosimplex	Methanosphaera	CandidatusWalczuchella
16	Methanohalobium	Methanolobus	Leptothrix
17	Palaeococcus	Geoglobus	Asaia
18	Thermofilum	Halorhabdus	Acidimicrobium
19	Vulcanisaeta	Halosimplex	Actinotignum
20	Mitsuaria	Methanohalobium	CandidatusAmoebophilus
21	Paucibacter	Methanosphaerula	Archangium
22	Rubrivivax	Vulcanisaeta	Martelella
23	Asaia	CandidatusIzimaplasma	Bdellovibrio
24	Kozakia	Mitsuaria	Geminocystis
25	Castellaniella	Nitratiruptor	Polaromonas
26	CandidatusAmoebophilus	Paucibacter	Aequorivita
27	Delftia	CandidatusCardinium	Croceibacter
28	Brachybacterium	Beutenbergia	Congregibacter
29	Desulfobacula	Nitrobacter	CandidatusEvansia
30	Petrimonas	Pragia	Persephonella
31	CandidatusIshikawaella	Dichelobacter	Rufibacter
32	Flammeovirga	Chlorobium	Filomicrobium
33	Sediminicola	Laribacter	Singulisphaera
34	Kosmotoga	Colwellia	Kangiella
35	Methylacidiphilum	Desulfurella	Tatlockia
36	Methylomicrobium	Endomicrobium	Methylovorus
37	Planktothrix	Blastococcus	Trichodesmium
38	Pirellula	Filomicrobium	Basfia
39	Sanguibacter	Leptolyngbya	Pirellula
40	Pseudopedobacter	Methylomicrobium	Leisingera
41	Sphingorhabdus	Moritella	Pannonibacter
42	Spongiibacter	CandidatusProtochlamydia	Thiobacimonas
43	Thiobacillus	Pirellula	Parvibaculum
44	Acaryochloris	Isoptericola	Haematospirillum
45	Halobacillus	Actinoalloteichus	Pararhodospirillum
46	Terribacillus	Defluviimonas	Sphingorhabdus
47	Ralstonia	CandidatusPhaeomarinobacter	Thermodesulfatator
48	Tetragenococcus	Solitalea	Thermodesulfobium
49	Edwardsiella	Zhongshania	Acholeplasma
50	Pleurocapsa	Steroidobacter	Arcanobacterium
51	Pediococcus	Tsukamurella	Thermosulfidibacter

52	Weissella	Acaryochloris	Bacillus
53	Mycobacterium	Acholeplasma	Jeotgalibaca
54	Neisseria	Bordetella	Chamaesiphon
55	Aeromicrobium	Agarivorans	CandidatusArthromitus
56	Chelativorans	Halobacillus	Comamonas
57	Saccharopolyspora	Salimicrobium	Hydrogenophaga
58	Megamonas	Terribacillus	Citrobacter
59	Sphaerochaeta	Proteiniphilum	Melissococcus
60	Salinicoccus	Gordonia	Psychroflexus
61	Jonquetella	Sulfuricurvum	Pseudobutyrivibrio
62	Aquifex	Isosphaera	Thermobifida
63	CandidatusRuthia	Gemella	Desulfitobacterium
64	Halothece	Odoribacter	Peptostreptococcus
65	Obesumbacterium	Dehalobacter	Chelativorans
66	Histoplasma	Frateuria	Chondromyces
67	Pochonia	Sphingobium	Xanthomonas
68	Gloeophyllum	Sphingopyxis	Obesumbacterium
69	Magnaporthe	Spirochaeta	Neofusicoccum
70	Malassezia	Wenzhouxiangella	Tsuchiyaea
71	Phycomyces	Pajaroellobacter	Babjeviella
72	Tetrapisispora	Thioalkalimicrobium	Gloeophyllum
73	Aureobasidium	Cyphellophora	Scedosporium
74	Sclerotinia	Colletotrichum	Arthrobotrys
75	Anthracocystis	Rhinocladiella	Phanerochaete
76	Pseudogymnoascus	Tilletiaria	Bipolaris
77	Batrachochytrium	Anthracocystis	K07138.K01783
78	Saprolegnia	Phialocephala	K05896.K01783
79	Trypanosoma	Pseudogymnoascus	K03702.K01783
80	Thecamonas	Nannochloropsis	K03723.K01783
81	K03723.K01783	Thecamonas	K07164.K01783
82	K06330.K01783	K03704.K01783	K08978.K01783
83	K12962.K01783	K03723.K01783	K15539.K01783
84	K02474.K01783	K07173.K01783	K01662.K01783
85	K01153.K01783	K01119.K01783	K01673.K01783
86	K01620.K01783	K12960.K01783	K00826.K01783
87	K01639.K01783	K16850.K01783	K00390.K01783
88	K05515.K01783	K12998.K01783	K08177.K01783
89	K14260.K01783	K02492.K01783	K00872.K01783
90	K22132.K01783	K02956.K01783	K00873.K01783
91	K06871.K01783	K00335.K01783	K06001.K01783
92	K01673.K01783	K00338.K01783	K03929.K01783
93	K23977.K01783	K05540.K01783	K08681.K01783
94	K00857.K01783	K09457.K01783	K03932.K01783
95	K02109.K01783	K17320.K01783	K03977.K01783
96	K07322.K01783	K23977.K01783	K19166.K01783

97	K02111.K01783	K05592.K01783	K00057.K01783
98	K09976.K01783	K18697.K01783	K09157.K01783
99	K06041.K01783	K00872.K01783	K01858.K01783
100	K01303.K01783	K02111.K01783	K01868.K01783
101	K07387.K01783	K02112.K01783	K03106.K01783
102	K07391.K01783	K03932.K01783	K03116.K01783
103	K02651.K01783	K09124.K01783	K07010.K01783
104	K19159.K01783	K15773.K01783	K07035.K01783
105	K19166.K01783	K00058.K01783	K03151.K01783
106	K01813.K01783	K08303.K01783	K03168.K01783
107	K00059.K01783	K01867.K01783	K07063.K01783
108	K01874.K01783	K21023.K01783	K16710.K01783
109	K01887.K01783	K01881.K01783	K07075.K01783
110	K13653.K01783	K01887.K01783	K07082.K01783
111	K03631.K01783	K03617.K01783	K07090.K01783
112	K06207.K01783	K04487.K01783	K07096.K01783
113	K03655.K01783	K06207.K01783	K03665.K01783
114	K11068.K01783	K12373.K01783	K19304.K01783
115	K01012.K01783	K19304.K01783	K02822.K01783
116	K02358.K01783	K01534.K01783	K07590.K01783
117	K03695.K01783	K02013.K01783	K15915.K01783
118	K18013.K01783	K02038.K01783	K18013.K01783
119	K19802.K01783	K07271.K01783	K14155.K01783
120	K00287.K01783	K20444.K01783	K12573.K01783
121	K03305.K01783	K01243.K01783	K01259.K01783
122	K02004.K01783	K01246.K01783	K06926.K01783
123	K07248.K01783	K01251.K01783	K13018.K01783
124	K02037.K01783	K01258.K01783	K07404.K01783
125	K02038.K01783	K06987.K01783	K03086.K01783
126	K03830.K01783	K03574.K01783	K03551.K01783
127	K01709.K01783	K10117.K01783	K03570.K01783
128	K01259.K01783	K00113.K01783	K07462.K01783
129	K22205.K01783	K01486.K01783	K07487.K01783
130	K01752.K01783	K01955.K01783	K02279.K01783
131	K00941.K01783	K01972.K01783	K07495.K01783
132	K22293.K01783	K13283.K01783	K01425.K01783
133	K07443.K01783	K21556.K01783	K05366.K01783
134	K03573.K01783	K04798.K01783	K18908.K01783
135	K21498.K01783	K13663.K01783	K04062.K01783
136	K02283.K01783	K14113.K01783	K00641.K01783
137	K16692.K01783	K07254.K01783	K01972.K01783
138	K01468.K01783	K02626.K01783	K00652.K01783
139	K01937.K01783	K04795.K01783	K03217.K01783
140	K01939.K01783	K02922.K01783	K00666.K01783
141	K00625.K01783	K09154.K01783	K13283.K01783

142	K01494.K01783	K02103.K01783	K10218.K01783
143	K01953.K01783	K06518.K01783	K04798.K01783
144	K04072.K01783	K02028.K01783	K11130.K01783
145	K09748.K01783	K19048.K01783	K06961.K01783
146	K00652.K01783	K07442.K01783	K09163.K01783
147	K22928.K01783	K07158.K01783	K03313.K01783
148	K13017.K01783	K03540.K01783	K11719.K01783
149	K02866.K01783	K16874.K01783	K19575.K01783
150	K01286.K01783	K03265.K01783	K00324.K01783
151	K06402.K01783	K07141.K01783	K07108.K01783
152	K06518.K01783	K12452.K01783	K04484.K01783
153	K01818.K01783	K06077.K01783	K12996.K01783
154	K02856.K01783	K13658.K01783	K10540.K01783
155	K03313.K01783	K00579.K01783	K09740.K01783
156	K02028.K01783	K13665.K01783	K06878.K01783
157	K07718.K01783	K02482.K01783	K03667.K01783
158	K01787.K01783	K18814.K01783	K07323.K01783
159	K07402.K01783	K00100.K01783	K15888.K01783
160	K03071.K01783	K21472.K01783	K05820.K01783
161	K20151.K01783	K15523.K01783	K03433.K01783
162	K18446.K01783	K14340.K01783	K02319.K01783
163	K22452.K01783	K18855.K01783	K01730.K01783
164	K13288.K01783	K03243.K01783	K07746.K01783
165	K07442.K01783	K16306.K01783	K03763.K01783
166	K03540.K01783	K09960.K01783	K23541.K01783
167	K10540.K01783	K11051.K01783	K01623.K01783
168	K19955.K01783	K07800.K01783	K01150.K01783
169	K07721.K01783	K01621.K01783	K09940.K01783
170	K17744.K01783	K12349.K01783	K04764.K01783
171	K17290.K01783	K05772.K01783	K05772.K01783
172	K05820.K01783	K02796.K01783	K02250.K01783
173	K01186.K01783	K00401.K01783	K01596.K01783
174	K15429.K01783	K03809.K01783	K11144.K01783
175	K23121.K01783	K08984.K01783	K11251.K01783
176	K18814.K01783	K13640.K01783	K09792.K01783
177	K16923.K01783	K09792.K01783	K21405.K01783
178	K00728.K01783	K15580.K01783	K01816.K01783
179	K09741.K01783	K11936.K01783	K04076.K01783
180	K00956.K01783	K21345.K01783	K19003.K01783
181	K11089.K01783	K00219.K01783	K00065.K01783
182	K15669.K01783	K04076.K01783	K08092.K01783
183	K14340.K01783	K07305.K01783	K03736.K01783
184	K06982.K01783	K01581.K01783	K01729.K01783
185	K21065.K01783	K12067.K01783	K03735.K01783
186	K09974.K01783	K09137.K01783	K03189.K01783

187	K02044.K01783	K01295.K01783	K03188.K01783
188	K00033.K01783	K03720.K01783	K00303.K01783
189	K22699.K01783	K01066.K01783	K15778.K01783
190	K19134.K01783	K22226.K01783	K09859.K01783
191	K08301.K01783	K17234.K01783	K17234.K01783
192	K21399.K01783	K00880.K01783	K02414.K01783
193	K23975.K01783	K12286.K01783	K09902.K01783
194	K03234.K01783	K03711.K01783	K22905.K01783
195	K00853.K01783	K01793.K01783	K09159.K01783
196	K15855.K01783	K01202.K01783	K08222.K01783
197	K04032.K01783	K05311.K01783	K11616.K01783
198	K02973.K01783	K11381.K01783	K14081.K01783
199	K20885.K01783	K15045.K01783	K07112.K01783
200	K00065.K01783	K18578.K01783	K08516.K01783
201	K21721.K01783	K05757.K01783	K01565.K01783
202	K08092.K01783	K23475.K01783	K14125.K01783
203	K03410.K01783	K02245.K01783	K13583.K01783
204	K03306.K01783	K01565.K01783	K10411.K01783
205	K22844.K01783	K00737.K01783	K02526.K01783
206	K13907.K01783	K01640.K01783	K09527.K01783
207	K06286.K01783	K12276.K01783	K14591.K01783
208	K07177.K01783	K09778.K01783	K00196.K01783
209	K17255.K01783	K03801.K01783	K09962.K01783
210	K09928.K01783	K07235.K01783	K16176.K01783
211	K03720.K01783		K00557.K01783
212	K01066.K01783		
213	K15778.K01783		
214	K22226.K01783		
215	K01753.K01783		
216	K14054.K01783		
217	K01040.K01783		
218	K00194.K01783		
219	K04568.K01783		
220	K03045.K01783		
221	K05311.K01783		
222	K00433.K01783		
223	K01761.K01783		
224	K09773.K01783		
225	K23253.K01783		
226	K02526.K01783		
227	K01640.K01783		
228	K20904.K01783		
229	K00054.K01783		
230	K17108.K01783		