STUDY OF THE HOST GENETIC CONTROL OVER THE RUMINAL MICROBIOTA AND THEIR RELATIONSHIPS WITH METHANE EMISSIONS IN DAIRY CATTLE

Ph.D. Thesis by

Alejandro Saborío Montero

Supervisor: **Dr. Oscar González Recio**

Institute for Animal Science and Technology

Universitat Politècnica de València

Valencia, 1st July 2021

STUDY OF THE HOST GENETIC CONTROL OVER THE RUMINAL MICROBIOTA AND THEIR RELATIONSHIPS WITH METHANE EMISSIONS IN DAIRY CATTLE

This thesis has been submitted in fulfilment of the requirements for the degree of Doctor at the Universitat Politècnica de València.

Esta tesis ha sido escrita y presentada como uno de los requisitos para optar al grado de Doctor por la Universitat Politècnica de València.

By

Alejandro Saborío Montero

_____

Thesis Supervisor

Dr. Oscar González Recio

_____

Valencia, 1st July 2021

"Cuando sea viejo y con melancolía
allá por las tardes me ponga a soñar
soltaré los canes de mi fantasía
y a lejanos bosques me iré a cazar.

No tendré ya perros ni el arco de Diana
ni la fuerza inquieta de la juventud
que corran conmigo en hora temprana
por llanos y montes de esta latitud.

Seguiré soñando y seguiré riendo
y que los ladridos de mi corazón
sean esa jauría que sigue corriendo
tras la presa viva, que es la ilusión."

Autor: Alejandro Saborío Villegas, mi padre.
"El Cazador", 1963.

ACKNOWLEDGMENTS

I would like to thank the University of Costa Rica (UCR) for the scholarship for course doctorate studies and to my colleagues at the Zootechnics School from UCR for trust in me and for their support. As well as to the Department of Animal Breeding of the National Institute of Agri-food Research and Technology (INIA) for host my research, and to all the people involved within the METALGEN project.

A deserved thanks to my advisor Ph.D. Oscar González Recio for been such an inspiring leader, who has given me the keys to open the door of a new academic future, with outstanding patience and kindness in the process. His enthusiastic nature and clear direction of the research has been essential through the development of this doctoral thesis.

I thank to Agustin Blasco for its role as tutor during the progress of this thesis, being the link between the Universitat Politècnica de València and me.

Special thanks to my parents Alejandro and Marta, and also to my sisters Gabriela, Priscilla, Rebeca and Graciela for all their support, and to my nephews Alejandro and Isaac for being a source of proud and joy in my life.

And finally, I would like to give the most special thanks to my wife Sinaí Mariela for being always there for me and for her unconditional support, also to our children Andrés, Priscilla and David for making sense of all my efforts and for being my source of happiness.

AGRADECIMIENTOS

Agradezco a la Universidad de Costa Rica (UCR) por la beca para realizar los estudios de doctorado y a mis compañeros de la Escuela de Zootecnia de la UCR por la confianza en mí y por su apoyo. Así como al Departamento de Mejora Genética del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) por acoger mi investigación, y para todas las personas involucradas dentro del proyecto METALGEN.

Un merecido agradecimiento a mi director Dr. Oscar González Recio por ser un líder tan inspirador, que me ha dado las llaves para abrir la puerta de un nuevo futuro académico, con destacada paciencia y amabilidad en el proceso. Su carácter entusiasta y clara dirección de la investigación ha sido fundamental para el desarrollo de esta tesis doctoral.

Agradezco a Agustín Blasco su papel de tutor durante el desarrollo de esta tesis, siendo el nexo entre la Universitat Politècnica de València y yo.

Un agradecimiento especial a mis padres Alejandro y Marta, y también a mis hermanas Gabriela, Priscilla, Rebeca y Graciela por todo su apoyo, y a mis sobrinos Alejandro e Isaac por ser una fuente de orgullo y alegría en mi vida.

Y finalmente, quisiera agradecer muy especialmente a mi esposa Sinaí Mariela por estar siempre ahí para mí y por su apoyo incondicional, también a nuestros hijos Andrés, Priscilla y David por dar sentido a todos mis esfuerzos y por ser mi fuente de felicidad.

LIST OF ABREVIATIONS

AIC	Akaike Information Criteria
Aitc	Aitchison
AMS	Automatic Milking System
ANOVA	Analysis of Variance
BGLR	Bayesian Generalized Linear Regression
BIC	Bayesian Information Criteria
BrCr	Bray-Curtis
CA	Correspondence Analysis
Canb	Canberra
CCA	Constrained Correspondence Analysis
CH_4	Methane
CLR	Centered Log-Ratio
COG	Cluster of Orthologous Groups
CONAFE	National Confederation of Associations of Spanish Friesian
CrPr	Cross-Product
CV	Coefficient of Variation
DCA	Detrended Correspondence Analysis
DIAMOND	Double Index Alignment of Next-Generation Sequencing Data
DIC	Deviance Information Criteria
DIM	Days In Milk
DNA	Deoxyribonucleic Acid
DT	Detrending
EGMV	Estimated Genomic-Microbiome Value
EMV	Estimated Microbiome Value
Eucl	Euclidean
F	Standard Normal Cumulative Distribution Function
G, GRM	Genomic Relationship Matrix
GBLUP	Genomic Best Linear Unbiased Predictor
GEBVs	Genomic Estimated Breeding Values
GHG	Greenhouse Gas
h^2	Heritability
HBLUP	Hologenomic Best Linear Unbiased Predictor
H_iBLUP	Hologenomic Best Linear Unbiased Predictor with host interaction
ho^2	Holobiability
HSD	Honestly Significant Difference
INIA	National Institute of Agri-food Research and Technology
Jacc	Jaccard
KEGG	Kyoto Encyclopedia of Genes and Genomes
Ks	Microbiota Relationship Matrices

MAF	Minor Allele Frequency
Maha	Mahalanobis
MBLUP	Microbiomic Best Linear Unbiased Predictor
MCMC	Markov Chains Monte Carlo
MDS	Multidimensional Scaling
MEGAN	Metagenome Analyzer
MET	Methane Concentration (ppm CH4)
METALGEN	Research project (RTA2015-00022-CO3) to improve feed efficiency and mitigate greenhouse gases emissions in dairy cattle
MI	Methane Intensity (ppm CH4 / kg milk)
MinION	Nanopore sequencer device from Oxford Nanopore Technologies.
MR	Monotone Regression
NCBI	National Center for Biotechnology Information
NEIKER-BRTA	Basque Institute for Agricultural Research and Development- Basque Research and Technology Alliance
NMDS	Non-Metric Multidimensional Scaling
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame
OTU	Operational Taxonomic Unit
PC1	First principal component
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCs	Principal components
PI	Probability Intervals
PSE	Standard Error of the Posterior Distribution
QTL	Quantitative Trait Loci
RA	Relative Abundance
RDA	Redundancy Analysis
SD	Standard Deviation
SE	Standard Error
SEMs	Structural Equation Models
SEP	Standard Error of Prediction
SNP	Single Nucleotide Polymorphism
ssGBLUP	Single Step Genomic Best Linear Unbiased Predictor
TBV	True Breeding Values
TM	Threshold Model Package
TMV	True Microbiome Value
95%HPD	95% highest posterior density interval

CONTENTS

ABSTRACT

The analysis of the host genetic control over its microbiota has recently been pointed out as a promising theme in different fields of study. The relationship between the host-microbiome holobiont and phenotypes in dairy cattle could lead to new insights in breeding programs. The analysis of microbiota implies dealing with sparce compositional data frames, among other issues concerning inference about causality of microbes over traits of interest, single microbe effect *vs* whole microbiota effect on relevant characters, or aggregation of whole microbiota to approach implementation in breeding selection programs. Optimal technics to analyze this kind of data are not well defined yet. Therefore, different statistical approaches to better deal with this sort of data are to be used. Within this Ph.D. thesis, estimation and analysis through different statistical approaches were performed aiming to unravel the host genetic control over the microbiota in dairy cattle. Besides, methane concentration trait was analyzed as a potential phenotype to be included in the Spanish dairy cattle breeding program. A total of 437 cows from 14 commercial dairy cattle farms at the northern region of Spain (País Vasco, Cantabria, Navarra and Gerona), were involved in the METALGEN project, which was articulated on three sub-projects. This thesis was developed within the sub-project one (The (meta)-genomic base of the animal-microbiome binomy, and its relationship with feed efficiency and methane emissions in the Holstein-Friesian breed) of the METALGEN project, leaded by INIA. The animals were genotyped by CONAFE using the low-density chips EuroG 10K and EuroG LD 12k (Illumina, San Diego, California, USA), and then imputed to the Bovine 50k SNP chip (Illumina, San Diego, California, USA) using its reference population to obtain genotypes containing 54,609 SNP. Microbial composition from each cow was obtained from whole metagenome sequencing of ruminal content samples using a MinION device from Oxford Nanopore Technologies. Methane concentration was measured with Guardian® NG infrared gas monitor from Edinburgh Sensors during cow's visits to the milking automated system. Methane was weekly averaged from the peaks for each cow during 2-to-3-week period. Concentration (ppm CH_4) and intensity (ppm CH_4/kg milk) were used as phenotypes.

Risk factors for methane concentration and methane intensity were estimated using phenotypic variables of milk yield (kg), milk fat (%), milk protein (%), conformation traits;

methane concentration and methane intensity GEBVs scales previously generated from a ssGBLUP; as well as centered log-ratio transformed (CLR) microbiome composition at phylum (n=86) and genus (n=1240) level, as potential risk factors. Cows were classified in quartiles (low, middle-low, middle-high, and high), according to individual records of methane measurements, and averaged milk yield in the case of methane intensity. A threshold model approach, under a Markov chain Monte Carlo (MCMC) Bayesian framework, was used to determine risk factors for being classified as high for both traits. High yielding cows had lower probability of being classified in the high category for methane intensity. Body structure and capacity traits were positively correlated to increased probability of being in the higher category of methane concentration. Larger GEBVs were also protector factors for both traits, reducing the probability of being classified as high methane emitters. Higher relative abundance of most eukaryotes (mainly ciliate protozoa and fungi) and some archaea (*Methanobrevibacter spp*. *Methanothermus spp* and *Methanosphera spp*.) were risk factors for being classified in the high categories.

The risk factor approach, using GEBVs, allowed to associate the genetic effect of the host over the phenotypes, as well as the microbiome effect over the phenotypes; however, it lacked of information regarding the host genetic effect over the microbiome. In order to tackle this limitation, a set of structural equation models (SEMs) of a recursive type within a Markov chain Monte Carlo (MCMC) framework was proposed to jointly analyze the host-metagenome-phenotype relationship. Non-recursive models were set as benchmark. The relative abundance of rumen single microbes (RA) and CH_4 concentration were included as phenotypes into both sets of bivariate mixed animal models (recursive and non-recursive). The host genetics was included into the models through the genomic relationship matrix between animals, allowing to estimate variance components. Heritability of CH_4 was estimated at 0.12 ± 0.01 in both, the recursive and non-recursive, models. Likewise, heritability estimates for the relative abundance of the taxa overlapped between models and ranged between 0.08 and 0.48. Genetic correlations between the microbial composition and CH_4 ranged from -0.76 to 0.65 in the non-recursive bivariate models and from -0.68 to 0.69 in the recursive models. Regardless of the statistical model used, positive genetic correlations with methane were estimated consistently for the 7 genera pertaining to the Ciliophora

phylum, as well as for those genera belonging to the Euryarchaeota (*Methanobrevibacter sp*.), Chytridiomycota (*Neocallimastix sp.*) and Fibrobacteres (*Fibrobacter sp*.) phyla.

The analysis of single taxa informed about independent relationships between the phenotype and one microbe at a time, as well as concerning of the host-genetic effect over a given microorganism; However, it lacks comprehensive information about the whole microbiome effect over the phenotype and the host genetic effect over the whole microbiome. To tackle the whole microbiome effect over the phenotype, the microbiome must take a conformable form to be included into genetic animal models. For that, twelve microbiota relationship matrices (K) from different microbiome distance metrics were built, aiming to compare its performance within a variance component estimation framework for CH_4 and whole microbiome analysis on simulation (n = 1000, 25 replicates) and real data were performed, considering four possible models: an additive genomic model (GBLUP), a microbiome model (MBLUP), a genetic and microbiome effects model (HBLUP) and a genetic, microbiome and genetic × microbiome interaction effects model (H_iBLUP). All models were implemented within a Bayesian framework using the BGLR package in R. The Ks built from Multidimensional Scaling (MDS), Redundancy Analysis (RDA) and Constrained Correspondence Analysis (CCA) performed better in simulation to estimate heritability and microbiability. The same methods to build Ks were between the most plausible in real data, according to the Deviance Information Criteria (DIC). The DIC was also used to obtain the most plausible model, which happened to be the H_iBLUP. A new term "*Holobiability*" was defined to refer to the proportion of the phenotypic variance attributable to the host-microbiome holobiont effects. Estimates from real data using H_iBLUP varied depending on the K used and ranged between 0.15-0.17, 0.15-0.21 and 0.42-0.59 for heritability, microbiability and holobiability, respectively.

Despite the usage of Ks allowed to estimate the whole microbiome effect over the phenotype, there was still lack of information concerning the host genetic effect over the whole microbiome. To deal with this issue, the whole microbiome not only needed to be conformable, as in K form, but needed to be aggregated into vectors in order to be treated as a phenotype. The microbiome dataset was aggregated through Principal Component Analysis (PCA) into few principal components (PCs) that were used as proxies of the core metagenome. The PCA allowed condensing the huge and fuzzy taxonomical and functional

information from the metagenome into few PCs. Bivariate animal models were applied using these PCs and methane production as phenotypes. Part of the variability condensed in these PCs is controlled by the cow genome, with heritability estimates for the first PC (PC1) of ~0.30 at all taxonomic levels, with a large probability (>83%) of the posterior distribution being > 0.20 and with the 95% highest posterior density interval (95%HPD) not containing zero. Most genetic correlation estimates between PC1 and methane were large (≥0.70) at all taxonomic levels, with most of the posterior distribution (>82%) being >0.50 and with its 95%HPD not containing zero.

These results suggest that rumen's whole metagenome recursively regulate methane emissions in dairy cows, and that both CH_4 and the microbiota compositions are partially controlled by the host genotype. The risk factor approach contributed to understand the effect of microbiome and host genetics over methane emissions; However, it lacked the ability of describing the control exerted by the host genetic over microbiome. The CH_4 was positively associated with relative abundance (RA) of eukaryotes (protozoa and fungi) at Phylum, Class, Order, Family and Genus. Nanopore long reads allowed the characterization of the core rumen metagenome using whole metagenome sequencing, and the purposed aggregated variables (PCs) could be used in animal breeding programs to reduce methane emissions in future generations.

RESUMEN

El análisis del control genético del hospedador sobre su microbiota ha sido señalado recientemente como un tema prometedor en diferentes campos de estudio. La relación entre el holobionte hospedador-microbioma y los fenotipos en el ganado lechero podría conducir a nuevos conocimientos en los programas de selección genética. El análisis de la microbiota implica lidiar con bases de datos composicionales, entre otras problemáticas relacionadas con la inferencia sobre la causalidad de los microbios sobre los rasgos de interés, el efecto de un solo microbio frente al efecto de toda la microbiota en los caracteres relevantes, o la agregación de toda la microbiota para abordar la implementación en los programas de selección genética. Las técnicas óptimas para analizar este tipo de datos no están bien definidas aún. Por lo tanto, se deben utilizar diferentes enfoques estadísticos para tratar mejor este tipo de datos. Dentro de esta tesis doctoral, se realizó la estimación y análisis a través de diferentes enfoques estadísticos con el objetivo de desentrañar el control genético del hospedador sobre la microbiota en ganado lechero. Además, se analizó el rasgo de concentración de metano como un fenotipo potencial para ser incluido en el programa de mejora de ganado lechero español. Un total de 437 vacas de 14 granjas ganaderas comerciales de ganado lechero de la región Norte de España (País Vasco, Cantabria, Navarra y Gerona) participaron en el proyecto METALGEN, que se articuló en tres subproyectos. Esta tesis se desarrolló dentro del subproyecto uno (La base (meta)-genómica del binomio animal – microbioma, y su relación con la eficiencia alimentaria y emisiones de metano en la raza Holstein-Frisona) del proyecto METALGEN, liderado por INIA. Los animales fueron genotipados por CONAFE usando los chips de baja densidad EuroG 10K y EuroG LD 12k (Illumina, San Diego, California, EE. UU.), Y luego imputados al chip Bovine 50k SNP (Illumina, San Diego, California, EE. UU.) usando su población de referencia para obtener genotipos que contienen 54.609 SNP. La composición microbiana de cada vaca se obtuvo a partir de la secuenciación completa del metagenoma de muestras de contenido ruminal utilizando un dispositivo MinION de Oxford Nanopore Technologies. La concentración de metano se midió con el monitor de gas infrarrojo Guardian® NG de Edinburgh Sensors durante las visitas de las vacas al sistema automatizado de ordeño. El metano fue semanalmente promediado a partir de los picos de cada vaca durante un periodo de 2 a 3

semanas. Concentración (ppm CH_4) e intensidad (ppm CH4/ kg leche) fueron utilizados como fenotipos.

Se estimaron factores de riesgo para concentración de metano e intensidad de metano utilizando variables fenotípicas de producción de leche (kg), grasa láctea (%) proteína láctea (%), carácteres de conformación; escalas GEBVs para concentración de metano e intensidad de metano previamente generadas a partir de un ssGBLUP; así como la transformación logarítmica centrada (CLR) de la composición de la microbiota a nivel de filo (n=86) y género (n=1240), como factores de riesgo potenciales. Las vacas fueron clasificadas en cuartiles (baja, media-baja, media-alta y alta), según registros individuales de las mediciones de metano y la producción de leche promedio en el caso de intensidad de metano. Un abordaje de modelo umbral, bajo un marco Bayesiano de Cadenas de Markov Monte Carlo (MCMC), fue utilizado para determinar factores de riesgo para ser clasificada como alta para ambos carácteres. Las vacas de alta producción tuvieron menor probabilidad de ser clasificadas en la categoría alta para intensidad de metano. Los caracteres de estructura y capacidad corporal fueron positivamente correlacionados a una probabilidad incrementada de estar en la categoría más alta de concentración de metano. Mayores GEBVs también fueron factores protectores para ambos caracteres, reduciendo la probabilidad de ser clasificadas como altas emisoras de metano. Mayor abundancia relativa de la mayoría de los eucariotas (principalmente protozoos ciliados y hongos) y algunas arqueas (*Methanobrevibacter spp*. *Methanothermus spp. y Methanosphaera spp*.) fueron factores de riesgo para ser clasificadas en la categoría alta.

El abordaje de factores de riesgo, utilizando GEBVs, permitió asociar el efecto genético del hospedador sobre los fenotipos, así como el efecto de la microbiota sobre los fenotipos; sin embargo, careció de información con relación al efecto genético del hospedador sobre el microbioma. Con el objetivo de abordar esta limitación, se propuso un conjunto de modelos de ecuaciones estructurales (SEM) de tipo recursivo dentro de un marco de Cadenas de Markov Monte Carlo (MCMC) para analizar conjuntamente la relación hospedador-metagenoma-fenotipo. Se estableció un modelo bivariado no-recursivo como punto de referencia. La abundancia relativa de microorganismos ruminales (RA) y concentración de CH_4 fueron incluidos como fenotipos en ambos sets de modelos mixtos animales bivariados (recursivos y no-recursivos). La genética del hospedador fue incluida

dentro de los modelos a través de la matriz de relaciones genómicas entre animales, permitiendo estimar componentes de varianza. La heredabilidad de CH_4 se estimó en 0,12 ± 0,01 en ambos modelos, recursivo y no recursivo. Asimismo, las estimaciones de heredabilidad para la abundancia relativa de los taxones se superpusieron entre los modelos y variaron entre 0.08 y 0.48. Las correlaciones genéticas entre la composición microbiana y el CH_4 variaron de -0,76 a 0,65 en el modelo bivariado no recursivo y de -0,68 a 0,69 en el modelo recursivo. Independientemente del modelo estadístico utilizado, se estimaron consistentemente correlaciones genéticas positivas con metano para los 7 géneros pertenecientes al filo Ciliophora, así como para los géneros pertenecientes a los filos Euryarchaeota (*Methanobrevibacter sp.*), Chytridiomycota (*Neocallimastix sp.*) y Fibrobacteres (*Fibrobacter sp.*).

El análisis de taxones únicos informó acerca de relaciones independientes entre el fenotipo y un microorganismo a la vez, así como lo concerniente al efecto genético del hospedador sobre un microorganismo dado. Sin embargo, careció de información exhaustiva acerca del efecto de toda la microbiota sobre el fenotipo y el efecto genético sobre el microbioma completo. Para abordar el efecto conjunto de toda la microbiota sobre el fenotipo, la microbiota debe tomar una forma conformable para ser incluida dentro de modelos animales genéticos. Para esto, doce matrices de relación de microbiota (K) fueron construidas a partir de diferentes métricas de distancia del microbioma, con el objetivo de comparar su desempeño dentro de un marco de estimación de componentes de varianza para CH_4 y toda la microbiota. Análisis de simulación (n = 1000) y datos reales fueron desarrollados considerando cuatro modelos posibles: un modelo genómico aditivo (GBLUP), un modelo de microbioma (MBLUP), un modelo de efectos genéticos y microbioma (HBLUP) y un modelo de efectos de interacción genético, microbioma y genético × microbioma (HiBLUP). Todos los modelos se implementaron dentro de un marco Bayesiano utilizando el paquete BGLR en R. Las (Ks) de escalado multidimensional (MDS), el análisis de redundancia (RDA) y el análisis de correspondencia restringida (CCA) funcionaron mejor en la simulación para estimar la heredabilidad y la microbiabilidad. Los mismos métodos para construir Ks estuvieron entre los modelos más plausibles en los datos reales, de acuerdo con el criterio de información de desviación (DIC). El DIC también fue utilizado para obtener el modelo más plausible, que fue el H_iBLUP. Un nuevo término "*Holobiabilidad*" fue

definido para referirse a la proporción de la varianza atribuible a los efectos del holobionte hospedador-microbioma. Las estimaciones a partir de datos reales usando H_iBLUP variaron dependiendo de la K utilizada y estuvieron entre 0.15-0.17, 0.15-0.21 y 0.42-0.59 para heredabilidad, microbiabilidad y holobiabilidad, respectivamente

A pesar de que el uso de Ks permitió estimar el efecto del microbioma completo sobre el fenotipo, faltó información concerniente al efecto genético del hospedador sobre el microbioma completo. Para lidiar con este problema, el microbioma completo no solo necesitaba ser conformable, como en forma de K, sino que necesitaba estar agregado en vectores, para poder ser tratados como un fenotipo. El conjunto de datos de microbioma fue agregado a través de análisis de componentes principales (PCA), en pocos componentes principales (PCs) que fueron utilizados como aproximaciones del metagenoma central. El PCA permitió condensar la enorme y difusa información taxonómica y funcional del metagenoma en unos pocos PC. Se aplicaron modelos animales bivariados utilizando estos PC y la producción de metano como fenotipos. Parte de la variabilidad condensada en estos PC está controlada por el genoma de la vaca, con estimaciones de heredabilidad para el primer PC (PC1) de ~ 0,30 en todos los niveles taxonómicos, con una gran probabilidad (> 83%) de que la distribución posterior sea > 0,20 y con un intervalo de mayor densidad posterior al 95% (95% HPD) no conteniendo cero. La mayoría de las estimaciones de correlación genética entre PC1 y metano fueron grandes (≥0,70) en todos los niveles taxonómicos, con la mayor parte de la distribución posterior (> 82%) siendo > 0,50 y con su 95% HPD no conteniendo cero.

Estos resultados sugieren que todo el metagenoma del rumen regula recursivamente las emisiones de metano en las vacas lecheras, y que tanto el CH_4 como las composiciones de la microbiota están parcialmente controladas por el genotipo del hospedador. El CH_4 fue positivamente asociado con la abundancia relativa (RA) de eucariotas (protozoos y hongos) en Filo, Clase, Orden, Familia y Genero. Las lecturas largas con nanoporos permitieron la caracterización del metagenoma central del rumen usando secuenciación del metagenoma completo, y las variables agregadas (PC) propuestas podrían ser usadas en programas de mejora de animales para reducir las emisiones de metano en las generaciones futuras.

RESUM

L'anàlisi del control genètic de l'hoste sobre la seva microbiota s'ha assenyalat recentment com un tema prometedor en diferents camps d'estudi. La relació entre el holobiont hoste-microbioma i els fenotips en bovins de llet podria conduir a nous coneixements en els programes de cria. L'anàlisi de la microbiota implica tractar amb marcs de dades compositius escassos, entre altres qüestions relacionades amb la inferència sobre la causalitat dels microbis sobre els trets d'interès, l'efecte microbiòtic únic vs l'efecte microbiota sencer sobre els caràcters rellevants o l'agregació de tota la microbiota per abordar la implementació en programes de selecció de cria. Les tècniques òptimes per analitzar aquest tipus de dades encara no estan ben definides. Per tant, s'han d'utilitzar diferents enfocaments estadístics per tractar millor aquest tipus de dades. Dins d'aquest doctorat es van realitzar tesis, estimacions i anàlisis mitjançant diferents enfocaments estadístics amb l'objectiu de desentranyar el control genètic de l'hoste sobre la microbiota en bestiar lleter. A més, es va analitzar el tret de concentració de metà com a fenotip potencial a incloure en el programa espanyol de cria de bestiar lleter. Un total de 437 vaques de 14 explotacions comercials de boví lleter de la regió nord d'Espanya (País Basc, Cantàbria, Navarra i Girona) van participar en el projecte METALGEN, articulat en tres subprojectes. Aquesta tesi es va desenvolupar dins del subprojecte (La base (meta) genòmica de la bionomia microbioma animal i la seva relació amb l'eficiència alimentària i les emissions de metà de la raça Holstein-Friesian) del projecte METALGEN, liderat per INIA. Els animals van ser genotipats per CONAFE mitjançant xips de baixa densitat EuroG 10K i EuroG LD 12k (Illumina, San Diego, Califòrnia, EUA), i després van ser imputats al xip SNP boví 50k (Illumina, San Diego, Califòrnia, EUA) mitjançant el seu població de referència per obtenir genotips que contenen 54.609 SNP. La composició microbiana de cada vaca es va obtenir a partir de la seqüenciació de metagenomes sencers de mostres de contingut ruminal mitjançant un dispositiu MinION d'Oxford Nanopore Technologies. La concentració de metà es va mesurar amb el monitor de gas infraroig Guardian® NG de Edinburgh Sensors durant les visites de les vaques al sistema automatitzat de munyir. El metà es feia una mitjana setmanal a partir dels pics de cada vaca durant un període de 2 a 3 setmanes. La concentració (ppm CH4) i la intensitat (ppm CH4 / kg de llet) es van utilitzar com a fenotips

Els factors de risc per a la concentració de metà i la intensitat del metà es van estimar utilitzant variables fenotípiques de rendiment de la llet (kg), greix de la llet (%), proteïna de la llet (%), trets de conformació; concentracions de metà i intensitat de metà. Escales de GEBV generades prèviament a partir d'un ssGBLUP; així com la composició del microbioma centrat en la relació logarítmica (CLR) centrada a nivell de fil (n = 86) i gènere (n = 1240), com a possibles factors de risc. Les vaques es van classificar en quartils (baixa, mitjana-baixa, mitjana-alta i alta), segons registres individuals de mesures de metà, i el rendiment mitjà de llet en el cas de la intensitat del metà. Es va utilitzar un enfocament del model llindar, sota un marc bayesià de Monte Carlo (MCMC) de la cadena Markov, per determinar els factors de risc per classificar-se com a elevats per a tots dos trets. Les vaques amb alt rendiment van tenir una menor probabilitat de ser classificades en la categoria alta per intensitat de metà. Els trets de l'estructura corporal i de la capacitat es van correlacionar positivament amb l'augment de la probabilitat de situar-se en la categoria més alta de concentració de metà. Els GEBV més grans també van ser factors de protecció per a ambdós trets, reduint la probabilitat de ser classificats com a elevats emissors de metà. La major abundància relativa de la majoria dels eucariotes (principalment protozous i fongs ciliats) i algunes arquees (Methanobrevibacter spp. Methanothermus spp i Methanosphera spp.) Van ser factors de risc per classificar-se en les categories altes.

L'enfocament del factor de risc, mitjançant GEBV, va permetre associar l'efecte genètic de l'hoste sobre els fenotips, així com l'efecte microbioma sobre els fenotips; Tanmateix, va mancar d'informació sobre l'efecte genètic de l'hoste sobre el microbioma. Per tal d'afrontar aquesta limitació, es va proposar un conjunt de models d'equacions estructurals (SEM) de tipus recursiu dins d'un marc de cadena Markov Monte Carlo (MCMC) per analitzar conjuntament la relació hoste-metagenoma-fenotip. Es van establir models no recursius com a referència. L'abundància relativa de microbis rumen sols (RA) i concentració de CH4 es van incloure com a fenotips en ambdós conjunts de models animals mixtos bivariables (recursius i no recursius). La genètica de l'hoste es va incloure als models a través de la matriu de relacions genòmiques entre animals, permetent estimar els components de la variància. L'heretabilitat del CH4 es va estimar en 0,12 ± 0,01 en ambdós models, recursius i no recursius. De la mateixa manera, les estimacions d'heretabilitat de l'abundància relativa dels tàxons es van superposar entre models i van oscil·lar entre 0,08 i 0,48. Les correlacions

genètiques entre la composició microbiana i el CH4 van oscil·lar entre -0,76 i 0,65 en els models bivariables no recursius i de -0,68 a 0,69 en els models recursius. Independentment del model estadístic utilitzat, les correlacions genètiques positives amb el metà s'estimaren constantment per als 7 gèneres pertanyents al fil Ciliophora, així com per als gèneres pertanyents a l'Euryarchaeota (Methanobrevibacter sp.), Chytridiomycota (Neocallimastix sp.) I Fibrobacteres (Fibrobacter sp.) Filus.

L'anàlisi de tàxons individuals va informar sobre les relacions independents entre el fenotip i un microbi a la vegada, així com sobre l'efecte genètic de l'hoste sobre un determinat microorganisme; Tot i això, no té informació completa sobre l'efecte microbioma sencer sobre el fenotip i l'efecte genètic de l'hoste sobre tot el microbioma. Per fer front a tot l'efecte microbioma sobre el fenotip, el microbioma ha de tenir una forma compatible per incloure'l en models animals genètics. Per a això, es van construir dotze matrius de relació de microbiota (K) de diferents mètriques de distància de microbiomes, amb l'objectiu de comparar el seu rendiment dins d'un marc d'estimació de components de variància per CH4 i anàlisi de microbiomes sencers en simulació (n = 1000, 25 rèpliques) i es van realitzar dades reals , considerant quatre possibles models: un model genòmic additiu (GBLUP), un model de microbioma (MBLUP), un model d'efectes genètics i microbiomes (HBLUP) i un model d'efectes d'interacció genètics, microbiomes i genètics × microbiomes (HiBLUP). Tots els models es van implementar dins d'un marc bayesià mitjançant el paquet BGLR de R. Els Ks construïts a partir de l'escala multidimensional (MDS), l'anàlisi de redundància (RDA) i l'anàlisi de correspondència restringida (CCA) van tenir un millor rendiment en simulació per estimar l'herència i la microbilitat. Els mateixos mètodes per construir Ks estaven entre els més plausibles en dades reals, segons els criteris d'informació sobre desviacions (DIC). El DIC també es va utilitzar per obtenir el model més plausible, que va passar a ser el HiBLUP. Es va definir un nou terme "Holobiabilitat" per referir-se a la proporció de la variància fenotípica atribuïble als efectes holobiont del microbioma host. Les estimacions de dades reals mitjançant HiBLUP van variar en funció de la K utilitzada i van oscil·lar entre 0,15-0,17, 0,15-0,21 i 0,42-0,59 per heretabilitat, microbiabilitat i holobiabilitat, respectivament.

Tot i l'ús de Ks permès estimar l'efecte del microbioma sencer sobre el fenotip, encara hi havia manca d'informació sobre l'efecte genètic de l'hoste sobre el microbioma sencer. Per

tractar aquest problema, tot el microbioma no només necessitava ser conformable, com en forma K, sinó que calia agregar-lo en vectors per tractar-lo com un fenotip. El conjunt de dades de microbiomes es va agregar mitjançant l'anàlisi de components principals (PCA) en pocs components principals (PC) que es van utilitzar com a proxies del metagenoma principal. El PCA va permetre condensar la enorme i difusa informació taxonòmica i funcional del metagenoma en pocs PC. Es van aplicar models animals bivariants utilitzant aquests PCs i la producció de metà com a fenotips. Part de la variabilitat condensada en aquestes PC està controlada pel genoma de la vaca, amb estimacions d'heretabilitat per a la primera PC (PC1) de ~ 0,30 a tots els nivells taxonòmics, amb una gran probabilitat (> 83%) de la distribució posterior> 0,20 i amb un 95% més alt interval de densitat posterior (95% HPD) que no conté zero. La majoria de les estimacions de correlació genètica entre PC1 i metà eren grans (≥0,70) en tots els nivells taxonòmics, amb una gran part de la distribució posterior (> 82%)> 0,50 i amb un 95% de HPD que no contenia zero.

Aquests resultats suggereixen que tot el metagenoma del rumen regula recursivament les emissions de metà en vaques lleteres i que tant el CH4 com les composicions de microbiota estan parcialment controlades pel genotip de l'hoste. L'enfocament del factor de risc va contribuir a entendre l'efecte del microbioma i la genètica de l'hoste sobre les emissions de metà; Tot i això, no tenia la capacitat de descriure el control exercit per la genètica hoste sobre el microbioma. El CH4 es va associar positivament amb l'abundància relativa (RA) d'eucariotes (protozous i fongs) a Phylum, Class, Order, Family i Genus. Les lectures llargues de Nanopore van permetre caracteritzar el metagenoma del rumen bàsic mitjançant una seqüenciació completa de metagenomes, i les variables agregades proposades (PC) es podrien utilitzar en programes de cria d'animals per reduir les emissions de metà en les generacions futures.

CHAPTER 1

GENERAL INTRODUCTION

Alejandro Saborío-Montero[1,2]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain; [2]Universidad de Costa Rica, Escuela de Zootecnia, 11501 San Pedro de Montes de Oca, San José, Costa Rica.

Methane emissions in ruminants

Ruminants exhale methane (CH_4) as a by-product of fermentation inside their digestive tract. The enteric CH_4 is a colourless and odourless gas produced primarily in the rumen (87%), most of it is emitted by eructation (95%). The remanent CH_4 production (13%) takes place in the large intestines and is mostly evacuated through the lungs (89%), and in a smaller extent (11%) through the anus (Murray et al., 1976). There are two main relevant issues concerning methane emissions from ruminants: A) environmental effect and B) effect on animal feed efficiency.

Methane environmental effect. The livestock industry has been one under recent environmental concern. The main reason is because bovines are the domesticated ruminants that emit the largest amount of methane (Clauss et al., 2020). The global warming potential of methane is 28 times that from carbon dioxide (CO_2). However, this comparison should be nuanced, because it remains less time (12 years) in the atmosphere than CO_2 (100 years) (Myhre et al., 2013). Livestock methane emissions contribute with 8% to 12% of anthropogenic emissions (Eckard et al., 2010; Gerber et al., 2013). The amount of methane emitted by an adult Holstein cow in a daily basis has been previously estimated. Differences between studies remark variation of this phenotype. For instance, Garnsworthy et al., (2012) reported 369 g/d (range 278-456 g/d); Difford et al., (2018) informed of 395.8 g/d (SD=63.5), and Ramayo-Caldas et al., (2019) registered 506 g/d (SD=56). Differences among studies regarding methane concentration (ppm) have also been previously reported (Wu et al., 2018; Huhtanen et al., 2015).

Methane effect on feed efficiency. Methane operates as the most important sink of hydrogen (H_2) inside the rumen (Ungerfeld, 2020). Therefore, inhibiting the rumen methanogenesis could redirect H_2 in the direction of fermentation products nutritionally valuable for the animal, enhancing the cow feed efficiency. There is not a universal definition of feed efficiency, a simple one would be how efficiently animals convert feed into product (Seymour et al., 2020). The energy utilization in ruminants is presumably diminished by the loss generated by methane production pathway, which ranges between 2 and 12% of ingested gross energy (Johnson and Johnson, 1995). Theoretically, feed efficiency is negatively

affected by high rates of methane emissions, which waste energy to the environment rather than use it to generate the final product of the system. Previous studies inhibiting methanogenesis invitro resulted in accumulation of H_2 (Ungerfeld, 2015), while in vivo trails reducing methanogenesis also increased H_2 (Ungerfeld, 2018).

Holobiont control of methane emissions

The host-microbiota system is referred as "holobiont". The introduction of the "holobiont" term has been attributed to Lynn Margulis in 1991 (Margulis, 1991), and extended to describe a host and its associated communities of microorganisms (Simon et al., 2019). Nevertheless, there is controversy regarding the introduction of this concept, ascribing it to Adolf Meyer-Abich nearly 50 years before (Baedke et al., 2020). Ruminants are animals especially dependent on their microbiota because their source of energy comes from volatile fatty acids generated during enteric fermentation. Microbiota composition at the same time is presumably controlled by the specific characteristics of the host. Generating a self-regulated two ways feedback. Methane emissions from ruminants are dependent on both, host, and microbiome, therefore regulated by the holobiont.

Host phenotypic control of methane emissions. Concerning methane emissions, several phenotypic traits of the host have been associated to methane production. The rumen size has been hypothesized as a methane associated trait in sheep, with smaller size and faster turnover rate favoring some bacteria abundance, whose fermentative processes result in reduction of H_2, which consequently reduces the substrate for hydrogenotrophic methane formation (Kamke et al., 2016). Other studies in sheep support the association between smaller rumen and lower methane emissions (Goopy et al., 2014) and between reduction in mean retention time of digesta in the whole tract and lower methane emissions (Barnett et al., 2012). In large ruminants, association between rumen size or mean retention time and methane emissions is limited (Negussie et al., 2017). Body weight and linear conformation are related traits (Yan et al., 2009) which had been associated to enteric methane emissions in dairy and beef cattle (Moraes et al., 2014). In general there is a positive correlation between body weight and methane production (Moraes et al., 2014). Linear conformation traits had been associated to methane emissions (López-Paredes et al., 2020), a possible reason for this could be that some

conformation traits (*i.e.* stature, chest wide, body depth, angularity) might indicate larger rumen capacity, which in turn involves slower rumen passage rates of digesta, increasing production of methane. Other phenotypic traits like the rumen epithelial cell wall had been associated to methane production (Xiang et al., 2016), suggesting that the layers of the rumen wall respond adapting specifically to diet, and therefore influencing CH_4 production.

Microbiota control of methane emissions. Rumen microbiota is essential for ruminants. All three domains of life, Bacteria, Archaea and Eukaryote (Woese et al., 1990), also known in biological taxonomy as superkingdom, are present in the rumen (Kittelmann et al., 2013). Virus clades, which are not included in the three-domain system, can be also found in the rumen (Gilbert et al., 2020). In dairy cattle, the rumen exhibits a consistent and highly conserved abundance rank structure of microbiome across geographical locations, breeds and diets (Wallace et al., 2019). Crucial functions such as forage fiber fermentation, volatile fatty acid production or pH homeostasis are possible thanks to ruminal microorganisms.

Differences between microbiota of cows have been consistently evidenced (Mizrahi and Jami, 2018). There are also evidences of microbial community resistance to perturbation and capacity to recover from it (resilience), which provides stability of the digestive function for the host across a range of feeding and management conditions (Weimer, 2015). Most microorganisms in the rumen of adult cows are bacteria, the most abundant phylum being *Bacteroidetes* and *Firmicutes* (Chaucheyras-Durand and Ossa, 2014).

In ruminants, the enteric methane is produced by methanogenic microorganisms inside their digestive tract. All known methanogens belong to the archaea superkingdom (Lang et al., 2015). Archaeas that produce methane using hydrogen (H_2) as a reducing agent and carbon dioxide (CO_2) are called hydrogenotrophic methanogens, the chemical reaction for this pathway can be describes as:

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$$

There are other methanogenic archaeas that produce methane using different substrates. For instance, acetoclastic archaeas use acetate (CH_3COOH) through the following reaction:

$$CH_3COOH \rightarrow CO_2 + CH_4$$

While methylotrophic archaeas use methylated compounds (Vanwonterghem et al., 2016). Methanol (CH_3OH) is an example of a methylated substrate, as described Yin et al., (2019) with the following formula:

$$4CH_3OH \rightarrow 3CH_4 + CO_2 + 2H_20$$

The hydrogenotrophic pathway certainly occur in the rumen, acetoclastic and methylotrophic pathways might be also present in the rumen, although neither pathways had been experimentally demonstrated yet (Qiao et al., 2014). There are evidences of rumen methanogens using formate, acetate, methyl compounds and ethanol as substrates, but usually jointly with hydrogen (Greening et al., 2019).

Methanogens have been conventionally classified into seven orders: Methanobacteriales, Methanococcales, Methanomicrobiales, Methanosarcinales, Methanocellales, Methanopyrales and Methanomassiliicoccales (Vanwonterghem et al., 2016). Most of them make use of the hydrogenotrophic pathway, while there are evidences suggesting the presence of the acetoclastic and methylotrophic pathways in Methanosarcinales and Methanomassiliicoccales, respectively (Lang et al., 2015).

Ruminants exhale methane, which is originated as an end-product from the metabolism of methanogenic archaeas that inhabit inside their gastrointestinal tract. The usage of H_2 by hydrogenotrophic archaeas could be important for the host, due to its contribution over the control of intraruminal pressure exerted by H_2, which at high concentration inhibits rumen fermentation (Morgavi et al., 2010).

A plethora of microorganisms have been associated to complex traits of interest in dairy cattle (Schären et al., 2018). The concept of microbiota variability between individuals and the partial host microbiota stability, in addition to the association of microbiota with certain complex traits, give rise to an interesting hypothesis: if the microbiota is heritable at some extent, it could generate the opportunity to select individuals with optimal microbiotas, genetically correlated traits associated to optimal microbiotas would pass to the next

generations due to its partial genetic control, which could modulate those genetically correlated traits for specific breeding objectives.

Hologenome

The hologenome is the combination of the host genome and its associated microbial genomes (Rosenberg et al., 2007). It can also be defined as the collective genomes of a holobiont (Simon et al., 2019). Analogously, as the holobiont controls methane emissions from ruminants, there is presumably a partial regulation of methane emissions by the hologenome, which in this case could be partitioned into cow genome and metagenome.

Cow genome. Heritability is the term used to describe the proportion of the phenotypic variance attributable to the additive genetic effect in a population (Templeton, 2006). This genetic parameter is used in animal breeding to predict response to selection, which is the difference in the phenotypic mean between a population and its progeny. Heritability values range between 0 and 1, with higher values causing a faster response in the targeted phenotypes under selection (Oldenbroek and van der Waaij, 2014). Methane emissions is a heritable trait. Lassen et al. (2012) estimated heritability (±SE) of methane emissions (g/day) from 1745 Holstein cows in 0.21±0.06. Lassen and Løvendahl (2016) obtained larger but less accurate estimates 0.25±0.16 from a smaller sample (n=339) of cows of the same breed. Similar heritability estimates were informed by Manzanilla-Pech et al., (2016) 0.23±0.23 (n=205); Pszczola et al., (2017) 0.27±0.09 (n=485) and (Difford et al., 2018) 0.21±0.09 (n=750). A more recent study (Breider et al., 2019) informed of heritability estimates for methane emissions (g/day) from Holstein cows of 0.45±0.11. Other studies estimating heritabilities for methane emissions concentration (ppm/day) reported values of 0.11±0.02 (n=355) (van Engelen et al., 2018) and 0.21±0.11 (n=434) (Difford et al., 2019). In general, average estimates of heritability for methane emissions ranged from 0.05 to 0.45 regardless of the measurement units (g/day, mg/kg, ppm/day) (Lassen and Difford, 2020). If methanogenic archaea are the only rumen producers of methane (Hook et al., 2010) and methane is a heritable trait, it is reasonable to hypothesizes that the composition of methanogenic archaea and associated microbiome regulating methane emissions is also heritable.

Metagenome. Some authors (Huws et al., 2018) confer the first postulate of the host influence over its microbiome to Weimer et al., (2010), who found that cows restored their bacterial composition after near total exchange of the rumen content. In this study, cows also returned volatile fatty acid concentration and pH back to pre-exchange values. In a more recent study of near-total rumen content exchange, the hosts capacity to return the rumen bacterial composition to the original status was further evidenced (Weimer et al., 2017). Other investigations have evidenced the control that the genetic background of the animal exerts over its ruminal microbiota composition: Roehe et al., (2016) assessed host genetic control using sire progeny groups, they found consistent ranking of the sire progeny groups (overall and within diet) for relative abundance of archaea and methane emissions, proposing a genetic regulation of the host over these traits. Difford et al. (2018) found heritabilities for relative abundance of rumen bacteria OTU and rumen archaea OTU reaching 0.4 and 0.3, respectively. Li et al. (2019) reported heritability estimates ≥ 0.15 for rumen bacteria in beef cattle, concluding that some rumen microbial features could be influenced by host genetics and suggested the potential to modulate the heritable microbiota through genetic selection and breeding. Several studies reported high heritabilities estimates for ruminal microorganisms, for instance: Sasson et al., (2017) reported heritability estimates larger than 0.7 for 22 bacterial Operational Taxonomic Units (OTU) in dairy cattle. Wallace et al. (2019) showed heritability estimates for family (Lachnospiraceae) and genus (*Prevotella sp.*) larger than 0.6 and 0.4, respectively in 1016 lactating dairy cows of the Holstein and Nordic Red breed from 4 countries (United Kingdom, Italy, Sweden and Finland). Heritability estimates from different populations should not be compared directly, due to the frequency dependent nature of linear models in variance component estimation (Templeton, 2006). Therefore, heritability estimates should be obtained from the population of interest, in order to increase the reliability of this genetic parameter.

Genetic correlations. The genetic association between traits is frequently expressed through the genetic correlation parameter, which ranges between -1 and 1, and is defined as the correlation between the additive effects (Hazel, 1943). Genetic correlations between specific microorganisms and complex traits of interest had been estimated showing correlations

different from zero (Aliakbari et al., 2021). The influence of the microbiome on relevant complex traits in cattle has been well established (Wallace et al., 2019). Nevertheless, there is a plethora of challenges to determine the most appropriate characteristics of a host to pass on future generations, in order to generate a given microbiota on their offspring, to enhance the desired performance of animals. Other challenges rise on determining the collateral effects of these microbiota composition changes on correlated traits. The estimation of genetic correlations between microbiota and traits of interest, as methane emissions in dairy cattle, should be taken as a way to orientate future decisions; however, selecting by lower methane emission could affect other correlated traits, such as milk fatty acids, that had been positively correlated to methane emissions (Bougouin et al., 2019).

Compositional nature of microbiome

Data structured as proportions or with a constant or irrelevant sum, are designated as compositional data (Gloor et al., 2017). Compositional data contains information regarding the relationships between the parts (Aitchison, 1986). Sequencing processes of microbiome samples computed as relative abundances confer to microbiome data sets the nature of being compositional. These characteristic of the microbiome datasets poses the challenge of the statistical approach to analyze compositional data, besides its sparsity and complex interrelationships. To deal with compositionality, data can be transformed from the relative abundance of taxa to the centered log-ratios (CLR) between taxa (Gloor et al., 2017). However, there is the inconvenient of zero inflated microbiome data sets obtained from sequencing, limiting the computation of logarithms. To fill this gap, imputation of count zeros from compositional data has been purposed (Martín-Fernández et al., 2015). Assigning weights to taxa according to the relative abundance in CLR transformation could be misleading. The complexity of synergic or antagonist interrelations between microorganisms is still under study and might be more cautious to approach each taxon equally. In other words, a low relative abundance of a given microorganism could have a huge effect on other.

Causality

Studies addressing the effect of microbiome on phenotypes of interest have recently boosted (Huws et al., 2018); However, the extent whether those associations are causal rather

than casual have not been clarified yet (Newbold and Ramos-Morales, 2020). There are specific cases in which causation could be inferred based on biological facts. For instance: archaeas are the unique microorganism capable to carry out biological methane production (Enzmann et al., 2018; Lyu et al., 2018). All methane originated in a given ecological niche should be generated from archaeas. This biological phenomenon offers the opportunity to assess causality from the archaea-methane relationship. Concerning the enteric methane produced inside the gastro-intestinal tract of ruminants, an unidirectional effect (known as recursive effect) between traits is expected, in which relative abundance of archaea recursively cause methane, without any feedback from methane affecting archaea. A bidirectional simultaneous effect could be feasible if methane does not leave the rumen, causing a negative feedback over archaea methane production. However, methane does leave rumen, turning this latter scenario into a biological unplausible pathway direction. Specific statistical models could be valuable to infer causality from hypothesized models in this particular case or in similar cases.

Single vs whole approaches

Metagenomic studies are frequently focus on independent effect of singular taxa over phenotypic complex traits. Besides, rumen microbiota is frequently expressed in microbiome studies as the relative abundance of each taxonomically classified microorganism. Therefore, effects and heritabilities estimates of rumen microbiota are independently expressed for the relative abundance of single taxonomic classifications (Difford et al., 2018; Wallace et al., 2019). Nevertheless, it is necessary to incorporate the whole microbiome into statistical models to assess its entire association with complex traits. Host genetics and microbiome has been approached so far as univariate relationships instead of as host-metagenome holobiont; however, a more realistic approach would be to analyze the holobiont as whole, rather than as independent organisms. Aggregation of microbiome into microbiota relationship matrices could be feasible, but the performance of those matrices should be assessed, in terms of plausibility and ability to capture phenotypic variance.

Dimensionality reduction

Microbiome data analysis requires handling big data. Metagenomic data are usually composed by high dimension matrices, which in turn requires costume made statistical management. Metagenomic data implies complex relationship between microorganisms in continuous change. Those relationships are attempted to be deciphered from large sparse compositional data obtained from sequencing procedures. The best statistical approaches to analyze this kind of data are still unclear and remain under study. Multivariate methods could be useful to deal with metagenomic data. Principal Component Analysis (PCA) is a multivariate technic used to reduce dimensionality and condense most of the variance from interrelated variables. It transform the original data set into uncorrelated subjacent variables, the principal components (PCs), which are ordered so that the first few retain most of the variation present in all the original variables (Jolliffe, 2002). In PCA, the number of extracted PCs is equal or lower to the number of original variables; commonly, the number of PCs to be kept can be selected by: 1) The variance explained by the PCs, 2) The number of PCs with eigenvalues higher than 1, or 3) The interpretation of the PCs structure (subjacent variables with biological meaning). In PCA, the algorithm generates a linear function of the elements of the vector of random variables having maximum variance (PC1), next it finds the linear function uncorrelated to PC1 having maximum variance (PC2) an so on. If a set of variables are substantially correlated between each other, then the first few PCs will account for most of the variance from the original variables (Jolliffe, 2002). Metagenomic variables are usually correlated, which make PCA suitable to summarize the data of taxa into PCs. Specific coordinates from the PCA of microbiome taxa can be assigned to each animal, it allows shrinking huge data sets to few, even unique, synthetic subjacent variables (PCs), which become an opportunity to constrain metagenome into a single vector. If these coordinates from the PCA are heritable and genetically correlated to some complex traits, the PCA coordinates could be used as phenotypes in multi-trait variance components estimation models, or in genomic best linear unbiased prediction (GBLUP) animal models, in order to perform metagenomic predictions of correlated traits of interest in animal breeding programs.

Implementation

A trait can be included into an animal breeding program if the trait has an heritable component or is genetically correlated with another heritable trait. However, the microbiome is a difficult trait to measure. If the microbiome is to be included into animal breeding programs, a reference population containing microbiota information, as well as phenotypic and genotypic information, is to be generated in order to establish associations between microbiome, phenotype and genotype. One of the cumbersome issues of including microbiome in breeding programs is the sparsity and large dimension of microbiota information, which possess the challenge of aggregate microbiota into few variables in order to be able to use it in animal breeding programs. The approach to condense the whole microbiome information into a phenotype, capturing as much as possible of variance from the original variables might be the way of including microbiome as a phenotype, However, the heritability of these aggregated variables and its genetic correlations with traits of interest need to be evaluated.

REFERENCES

Aitchison, J. 1986. The Statistical Analysis of Compositional Data. Chapman and Hall.

Aliakbari, A., O. Zemb, Y. Billon, C. Barilly, I. Ahn, J. Riquet, and H. Gilbert. 2021. Genetic relationships between feed efficiency and gut microbiome in pig lines selected for residual feed intake. J. Anim. Breed. Genet. jbg.12539. doi:10.1111/jbg.12539.

Baedke, J., A. Fábregas-Tejeda, and A. Nieves Delgado. 2020. The holobiont concept before Margulis. J. Exp. Zool. Part B Mol. Dev. Evol. 334:149–155. doi:10.1002/jez.b.22931.

Barnett, M.C., J.P. Goopy, J.R. McFarlane, I.R. Godwin, J. V. Nolan, and R.S. Hegarty. 2012. Triiodothyronine influences digesta kinetics and methane yield in sheep. Anim. Prod. Sci. 52:572. doi:10.1071/AN11303.

Breider, I.S., E. Wall, and P.C. Garnsworthy. 2019. Short communication: Heritability of methane production and genetic correlations with milk yield and body weight in Holstein-Friesian dairy cows. J. Dairy Sci. 102:7277–7281. doi:10.3168/jds.2018-15909.

Chaucheyras-Durand, F., and F. Ossa. 2014. The rumen microbiome: Composition, abundance, diversity, and new investigative tools. Prof. Anim. Sci. 30:1–12. doi:10.15232/S1080-7446(15)30076-0.

Clauss, M., M.T. Dittmann, C. Vendl, K.B. Hagen, S. Frei, S. Ortmann, D.W.H. Müller, S. Hammer, A.J. Munn, A. Schwarm, and M. Kreuzer. 2020. Review: Comparative methane production in mammalian herbivores. Pages S113–S123 in Animal. Cambridge University Press.

Difford, G.F., D.W. Olijhoek, A.L.F. Hellwing, P. Lund, M.A. Bjerring, Y. de Haas, J. Lassen, and P. Løvendahl. 2019. Ranking cows' methane emissions under commercial conditions with sniffers versus respiration chambers. Acta Agric. Scand. A Anim. Sci. 68:25–32. doi:10.1080/09064702.2019.1572784.

Difford, G.F., D.R. Plichta, P. Løvendahl, J. Lassen, S.J. Noel, O. Højberg, A.-D.G. Wright, Z. Zhu, L. Kristensen, H.B. Nielsen, B. Guldbrandtsen, and G. Sahana. 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLOS Genet. 14:e1007580. doi:10.1371/journal.pgen.1007580.

Eckard, R.J., C. Grainger, and C.A.M. de Klein. 2010. Options for the abatement of methane and nitrous oxide from ruminant production: A review. Livest. Sci. 130:47–56. doi:10.1016/j.livsci.2010.02.010.

van Engelen, S., H. Bovenhuis, P.P.J. van der Tol, and M.H.P.W. Visker. 2018. Genetic background of methane emission by Dutch Holstein Friesian cows measured with infrared sensors in automatic milking systems. J. Dairy Sci. 101:2226–2234. doi:10.3168/jds.2017-13441.

Enzmann, F., F. Mayer, M. Rother, and D. Holtmann. 2018. Methanogens: biochemical

background and biotechnological applications. AMB Express 8:1. doi:10.1186/s13568-017-0531-x.

Garnsworthy, P., J. Craigon, J. Hernandez-Medrano, and N. Saunders. 2012. Variation among individual dairy cows in methane measurements made on farm during milking. J. Dairy Sci. 95:3181–3189. doi:10.3168/jds.2011-4606.

Gerber, PJ, Steinfeld, Henderson, Mottet, Opio, Dijkman, Falcucci, and Tempio. 2013. Tackling Climate Change Through Livestock. A Global Assessment of Emissions and Mitigation Opportunities. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Gloor, G.B., J.M. Macklaim, V. Pawlowsky-Glahn, and J.J. Egozcue. 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 8:2224. doi:10.3389/fmicb.2017.02224.

Goopy, J.P., A. Donaldson, R. Hegarty, P.E. Vercoe, F. Haynes, M. Barnett, and V. Hutton Oddy. 2014. Low-methane yield sheep have smaller rumens and shorter rumen retention time. Br. J. Nutr. 111:578–585. doi:10.1017/S0007114513002936.

Greening, C., R. Geier, C. Wang, L.C. Woods, S.E. Morales, M.J. McDonald, R. Rushton-Green, X.C. Morgan, S. Koike, S.C. Leahy, W.J. Kelly, I. Cann, G.T. Attwood, G.M. Cook, and R.I. Mackie. 2019. Diverse hydrogen production and consumption pathways influence methane production in ruminants. ISME J. 13:2617–2632. doi:10.1038/s41396-019-0464-2.

de Haas, Y., J. Windig, M. Calus, J. Dijkstra, M. de Haan, A. Bannink, and R. Veerkamp. 2011. Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. J. Dairy Sci. 94:6122–6134. doi:10.3168/jds.2011-4439.

Hazel, L.N. 1943. The Genetic Basis for Constructing Selection Indexes. Genetics 28:476–490.

Huhtanen, P., E.H. Cabezas-Garcia, S. Utsumi, and S. Zimmerman. 2015. Comparison of methods to determine methane emissions from dairy cows in farm conditions. J. Dairy Sci. 98:3394–3409. doi:10.3168/jds.2014-9118.

Huws, S.A., C.J. Creevey, L.B. Oyama, I. Mizrahi, S.E. Denman, M. Popova, R. Muñoz-Tamayo, E. Forano, S.M. Waters, M. Hess, I. Tapio, H. Smidt, S.J. Krizsan, D.R. Yáñez-Ruiz, A. Belanche, L. Guan, R.J. Gruninger, T.A. McAllister, C.J. Newbold, R. Roehe, R.J. Dewhurst, T.J. Snelling, M. Watson, G. Suen, E.H. Hart, A.H. Kingston-Smith, N.D. Scollan, R.M. do Prado, E.J. Pilau, H.C. Mantovani, G.T. Attwood, J.E. Edwards, N.R. McEwan, S. Morrisson, O.L. Mayorga, C. Elliott, and D.P. Morgavi. 2018. Addressing Global Ruminant Agricultural Challenges Through Understanding the Rumen Microbiome: Past, Present, and Future. Front. Microbiol. 9:1–33. doi:10.3389/fmicb.2018.02161.

Johnson, K.A., and D.E. Johnson. 1995. Methane emissions from cattle. J. Anim. Sci. 73:2483–2492.

Jolliffe, I.T. 2002. Principal Component Analysis, Second Edition. 2nd ed. Springer, Aberdeen.

Kamke, J., S. Kittelmann, P. Soni, Y. Li, M. Tavendale, S. Ganesh, P.H. Janssen, W. Shi, J. Froula, E.M. Rubin, and G.T. Attwood. 2016. Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. Microbiome 4. doi:10.1186/s40168-016-0201-2.

Kittelmann, S., H. Seedorf, W.A. Walters, J.C. Clemente, R. Knight, J.I. Gordon, and P.H. Janssen. 2013. Simultaneous Amplicon Sequencing to Explore Co-Occurrence Patterns of Bacterial, Archaeal and Eukaryotic Microorganisms in Rumen Microbial Communities. PLoS One 8. doi:10.1371/journal.pone.0047879.

Lang, K., J. Schuldes, A. Klingl, A. Poehlein, R. Daniel, and A. Brune. 2015. New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of "Candidatus Methanoplasma termitum". Appl. Environ. Microbiol. 81:1338–1352. doi:10.1128/AEM.03389-14.

Lassen, J., and G.F. Difford. 2020. Review: Genetic and genomic selection as a methane mitigation strategy in dairy cattle. Animal 1–11. doi:10.1017/S1751731120001561.

Lassen, J., and P. Løvendahl. 2016. Heritability estimates for enteric methane emissions from Holstein cattle measured using noninvasive methods. J. Dairy Sci. 99:1959–1967. doi:10.3168/jds.2015-10012.

Lassen, J., P. Løvendahl, and J. Madsen. 2012. Accuracy of noninvasive breath methane measurements using Fourier transform infrared methods on individual cows. J. Dairy Sci. 95:890–898. doi:10.3168/jds.2011-4544.

Li, F., C. Li, Y. Chen, J. Liu, C. Zhang, B. Irving, C. Fitzsimmons, G. Plastow, and L.L. Guan. 2019. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. Microbiome 92:1–17. doi:10.1186/s40168-019-0699-1.

López-Paredes, J., I. Goiri, R. Atxaerandio, A. García-Rodríguez, E. Ugarte, J.A. Jiménez-Montero, R. Alenda, and O. González-Recio. 2020. Mitigation of greenhouse gases in dairy cattle via genetic selection: 1. Genetic parameters of direct methane using noninvasive methods and proxies of methane. J. Dairy Sci. 103:7199–7209. doi:10.3168/jds.2019-17597.

Lyu, Z., N. Shao, T. Akinyemi, and W.B. Whitman. 2018. Methanogenesis. Curr. Biol. 28:R727–R732. doi:10.1016/j.cub.2018.05.021.

Manzanilla-Pech, C.I.V., Y. De Haas, B.J. Hayes, R.F. Veerkamp, M. Khansefid, K.A. Donoghue, P.F. Arthur, and J.E. Pryce. 2016. Genomewide association study of methane emissions in angus beef cattle with validation in dairy cattle. J. Anim. Sci. 94:4151–4166. doi:10.2527/jas.2016-0431.

Margulis, L. 1991. Symbiogenesis and Symbionticism. The MIT Press.

Mizrahi, I., and E. Jami. 2018. Review: The compositional variation of the rumen

microbiome and its effect on host performance and methane emission. Animal 12:220–232. doi:10.1017/S1751731118001957.

Moraes, L.E., A.B. Strathe, J.G. Fadel, D.P. Casper, and E. Kebreab. 2014. Prediction of enteric methane emissions from cattle. Glob. Chang. Biol. 20:2140–2148. doi:10.1111/gcb.12471.

Morgavi, D.P., E. Forano, C. Martin, and C.J. Newbold. 2010. Microbial ecosystem and methanogenesis in ruminants. Animal 4:1024–1036. doi:10.1017/S1751731110000546.

Murray, R.M., A.M. Bryant, and R.A. Leng. 1976. Rates of production of methane in the rumen and large intestine of sheep. Br. J. Nutr. 36:1–14. doi:10.1079/bjn19760053.

Myhre, G., D. Shindell, F. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, H. Zhang, D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley. 2013. Anthropogenic and natural radiative forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I 659–740.

Negussie, E., Y. de Haas, F. Dehareng, R.J. Dewhurst, J. Dijkstra, N. Gengler, D.P. Morgavi, H. Soyeurt, S. van Gastelen, T. Yan, and F. Biscarini. 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.. J. Dairy Sci. 100:2433–2453. doi:10.3168/jds.2016-12030.

Newbold, C.J., and E. Ramos-Morales. 2020. Review: Ruminal microbiome and microbial metabolome: Effects of diet and ruminant host. Pages S78–S86 in Animal. Cambridge University Press.

Oldenbroek, K., and L. van der Waaij. 2014. Textbook Animal Breeding Animal Breeding and Genetics for BSc Students. A. van Genderen, H. van Tartwijk, J. van Diepen, and L. Krijgsman, ed. Centre for Genetic Resources and Animal Breeding and Genomics Group, Wageningen University and Research Centre, the Netherlands., Wageningen.

Pszczola, M., K. Rzewuska, S. Mucha, and T. Strabel. 2017. Heritability of methane emissions from dairy cows over a lactation measured on commercial farms. J. Anim. Sci. 95:4813–4819. doi:10.2527/jas2017.1842.

Qiao, J., Z. Tan, and M. Wang. 2014. Potential and existing mechanisms of enteric methane production in ruminants. Sci. Agric. 71:430–440. doi:10.1590/0103-9016-2013-0423.

Ramayo-Caldas, Y., L. Zingaretti, M. Popova, J. Estellé, A. Bernard, N. Pons, P. Bellot, N. Mach, A. Rau, H. Roume, M. Perez-Enciso, P. Faverdin, N. Edouard, D. Ehrlich, D.P. Morgavi, and G. Renand. 2019. Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. J. Anim. Breed. Genet. 00:1–11. doi:10.1111/jbg.12427.

Roehe, R., R.J. Dewhurst, C.A. Duthie, J.A. Rooke, N. McKain, D.W. Ross, J.J. Hyslop, A. Waterhouse, T.C. Freeman, M. Watson, and R.J. Wallace. 2016. Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection

Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. PLoS Genet. 12:1–20. doi:10.1371/journal.pgen.1005846.

Rosenberg, E., O. Koren, L. Reshef, R. Efrony, and I. Zilber-Rosenberg. 2007. The role of microorganisms in coral health, disease and evolution. Nat. Rev. Microbiol. 5:355–362. doi:10.1038/nrmicro1635.

Sasson, G., S.K. Ben-Shabat, E. Seroussi, A. Doron-Faigenboim, N. Shterzer, S. Yaacoby, M.E.B. Miller, B.A. White, E. Halperin, and I. Mizrahi. 2017. Heritable bovine rumen bacteria are phylogenetically related and correlated with the cow's capacity to harvest energy from its feed. MBio 8:703–720. doi:10.1128/mBio.00703-17.

Schären, M., J. Frahm, S. Kersten, U. Meyer, J. Hummel, G. Breves, and S. Dänicke. 2018. Interrelations between the rumen microbiota and production, behavioral, rumen fermentation, metabolic, and immunological attributes of dairy cows. J. Dairy Sci. 101:4615–4637. doi:10.3168/jds.2017-13736.

Seymour, D., A. Canovas, T. Chud, J. Cant, V. Osborne, C. Baes, F. Schenkel, and F. Miglior. 2020. The dynamic behavior of feed efficiency in primiparous dairy cattle. j dairy Sci. 103:1528–1540. doi:10.3168/jds.2019-17414.

Simon, J.-C., J.R. Marchesi, C. Mougel, and M.-A. Selosse. 2019. Host-microbiota interactions: from holobiont theory to analysis. Microbiome 7:1–5. doi:10.1186/s40168-019-0619-4.

Templeton, A.R. 2006. Population Genetics and Microevolutionary Theory. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Ungerfeld, E.M. 2015. Shifts in metabolic hydrogen sinks in the methanogenesis-inhibited ruminal fermentation: A meta-analysis. Front. Microbiol. 6. doi:10.3389/fmicb.2015.00037.

Ungerfeld, E.M. 2018. Inhibition of rumen methanogenesis and ruminant productivity: A meta-analysis. Front. Vet. Sci. 5. doi:10.3389/fvets.2018.00113.

Ungerfeld, E.M. 2020. Metabolic Hydrogen Flows in Rumen Fermentation: Principles and Possibilities of Interventions. Front. Microbiol. 11:15. doi:10.3389/fmicb.2020.00589.

Vanwonterghem, I., P.N. Evans, D.H. Parks, P.D. Jensen, B.J. Woodcroft, P. Hugenholtz, and G.W. Tyson. 2016. Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. Nat. Microbiol. | 1:1–9. doi:10.1038/NMICROBIOL.2016.170.

Wallace, R.J., G. Sasson, P.C. Garnsworthy, I. Tapio, E. Gregson, P. Bani, P. Huhtanen, A.R. Bayat, F. Strozzi, F. Biscarini, T.J. Snelling, N. Saunders, S.L. Potterton, J. Craigon, A. Minuti, E. Trevisi, M.L. Callegari, F.P. Cappelli, E.H. Cabezas-Garcia, J. Vilkki, C. Pinares-Patino, K.O. Fliegerová, J. Mrázek, H. Sechovcová, J. Kopečný, A. Bonin, F. Boyer, P. Taberlet, F. Kokou, E. Halperin, J.L. Williams, K.J. Shingfield, and I. Mizrahi. 2019. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci. Adv. 5:1–12. doi:10.1126/sciadv.aav8391.

Weimer, P.J. 2015. Redundancy, resilience, and host specificity of the ruminal microbiota: Implications for engineering improved ruminal fermentations. Front. Microbiol. 6:296. doi:10.3389/fmicb.2015.00296.

Weimer, P.J., M.S. Cox, T. Vieira de Paula, M. Lin, M.B. Hall, and G. Suen. 2017. Transient changes in milk production efficiency and bacterial community composition resulting from near-total exchange of ruminal contents between high- and low-efficiency Holstein cows. J. Dairy Sci. 100:7165–7182. doi:10.3168/jds.2017-12746.

Weimer, P.J., D.M. Stevenson, H.C. Mantovani, and S.L.C. Man. 2010. Host specificity of the ruminal bacterial community in the dairy cow following near-total exchange of ruminal contents. J. Dairy Sci. 93:5902–5912. doi:10.3168/jds.2010-3500.

Woese, C.R., O. Kandler, and M.L. Wheelis. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. U. S. A. 87:4576–4579. doi:10.1073/pnas.87.12.4576.

Wu, L., P.W.G.G. Koerkamp, and N. Ogink. 2018. Uncertainty assessment of the breath methane concentration method to determine methane production of dairy cows. J. Dairy Sci. 101:1554–1564. doi:10.3168/jds.2017-12710.

Xiang, R., J. McNally, S. Rowe, A. Jonker, C.S. Pinares-Patino, V.H. Oddy, P.E. Vercoe, J.C. McEwan, and B.P. Dalrymple. 2016. Gene network analysis identifies rumen epithelial cell proliferation, differentiation and metabolic pathways perturbed by diet and correlated with methane production. Sci. Rep. 6:1–14. doi:10.1038/srep39022.

Yan, T., C.S. Mayne, D.C. Patterson, and R.E. Agnew. 2009. Prediction of body weight and empty body composition using body size measurements in lactating dairy cows. Livest. Sci. 124:233–241. doi:10.1016/j.livsci.2009.02.003.

Yin, X., W. Wu, M. Maeke, T. Richter-Heitmann, A.C. Kulkarni, O.E. Oni, J. Wendt, M. Elvert, and M.W. Friedrich. 2019. CO2 conversion to methane and biomass in obligate methylotrophic methanogens in marine sediments. ISME J. 13:2107–2119. doi:10.1038/s41396-019-0425-9.

CHAPTER 2

SCOPE OF THE STUDY

Alejandro Saborío-Montero[1,2]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain; [2]Universidad de Costa Rica, Escuela de Zootecnia, 11501 San Pedro de Montes de Oca, San José, Costa Rica.

Three research groups including the Department of Animal Production of the Basque Institute for Agricultural Research and Development (NEIKER)-Basque Research and Technology Alliance (BRTA), National Confederation of Associations of Spanish Friesian (CONAFE) and National Institution of Agri-Food Research and Technology (INIA) were involved in the RTA2015-00022-C03 (METALGEN) project; from the national plan of research, development, and innovation 2013-2020. Each group leaded the obtention of phenotypes (NEIKER-BRTA), genotypes (CONAFE), and rumen microbiota composition (INIA) from dairy cows of 14 commercial herds at four regions of the Northern Spain (País Vasco, Cantabria, Navarra, and Gerona). The project was focused on improvement of feed efficiency and mitigation of greenhouse gases emissions in dairy cattle. The current study approached the genotype-microbiota-phenotype relationship, through different statistical methodologies and through variance component estimation. Methane emission were used as the target phenotype, to clarify that complex interrelation. A partial genomic control of microbiota composition and methane emissions was evidenced, with the latter being also modulated by the microbiota composition. The findings of this thesis contribute to the understanding of genotype-microbiota-phenotype relationship. This study also provides information on the possible inclusion of microbiota composition as a proxy for methane emissions and other correlated traits of interest through genomic selection in animal breeding programs.

SPECIFIC OBJECTIVES OF THIS THESIS

1. To identify risk factors and quantify their effects on high methane emissions in dairy cattle.
2. To develop statistical models to jointly analyze host genotype and ruminal microbiota information in bovines.
3. Variance components estimation for the genotype and microbiota in complex traits of productive interest.
4. To tackle the problem of compositional data in metagenomics.
5. To propose the implementation of genetic evaluations including microbiota information.

CHAPTER 3

RUMEN EUKARYOTES ARE THE MAIN PHENOTYPIC RISK FACTORS FOR LARGER METHANE EMISSIONS IN DAIRY CATTLE

Alejandro Saborío-Montero[1,2], Idoia Goiri[3], Raquel Atxaerandio[3], Aser García-Rodriguez[3], Javier López-Paredes[4], José Antonio Jiménez-Montero[4], and O. Gonzalez-Recio[2,5]

[1]Escuela de Zootecnia y Centro de Investigación en Nutrición Animal, Universidad de Costa Rica, 11501, San José, Costa Rica; [2]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain; [3]Department of Animal Production, NEIKER – Basque Institute for Agricultural Research and Development. Basque Research and Technology Alliance (BRTA). Campus Agroalimentario de Arkaute s/n, 01192; [4]Spanish Holstein Association (CONAFE). Ctra. De Andalucía km 23600 Valdemoro, 28340 Madrid, Spain; [5]Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

ABSTRACT

Mitigation of methane emissions from dairy cattle is relevant to reduce environment impact and increase profitability through improvement of energy usage. The objective of this study was to estimate how microbiome composition determines large methane concentration (MET) and methane intensity (MI, ppm CH4/kg Milk) in comparison to more traditional proxies (i.e. milk yield and conformation traits). A total of 1359 Holstein cows from 17 herds in 4 northern regions of Spain were included in this study, the microbiome composition data was a subset containing 437 cows from 14 herds. Cows were classified in quartiles for MET and MI, according to individual records of methane measurements during the cow's visit to the automatic milking system unit. A probit approach under a Markov chain Monte Carlo (McMC) Bayesian framework was used to determine risk factors for high MET and high MI. Genetic merit for methane concentration and microbiome composition (86 phylum and 1240 genus) were the main drivers for a cow to be classified as high MET and MI. Reducing MET and MI genetic merit by unit of standard deviation (SD) reduced the probability of being classified in the upper quartile by 35.2% (33.9% to 36.4%) and 28.8% (27.6% to 29.6%), respectively. A reduction in probabilities was observed as the relative abundance of most bacteria increased (*i.e.*, Firmicutes 9.9% (8.3 to 11.3) for MET and 7.1% (6.2 to 8.2) for MI, per unit of SD). An opposite effect occurred with Eukaryotes. Larger abundance of most eukaryote became a risk factor to be classified as a high emitter animal (i.e., Oomycetes 14.2% (11.7% to 16.4%) for MET and 11.8% (9.4% to 14.0%) for MI, per unit of SD). An increment of one unit of SD in milk yield increased the probability of being classified in the upper quartile for MET by 3.7% (2.3% to 4.2%) and reduced the probability for MI by 12.6% (12.2% to 13.3%). Structure and capacity traits were not main drivers of being classified in the higher quartile of methane emission and intensity, with risk odds lower than 2% per unit of SD. After genetic merit, microbiome composition was the most relevant risk factor for larger methane emissions. This study suggests that mitigation of MET and MI could be addressed through animal breeding programs including genetic merits and strategies that modulate the microbiome.

Keywords: genomic selection, methane emission, microbiome, risk factor.

INTRODUCTION

Methane (CH_4) is a short live (8 years) climate greenhouse gas GHG (Allen et al., 2018). It has a short lifetime in the atmosphere, which might allow evaluating the effectiveness of directed actions towards its mitigation in a short period of time, compared to other GHGs. A rapid response of CH_4 levels is to be expected from successful approaches that preclude possible rising dynamics and even turn the slope in the desired downward direction. Every sector dealing with methane emissions should be capable of evaluate approaches to reduce its emissions, and consequently, its impact on environment.

Ruminants are methane emitters, with higher rates in large ruminants (Clauss et al., 2020), therefore it is an ethical duty for people involved in livestock to tackle methane emissions, leading to more sustainable productive systems. In the dairy industry, the productive performance of cows is of special interest because of its impact on economic retribution to the farmer; while methane emission reduction gives none currently economic revenue, other than the expectancy of increase productive performance trough better pathways in the usage of energy (Johnson and Johnson, 1995). This remarks the importance of approach the methane emission reduction through technics that allows combining an environmentally responsible production with competitive productive performance of the animals.

There are mainly two different thinking pathways about methane emissions in livestock. The first is a strictly "environmental" approach. This one is focused on total methane production, regardless its relationship with yield. A second approach is based on methane production per unit of product, which evaluates MI regarding to the production of the animal. The latter approach is often considered more balanced, as it takes into account the aim of domesticated ruminants. The variability for methane production between cows provides some potential in reducing methane emissions with management and genetic selection. The evidence supporting that methane emissions are partially driven by the cow's additive genetic effect (Difford et al., 2018; Saborío-Montero et al., 2020) could be used as an advantage to mitigate this problem. Genomic selection might be a sustainable approach to maintain high productive performance while a reduction of methane emissions takes place.

The epidemiological approach of risk factors determination has been applied as a tool to identify, describe and quantify disease causative variables in dairy cattle (Saborío-Montero et al., 2017). Disease definition from a health management perspective has increased its boundaries, to consider any factor limiting animal or herd performance as a component of disease (LeBlanc et al., 2006). High methane emissions reduce animal performance through energy lost (Johnson and Johnson, 1995). Therefore, we considered methane emissions as a trait that can be analyzed under a risk factor approach to identify and measure the effect of some potential factors. A plethora of factors have been associated as proxies for enteric methane emissions in livestock (Knapp et al., 2014; Negussie et al., 2017). Metagenomic profiles are a most recent topic associated to methane emissions, that has become popular in the last decade, due to advances like nanopore methodologies, which enables the determination of microorganisms from all taxonomic domains (Saborío-Montero et al., 2020).

At the best of our knowledge, a risk factors approach on methane emissions has not been implemented yet. Here, we considered some low-cost proxies that are easy to measure, commonly recorded traits in dairy production systems (milk yield and composition, conformation traits), as well as some more complex traits (rumen metagenome, MET and MI genetic merits) with larger associated costs. The aim of this study was to contribute to the identification of methane emissions risk factors in dairy cattle and quantify their effect on the probability of a cow being classified as high methane emitter.

MATERIALS AND METHODS

Ethical and animal welfare statement

This study was approved by the Basque Institute for Agricultural Research and Development Ethics Committee (Neiker-OEBA-2017-004) on March 28, 2017. This study was carried out in accordance with Spanish Royal Decree 53/2013 for the protection of animals used for experimental and other scientific purposes.

Study design, population, and data

An observational study with a total of 1359 lactating Holstein cows was developed to determine risk factors associated to methane emissions. Cows were from 17 commercial herds in four regions of northern Spain (Basque Country, Cantabria, Girona and Navarre). The microbiome data was a subset containing 437 cows (≤ second lactation) from 14 herds. Methane concentration (ppm CH_4) was measured using Guardian® NG infrared gas monitor (Edinburgh Sensors; measure range 0-1%), a non-dispersive infrared methane detector as described by Rey et al. (2019). Briefly, the device was installed within the feed bin of the automatic milking system, allowing the measurements for each cow, at every milking. MET in breath samples was measured individually for each cow during 2 to 3 weeks period. A single methane record per cow was obtained by averaging the eructation peaks recorded. Traits related to milk yield and composition were available for all herds at this sample period, through a test-day recording, as part of the official test day recording scheme. Conformation traits were facilitated by the Spanish Holstein Association (CONAFE). Cows were scored (1 to 9 points categorical scale, with one-unit increments), by CONAFE's official qualifiers, during their first lactation. Information regarding rumen content samples acquisition and sequencing are described in Saborío-Montero et al. (2021a, 2021b).

Single step models for genomic BLUP (ssGBLUP)

A single-step genomic BLUP was performed for estimating solutions for methane production (ppm CH_4) and MI (ppm CH_4/kg milk) combining pedigree and genomic relationships between animals following the model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where **y** was a $n \times 1$ vector of phenotypic records either for MET or MI, **b** denoted the vector of systematic effects with incidence matrix **X**, **u** was the vector of direct additive genetic effects assumed distributed as $\mathbf{u} \sim N(0, \mathbf{H}\sigma_u^2)$ with **H** being the numerator relationship matrix based on a combination of pedigree and genomic relationships, as suggested by Legarra et al. (2014). Then σ_u^2 was the additive genetic variance, **Z** the

corresponding incidence matrix. Finally, **e** was a 1 × n vector of residuals **e** ~N (0, $\mathbf{I}\sigma_e^2$), where **I** was an identity matrix of appropriate order and σ_e^2 the residual variance.

Cows were genotyped using the EuroG 10K and EuroGMD 12k SNP chip from Illumina (Illumina, San Diego, California, USA), and then imputed to 54,609 SNPs (Bovine 50k SNP chip, Illumina, San Diego, California, USA). The BEAGLE software (Browning et al., 2018) was used for imputation, using 3669 animals from a Spanish reference population provided by CONAFE (Jiménez-Montero et al. 2013). Monomorphic SNPs and those with MAF<0.05 were filtered out from the analysis, resulting in 42,372 SNPs left for the analyses. A genomic relationship matrix (**G**) was built following method 2 of VanRaden (2008). Pedigree records for all animals were also traced back to a final dataset of 2,619 animals.

Results from the ssGBLUP were transformed to a ranking scale (GEBVs) for MET and MI, with lower genetic merit values for animals with higher MET and MI.

Statistical analysis

Methane phenotypes were categorized in quartiles (low, middle-low, middle-high, and high). Descriptive statistics of means and standard deviations were obtained for every factor according to response variable category.

The statistical analyses were performed using a threshold probit model approach under a Markov chain Monte Carlo (MCMC) Bayesian framework (Gianola, 1982; Gianola y Foulley, 1983). Threshold models assume that a categorical response variable has a subjacent continuous distribution for a random variable named liability (λ). Likewise, the observed response *y* takes the value *j* when λ is greater than or equal to T_{j-1} and less than T_j where T_{j-1} and T_j are thresholds, with T_0 and T_4 being $-\infty$ and ∞, respectively.

$$y = \begin{cases} 1, if\ T_0 < \lambda_i < T_1 \\ 2, if\ T_1 \leq \lambda_i < T_2 \\ 3, if\ T_2 \leq \lambda_i < T_3 \\ 4, if\ T_3 \leq \lambda_i < T_4 \end{cases}$$

Variance of the liability distribution was set to an arbitrary value of one: λ ~ N (μ+**X**b, 1). The first and second thresholds were arbitrarily set to zero and 1 for the model to be identifiable, while the third threshold was estimated.

A set of potential risk factors were independently included into a threshold-liability model (Gianola, 1982), to analyze and compare their effects on classifying a cow in the upper quartile of MET and MI. The statistical model for MET and MI outcome was described as:

$$\lambda_{ijklmn} = \mu + \mathrm{RF}_j + \mathrm{DIM}_k + \mathrm{DIM2}_l + \mathrm{HR}_m + \mathrm{e}_{ijklmn}$$

Where $\lambda_{ijklmno}$ was the corresponding liability of the observation, either for MET or MI, μ was the population mean, RF_j was the risk factor of interest, DIM_k was a covariate of days in milk, $DIM2_l$ was a covariable of the quadratic effect of days in milk, HR_m was the effect of herd-robot of methane recording (24 levels) and e_{ijklmn} was the randomly distributed residual error with mean zero and residual variance set to 1. Each risk factor was independently analysed, and were: milk yield (kg/d), fat (%) and protein (%), conformation traits of overall structure and capacity value (summatory scores from the variables conforming the index), stature (1 to 9 scale), chest width (1 to 9 scale), loin strength (1 to 9 scale), body depth (1 to 9 scale), body weight (kg), body condition score (BCS) (1 to 9 scale), genetic merit (GEBVs) for MET and MI (mean = 100, SD =10 scale; higher values represent lower MET or MI) from the ssGBLUP described above, and finally the centered log-ratio transformed relative abundances (%) of 86 phylum (replaced by class when no phylum assigned in NCBI) and 1240 features at genus taxonomic level. The described models were implemented using the TM package (Legarra et al., 2011). Each analysis consisted of a single chain of 60,000 iterations, discarding the first 10,000 samples. Burn-in period and chain length were determined by the convergence observed in trace plots. The lag period (thin) was 10 samples, keeping 5000 samples for final inferences.

Probability estimation

The probability for a cow of being classified in the upper quartile for MET or MI was based on the probability of overpassing the threshold separating categories 3 and 4 on the observed scale (T_3), and was estimated as follows:

$$P = 1 - F(T_3 - \beta_i)$$

Where, F was the standard normal cumulative distribution function, and β_i was the posterior mean on the underlying scale for each analyzed risk factor. The lower and upper bounds of the probability intervals (*PI*) of being classified as high for MET or MI were calculated as:

$$PI = 1 - F(T_3 - (\beta_i \pm PSE))$$

where, *PSE* was the standard error of the posterior distribution, after burn-in, associated to the solution (β_i) on the underlying scale. Probability change, expressed in percentage, was calculated as the difference between the probability estimated by the model including the potential risk factor and the probability estimated by the benchmark model, and then multiplied by a 100 factor. An standardization was also applied to the estimated probability differences, in order to remove the noise from different scales of measurement. The standard deviation from each variable was multiplied by the corresponding probability difference, leading to probability differences per unit of standard deviation (SD) of the risk factor.

RESULTS AND DISCUSSION

Descriptive statistics

The descriptive statistics according to group for MET and MI are shown in Table 1 and Table 2, respectively.

Table 1. Descriptive summary of means and their standard deviation of potential risk factors according to quartile group for methane concentration (MET) from 1359 dairy cows.

Variables*	Group of MET (ppm CH_4)			
	Low	Middle-Low	Middle-High	High
Milk yield (kg)	36.78 (9.93)	36.13 (9.52)	37.85 (9.74)	38.04 (10.22)
Milk Fat (%)	3.55 (0.96)	3.64 (0.84)	3.59 (0.87)	3.66 (0.90)
Milk Protein (%)	3.23 (0.36)	3.29 (0.49)	3.28 (0.38)	3.24 (0.41)
Overall Structure and Capacity (Score 1 - 100)	81.48 (2.94)	81.82 (3.26)	81.73 (3.07)	81.86 (3.35)
Stature (Score 1 - 9)	5.79 (1.49)	5.77 (1.49)	5.70 (1.41)	5.82 (1.57)
Chest Width (Score 1 - 9)	5.33 (1.08)	5.50 (1.20)	5.44 (1.05)	5.52 (1.14)
Loin Strength (Score 1 - 9)	5.36 (1.16)	5.39 (1.23)	5.40 (1.24)	5.43 (1.16)
Body Depth (Score 1 - 9)	5.38 (1.11)	5.54 (1.08)	5.55 (1.01)	5.53 (1.09)
Body Weight (kg)	597.72 (27.76)	601.10 (29.95)	600.03 (27.26)	602.66 (30.86)
BCS (Score 1 - 9)	5.10 (1.01)	5.38 (1.07)	5.23 (0.95)	5.27 (0.99)
Methane (ppm)	850.59 (132.70)	1128.13 (63.48)	1364.07 (77.08)	1807.74 (287.92)
GEBV (MET) (Scale centered in 100 (SD = 10))	109.46 (7.00)	103.17 (6.39)	98.76 (5.65)	88.79 (8.44)

*Each group has 340 observations except for group Middle-High with 339 observations. GEBVs are in a scale centered in 100 (SD = 10) with lower values for high emitting animals. The measurement units for the overall variables are the summatory scores from the variables conforming the index.

Table 2. Descriptive summary of means and their standard deviation of potential risk factors according to quartile group for methane intensity (MI) from 1359 dairy cows.

Variables*	Group of MI (ppm CH_4 / kg milk)			
	Low	Middle-Low	Middle-High	High
Milk yield (kg)	44.77 (8.97)	40.04 (7.41)	35.34 (7.97)	28.65 (7.06)
Milk Fat (%)	3.32 (0.86)	3.48 (0.78)	3.63 (0.89)	4.01 (0.90)
Milk Protein (%)	3.12 (0.28)	3.21 (0.28)	3.26 (0.45)	3.45 (0.50)
Overall Structure and Capacity (Score 1 - 100)	81.75 (3.08)	81.80 (3.09)	81.53 (3.05)	81.81 (3.40)
Stature (Score 1 - 9)	5.91 (1.39)	5.76 (1.54)	5.66 (1.50)	5.75 (1.53)
Chest Width (Score 1 - 9)	5.37 (1.14)	5.45 (1.16)	5.43 (1.02)	5.54 (1.15)
Loin Strength (Score 1 - 9)	5.35 (1.25)	5.36 (1.16)	5.42 (1.22)	5.47 (1.16)
Body Depth (Score 1 - 9)	5.48 (1.12)	5.51 (1.03)	5.47 (1.06)	5.54 (1.07)
Body Weight (kg)	599.71 (29.15)	600.28 (29.07)	599.46 (28.31)	602.05 (29.59)
BCS (Score 1 - 9)	5.19 (0.97)	5.20 (1.06)	5.34 (1.00)	5.25 (1.02)
Methane (ppm/kg milk)	21.36 (3.73)	30.22 (1.98)	38.60 (3.06)	58.69 (15.62)
GEBV (MI) (Scale centered in 100 (SD = 10))	107.09 (7.46)	101.58 (7.54)	98.20 (9.09)	91.48 (10.94)

*Each group has 340 observations except for group Middle-High with 339 observations. GEBVs are in a scale centered in 100 (SD = 10) with lower values for high emitting animals. The measurement units for the overall variables are the summatory scores from the variables conforming the index.

The relative abundance of the microbial superkingdom in the rumen microbiota are described in Table 3 for each quartile group of MET and MI.

Table 3. Descriptive summary of relative abundance (%) of microbiota superkingdom according to quartile group of methane concentration (MET) and methane intensity (MI) from 437 dairy cows.

	Group of MET (ppm CH_4)			
Superkingdom	Low	Middle-Low	Middle-High	High
Archaea	0.19	0.22	0.21	0.21
Bacteria	94.16	93.39	92.58	92.48
Eukaryote	5.65	6.39	7.21	7.31
	Group of MI (ppm CH_4 / kg milk)			
Superkingdom	Low	Middle-Low	Middle-High	High
Archaea	0.22	0.19	0.21	0.22
Bacteria	94.25	93.34	92.59	92.38
Eukaryote	5.53	6.47	7.20	7.40

The relative abundance of eukaryotes showed an uplifting trend as groups increased from low to high MET or MI. Consequently, the average abundance of bacteria decreased proportionally from low to high MET and MI, which implies larger Eukaryote/Bacteria ratio

in groups with larger MET or higher MI. The means in archaea superkingdom by quartile remained close to a modest value of 0.20%, regardless the group of MET or MI.

The correlation for genetic merit scores between MET and MI was high and positive (0.84). The genetic merit for MET and MI allowed to evaluate, at a genetic level, the potential effect of the animal's genetic background as a risk factor for both traits.

Threshold models

The use of non-linear mixed models based in threshold theory has been proposed as more appropriate alternative to linear models for the analysis of categorical traits (Gianola, 1982; Gianola y Foulley, 1983) and extensively compared between both approaches in posterior studies (Kadarmideen et al., 2000; González-Recio y Alenda, 2005; Saborío-Montero et al., 2018). Results from the threshold model for the potential risk factors analysis are described in subsequent sections.

Milk yield and composition

Results from the threshold model showed that an increment in milk yield was associated to larger MET, with slight increments in the probability of being classified in the upper quartile (Figure 1A). The slight increase of probability agreed with previous studies, which stated that methane production in dairy cattle is not well predicted by milk yield alone (Negussie et al., 2017; Niu et al., 2018). Other studies reported low to moderate positive genetic correlations between those traits (Breider et al., 2019; López-Paredes et al., 2020), pointing milk yield as a poor predictor of methane emissions.

Milk fat percentage was positively associated to MET, with increased probability of being classified in the high methane emissions group when milk fat percentage increased (Figure 1A). This is consistent with previous studies that showed positive correlations between methane emissions and milk fat content (Moate et al., 2018; Bougouin et al., 2019). Rumen acetate enhance *de novo* milk fat synthesis; this process requires a carbon source, mainly acetate and β-hydroxybutyrate, as well as $NADPH+H^+$, primarily from glucose and acetate (Bauman and Davis, 1974). Rumen acetate also enhance hydrogenogenesis (Moraes et al., 2014). The latter process is biologically linked to methane emissions through hydrogen

supply for hydrogenotrophic methanogens (Moraes et al., 2014; Danielsson et al., 2017). Those biological relationships support the finding of positive association between milk fat percentage and methane emissions. Conversely, in the MI (ppm CH_4/kg milk) analysis, milk yield was a protector factor. The probability of being classified in the upper quartile was lower per unit of SD for milk yield, showing a large protective effect (Figure 1B). Other studies also showed negative associations between milk yield and MI (Kandel et al., 2017). Jiao et al. (2014) showed that MI decreased with larger milk yield level in grazing dairy cattle fed with higher doses of concentrate. A low forage/concentrate diet reduce rumen pH, affecting microbial fermentation process (Bauman and Griinari, 2001). These results in an acetate to propionate ratio reduction (Danielsson et al., 2017), causing a diminish to the overall MI due to a lack of substrate (acetate) for hydrogenotrophic methanogens (Moraes et al., 2014; Danielsson et al., 2017). It can also affect the protozoal composition through inhibiting methanogens growth and activity (Martin et al., 2010). Besides, there is a direct effect on MI reduction due to the increase in milk yield, that straightforward influences the calculation of MI.

Percentages of milk protein and milk fat increased the probability of a cow being classified in the upper quartile for MI by 7.03% per unit of SD in milk protein percentage and by 6.45% per unit of SD in milk fat percentage (Figure 1B). These results are consistent with a dilution effect on milk components, reducing fat and protein percentages, in high producing cows, and a concentration effect, causing increments on milk fat and protein percentages on low producing cows, as evidenced in Table 2. Volatile fatty acids are related to milk components (Liu et al., 2018) and methane emissions (Williams et al., 2019), with acetate positively correlated to milk fat content (Urrutia and Harvatine, 2017), as well as with milk protein synthesis (Zhao et al., 2019) and methanogenesis (Lopes et al., 2016). Acetate and leucine interact synergically in the up-regulation of milk protein synthesis (Zhao et al., 2019), the positive association of acetate with milk fat and milk protein, as well as with methane emissions, might explain the findings of milk fat and milk protein as risk factors for MI in the present study.

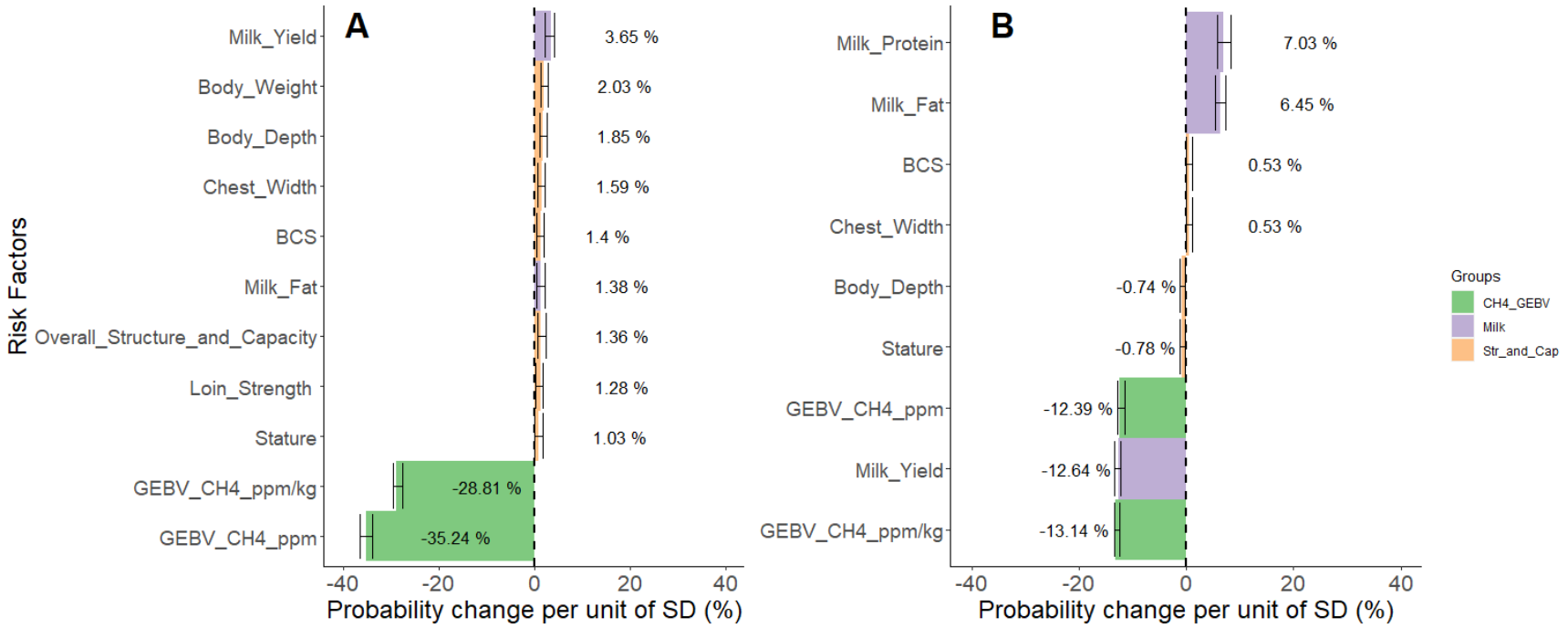

Figure 1. Change in the probability of a cow being classified in the upper quartile for methane concentration (ppm CH_4) and methane intensity (ppm CH_4 / kg milk) per unit of increment in the risk factor corrected for standard deviation for methane concentration (**A**) and methane intensity (**B**). Black dashed lines indicate the baseline probability of being classified in the upper quartiles without any variable effect. Probability intervals are depicted for each risk factors. BCS=Body condition score, GEBV_CH4_ppm/kg= Genetic merit for methane intensity (MI), GEBV_CH4_ppm=Genetic merit for methane concentration (MET), CH4_GEBV=Genetic merit for methane traits, Milk=Milk related traits, Str_and_Cap= Structure and capacity related traits.

Conformation traits

Seven conformation traits, related to structure and capacity, were associated to larger MET (Figure 1A). An increment of one unit of SD in the score of these traits increased the probability of a cow being classified in the upper quartile for MET (body weight (2.03%), body depth (1.85%), chest width (1.59%), BCS (1.40%), overall structure and capacity (1.36%), loin strength (1.28%) and stature (1.03%)). Larger cows have higher energy demands and intake levels (Li et al., 2018), hence larger probability of being classified in the upper quartile for MET (Figure 1). Similar results were observed in previous studies that studied the phenotypic (Yin et al., 2015) or genetic (Pszczola et al., 2019; López-Paredes et al., 2020) (López-Paredes et al., 2020; Pszczola et al., 2019) association between conformation traits and methane.

The probability of being classified in the upper quartile for MI increased with the increment by one unit of SD for BCS (0.53%) and chest width (0.53%). A reduction in that probability was observed per unit of body depth (0.74%) and stature (0.78%), presumably by a correlated response on milk yield and feed intake in larger animals. The increase in milk yield attenuates the emission per kg of milk by increasing the denominator of the response variable. Structure and capacity traits increased the risk for high MET, with a lesser effect over high MI. These traits could be modified through genetic selection, with an impact on methane emissions. However, the effect over correlated traits of interest such as milk yield and composition should be analyzed before including these traits into breeding programs. The extent to which structure and conformation affect both methane traits was marginal compared to the genetic merit.

Additive genetic effect

Larger values for genetic merits decreased the risk of being classified in the upper quartile for MET and MI. The genetic merit reduced the probabilities of being classified in the upper quartile for MET by 35.2% (Figure 1A) and by 13.1% for MI (Figure 1B) per unit of SD. Conversely, cows with lower genetic merit for MET and MI were more likely classified in the higher categories, as expected.

The genetic merit was the main driver for reducing the probability of being classified in the upper quartile for MET and MI, compared to all productive and conformation phenotypic traits analyzed, when measurement scales were standardized. In dairy cattle, methane emissions are partially controlled by the additive genetic effect of the cow (Difford et al., 2018; Saborío-Montero et al., 2020), plus external inherent conditions affecting each individual (Knapp et al., 2014). Direct genetic merit of the cow was the variable that best explained the methane production level of a cow. Our results provide evidence of the existence of a genetic effect of the animal over its methane emissions, it also allows the comparison of the relative impact of the genetic and environmental factors over this complex trait.

Microbiome

Microbiome showed the largest phenotypic risk factor effect in this study. Ciliate protozoa and fungi were among the top 5 taxonomic features with the highest increment in the probability of a cow being classified in the upper quartile for MET and MI, according to an increment of one unit of SD in their relative abundances. For instance, a one unit of SD in Chytridiomycota relative abundance, increased that probability in 13.42% for MET and 12.33% for MI (Figure 2). This phylum includes the Neocallimastigomycetes class (Spatafora et al., 2016) which in turn, contain *Pyromyces* spp. and *Neocallimastix* spp., a couple of genera associated to the presence of *Methanobrevibacter spp*. (Jin et al., 2011). The utility of Chytridiomycota phylum and other taxonomic features (*i.e.* Oomycetes, Mucoromycota, Heterotrichea, Oligohymenophorea and Spirotrichea), as rumen microbiota markers for methane emissions should be further studied given its high influence as a risk factor for MET and MI obtained in this study, as well as previous evidence of genera from this phylum associated to methanogenic archaea. The phyla with the largest reduction in the probability of a cow being classified in the upper quartile for both traits per unit of SD were Proteobacteria and Firmicutes, with a 10.02% and 9.91% reduction for MET, and a 7.08% and 7.12% reduction for MI, respectively (Figure 2).

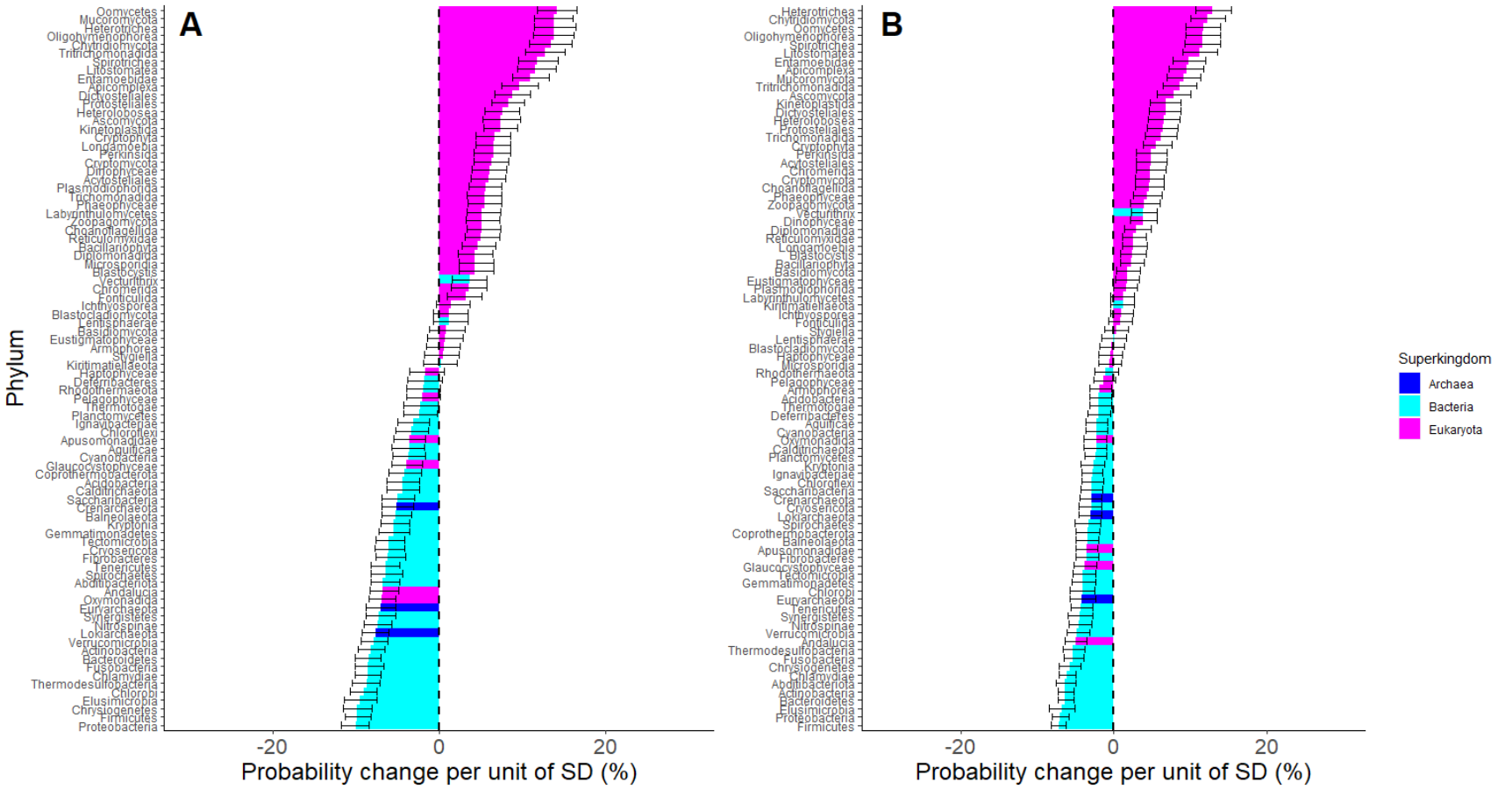

Figure 2. Change in the probability of a cow being classified in the upper quartile for (**A**) methane concentration (ppm CH_4) and (**B**) methane intensity (ppm CH_4/kg milk) per standard deviation unit of 86 phyla colored by superkingdom. Black dashed line indicates the baseline probability of being classified in the upper quartiles without any phylum effect. Probability intervals based on posterior standard deviations are depicted for all phyla.

The results of the microbiota analysis at the phylum level indicated that the probability of being classified as high methane emitter for MET and MI was larger in cows with larger relative abundance of Eukaryote phyla (Figure 2). Previous studies showed associations between rumen protozoa and methane emissions (Guyader et al., 2014). There is a plausible biological mechanism underlying this relationship as some eukaryote microbes host archaea inside their cells. Archaea living in Eukaryotes are surrounded by mitochondria-like structures called "hydrogenosomes", these structures supply archaeas with hydrogen, which is then transformed into methane as an end-product from carbon dioxide reduction that takes place inside the archaea (Shinzato et al., 2010). Therefore, the larger the abundance of eukaryotes, the larger substrate places for methanogenic archaea into the rumen microbiota. Most of the bacteria phyla reduced the probability of being classified as high methane emitter, both for MET and MI per unit of SD in its relative abundance. Similar results were observed in a previous study that associated bacteria with methane emissions (Aguilar-Marin et al., 2020). This study proposed that higher abundances of ruminal *Prevotella spp.* compete with methanogens for hydrogen utilization, redirecting it for propionic acid production, and therefore reducing the amount of hydrogen available for methanogenesis.

A similar pattern was observed for the 1240 genera analysed (Figure 3), with higher probabilities of being classified in the upper quartile for MET and MI as the SD of relative abundance of Eukaryotes increased. Lower probabilities of being classified in the upper quartile for MET and MI were found as the relative abundances of genera standardized for SD from most bacteria increased (Figure 3).

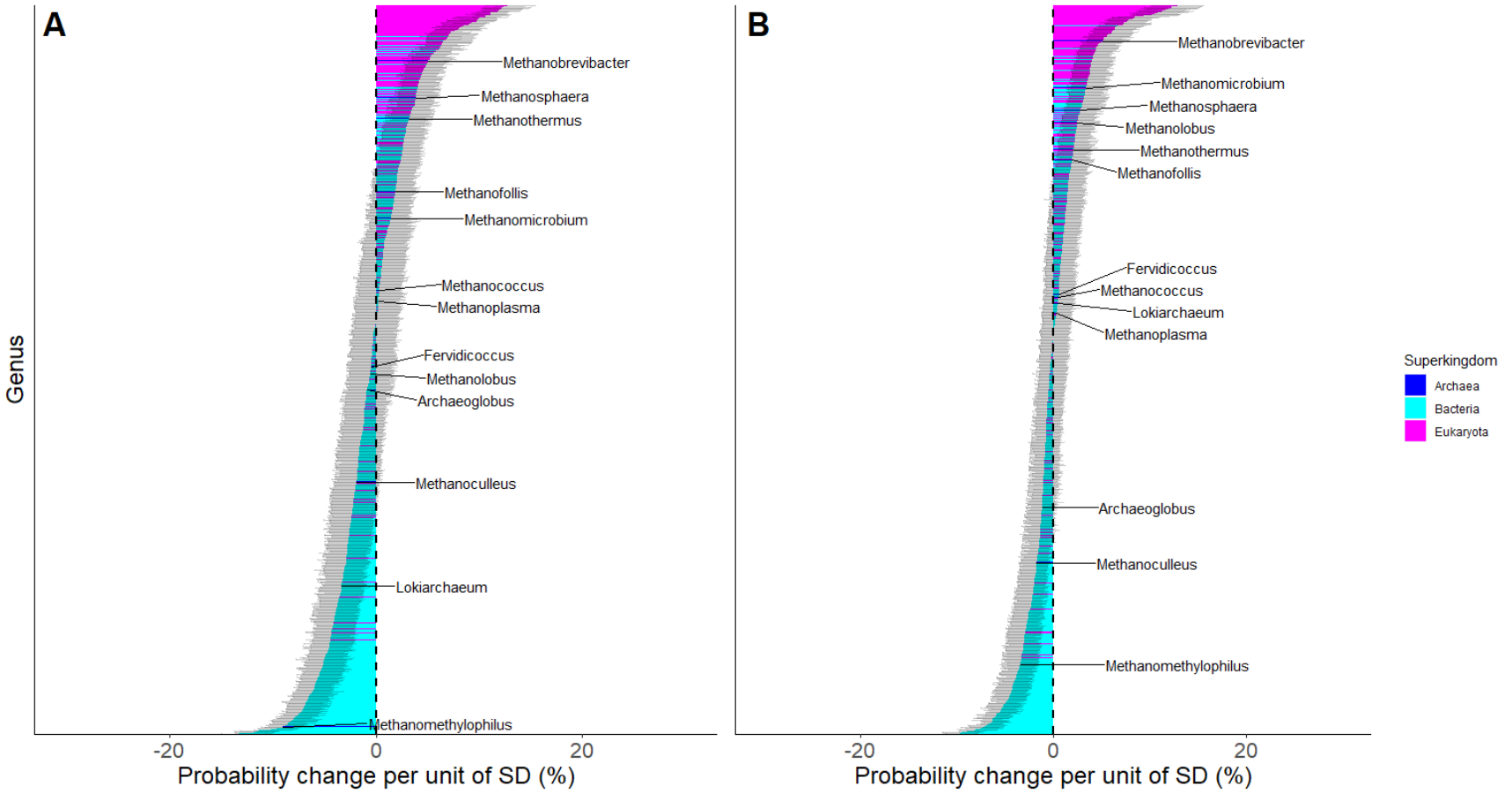

Figure 3. Change in the probability of being classified in the upper quartile for (**A**) methane concentration (ppm CH_4) and (**B**) methane intensity (ppm CH_4/kg milk) per unit of increment in relative abundance (%) of 1240 genera colored by superkingdom. Black dashed line indicates the baseline probability of being classified in the upper quartiles without any genus effect. All the archaea genera are explicitly indicated. Probability intervals based on posterior standard deviations are depicted in gray for all genera.

The top 10 eukaryote genera that caused larger increment in the probability of being classified in the upper quartile for MET and MI per SD unit of increment in their relative abundance were ciliate protozoa from the Ciliophora phylum and two fungi (Neocallimastix and Piromyces) from the Neocallimastigomycetes class (Table 4).

Table 4. Increment of probability of a cow being classified in the upper quartile for methane concentration (MET) and methane intensity (MI) according to a unit of increment in the SD of the relative abundance of the top 10 risk factors from eukaryote genera.

MET (ppm CH_4)			
Genera	Mean (%)	LB[1] (%)	UB[2] (%)
Paramecium spp.	12.7	10.6	15.5
Tritrichomonas spp.	12.6	11.5	14.4
Stentor spp.	12.3	10.4	15.0
Oxytricha spp.	12.2	10.3	14.8
Neocallimastix spp.[3]	12.1	10.2	14.2
Pseudocohnilembus spp.	12.1	10.5	14.2
Tetrahymena spp.	12.0	10.1	13.9
Ichthyophthirius spp.	11.8	10.3	14.0
Entamoeba spp.	11.8	9.8	12.9
Planoprotostelium spp.	11.2	9.5	12.7
MI (ppm CH_4 / kg milk)			
Genera	Mean (%)	LB (%)	UB (%)
Stentor spp.	12.9	10.2	15.6
Oxytricha spp.	12.5	10.0	15.5
Pseudocohnilembus spp.	12.3	9.4	15.3
Paramecium spp.	12.2	9.7	15.1
Ichthyophthirius spp.	11.9	9.0	14.8
Plasmodium spp.	11.7	9.4	13.9
Entodinium spp.	11.6	8.9	14.2
Neocallimastix spp.[3]	11.4	8.6	14.5
Piromyces spp.[3]	11.3	9.0	13.5
Stylonychia spp.	11.0	8.4	14.1

[1]LB=Lower bound, [2]UB=Upper bound, [3]Fungi from the Neocallimastigomycetes class.

Interestingly, most of the eukaryote from the top 10 genera that caused larger increment in the probability of being classified in the upper quartile for MET were common for MI, with slight differences in their ranking order. This group of microorganisms could point to the joint reduction of both MET and MI, through rumen control approach of ciliate protozoa and fungi abundances.

There are evidences of endosymbiotic associations between methanogenic archaea and anaerobic ciliates protozoa, such as *Entodinium spp*. (Finlay et al., 1994). There is also information regarding hydrogenosomes in *Neocallimastix spp*., suggesting those structures as an intracellular hydrogen source for methanogenic archaea (Shinzato et al., 2010). This associations between eukaryote and archaea support the findings of this study, relating larger probabilities of being in the high category of MET and MI as eukaryote relative abundance increase. The protective factors that reduced the probability of a cow being classified in the upper quartile for MET and MI were mainly bacteria genera from the Firmicutes, Bacteroidetes and Proteobacteria phyla (Table 5).

Table 5. Reduction of probability of a cow being classified in the upper quartile for methane concentration (MET) and methane intensity (MI) according to a unit of increment in the SD of the relative abundance of the top 10 protective factors from bacteria genera.

MET (ppm CH_4)			
Genera	Mean (%)	LB[1] (%)	UB[2] (%)
Dialister spp.	13.2	10.7	15.0
Mitsuokella spp.	11.8	9.9	13.7
Sutterella spp.	11.6	9.8	13.6
Oribacterium spp.	11.5	9.3	13.7
Megasphaera spp.	11.3	9.5	13.2
Anaerobiospirillum spp.	10.2	8.1	11.8
Lactobacillus spp.	9.6	7.5	11.3
Vibrio spp.	9.5	7.6	11.4
Hespellia spp.	9.0	7.2	10.7
Halomonas spp.	8.9	6.7	10.4
MI (ppm CH_4 / kg milk)			
Genera	Mean (%)	LB[1] (%)	UB[2] (%)
Megasphaera spp.	9.7	8.1	11.1
Mitsuokella spp.	9.4	7.7	11.4
Dialister spp.	9.2	7.2	11.6
Oribacterium spp.	8.8	7.7	9.8
Citrobacter spp.	8.1	7.1	9.0
Acidaminococcus spp.	8.0	5.8	9.6
Sutterella spp.	7.6	5.8	10.0
Colwellia spp.	7.5	5.2	9.0
Halomonas spp.	7.3	5.6	9.6
Aggregatibacter spp.	7.1	5.1	8.7

[1]LB=Lower bound, [2]UB=Upper bound.

The top 10 protective factors for MET and MI overlapped for 6 genera. Lower relative abundances of *Megasphaera spp*, *Dialister spp*. and *Mitsuokella spp*. had been previously described in higher methane emitter animals (Wallace et al., 2015). While *Oribacterium spp*. has been reported as a potential H_2 sink in low methane emitter animals (Greening et al., 2019). The *Halomonas spp*. genus has been qualified as a methanotroph able to degrade methane to ectoines (metabolites produced by some bacteria to resist salinity stress), with higher ectoines yields than those reported for the already known ectoines producer methanotrophs (Cantera et al., 2019). Concordance between bacteria protective factors might indicate that those genera can improve efficiency by jointly reducing MET and MI.

The *Methanobrevibacter spp*., was the most relevant risk factor among archaea, for a cow to be classified in the upper quartile for MET by 5.05% probability increment and MI by 5.19% probability increment. This is a recognized hydrogenotrophic methanogenic archaeas previously described in rumen (Janssen and Kirs, 2008; Martínez-Álvaro et al., 2020). In contrast, *Methanomethylophylus spp*. was the archaea genus with the largest reduction for the probability of a cow being classified in the upper quartile for MET (9.0%) and MI (3.4%). *Methanomethylophylus spp*. is a methylotrophic archaea negatively correlated with methane emissions in a previous study (Martínez-Álvaro et al., 2020), belonging to the Termoplasmata class, which has been implicated in reduced methane emissions from bovine rumen (Poulsen et al., 2013).

At the genus taxonomic level, an increase in relative abundance of some archaea increased the probabilities of being classified as high methane emitter. The probabilities linked to archaea at the genus level differed from those observed at phylum level, in which increments in all archaea phyla had a reductive effect of probabilities. A plausible explanation for this would be that there is a hidden effect of some methanogenic archaea by some other non-methanogenic archaea within the same phylum, resulting in an overall effect which is confounded at higher taxonomic levels and unmasked at more specific ones. These results may explain why archaea relative abundances have shown weaker or lack of associations with methane in previous studies (Negussie et al., 2017).

CONCLUSION

Direct genetic merit of the cow was the variable that best explained the methane production level of a cow. Genetic merit was the most relevant protective factor that reduced the probability of being classified in the upper quartile for MET (35.2%) and MI (13.1%) per SD unit. These results evidenced its relevance as a proxy for methane emission reduction. The microbiota was the main phenotypic factor influencing the probability of being classified in the upper quartile for both methane traits. A reduction in that probability was observed as the relative abundance per SD unit of most bacteria increased (*i.e.,* ≈10% for Firmicutes and Proteobacteria). An opposite effect occurred with Eukaryotes. Larger abundance of most eukaryote became a risk factor to be classified as a high emitter animal (*i.e.,* ≈15% for ciliate protozoa and fungi). A similar pattern was observed at the genus level, showing reduced probabilities as the relative abundances per SD unit of bacteria increased (*i.e.,* ≈10% for *Dialister spp., Mitsuokella spp., Sutterella spp.*) and relative abundances of eukaryote increased (*i.e.,* ≈12% for *Stentor spp., Paramecium spp. and Oxytricha spp.*). Some archaea genera were risk factors (*i*.e., ≈5% for *Methanobrevibacter spp., Methanothermus spp. and Methanosphaera spp.*) but other were protective factors (*i.e.,* ≈10% *Methanomethylophilus spp.*) influencing the probability of being classified in the upper quartile for MET and MI, depending on the genus analyzed. Archaea should not be taken as a single group when analyzing their association with MET and MI. Milk yield increments slightly increased the probability of being classified in the upper quartile for MET (3.6%) and reduced that probability for MI (12.6%) per unit of standard deviation. The structure and capacity trait analysis indicated that bigger cows had larger probabilities of being classified in the upper quartile for MET, not so for MI, although they were not the main drivers of larger methane emissions. The findings from this study suggest that mitigation of MET and MI could be addressed from multifactorial approaches, with promising responses for genetic merits and microbiome modulation through nutritional and genetic management.

ACKNOWLEGDGMENT

This research was financed by RTA2015-00022-C03 (METALGEN) project from the national plan of research, development, and innovation 2013-2020. The first author of this paper was granted a scholarship from Universidad de Costa Rica for course doctorate studies which partially conducted to the progress of this study.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest regarding the topics approached within this study.

REFERENCES

Aguilar-Marin, S.B., C.L. Betancur-Murillo, G.A. Isaza, H. Mesa, and J. Jovel. 2020. Lower methane emissions were associated with higher abundance of ruminal Prevotella in a cohort of Colombian buffalos. BMC Microbiol. 20:364. doi:10.1186/s12866-020-02037-6.

Allen, M.R., K.P. Shine, J.S. Fuglestvedt, R.J. Millar, M. Cain, D.J. Frame, and A.H. Macey. 2018. A solution to the misrepresentations of CO2-equivalent emissions of short-lived climate pollutants under ambitious mitigation. npj Clim. Atmos. Sci. 1:16. doi:10.1038/s41612-018-0026-8.

Bauman, D.E., and C.L. Davis. 1974. Biosynthesis of milk fat. In:Lactation,Vol.II. B.L. Larson and V.R. Smith, ed. Academic Press., New York, USA. pp.31-75.

Bauman, D.E., and J.M. Griinari. 2001. Regulation and nutritional manipulation of milk fat: Low-fat milk syndrome. Livest. Prod. Sci. 70:15–29. doi:10.1016/S0301-6226(01)00195-6.

Bougouin, A., J.A.D.R.N. Appuhamy, A. Ferlay, E. Kebreab, C. Martin, P.J. Moate, C. Benchaar, P. Lund, and M. Eugène. 2019. Individual milk fatty acids are potential predictors of enteric methane emissions from dairy cows fed a wide range of diets: Approach by meta-analysis. J. Dairy Sci. 102:10616–10631. doi:10.3168/jds.2018-15940.

Breider, I.S., E. Wall, and P.C. Garnsworthy. 2019. Short communication: Heritability of methane production and genetic correlations with milk yield and body weight in Holstein-Friesian dairy cows. J. Dairy Sci. 102:7277–7281. doi:10.3168/jds.2018-15909.

Browning, B.L., Y. Zhou, and S.R. Browning. 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am. J. Hum. Genet. 103:338–348. doi:10.1016/j.ajhg.2018.07.015.

Cantera, S., I. Sánchez-Andrea, L.J. Sadornil, P.A. García-Encina, A.J.M. Stams, and R. Muñoz. 2019. Novel haloalkaliphilic methanotrophic bacteria: An attempt for enhancing methane bio-refinery. J. Environ. Manage. 231:1091–1099. doi:10.1016/j.jenvman.2018.11.017.

Clauss, M., M.T. Dittmann, C. Vendl, K.B. Hagen, S. Frei, S. Ortmann, D.W.H. Müller, S. Hammer, A.J. Munn, A. Schwarm, and M. Kreuzer. 2020. Review: Comparative methane production in mammalian herbivores. Pages S113–S123 in Animal. Cambridge University Press.

Danielsson, R., J. Dicksved, L. Sun, H. Gonda, B. Müller, A. Schnürer, and J. Bertilsson. 2017. Methane Production in Dairy Cows Correlates with Rumen Methanogenic and Bacterial Community Structure. Front. Microbiol. 8:226. doi:10.3389/fmicb.2017.00226.

Difford, G.F., D.R. Plichta, P. Løvendahl, J. Lassen, S.J. Noel, O. Højberg, A.-D.G. Wright, Z. Zhu,

L. Kristensen, H.B. Nielsen, B. Guldbrandtsen, and G. Sahana. 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLOS Genet. 14:e1007580. doi:10.1371/journal.pgen.1007580.

Finlay, J., G. Esteban, K.J. Clarke, A.G. Williams, T.M. Embley, and R.P. Hirt. 1994. Some rumen ciliates have endosymbiotic methanogens. FEMS Microbiol. Lett. 117:157–162. doi:10.1111/j.1574-6968.1994.tb06758.x.

Gianola, D. 1982. Theory and Analysis of Threshold Characters. J. Anim. Sci. 54:1079. doi:10.2527/jas1982.5451079x.

Gianola, D., and J.L. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. Genet Sel Evol 15:201–224.

González-Recio, O., and R. Alenda. 2005. Genetic Parameters for Female Fertility Traits and a Fertility Index in Spanish Dairy Cattle.

Greening, C., R. Geier, C. Wang, L.C. Woods, S.E. Morales, M.J. McDonald, R. Rushton-Green, X.C. Morgan, S. Koike, S.C. Leahy, W.J. Kelly, I. Cann, G.T. Attwood, G.M. Cook, and R.I. Mackie. 2019. Diverse hydrogen production and consumption pathways influence methane production in ruminants. ISME J. 13:2617–2632. doi:10.1038/s41396-019-0464-2.

Guyader, J., M. Eugène, P. Nozière, D.P. Morgavi, M. Doreau, and C. Martin. 2014. Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. Animal 8:1816–1825. doi:10.1017/S1751731114001852.

Janssen, P.H., and M. Kirs. 2008. Structure of the archaeal community of the rumen. Appl. Environ. Microbiol. 74:3619–3625. doi:10.1128/AEM.02812-07.

Jiao, H.P., A.J. Dale, A.F. Carson, S. Murray, A.W. Gordon, and C.P. Ferris. 2014. Effect of concentrate feed level on methane emissions from grazing dairy cows. J. Dairy Sci. 97:7043–7053. doi:10.3168/jds.2014-7979.

Jiménez-Montero, J., D. Gianola, K. Weigel, R. Alenda, and O. González-Recio. 2013. Assets of imputation to ultra-high density for productive and functional traits. J. Dairy Sci. 96:6047–6058. doi:10.3168/jds.2013-6793.

Jin, W., Y.F. Cheng, S.Y. Mao, and W.Y. Zhu. 2011. Isolation of natural cultures of anaerobic fungi and indigenously associated methanogens from herbivores and their bioconversion of lignocellulosic materials to methane. Bioresour. Technol. 102:7925–7931. doi:10.1016/j.biortech.2011.06.026.

Johnson, K.A., and D.E. Johnson. 1995. Methane emissions from cattle. J. Anim. Sci. 73:2483–2492.

Kadarmideen, H.N., R. Thompson, and G. Simm. 2000. Linear and threshold model genetic

parameters for disease, fertility and milk production in dairy cattle. Anim. Sci. 71:411–419. doi:10.1017/S1357729800055338.

Kandel, P.B., M.L. Vanrobays, A. Vanlierde, F. Dehareng, E. Froidmont, N. Gengler, and H. Soyeurt. 2017. Genetic parameters of mid-infrared methane predictions and their relationships with milk production traits in Holstein cattle. J. Dairy Sci. 100:5578–5591. doi:10.3168/jds.2016-11954.

Knapp, J.R., G.L. Laur, P.A. Vadas, W.P. Weiss, and J.M. Tricarico. 2014. Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. J. Dairy Sci. 97:3231–3261. doi:10.3168/JDS.2013-7234.

LeBlanc, S.J., K.D. Lissemore, D.F. Kelton, T.F. Duffield, and K.E. Leslie. 2006. Major advances in disease prevention in dairy cattle. J. Dairy Sci. 89:1267–1279. doi:10.3168/jds.S0022-0302(06)72195-6.

Legarra, A., O.F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65. doi:10.1016/j.livsci.2014.04.029.

Legarra, A., L. Varona, and E. Lopez de Maturana. 2011. Threshold Model 1–33.

Li, B., W.F. Fikse, P. Løvendahl, J. Lassen, M.H. Lidauer, P. Mäntysaari, and B. Berglund. 2018. Genetic heterogeneity of feed intake, energy-corrected milk, and body weight across lactation in primiparous Holstein, Nordic Red, and Jersey cows. J. Dairy Sci. 101:10011–10021. doi:10.3168/jds.2018-14611.

Liu, Q., C. Wang, G. Guo, W.J. Huo, S.L. Zhang, C.X. Pei, Y.L. Zhang, and H. Wang. 2018. Effects of branched-chain volatile fatty acids on lactation performance and mRNA expression of genes related to fatty acid synthesis in mammary gland of dairy cows. Animal 12:2071–2079. doi:10.1017/S1751731118000113.

Lopes, J.C., L.F. de Matos, M.T. Harper, F. Giallongo, J. Oh, D. Gruen, S. Ono, M. Kindermann, S. Duval, and A.N. Hristov. 2016. Effect of 3-nitrooxypropanol on methane and hydrogen emissions, methane isotopic signature, and ruminal fermentation in dairy cows. J. Dairy Sci. 99:5335–5344. doi:10.3168/jds.2015-10832.

López-Paredes, J., I. Goiri, R. Atxaerandio, A. García-Rodríguez, E. Ugarte, J.A. Jiménez-Montero, R. Alenda, and O. González-Recio. 2020. Mitigation of greenhouse gases in dairy cattle via genetic selection: 1. Genetic parameters of direct methane using noninvasive methods and proxies of methane. J. Dairy Sci. 103:7199–7209. doi:10.3168/jds.2019-17597.

Martin, C., D.P. Morgavi, and M. Doreau. 2010. Methane mitigation in ruminants: from microbe to the farm scale. Animal 4:351–365. doi:10.1017/S1751731109990620.

Martínez-Álvaro, M., M.D. Auffret, R.D. Stewart, R.J. Dewhurst, C.-A. Duthie, J.A. Rooke, R.J.

Wallace, B. Shih, T.C. Freeman, M. Watson, and R. Roehe. 2020. Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. Front. Microbiol. 11:659. doi:10.3389/fmicb.2020.00659.

Moate, P.J., J.L. Jacobs, M.C. Hannah, G.L. Morris, K.A. Beauchemin, P.S. Alvarez Hess, R.J. Eckard, Z. Liu, S. Rochfort, W.J. Wales, and S.R.O. Williams. 2018. Adaptation responses in milk fat yield and methane emissions of dairy cows when wheat was included in their diet for 16 weeks. J. Dairy Sci. 101:7117–7132. doi:10.3168/jds.2017-14334.

Moraes, L.E., A.B. Strathe, J.G. Fadel, D.P. Casper, and E. Kebreab. 2014. Prediction of enteric methane emissions from cattle. Glob. Chang. Biol. 20:2140–2148. doi:10.1111/gcb.12471.

Negussie, E., Y. de Haas, F. Dehareng, R.J. Dewhurst, J. Dijkstra, N. Gengler, D.P. Morgavi, H. Soyeurt, S. van Gastelen, T. Yan, and F. Biscarini. 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.. J. Dairy Sci. 100:2433–2453. doi:10.3168/jds.2016-12030.

Niu, M., E. Kebreab, A.N. Hristov, J. Oh, C. Arndt, A. Bannink, A.R. Bayat, A.F. Brito, T. Boland, D. Casper, L.A. Crompton, J. Dijkstra, M.A. Eugène, P.C. Garnsworthy, M.N. Haque, A.L.F. Hellwing, P. Huhtanen, M. Kreuzer, B. Kuhla, P. Lund, J. Madsen, C. Martin, S.C. McClelland, M. McGee, P.J. Moate, S. Muetzel, C. Muñoz, P. O'Kiely, N. Peiren, C.K. Reynolds, A. Schwarm, K.J. Shingfield, T.M. Storlien, M.R. Weisbjerg, D.R. Yáñez-Ruiz, and Z. Yu. 2018. Prediction of enteric methane production, yield, and intensity in dairy cattle using an intercontinental database. Glob. Chang. Biol. 24:3368–3389. doi:10.1111/gcb.14094.

Poulsen, M., C. Schwab, B. Borg Jensen, R.M. Engberg, A. Spang, N. Canibe, O. Højberg, G. Milinovich, L. Fragner, C. Schleper, W. Weckwerth, P. Lund, A. Schramm, and T. Urich. 2013. Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. Nat. Commun. 4:1–7. doi:10.1038/ncomms2432.

Pszczola, M., M.P.L. Calus, and T. Strabel. 2019. Short communication: Genetic correlations between methane and milk production, conformation, and functional traits. J. Dairy Sci. 102:5342–5346. doi:10.3168/jds.2018-16066.

Rey, J., R. Atxaerandio, R. Ruiz, E. Ugarte, O. González-Recio, A. Garcia-Rodriguez, and I. Goiri. 2019. Comparison Between Non-Invasive Methane Measurement Techniques in Cattle. Animals 9:1–9. doi:10.3390/ani9080563.

Saborío-Montero, A., M. Gutiérrez-Rivas, A. García-Rodríguez, R. Atxaerandio, I. Goiri, E. López de Maturana, J.A. Jiménez-Montero, R. Alenda, and O. González-Recio. 2020. Structural

equation models to disentangle the biological relationship between microbiota and complex traits: Methane production in dairy cattle as a case of study. J. Anim. Breed. Genet. 137. doi:10.1111/jbg.12444.

Saborío-Montero, A., M. Gutiérrez-Rivas, A. López-García, A. García-Rodríguez, R. Atxaerandio, I. Goiri, J.A. Jiménez-Montero, and O. González-Recio. 2021a. Holobiont effect accounts for more methane emission variance than the additive and microbiome effects on dairy cattle. Livest. Sci. 250:104538. doi:10.1016/j.livsci.2021.104538.

Saborío-Montero, A., A. López-García, M. Gutiérrez-Rivas, R. Atxaerandio, I. Goiri, A. García-Rodriguez, J.A. Jiménez-Montero, C. González, J. Tamames, F. Puente-Sánchez, L. Varona, M. Serrano, C. Ovilo, and O. González-Recio. 2021b. A dimensional reduction approach to modulate the core ruminal microbiome associated with methane emissions via selective breeding. J. Dairy Sci. 104:8135–8151. doi:10.3168/jds.2020-20005.

Saborío-Montero, A., B. Vargas-Leitón, J. Romero-Zúñiga, and J. Camacho-Sandoval. 2018. Additive genetic and heterosis effects for milk fever in a population of Jersey, Holstein x Jersey, and Holstein cattle under grazing conditions. J. Dairy Sci. 101:9128–9134. doi:10.3168/jds.2017-14234.

Saborío-Montero, A., B. Vargas-Leitón, J.J. Romero-Zúñiga, and J.M. Sánchez. 2017. Risk factors associated with milk fever occurrence in grazing dairy cattle. J. Dairy Sci. 100:9715–9722. doi:10.3168/jds.2017-13065.

Shinzato, N., K. Takeshita, and Y. Kamagata. 2010. Methanogenic and Bacterial Endosymbionts of Free-Living Anaerobic Ciliates. J.H.P. Hackstein, ed. (Endo)symbiotic Methanogenic Archaea. Springer Berlin Heidelberg., Berlin, Heidelberg. 37-54.

Spatafora, J.W., Y. Chang, G.L. Benny, K. Lazarus, M.E. Smith, M.L. Berbee, G. Bonito, N. Corradi, I. Grigoriev, A. Gryganskyi, T.Y. James, K. O'Donnell, R.W. Roberson, T.N. Taylor, J. Uehling, R. Vilgalys, M.M. White, and J.E. Stajich. 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia 108:1028–1046. doi:10.3852/16-042.

Urrutia, N.L., and K.J. Harvatine. 2017. Acetate Dose-Dependently Stimulates Milk Fat Synthesis in Lactating Dairy Cows. J. Nutr. 147:763–769. doi:10.3945/jn.116.245001.

VanRaden, P.M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

Wallace, R.J., J.A. Rooke, N. McKain, C.A. Duthie, J.J. Hyslop, D.W. Ross, A. Waterhouse, M. Watson, and R. Roehe. 2015. The rumen microbial metagenome associated with high methane production in cattle. BMC Genomics 16:1–14. doi:10.1186/s12864-015-2032-0.

Williams, S.R.O., M.C. Hannah, J.L. Jacobs, W.J. Wales, and P.J. Moate. 2019. Volatile Fatty Acids in Ruminal Fluid Can Be Used to Predict Methane Yield of Dairy Cows. Animals 9:1006. doi:10.3390/ani9121006.

Yin, T., T. Pinent, K. Brügemann, H. Simianer, and S. König. 2015. Simulation, prediction, and genetic analyses of daily methane emissions in dairy cattle. J. Dairy Sci. 98:5748–5762. doi:10.3168/jds.2014-8618.

Zhao, Y., S. Yan, L. Chen, B. Shi, and X. Guo. 2019. Effect of interaction between leucine and acetate on the milk protein synthesis in bovine mammary epithelial cells. Anim. Sci. J. 90:81–89. doi:10.1111/asj.13125.

CHAPTER 4

STRUCTURAL EQUATION MODELS TO DISENTANGLE THE BIOLOGICAL RELATIONSHIP BETWEEN MICROBIOTA AND COMPLEX TRAITS: METHANE PRODUCTION IN DAIRY CATTLE AS A CASE OF STUDY

Alejandro Saborío-Montero[1,2], Mónica Gutiérrez-Rivas[1], Aser García-Rodríguez[3], Raquel Atxaerandio[3], Idoia Goiri[3], Evangelina López de Maturana[4], José Antonio Jiménez-Montero[5], Rafael Alenda[6] and Oscar González-Recio[1,6]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain; [2]Universidad de Costa Rica, Escuela de Zootecnia, 11501 San Pedro de Montes de Oca, San José, Costa Rica; [3]Departamento de Producción Animal, NEIKER-Tecnalia, Vitoria-Gasteiz, Spain; [4]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), and CIBERONC, Melchor Fernández Almagro 3, 28029 Madrid, Spain. [5]Departamento técnico de Confederación de Asociaciones de Frisona Española (CONAFE) Ctra. de Andalucía, km. 23,600 - 28340 Valdemoro, Madrid, Spain. [6]Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain.

Published in Journal of Animal Breeding and Genetics.
DOI: https://doi.org/10.1111/jbg.12444

ABSTRACT

The advent of metagenomics in animal breeding poses the challenge of statistically modeling the relationship between the microbiome, the host genetics and relevant complex traits. A set of structural equation models (SEMs) of a recursive type within a Markov chain Monte Carlo (MCMC) framework was proposed here to jointly analyze the host-metagenome-phenotype relationship. A non-recursive bivariate model was set as benchmark to compare the recursive model. The relative abundance of rumen microbes (RA), methane concentration (CH_4) and the host genetics was used as a case of study. Data were from 337 Holstein cows from 12 herds in the north and north-west of Spain. Microbial composition from each cow was obtained from whole metagenome sequencing of ruminal content samples using a MinION device from Oxford Nanopore Technologies. Methane concentration was measured with Guardian® NG infrared gas monitor from Edinburgh Sensors during cow's visits to the milking automated system. A quarterly average from the methane eructation peaks for each cow was computed and used as phenotype for CH_4. Heritability of CH_4 was estimated at 0.12 ± 0.01 in both, the recursive and bivariate, models. Likewise, heritability estimates for the relative abundance of the taxa overlapped between models and ranged between 0.08 and 0.48. Genetic correlations between the microbial composition and CH_4 ranged from -0.76 to 0.65 in the non-recursive bivariate model and from -0.68 to 0.69 in the recursive model. Regardless of the statistical model used, positive genetic correlations with methane were estimated consistently for the 7 genera pertaining to the Ciliophora phylum, as well as for those genera belonging to the Euryarchaeota (*Methanobrevibacter sp*.), Chytridiomycota (*Neocallimastix sp.*) and Fibrobacteres (*Fibrobacter sp*.) phyla. These results suggest that rumen's whole metagenome recursively regulate methane emissions in dairy cows, and that both CH_4 and the microbiota compositions are partially controlled by the host genotype. Recursive models are an interesting approach to disentangle this complex relationship.

Keywords: Methane emissions, recursive model, structural equations models, metagenome, genetic correlation.

INTRODUCTION

The collection of genes and genomes from the members of a microbiota are defined as the metagenome. Its association with complex traits of interest has been previously suggested (Marchesi and Ravel, 2015). The metagenome is gaining a crescent interest in animal production worldwide for different purposes on several species (Thompson et al., 2017). Previous studies have analyzed the association between the metagenome in ruminants (sheep, goat, beef cattle, dairy cattle, tammar wallaby) and different traits such as methane emissions or feed efficiency (Tapio et al., 2017). The relationship between the metagenome and the host genotype has also been proposed (Roehe et al., 2016; Camarinha-Silva et al., 2017; Gonzalez-Recio, Zubiria, García-Rodríguez, Hurtado, & Atxaerandio, 2017). Previous studies have analyzed the relationship between the metagenome and the phenotype but ignoring any recursive phenotypic relationship that might exists between the involved traits (Difford et al., 2018; Buitenhuis et al., 2019). Analyzing the host effect on the microbiota and a complex trait of interest simultaneously, accounting for a recursive plausible effect between them, might be useful because it could lead to more realistic estimates, hence refining the conclusions drawn from these studies. Statistical complexity and limited availability of software capable to deal with this kind of relationship are among the reasons of the lack of studies considering this joint relationship. Thus, there is the need to consider simultaneously the host-metagenome-phenotype relationships. Structural equation models (SEM) could fill the gap in this area. They allow representing recursive or simultaneous mechanisms among phenotypic traits, while provide inference on the magnitude of those relationships (Rosa et al., 2011; Valente and de Magalhães Rosa, 2013). In SEM a trait can be used as a predictor of some other trait, assuming a functional link between them (Rosa et al., 2011).

Applying SEM in this context may contribute in elucidating the host-metagenome-phenotype relationships. They also assist on clarifying the direct and indirect effects of the microbiota on relevant complex traits. This relationship can be either recursive or simultaneous: a recursive system describes an unidirectional effect between traits, with one trait affecting the other without feedback from the latter. A simultaneous system assumes a feedback effect between variables, with changes of a quantity indirectly influencing the quantity itself (Gianola and Sorensen, 2004). Although SEM has been extensively used in

animal breeding (De Los Campos et al., 2006; Wu et al., 2007; Konig et al., 2008; López de Maturana et al., 2008; Heringstad et al., 2009; Rosa et al., 2011; De Los Campos et al., 2014), their application to host-metagenome studies has not been yet referenced.

We hypothesized that there is a recursive relationship in which the microbial composition affects the methane production (CH_4) and both are under the host genetic control. Here enteric methane is used as a case of study, which is affected by the microbial digestion fermentation that occur in the rumen. Both CH_4 and the relative abundance (RA) of the rumen microbes are partially controlled by the host genetics. Hence, it is rational to modelize a recursive effect from RA on CH_4, without a feedback from CH_4 over the microbial RA. The two traits recursive system include environmental (*E*) and genetic (*U*) effects and could be represented schematically as in Figure 1, where a structural coefficient $\lambda_{CH_4 \leftarrow RA}$ describes the rate of change of methane emissions with respect to the RA of a given genus in the metagenome. Under this assumption, RA has a direct recursive effect on CH_4, which corresponds to the structural coefficient $\lambda_{CH_4 \leftarrow RA}$. (López De Maturana et al., 2008).

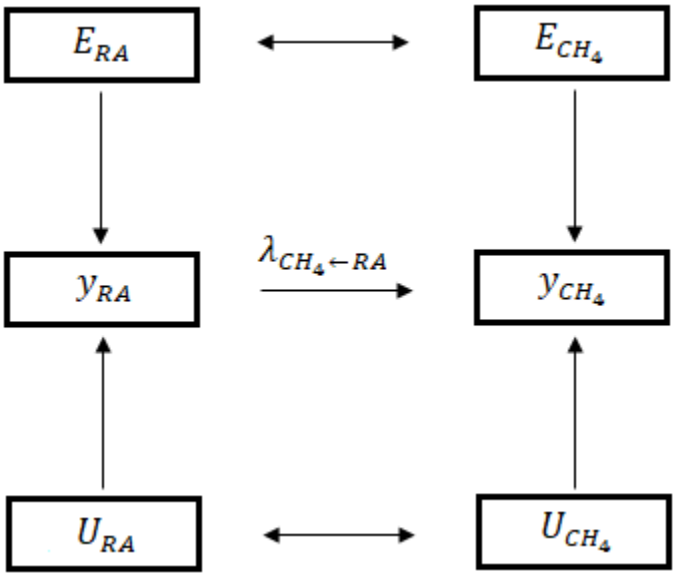

Figure 1. Recursive model in a two trait system between metagenome relative abundance (y_{Micr}) and methane emissions(y_{CH_4}) phenotypes. U_{Micr} and U_{CH_4} are additive genetic effects acting on the system; E_{Micr} and E_{CH_4} are residual effects. The single head arrows indicate which variable is being affected (e.g., $y_{Micr} \rightarrow y_{CH_4}$ indicates that y_{CH_4} is affected by y_{Micr}). The $\lambda_{CH_4 \leftarrow Micr}$ is a structural coefficient that indicate the rate of change of variable y_{CH_4} with respect to variable y_{Micr}. Double headed arrows denote correlations between pairs of variables.

This study plains to implement recursive SEM to analyze the host-metagenome-phenotype relationship, and to evaluate their behavior using methane emissions in dairy cattle as a case of study. Traditional bivariate models were used as benchmark.

MATERIALS AND METHODS

This study was carried out in accordance with Spanish Royal Decree 53/2013 for the protection of animals used for experimental and other scientific purposes, and was approved by the Basque Institute for Agricultural Research and Development Ethics Committee (Neiker-OEBA-2017-004) on March 28, 2017.

Data

Four hundred and sixteen lactating Friesian Holstein cows (primiparous or second lactation) from 12 commercial farms at regions of País Vasco, Gerona and Navarra in Spain were included in this study. Methane emissions (ppm) were measured using a non-dispersive infrared methane detector (The Guardian® NG infrared gas monitor from Edinburg Sensors; measure range 0-1%), installed within the feed bin of the automatic milking system (AMS). Methane concentration in breath samples was measured individually for each cow, during the milking time at each cow visit to the AMS for 2-3 weeks periods. Eructation peaks recorded were averaged to obtain a single record per cow. During this sample period herds underwent a test-day recording as part of the official test day recording scheme. All traits related to milk yield and composition were also available.

During the ruminal content sampling, cows were placed in individual stalls. A custom made mechanical device was used to raise the snout of the animal. Samples of ruminal content (approximately 100 ml) were extracted from each cow by introducing a stomach tube (18 mm diameter and 160 mm long) orally through the esophagus connected to a mechanical pumping unit (Vacubrand ME 2SI, Wertheim, Germany) with a 1000 ml Erlenmeyer trapped in-between. Samples were then stored in a sterilized container. Hose and all material in contact with the samples were systematically washed between cows. Samples were filtered through 4 layers of sterile cheesecloth, in order to remove the solid fraction and the filtered

fraction was frozen in liquid nitrogen (N_2) vapors immediately after. Then frozen samples were transported to the laboratory in liquid N_2 containers and stored at −80 °C until analysis.

The samples were thawed, and then homogenized in a blender. The DNA extraction was performed using 250 µl from the homogenized samples with the commercial "DNeasy Power Soil Kit" (QIAGEN, Valencia, CA, USA). The genomic DNA concentrations and their purity were measured by spectrophotometry using a Nanodrop ND-1000 UV/Vis spectrophotometer (Nanodrop Technologies Inc., DE, USA) and also the Qubit fluorometer (ThermoFisher Scientific, 150 Waltham, MA, USA). 1 µg of DNA sample were analyzed following the 1D Native barcoding genomic DNA (with EXP-NBD104, EXP-NBD114) and ligation sequencing kit (SQK-LSK109) protocol from Oxford Nanopore, using the MinION sequencer. This device uses the variation in an ionic current to perform the base calling. This current is distinctively generated depending on which nucleotide from a DNA simple strand passes through the protein-based nanopores in a flow cell.

Sequences were analyzed using the open-source program software DIAMOND (double index alignment of next-generation sequencing data), to exhaustively find all significant alignments. This software is based on double indexing to determine the list of all seeds and their locations in both the query and the protein reference sequences database; the two lists are sorted lexicographically and traversed together to obtain all matching seeds and their corresponding locations (Buchfink et al., 2014). Next, MEGAN software was used to compute the taxonomical content of the data set, using the National Center for Biotechnology Information (NCBI) taxonomy database to assign reads to a taxonomic category (Huson et al., 2007).

Cows were genotyped using the EURO12K SNP chip from Illumina, and then imputed to 54609 SNPs (Bovine 50k SNP chip, Illumina, San Diego, California, USA) using BEAGLE software (http://faculty.washington.edu/browning/beagle/beagle.html), using the Spanish reference population provided by CONAFE (Spanish Friesian Associations Confederation), as described by Jiménez-Montero, Gianola, Weigel, Alenda, & González-Recio (2013). Monomorphic SNPs and those with MAF<0.05 were filtered out from the analysis.

Grouping and filtering of zero inflated compositional data

The initial metagenome dataframe contained nucleotide counts linked to 627 taxonomically classified taxa in the 416 samples. Data were summarized at the genus level, resulting in the clustering of 134 genera. Counts in each sample were closured to 100 in order to obtain a relative abundance (%) at the genus level. Then, genera that were not present in at least 70% of the samples were excluded from the analysis, as well as samples that lacked reads from more than 50% of the genera. This hard filter reduced the dataframe to 24 genus plus a taxonomically unknown group in 337 samples. The remaining zeros were assumed to be count zeros and were then imputed using a Bayesian-multiplicative replacement of count zeros through a geometric Bayesian Multiplicative method with the *cmultRepl* function of the *zCompositions* package (https://cran.r-project.org/web/packages/zCompositions/zCompositions.pdf) in R environment.

Genomic relationship matrix

A genomic relationship matrix (GRM), between individuals *j* and *k* was built following method 2 of VanRaden (2008) and Yang et al. (2010) with the following formula:

$$G_{jk} = \begin{cases} \frac{1}{N}\sum_i \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1-p_i)}, j \neq k \\ 1 + \frac{1}{N}\sum_i \frac{x_{ij}^2 - (1+2p_i)\,x_{ij} + 2p_i^2}{2p_i(1-p_i)}, j = k \end{cases} \quad \textbf{[1]}$$

Where x_i refers to the AA, Aa and aa SNP genotypes, coded as 2, 1, and 0, respectively, of individual *j* or *k* at locus *i* (*i* = 1, …, *N*), with *N* being the number of SNP (42372) and p_i being the SNP allele frequency in the whole genotyped population. This matrix depicted the genetic relationships within individuals (diagonal elements) and between individuals (out-diagonal elements), where elements in the diagonal were close to one with some deviation according to the difference between the observed and the expected consanguinity. Out diagonal elements varied according to the genomic similarity between individuals.

Non-recursive bivariate model

A non-recursive bivariate model within a Bayesian framework was set as benchmark for variance components estimation for RA and CH_4 as follow:

$$\mathbf{y = Xb + Z_{(h)}h + Z_{(u)}u + e}, \qquad \textbf{[2]}$$

where: **y** is a 2 × n vector containing the observed RA and the observed CH_4 concentration for the *i*th individual (cow); **b** is a vector including effects on parity (2 levels), **h** is a vector of herd-robot effect (20 levels); **u** is a 2 × n vector of genetic effects (337 levels); and **e** is a 2 × n vector of residuals. Then, **X** and **Z** are incidence matrices of appropriate order, with **b** distributed as uniform (-9999, 9999), **h** ~N (0, $\mathbf{I}\sigma_h^2$), **u** ~N(0, $\mathbf{G}\sigma_u^2$), and **e** ~N (0,$\mathbf{I}\sigma_e^2$), where **I** is an identity matrix of appropriate order and **G** is the genomic relationship matrix between cows.

The heritability for each trait (RA or CH_4) was calculated as:

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_h^2 + \sigma_e^2} \qquad \textbf{[3]}$$

Where σ_u^2 is the additive genetic variance of the analysed trait, $\sigma_h^2$is the herd-robot variance and σ_e^2 is the residual variance.

The genetic correlation between RA and CH_4 was computed as:

$$Corr_{u_{RA}u_{CH_4}} = \frac{\sigma_{u_{RA}u_{CH_4}}}{\sqrt{\sigma_{u_{RA}} * \sigma_{u_{CH_4}}}} \qquad \textbf{[4]}$$

Where $\sigma_{u_{RA}u_{CH_4}}$ is the additive genetic covariance between RA and CH_4, $\sigma_{u_{RA}}$ is the additive genetic variance for RA and $\sigma_{u_{CH_4}}$ is the additive genetic variance for CH_4.

Recursive Gaussian Structural Equation Models

A set of recursive Gaussian SEM was used to analyze the relationship between RA for each taxa and CH_4. Assuming a joint multivariate normal distribution for RA and CH_4, the observed data were modeled as

$$\mathbf{\Lambda y = Xb + Z_{(h)}h + Z_{(u)}u + e,} \quad \textbf{[5]}$$

where

$$\mathbf{y \mid \Lambda, b, h, u, R_0 \sim \mathit{N}(\Lambda^{-1}(Xb + Z_{(h)}h + Z_{(u)}u), \Lambda^{-1} R_0 \Lambda'^{-1}).} \quad \textbf{[6]}$$

Here, $\mathbf{\Lambda}$ is the matrix of the structural coefficient for RA, and it takes the form

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ -\lambda_{CH_4 \leftarrow RA} & 1 \end{bmatrix} \quad \textbf{[7]}$$

In [5] above, all other terms are as described in [2].

Genetic effects for the two traits were assumed to be distributed, *a priori*, as a multivariate normal, with a null mean vector and a (co)variance matrix $\mathbf{K_0} \otimes \mathbf{G}$, where **G** is a GRM between individuals of order 337 and

$$\mathbf{K_0} = \begin{pmatrix} \sigma^2_{u_{RA}} & \sigma_{u_{RA}u_{CH_4}} \\ symmetric & \sigma^2_{u_{CH_4}} \end{pmatrix} \quad \textbf{[8]}$$

In [8], $\sigma^2_{u_{RA}}$ is the genetic variance for the RA of a given taxa, $\sigma^2_{u_{CH_4}}$ is the variance between animals for CH_4, and $\sigma_{u_{RA}u_{CH_4}}$ is the additive genetic covariance between CH_4 and RA. A random effect of herd-robot combination (**h**) on the two traits were assumed to follow a multivariate normal distribution with null mean vector and (co)variance matrix $\mathbf{H_0} \otimes \mathbf{I}$, where

$$\mathbf{H_0} = \begin{pmatrix} \sigma^2_{h_{RA}} & \sigma_{h_{RA}h_{CH_4}} \\ symmetric & \sigma^2_{h_{CH_4}} \end{pmatrix} \quad \textbf{[9]}$$

Above $\sigma^2_{h_{RA}}$ is the herd-robot variance on RA, $\sigma^2_{h_{CH_4}}$ is the herd-robot variance on CH_4 and $\sigma_{h_{RA}h_{CH_4}}$ is the covariance between herd-robot effects on CH_4. **I** is an identity matrix of corresponding order.

The residual vector **e** was assumed to follow a multivariate normal distribution with mean vector 0 and (co)variance matrix $\mathbf{R_0} \otimes \mathbf{I}$, where

$$\mathbf{R_0} = \begin{pmatrix} \sigma^2_{e_{RA}} & \sigma_{e_{RA}e_{CH_4}} \\ symmetric & \sigma^2_{e_{CH_4}} \end{pmatrix} \quad \textbf{[10]}$$

And **I** is an identity matrix of corresponding order.

Markov chain Monte Carlo (MCMC) and posterior analysis

Denote $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \mathbf{b}, \mathbf{h}, \mathbf{u}, \mathbf{K}_0, \mathbf{R_0})$, let $\mathbf{y}_{CH_4}$ be the vector denoting the observed CH_4 for the 337 cows and for the sake of simplicity ignore hyperparameters in the notation. The joint posterior density of $\boldsymbol{\theta}$ and CH_4 is given by

$$p(\mathbf{y}_{CH_4}, \boldsymbol{\theta} \mid \mathbf{y}_{RA}) \propto p(\mathbf{y}_{CH_4}, \mathbf{y}_{RA}, \boldsymbol{\theta}) = p(\mathbf{y}_{RA}, \mathbf{y}_{CH_4} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}). \quad \textbf{[11]}$$

The last term is the prior density of the unknown parameters, and it can be factorized as:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\Lambda})\, p(\mathbf{b})\, p(\mathbf{h} \mid \mathbf{H_0})\, p(\mathbf{h})\; p(\mathbf{u} \mid \mathbf{K}_0)\, p(\mathbf{u})\, p(\mathbf{R_0}). \quad \textbf{[12]}$$

A uniform prior distribution (-9999, 9999) was assigned to the structural coefficient vector and the systematic effect **b**. Herd-robot effect was assumed to be distributed as, $\mathbf{h} \sim N(0, \mathbf{I} \otimes \mathbf{H_0})$; and animal genetic effect $\mathbf{u} \sim N(0, \mathbf{G} \otimes \mathbf{K}_0)$ so that their fully conditional posterior distribution were also normal.

Inferences on the parameters in the model were through an MCMC approach using a modified version of the TM package (Legarra et al., 2011). After preliminary runs, visual examination of trace plots, and additional diagnostic assessments, the length of the chain was set to 300,000 iterations with a burn-in of 100,000 iterations; subsequently one of each 10 successive samples was retained in order to reduce autocorrelation between samples. A total of 20,000 samples were used to infer the posterior distributions of the unknown parameters. The autocorrelation between samples, the effective sample size and convergence diagnosis were obtained using the *coda* package (https://cran.r-project.org/web/packages/coda/coda.pdf) in R environment.

Probability intervals were estimated as

$$\overline{\boldsymbol{\theta}} \pm 1.96\, \sigma/\sqrt{m}\,, \qquad \textbf{[13]}$$

where $\overline{\boldsymbol{\theta}}$ is the estimated mean from the *a posteriori* distribution, σ is the standard deviation from the *a posteriori* distribution, and m is the effective sample size corrected by autocorrelation between samples.

Genetic and structural parameters

In recursive models, location and dispersion parameters need specific calculations to be done in order to be comparable with spatial mixed models. In our case, a transformation to the $\mathbf{K}_0$, $\mathbf{H_0}$ and $\mathbf{R_0}$ matrices was implemented using the matrix of structural coefficients $\boldsymbol{\Lambda}$ (Sorensen and Gianola, 2002):

$$\mathbf{K}^* = \boldsymbol{\Lambda}^{-1}\mathbf{K}_0\boldsymbol{\Lambda}^{'-1} \qquad \textbf{[14]}$$

$$\mathbf{H}^* = \boldsymbol{\Lambda}^{-1}\mathbf{H}_0\boldsymbol{\Lambda}^{'-1} \qquad \textbf{[15]}$$

$$\mathbf{R}^* = \boldsymbol{\Lambda}^{-1}\mathbf{R}_0\boldsymbol{\Lambda}^{'-1} \qquad \textbf{[16]}$$

Heritabilities and genetic correlations were computed based on $\mathbf{K}^*$, $\mathbf{H}^*$ and $\mathbf{R}^*$.

Posterior analysis of structural coefficients

The rate of change of CH_4 with respect to RA [$\lambda_{CH_4 \leftarrow RA}$] indicates the expected change in CH_4 by the increment of 1 unit in RA. If $\lambda_{CH_4 \leftarrow RA}$ is positive, an increase of RA in 1% will increase CH_4 by $\lambda_{CH_4 \leftarrow RA}$ units.

RESULTS

Descriptive statistics

The descriptive statistics for days in milk (DIM), body weight, milk yield, milk fat, milk protein and milk lactose from the Spanish official control of dairy performance as well as methane emissions from cows included in the study are summarized in Table 1.

Table 1. Descriptive statistics of quantitative variables from 337 Spanish Holstein cows.

Variable		Mean	S.D. †	C.V. ‡ (%)
DIM§	(days)	164	80	48.8
Body weight	(kg)	592	37	6.3
Milk yield	(kg)	33.6	7.6	22.6
Milk fat	(%)	3.59	0.84	23.4
Milk protein	(%)	3.29	0.42	12.8
Lactose	(%)	4.74	1.07	22.6
CH_4¶	(ppm)	853	278	32.6

†S.D. = Standard deviation, ‡ C.V. = Coefficient of variation, §DIM = Days in milk, ¶CH_4 = Methane concentration.

Recorded values of dairy performance are within normal records for the Holstein breed in Spain. Methane concentration was estimated at 853 ppm (±278).

Microbiota recursive effect on methane

In total, taxa from this study belonged to 8 phyla. The highest magnitude values for $\lambda_{CH_4 \leftarrow RA}$were in genus with low relative abundances. In general, lower values for $\lambda_{CH_4 \leftarrow RA}$ were obtained for genera with larger RA (Table 2). The 7 genera from the Ciliophora phylum

had positive $\lambda_{CH_4 \leftarrow RA}$ as well as some genera from the Euryarchaeota (*Methanobrevibacter sp.*), Chytridiomycota (*Neocallimastix sp.*) and Fibrobacteres (*Fibrobacter sp.*) phyla. The Proteobacteria and Spirochaetes phyla had negative $\lambda_{CH_4 \leftarrow RA}$. Most genera from Firmicutes phylum had positive $\lambda_{CH_4 \leftarrow RA}$. Two genera (*Prevotella* and *Paraprevotella sp.*), from the *Bacteroidetes* phylum, had positive $\lambda_{CH_4 \leftarrow RA}$. *Bacteroides* and *Alistipes sp.* resulted in negative $\lambda_{CH_4 \leftarrow RA}$.

Table 2. Phylum, genus, relative abundance of genus (RA), rate of change of methane (ppm) with respect to RA $[\lambda_{CH_4 \leftarrow RA}]$ and standard error of prediction (SEP) of $[\lambda_{CH_4 \leftarrow RA}]$ according to genus from the rumen content of 337 Holstein cows.

Phylum	Genus	RA (%)	$[\lambda_{CH_4 \leftarrow RA}]$(ppm)	SEP (ppm)
Bacteroidetes	*Paraprevotella sp.*	0.19	284	14.5
Ciliophora	*Pseudocohnilembus sp.*	0.29	138	8.5
Chytridiomycota	*Neocallimastix sp.*	0.44	70	7.6
Ciliophora	*Ichthyphthirius sp.*	0.38	61	8.2
Ciliophora	*Stylonychia sp.*	0.80	60	6.7
Ciliophora	*Tetrahymena sp.*	0.78	59	6.5
Ciliophora	*Oxytricha sp.*	0.81	58	6.7
Firmicutes	*Butyrivibrio sp.*	0.82	49	12.2
Euryarchaeota	*Methanobrevibacter sp.*	0.21	46	12.6
Ciliophora	*Paramecium sp.*	0.74	46	6.9
Ciliophora	*Stentor sp.*	0.93	44	6.3
Firmicutes	*Succiniclasticum sp.*	0.65	28	7.8
Firmicutes	*Mycoplasma sp.*	0.25	20	12.7
Firmicutes	*Ruminococcus sp.*	1.85	10	9.6
Fibrobacteres	*Fibrobacter sp.*	3.02	6	4.4
Bacteroidetes	*Prevotella sp.*	66.87	1	2.8
Firmicutes	*Selenomonas sp.*	0.94	-4	8.2
Firmicutes	*Clostridium sp.*	3.81	-7	4.6
Proteobacteria	*Succinimonas sp.*	0.18	-8	10.9
Unknow	*Unknow sp.*	8.35	-10	4.3
Spirochaetes	*Treponema sp.*	1.79	-13	5.5
Firmicutes	*Eubacterium sp.*	0.63	-21	10.0
Bacteroidetes	*Bacteroides sp.*	4.36	-25	6.9
Bacteroidetes	*Alistipes sp.*	0.68	-28	14.9
Proteobacteria	*Succinivibrio sp.*	0.23	-63	10.2

Host genetic effect on methane production

The non-recursive bivariate models were set as the benchmarks to compare the parameters estimation obtained with the recursive models across all the genera. The CH_4 heritability ($h^2_{CH_4}$), calculated as the average of the posterior means and the average Monte Carlo standard errors across all the models, was estimated at 0.12 ± 0.01 for the non-recursive bivariate model. The same $h^2_{CH_4}$ (0.12 ± 0.01) was estimated for the recursive model. Individual estimates obtained in each bivariate analysis are provided in Figure 2.

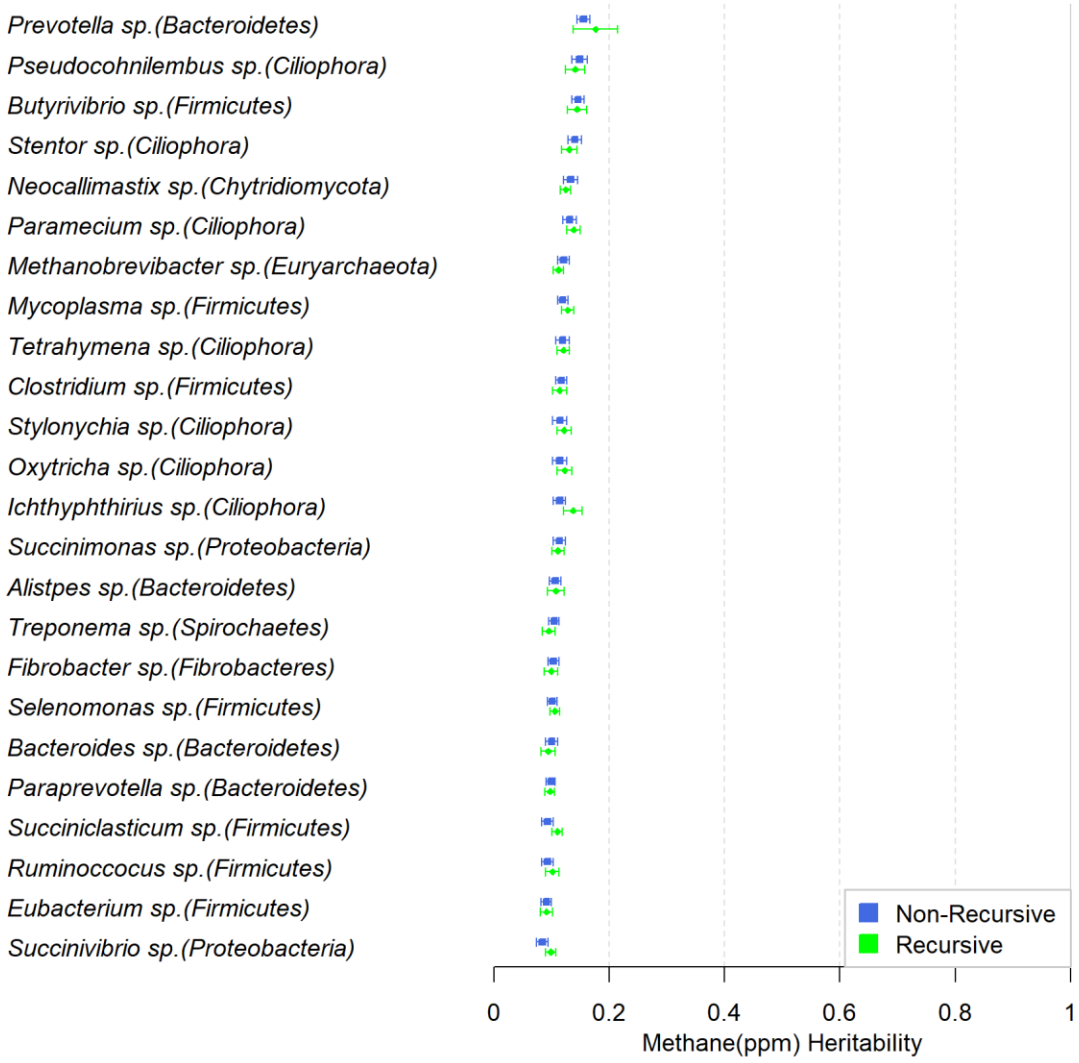

Figure 2. Heritability of methane concentration (ppm) according to statistical model (Non-Recursive vs Recursive) for each bi-trait model according to genus (phylum within parenthesis), estimated from the mean of the stationary posterior distribution. The bars show the 95% probability intervals around the heritability estimates using the Monte Carlo standard error.

Similar $h^2_{CH_4}$ were obtained across all the set of models from both statistical approaches, ranging from 0.08 to 0.16 in the non-recursive bivariate model and from 0.09 to 0.18 in the recursive model. In this case the recursive model did not resulted in different $h^2_{CH_4}$ estimate.

Host genetic effect on the microbiota composition

The RA heritability (h^2_{RA}) estimates, obtained from the posterior mean and the Monte Carlo standard error, are shown in Figure 3. They ranged between 0.08 and 0.46, using the non-recursive bivariate model, with an average of 0.25±0.01 and between 0.08 and 0.48 for the recursive model which also averaged 0.25±0.01. As expected similar heritabilities for RA were obtained as the structural parameter do not affect the heritability estimate of the recursive trait. The larger h^2_{RA} was estimated for *Prevotella sp., Butyrivibrio sp*. and *Mycoplasma sp*. (0.34 - 0.48). The lowest h^2_{RA} were obtained for *Treponema sp*. and *Fibrobacter sp*. (0.08 - 0.10).

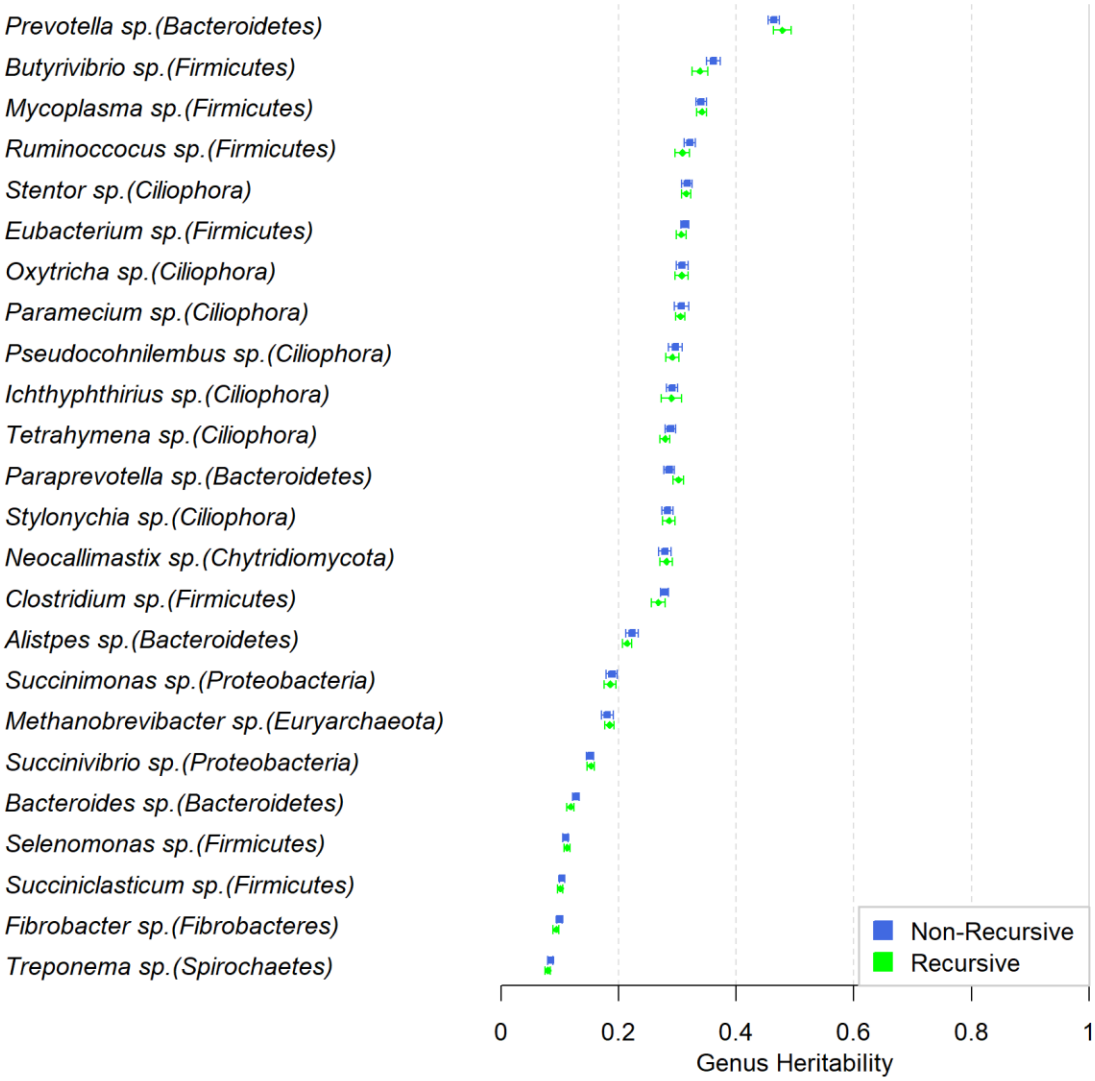

Figure 3. Heritability of microbial genus relative abundance (%) according to statistical model (Non-Recursive vs Recursive) from each bi-trait model according to genus (phylum within parenthesis), estimated from the mean of the stationary posterior distribution. The bars show the 95% probability intervals around the heritability estimates using the Monte Carlo standard error.

The analyzed phyla, sorted from higher to lower averages within phyla of the h^2_{KA}estimates in both models were: Ciliophora (0.30), Chytridiomycota (0.28), Bacteroidetes (0.28), Firmicutes (0.26), Euryarchaeota (0.18), Proteobacteria (0.17), Fibrobacteres (0.10) and Spirochaetes (0.08).

Genetic correlations between CH4 and microbiota composition

The genetic correlation estimates between CH_4 and RA of rumen microbes at the genus level are shown in Figure 4. They ranged from -0.76 to 0.65 using the non-recursive bivariate model and from -0.68 to 0.69 using the recursive model.

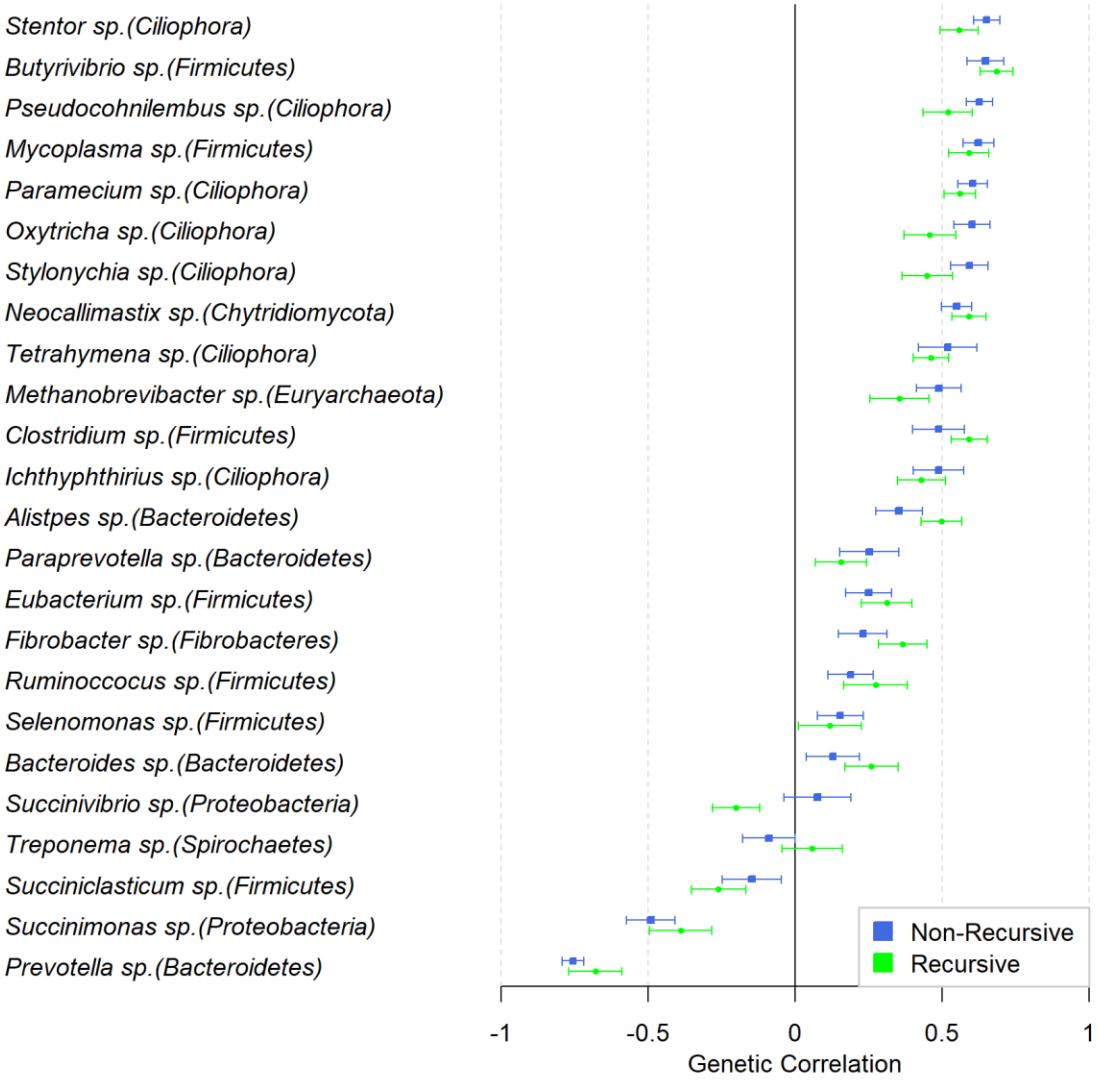

Figure 4. Genetic correlation between methane concentration (ppm) and microbial genus relative abundance (%) according to statistical model (Non-Recursive vs Recursive) from each bi-trait model according to genus (phylum within parenthesis), estimated from the mean of the stationary posterior distribution. The bars show the 95% probability intervals around the heritability estimates using the Monte Carlo standard error.

Positive genetic correlations with methane were estimated consistently for the 7 genera from the Ciliophora phylum both with the non-recursive bivariate model (0.49 to 0.65) as well as with the recursive model (0.43 to 0.56). Positive genetic correlations were also obtained for the Euryarchaeota (*Methanobrevibacter sp*.), Chytridiomycota (*Neocallimastix sp*.) and Fibrobacteres (*Fibrobacter sp*.) phyla, in both, non-recursive bivariate and recursive models.

All genera in this case, excepting (*Succinivibrio sp*.) from the Proteobacteria pylum, resulted in overlapped genetic correlations between the non-recursive bivariate model and the recursive model. However, high differences were observed. *Succinivibrio sp*. showed the largest disagreement changing from positively correlated (0.08) in the non-recursive bivariate model to negatively correlated (-0.20) in the recursive model. In the non-recursive bivariate model the two genera (*Succinivibrio sp*. and *Succinimonas sp*.) from the Proteobacteria phylum showed opposite genetic correlations (0.08 and -0.49, respectively), whereas within the recursive model, both genera showed negative genetic correlations (-0.20 and -0.38, respectively), grouping both genus from the same phylum within the same direction of genetic correlation. For the *Treponema sp*. genus a change from negatively correlated (-0.09) in the non-recursive bivariate model to positively correlated (0.06) in the recursive model was obtained, but in this case probability intervals overlapped.

The *Stentor sp*. and *Butyrivibrio sp*. genera showed the largest genetic correlation (0.65) with CH_4, using the non-recursive bivariate model, while for the recursive model, *Butyrivibrio sp. and Neocallimastix sp.* (an obligate anaerobic rumen fungi) genera showed the largest genetic correlation (0.69 and 0.59, respectively) with CH_4. Conversely, *Prevotella sp.* (Bacteroidetes phylum) showed the largest antagonistic genetic correlation with CH_4 (-0.76 and -0.68, respectively) for the non-recursive bivariate model and the recursive model. The RA of *Methanobrevibacter sp*. in the rumen showed a positive genetic correlation with CH_4 (0.49 and 0.36, respectively) within the non-recursive bivariate model and the recursive model.

DISCUSSION

Methane is the second most relevant greenhouse gas (GHG) after carbon dioxide (CO_2). It is relevant for global warming because it has a global warming potential ≈ 28 times greater than CO_2 (Myhre et al., 2013), although it remains less time in the atmosphere than CO_2 (12 years *vs* 100 years). Ruminants emit enteric CH_4 as a by-product from feed degradation and fermentation processes. In addition to the detriment caused by methane on the environment, the loss of energy associated with the production of methane is an undesirable metabolic route for the productive system, since this energy could be used in other ways directly related to the final product of the system. The CH_4 produced is mainly exhaled through breath and eructation to the atmosphere and not seize by the host. It is estimated that the loss of energy for the animal due to methane production varies between 2 and 12% of gross energy consumption (Johnson et al., 1993).

The daily CH_4 production of an adult cow, of the Holstein breed, has previously been estimated at 369 g/d (range 278 to 456 g/d CH_4) (Garnsworthy et al., 2012) and more recently at 395.8 g/d (SD = 63.5) (Difford et al., 2018). Values as high as 506 g/d (SD =56) had been reported for this trait in a latter study (Ramayo-Caldas et al., 2019). A recent result from 28 heifers averaged 267 g/d of CH_4 (Flay et al., 2019) indicated that younger cows are in average less CH_4 emitters than adult cows. Regarding CH_4 concentration, a recent study (Wu, Koerkamp, & Ogink, 2018) reported a range between 543 ppm (SD, 33) to 1,100 ppm (SD, 92), using a different infrared spectroscopy gas analyzer. Those values are in accordance to our estimations (853 ppm, SD = 278). Another study (Huhtanen et al., 2015) reported CH_4 concentrations of 1405 ppm (SD=247) using a sniffer method. A possible reason for our lower values compared to the latter study could be due to the measurement instrument used. Another reason for lower estimation of CH_4 emissions in this study could be due to parity; as animals in our study were first and second calving cows instead of adult cows.

Methane heritability estimates (0.09 to 0.18) were relatively homogeneous across the set of recursive Gaussian SEM. The estimates for $h^2_{CH_4}$ in this study is in agreement with previous studies (Difford et al., 2018; Breider, Wall, & Garnsworthy, 2019) and further support that methane production is a heritable trait, and breeding strategies can be implemented to reduce the amount of methane produced in dairy cattle. The estimates for

$h^2_{CH_4}$ in this study were slightly lower than 0.19 ± 0.09 (mean ± SE) previously reported by Difford et al. (2018) and 0.21 ± 0.06 (h^2 ± SE) reported in lactating Holstein cows in Denmark (Lassen and Løvendahl, 2016).

Methanogenic microorganisms in the rumen belong to the archaea domain and most of them use hydrogen (H_2) to reduce CO_2 into CH_4 controlling H_2 concentrations and intra-ruminal pressure (hydrogenotrophic methanogens); Bacterial fermentation could be inhibited at high concentrations of H_2 (Mizrahi and Jami, 2018); however metagenome interrelationships contributing to methane emissions, including non-methanogens microorganisms, are currently under study and still to be deciphered. In dairy cattle, there are many studies associating the metagenome of the host with phenotypic traits. A meta-analysis (Guyader et al., 2014) with 28 experiments on ruminants analyzed the relationship between methane emissions and the number of protozoa. They found a significant linear relationship (R^2 = 0.90, P <0.05) in the association of these traits. In a recent study (Delgado et al., 2018) found that Holstein cows with better feed efficiency had higher abundance of Bacteroidetes, Prevotella, and lower abundance of Methanobacteria and Methanobrevibacter in the rumen microbiota than less efficient cows, concluding that more efficient animals showed different metagenomes than less efficient animals. These findings provide evidence about the relationship between the animal metagenome and the phenotypes expressed by its host.

The range of h^2_{RA} obtained from both models is in accordance with previous estimates of this parameter from previous studies. For instance, Li et al., (2019) reported h^2_{RA} for Firmicutes, (0.16 ± 0.08; mean ± SE), unclassified Clostridiales (0.25 ± 0.09) and *Blautia* (0.18 ± 0.08), as well as for *Methanobacterium* (0.23 ± 0.08), *Mbb. ruminantium* (0.18 ± 0.08), and *Methanobacterium alkaliphilum* (0.23 ± 0.08). Another study (Wallace et al., 2019) informed about heritable microbial OTUs with heritability estimates ranging from 0.20 to 0.60.

The study of the relationship between metagenome and methane production using SEM allows to account for biologically plausible effects between traits, and hence, modelize a more realistic scenario, which could yield to better estimation of parameters and a better fitting of the statistical models for this biological problem. Non-zero estimates for $\lambda_{CH_4 \leftarrow RA}$ indicated that there exist a biological recursive effect of microbes RA on CH_4. While some taxa, like ciliate protozoa or *Methanobrevibacter sp.*, increased the CH_4 emissions (positive

values of $\lambda_{CH_4 \leftarrow RA}$), others like *Succinivibrio sp*. from Proteobacteria phylum decreased it (negative value of $\lambda_{CH_4 \leftarrow RA}$.These variation between $\lambda_{CH_4 \leftarrow RA}$ could be useful to generate a ranking of ruminal microbial taxa linked to CH_4 emissions, which added to the relative (or total) abundance of a given genus in the metagenome, could contribute as a proxy for CH_4 predictions at the phenotypic level.

Biological implications of genetic parameters for RA and CH_4

Among taxa analyzed, Ciliate protozoa (7 genera) showed moderate heritabilities (0.28-0.32) as well as positive genetic correlations (0.43-0.56) with CH_4, regardless of the model used. Protozoa are known to be associated to CH_4 emissions, especially through their ability to produce H_2 in their hydrogenosomes and their ability to play a symbiotic role with methanogenic archaea while protecting them from the toxicity of oxygen (Suen et al., 2015; Belanche, De La Fuente, & Newbold, 2014). On the other hand, the most abundant genera from Bacteroidetes phylum was *Prevotella sp.* which resulted in negative correlation with CH_4. This is congruent to previous studies reporting that *Prevotella sp*. is associated with improved feed efficiency in dairy cows (Bach et al., 2019).

The two genera from the Proteobacteria phylum (*Succinivibrio sp*. and *Succinimonas sp*.) were only negatively correlated to CH_4 in the recursive model, which is congruent to previous studies reporting that lower abundances of Proteobacteria seem to be associated with high methane emissions (Tapio et al., 2017). Those two genera had been negatively correlated to CH_4 emissions previously (Granja-Salcedo et al., 2019) trough increasing fumarate reductase activity (Asanuma and Hino, 2000), resulting in less free H_2 and therefore lower CH_4. In the non-recursive bivariate model one genus from the Proteobacteria phylum (*Succinivibrio sp*.) showed positive genetic correlation, while the other (*Succinimonas sp*.) was negatively correlated. The genetic correlation between CH_4 and the RA of *Methanobrevibacter sp*. was also moderate positive. There are some inconsistencies regarding the association in the literature between the relative abundance of archaea and methane (Negussie et al., 2017). A possible explanation to this phenomenon could be based on the endosymbiotic relationship from methanogenic archaea to anaerobic ciliates (van Hoek et al., 2000), in which ciliates provide to methanogenic archaea with an intracellular

source of hydrogen as the basis for a stable association (Finlay et al., 1994). The archaea cell walls are structurally different from that of bacteria. It confers archaea more resistance to lytic protocols that work well for bacterial cells (Roopnarain et al., 2017). This may affect and difficult their DNA extraction. Consequently, the archaea relative abundance could be an unreliable indicator of methane synthesis, while a higher relative abundance of ciliates could indicate a favorable ciliates-archaea symbiosis and therefore higher production of CH_4. The RA of Ciliate in the rumen is larger than that of archaea, and it may assist on a more reliable estimate of the genetic correlation with CH_4.

New strategies to jointly analyze host-metagenome-phenotype relationship, as the one presented here, are to be developed, with the purpose of integrating the complex biological interrelationships between microorganisms and its link to genotype and phenotype. The magnitude of h^2_{RA} and $h^2_{CH_4}$ and genetic correlations obtained here indicates that there could be an opportunity to include CH_4 emissions in animal breeding programs. In the same way as the $\lambda_{CH_4 \leftarrow RA}$ could be useful to generate a ranking of ruminal microbial taxa linked to CH_4 emissions, to be used as proxies; genetic parameters of rumen microbiota composition, including the genetic correlations with CH_4, could be used to select individuals that are able to modulate a favorable microbiota composition that pass their genes onto the next generations for lower CH_4. Future genetic trends in CH_4 could be modified through direct selection of animals with favorable genetic breeding values for RA of relevant microbes. For instance, selecting animals with low relative abundance of ciliates could reduce the substrate for methanogenic microorganisms in next generations. Based on the high heritabilities found for those genera, a decrease of CH_4 emissions from the progeny of selected cows would be expected. This kind of strategies could enhance the genetic progress of lower values for CH_4 in the future populations. Ross, Moate, Marett, Cocks, & Hayes (2013) argued that it is possible that metagenomic predictions could aid in the reduction of enteric methane production from ruminants, if increased accuracy in the prediction of enteric methane production level is achieved. Genetic selection using estimation of additive breeding values is potentially the most sustainable way of reducing enteric methane emission from ruminant (Pickering et al., 2015).

CONCLUSIONS

This study applied SEM as a tool to integrate genomic, metagenomic and phenotypic information in order to jointly analyze plausible biological relationships. Ciliate protozoa (7 genus) showed moderate heritabilities and consistent positive genetic correlation to CH_4 in both statistical model approaches (non-recursive and recursive). Genetic correlation estimates revealed differences according to the usage of non-recursive and recursive models, with a more biologically supported result for the recursive model estimation. Based on the heritabilities and genetic correlations obtained from this study we conclude that, methane emissions could be included in genetic evaluations for dairy cattle, in order to obtain more cost effective animals while diminishing their environmental footprint. SEM could be used to also include metagenomic information into genetic evaluations analysis accounting for the recursive relationship between traits, and with the opportunity to increase reliability.

ACKNOWLEDGEMENTS

This research was financed by RTA2015-00022-C03 (METALGEN) project from the national plan of research, development and innovation 2013-2020. The first author of this paper was granted a scholarship from Universidad de Costa Rica to course doctorate studies which partially conducted to the progress of this study. We acknowledge to Adrián López-García for processing of sequences and its conversion into counts within taxa.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest regarding the topics approached within this study.

REFERENCES

Asanuma, N., and T. Hino. 2000. Activity and properties of fumarate reductase in ruminal bacteria.. J. Gen. Appl. Microbiol. 46:119–125.

Bach, A., A. López-García, O. González-Recio, G. Elcoso, F. Fàbregas, F. Chaucheyras-Durand, and M. Castex. 2019. Changes in the rumen and colon microbiota and effects of live yeast dietary supplementation during the transition from the dry period to lactation of dairy cows. J. Dairy Sci. Sci 102:6180–6198. doi:10.3168/jds.2018-16105.

Belanche, A., G. De La Fuente, and C.J. Newbold. 2014. Study of methanogen communities associated with different rumen protozoal populations. doi:10.1111/1574-6941.12423.

Breider, I.S., E. Wall, and P.C. Garnsworthy. 2019. Short communication: Heritability of methane production and genetic correlations with milk yield and body weight in Holstein-Friesian dairy cows. J. Dairy Sci. 102:7277–7281. doi:10.3168/jds.2018-15909.

Buchfink, B., C. Xie, and D.H. Huson. 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59–60. doi:10.1038/nmeth.3176.

Buitenhuis, B., J. Lassen, S.J. Noel, D.R. Plichta, P. Sørensen, G.F. Difford, and N.A. Poulsen. 2019. Impact of the rumen microbiome on milk fatty acid composition of Holstein cattle. Genet. Sel. Evol. 51:1–8. doi:10.1186/s12711-019-0464-8.

Camarinha-Silva, A., M. Maushammer, R. Wellmann, M. Vital, S. Preuss, and J. Bennewitz. 2017. Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. doi:10.1534/genetics.117.200782.

Delgado, B., A. Bach, I. Guasch, C. González, G. Elcoso, J.E. Pryce, and O. Gonzalez-Recio. 2018. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. Sci. Rep. 9:1–13. doi:10.1038/s41598-018-36673-w.

Difford, G.F., D.R. Plichta, P. Løvendahl, J. Lassen, S.J. Noel, O. Højberg, A.-D.G. Wright, Z. Zhu, L. Kristensen, H.B. Nielsen, B. Guldbrandtsen, and G. Sahana. 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLOS Genet. 14:e1007580. doi:10.1371/journal.pgen.1007580.

Finlay, J., G. Esteban, K.J. Clarke, A.G. Williams, T.M. Embley, and R.P. Hirt. 1994. Some rumen ciliates have endosymbiotic methanogens. FEMS Microbiol. Lett. 117:157–162. doi:10.1111/j.1574-6968.1994.tb06758.x.

Flay, H., B. Kuhn-Sherlock, K. Macdonald, M. Camara, N. Lopez-Villalobos, D. Donaghy, and J. Roche. 2019. Hot topic: Selecting cattle for low residual feed intake did not affect daily methane production but increased methane yield. J. Dairy Sci.. doi:10.3168/jds.2018-15234.

Garnsworthy, P., J. Craigon, J. Hernandez-Medrano, and N. Saunders. 2012. Variation among individual dairy cows in methane measurements made on farm during milking. J. Dairy Sci. 95:3181–3189. doi:10.3168/jds.2011-4606.

Gianola, D., and D. Sorensen. 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. Genetics 167:1407–1424. doi:10.1534/genetics.103.025734.

Gonzalez-Recio, O., I. Zubiria, A. García-Rodríguez, A. Hurtado, and R. Atxaerandio. 2017. Short communication: Signs of host genetic regulation in the microbiome composition in 2 dairy breeds: Holstein and Brown Swiss.. J. Dairy Sci. 101:1–8. doi:10.3168/jds.2017-13179.

Granja-Salcedo, Y.T., R.M. Fernandes, R.C. de Araujo, L.T. Kishi, T.T. Berchielli, F.D. de Resende, A. Berndt, and G.R. Siqueira. 2019. Long-Term Encapsulated Nitrate Supplementation Modulates Rumen Microbial Diversity and Rumen Fermentation to Reduce Methane Emission in Grazing Steers. Front. Microbiol. 10:1–12. doi:10.3389/fmicb.2019.00614.

Guyader, J., M. Eugène, P. Nozière, D.P. Morgavi, M. Doreau, and C. Martin. 2014. Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. Animal 8:1816–1825. doi:10.1017/S1751731114001852.

Heringstad, B., X.-L. Wu, and D. Gianola. 2009. Inferring relationships between health and fertility in Norwegian Red cows using recursive models. J. Dairy Sci 92:1778–1784. doi:10.3168/jds.2008-1535.

van Hoek, A.H.A.M., T.A. van Alen, V.S.I. Sprakel, J.A.M. Leunissen, T. Brigge, G.D. Vogels, and J.H.P. Hackstein. 2000. Multiple Acquisition of Methanogenic Archaeal Symbionts by Anaerobic Ciliates. Mol. Biol. Evol. 17:251–258. doi:10.1093/oxfordjournals.molbev.a026304.

Huhtanen, P., E.H. Cabezas-Garcia, S. Utsumi, and S. Zimmerman. 2015. Comparison of methods to determine methane emissions from dairy cows in farm conditions. J. Dairy Sci. 98:3394–3409. doi:10.3168/jds.2014-9118.

Huson, D.H., A.F. Auch, J. Qi, and S.C. Schuster. 2007. MEGAN analysis of metagenomic data. Genome Res. 17:377–386. doi:10.1101/gr.5969107.

Jiménez-Montero, J., D. Gianola, K. Weigel, R. Alenda, and O. González-Recio. 2013. Assets of imputation to ultra-high density for productive and functional traits. J. Dairy Sci. 96:6047–6058. doi:10.3168/jds.2013-6793.

Konig, S., X.L. Wu, D. Gianola, B. Heringstad, and H. Simianer. 2008. Exploration of relationships between claw disorders and milk yield in Holstein cows via recursive linear and threshold models. J. Dairy Sci. 91:395–406. doi:DOI: 10.3168/jds.2007-0170.

Lassen, J., and P. Løvendahl. 2016. Heritability estimates for enteric methane emissions from Holstein cattle measured using noninvasive methods. J. Dairy Sci. 99:1959–

1967. doi:10.3168/jds.2015-10012.

Legarra, A., L. Varona, and E. Lopez de Maturana. 2011. Threshold Model 1–33.

Li, F., C. Li, Y. Chen, J. Liu, C. Zhang, B. Irving, C. Fitzsimmons, G. Plastow, and L.L. Guan. 2019. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. Microbiome 92:1–17. doi:10.1186/s40168-019-0699-1.

López De Maturana, E., X.-L. Wu, D. Gianola, K.A. Weigel, and G.J.M. Rosa. 2008. Exploring Biological Relationships Between Calving Traits in Primiparous Cattle with a Bayesian Recursive Model. Genetics 181:277–287. doi:10.1534/genetics.108.094888.

De Los Campos, G., D. Gianola, P. Boettcher, and P. Moroni. 2014. A structural equation model for describing relationships between somatic cell score and milk yield in dairy goats. J. Anim. Sci. 84:2934–2941. doi:10.2527/jas.2006-016.

De Los Campos, G., D. Gianola, and B. Heringstad. 2006. A Structural Equation Model for Describing Relationships Between Somatic Cell Score and Milk Yield in First-Lactation Dairy Cows. J. Dairy Sci. 89:4445–4455. doi:10.3168/jds.S0022-0302(06)72493-6.

Marchesi, J.R., and J. Ravel. 2015. The vocabulary of microbiome research: a proposal. Microbiome 3. doi:10.1186/s40168-015-0094-5.

Mizrahi, I., and E. Jami. 2018. Review: The compositional variation of the rumen microbiome and its effect on host performance and methane emission. Animal 12:220–232. doi:10.1017/S1751731118001957.

Myhre, G., D. Shindell, F. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, H. Zhang, D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley. 2013. Anthropogenic and natural radiative forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I 659–740.

Negussie, E., Y. de Haas, F. Dehareng, R.J. Dewhurst, J. Dijkstra, N. Gengler, D.P. Morgavi, H. Soyeurt, S. van Gastelen, T. Yan, and F. Biscarini. 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.. J. Dairy Sci. 100:2433–2453. doi:10.3168/jds.2016-12030.

Newbold, C.J., G. De La Fuente, A. Belanche, E. Ramos-Morales, and N.R. McEwan. 2015. The Role of Ciliate Protozoa in the Rumen. Front. Microbiol. 6:1–14. doi:10.3389/fmicb.2015.01313.

Pickering, N.K., V.H. Oddy, J. Basarab, K. Cammack, B. Hayes, R.S. Hegarty, J. Lassen, J.C. Mcewan, S. Miller, C.S. Pinares-Patiño, and Y. De Haas. 2015. Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants. Animal 9:1431–1440. doi:10.1017/S1751731115000968.

Ramayo-Caldas, Y., L. Zingaretti, M. Popova, J. Estellé, A. Bernard, N. Pons, P. Bellot, N.

Mach, A. Rau, H. Roume, M. Perez-Enciso, P. Faverdin, N. Edouard, D. Ehrlich, D.P. Morgavi, and G. Renand. 2019. Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. J. Anim. Breed. Genet. 00:1–11. doi:10.1111/jbg.12427.

Roehe, R., R.J. Dewhurst, C.A. Duthie, J.A. Rooke, N. McKain, D.W. Ross, J.J. Hyslop, A. Waterhouse, T.C. Freeman, M. Watson, and R.J. Wallace. 2016. Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. PLoS Genet. 12:1–20. doi:10.1371/journal.pgen.1005846.

Roopnarain, A., M. Mukhuba, R. Adeleke, and M. Moeletsi. 2017. Biases during DNA extraction affect bacterial and archaeal community profile of anaerobic digestion samples. Biotech 7:375:1–12. doi:10.1007/s13205-017-1009-x.

Rosa, G.J., B.D. Valente, G. de los Campos, X.-L. Wu, D. Gianola, and M.A. Silva. 2011. Inferring causal phenotype networks using structural equation models. Genet. Sel. Evol. 43:6. doi:10.1186/1297-9686-43-6.

Ross, E.M., P.J. Moate, L.C. Marett, B.G. Cocks, and B.J. Hayes. 2013. Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle. PLoS One 8:1–8. doi:10.1371/journal.pone.0073056.

Sorensen, D., and D. Gianola. 2002. Likelihood of Bayesian, and MCMC Methods in Quantitative Genetics. 1st ed. K. Dietz, M. Gail, K. Krickeberg, J. Samet, and A. Tsiatis, ed. Springer-Verlag New York, Inc, New York.

Tapio, I., T.J. Snelling, F. Strozzi, and R.J. Wallace. 2017. The ruminal microbiome associated with methane emissions from ruminant livestock. J. Anim. Sci. Biotechnol. 8:1–11. doi:10.1186/s40104-017-0141-0.

Thompson, L., J.G. Sanders, D. McDonald, A. Amir, J. Ladau, K.J. Locey, robert J. Prill, A. Tripathi, S. Gibbons, G. Ackermann, J.A. Navas-molina, S. Janssen, E. Kopylova, Y. Vázquez-baeza, A. González, J.T. Morton, S. Mirarab, Z. Zech Xu, L. Jiang, mohamed F. Haroon, J. Kanbar, Q. Zhu, S. Jin Song, T. Kosciolek, nicholas A. Bokulich, J. Lefler, colin J. Brislawn, and G. Humphrey. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Springer Nat. 551:457–463. doi:10.1038/nature24621.

Valente, B.D., and G.J. de Magalhães Rosa. 2013. Mixed Effects Structural Equation Models and Phenotypic Causal Networks. In: Methods in Molecular Biology. Humana Press, Totowa, NJ.

VanRaden, P.M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

Wallace, R.J., G. Sasson, P.C. Garnsworthy, I. Tapio, E. Gregson, P. Bani, P. Huhtanen, A.R. Bayat, F. Strozzi, F. Biscarini, T.J. Snelling, N. Saunders, S.L. Potterton, J. Craigon, A. Minuti, E. Trevisi, M.L. Callegari, F.P. Cappelli, E.H. Cabezas-Garcia, J.

Vilkki, C. Pinares-Patino, K.O. Fliegerová, J. Mrázek, H. Sechovcová, J. Kopečný, A. Bonin, F. Boyer, P. Taberlet, F. Kokou, E. Halperin, J.L. Williams, K.J. Shingfield, and I. Mizrahi. 2019. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci. Adv. 5:1–12. doi:10.1126/sciadv.aav8391.

Wu, L., P.W.G.G. Koerkamp, and N. Ogink. 2018. Uncertainty assessment of the breath methane concentration method to determine methane production of dairy cows. J. Dairy Sci. 101:1554–1564. doi:10.3168/jds.2017-12710.

Wu, X., B. Heringstad, Y. Chang, G.D.L. Campos, and D. Gianola. 2007. Inferring Relationships Between Somatic Cell Score and Milk Yield Using Simultaneous and Recursive Models. J. Dairy Sci. 90:3508–3521. doi:10.3168/jds.2006-762.

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569. doi:10.1038/ng.608.

CHAPTER 5

HOLOBIONT EFFECT ACCOUNTS FOR MORE METHANE EMISSION VARIANCE THAN THE ADDITIVE AND MICROBIOME EFFECTS ON DAIRY CATTLE

Alejandro Saborío-Montero[1,2], Mónica Gutiérrez-Rivas[1], Adrián López-García[1], Aser García-Rodríguez[3], Raquel Atxaerandio[3], Idoia Goiri[3], Evangelina López de Maturana[4], José Antonio Jiménez-Montero[5], Rafael Alenda[6] and Oscar González-Recio[1,6]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain; [2]Centro de Investigación en Nutrición Animal y Escuela de Zootecnia, Universidad de Costa Rica, 11501 San José, Costa Rica; [3]Department of Animal Production, NEIKER - Basque Institute for Agricultural Research and Development. Basque Research and Technology Alliance (BRTA). Campus Agroalimentario de Arkaute s/n, 01192; [4]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), and CIBERONC, Melchor Fernández Almagro 3, 28029 Madrid, Spain; [5]Departamento técnico de Confederación de Asociaciones de Frisona Española (CONAFE) Ctra. de Andalucía, km. 23,600 - 28340 Valdemoro, Madrid, Spain; [6]Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain.

Published in Livestock Science.
DOI: https://doi.org/10.1016/j.livsci.2021.104538

ABSTRACT

Rumen microbiota has been previously related to phenotypic complex traits of relevance in dairy cattle. The joint analysis of the host's genetic background and its microbiota can be statistically modelled using similarity matrices between microorganism communities in the different hosts. Microbiota relationship matrices (**K**) enable considering the whole microbiota and the cumbersome interrelations between taxa, rather than analyzing single taxa one at the time. Several methods have been proposed to ordinate these matrices. The aim of this study was to compare the performance of twelve K built from different microbiome distance metrics, within a variance component estimation framework for methane concentration in dairy cattle. Phenotypic, genomic and rumen microbiome information from simulations (n = 1000) and real data (cows = 437) were analyzed. Four models were considered: an additive genomic model (GBLUP), a microbiome model (MBLUP), a genetic and microbiome effects model (HBLUP) and a genetic, microbiome and genetic × microbiome interaction effects model (H_iBLUP). Results from simulation were obtained from 25 replicates. Results from simulated data suggested that Ks with flattened off-diagonal elements were more accurate in variance components estimation for all compared models that included Ks information (MBLUP, HBLUP and H_iBLUP). Multidimensional scaling (MDS), redundancy analysis (RDA) and constrained correspondence analysis (CCA) performed better in simulation to estimate heritability and microbiability. The models including Ks from the MDS, RDA and CCA methods were also between the most plausible models in the real data set, according to the deviance information criteria (DIC). Real data was analyzed under the same framework as in the simulation. The most plausible model in real data was H_iBLUP. Estimates variated depending on K; methane heritability (0.15-0.17) and microbiability (0.15-0.21) were lower than the proportion of the phenotypic variance attributable to the host-microbiome holobiont effect (0.42-0.59), which we have defined here as "holobiability". The holobiability including the genomic × microbiome interaction from the H_iBLUP was between 0.01 and 0.15 larger than the holobiability explained from the sum of the genetic and microbiome effects without interaction between them, from the HBLUP, depending on K. The findings in this study support the potential of the joint analysis of genomic and microbiome information. Accounting for the hologenome effect (genomic and

microbiome) could improve the accuracy in variance component estimation of complex traits relevant in livestock science.

Keywords: Heritability, holobiability, methane, microbiability, microbiota relationship matrix.

INTRODUCTION

The microbiome contributes to the phenotypic variability of complex traits (Difford et al., 2018; Buitenhuis et al., 2019). There is also extensive support that digestive microbiome is partially controlled by host genetic variation in several species, such as: humans (Blekhman et al., 2015; Zoetendal et al., 2001), mice (Benson et al., 2010; McKnite et al., 2012), poultry (Zhao et al., 2013), pigs (Camarinha-Silva et al., 2017) and cattle (Roehe et al., 2016; Gonzalez-Recio et al., 2017; Difford et al., 2018). In animal breeding, a host's complex traits prediction that incorporates information regarding its microbiome is a promising field. For instance, the microbiability (m^2), defined as the proportion of the phenotypic variance attributable to the microbiome (Difford et al. 2016), has been previously estimated with higher values than narrow-sense heritability for feed efficiency and feed intake in pigs (Camarinha-Silva et al., 2017). Thus, considering the microbiome as a source of information at predicting complex traits is a rational practice and could lead to improved accuracy in genomic prediction. Sequencing the metagenome is now the method of choice to characterize the microbiota. Affordable whole genome sequencers (~ $50 USD per sample) have been recently developed and commercialized in a more compact format than previous devices (Lu et al., 2016; Santos et al., 2020). These characteristics facilitate its acquisition, transportation, and *in-situ* handling in a more versatile way. Those sequencers are now commonly used to estimate relative abundance of microbial taxa, allowing to analyze microbiome at a compositional level (Saborío-Montero et al., 2020). Third generation sequencers have enlarged the boundaries of microbial identification, previously restricted to culture methods or amplicon-based studies (Ciuffreda et al., 2021).

Many traits of economical and sustainability importance in livestock have been previously associated to microbiome composition (Roehe et al., 2016; Tapio et al., 2017). For instance, methane emissions from livestock contribute with approximately 8% to 12% of anthropogenic emissions (Eckard et al., 2010) and is an area of concern because of its potential as a greenhouse gas. Previous studies have reported its potential effects in the global warning context as 28-fold more harmful than carbon dioxide (Myhre et al., 2013). The joint analysis of the effect of host genetics and the ruminal microbiota on complex traits can be conducted through incorporating a microbiota relationship matrix (**K**) into the statistical models. This matrix considers the microbiota as a whole instead of individual assessment of single microorganisms.

The metagenomic relationship between a group of samples (**K)** can be described by a square matrix of dimensions $n \times n$ samples, generated through the cross-product of a matrix (**X**) of the centred log-ratio transformation of samples by taxa contingency table, divided by the number of taxa (p) as $(1/p)\mathbf{XX^T}$ (hereafter "CrPr"). This approach has been applied to incorporate microbiome effect into mixed models in several studies (Camarinha-silva et al., 2017; Hadfield and Nakagawa, 2010; Khanal et al., 2020a, 2020b; Ross et al., 2013; Weishaar et al., 2020). Other approaches to build the K matrix include the Jensen-Shannon distance matrix (Maltecca et al., 2019), or the usage of an exponential product on a function of the Euclidean distance (Pérez and De Los Campos, 2014). Similarly, it is possible to include the interaction between the genomic and the microbiota relationship matrices (Jarquín et al., 2014; Saborío-Montero, 2018). Variances from genomic information, microbiome information and its interaction can be extracted from the multi-kernel analysis, allowing the estimation of heritability, microbiability and holobiability. We have defined the latter term as the proportion of the phenotypic variance attributable to whole animal-microbiome holobiont organism.

In this study we compared several methods of ordination (Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA)) and distance matrices (Euclidean, Bray-Curtis, Canberra, Jaccard, Mahalanobis and Aitchison). The objectives of this study were to disentangle the most appropriate distance metric to estimate variance component for methane emission in dairy cattle, as well as to compare

different models including the additive and microbiome effects using simulated and real data. A deeply review on these subjects will be addressed in this study, to understand the state of the art focusing on the development and evaluation of microbiome relationship matrices and the holobiability concept, innovative in prediction of complex traits.

MATERIALS AND METHODS

This study was developed in the Department of Animal Breeding and Genetics of the National Agricultural and Food Research and Technology Institute (INIA), Madrid, Spain. This research was carried out in accordance with Spanish Royal Decree 53/2013 for the protection of animals used for experimental and other scientific purposes and was approved by the Basque Institute for Agricultural Research and Development Ethics Committee (Neiker-OEBA-2017-004) on March 28, 2017.

Variance components, heritability, microbiability and holobiability, regarding methane production in dairy cows, was estimated using two data subsets: a simulated dataset and a real dataset.

Simulated data

Simulations were generated using the observed data structure in the real data set. A data frame of real genotypes from 1000 Holstein animals was used. A total of 9244 SNPs loci were randomly selected from a 50K SNP bead chip for subsequent analysis. Additive genetic effects were determined by 1000 QTL randomly selected along the genome. QTL effects were generated based on a normal distribution $N \sim (0, 1)$. True breeding values (**u**) were calculated by summing all QTL effects and were subsequently scaled to a realized genetic variance of σ_u^2.

The relative abundances of phylum *j* (j=1-86) in animal *i* (i=1-1000) were simulated from real data (samples of 437 animals and 86 phyla), following the steps described below:

1) The covariance matrix (86 × 86) from the real relative abundance of phyla was obtained.

2) The resulting symmetric matrix was then converted to the nearest positive-definite matrix (86 × 86) to ensure it was a valid covariance matrix.
3) A total of 86 vectors of 1000 random values, replicating relative abundance of phyla for 1000 cows, were sampled from a normal distribution (N ~ (0.5, 0.1)). The used parametric values were arbitrarily chosen under the assumption that mean relative abundance should be between 0 and 1 and that a narrow variance would improve simulation results.
4) The cross product of the Cholesky factorized matrix of the positive-definite matrix of (co)variances, calculated in step 2, was then multiplied by the matrix (86 × 1000) built from the random values obtained in step 3, creating the desired final matrix of 1000 simulated microbiotas with 86 phyla each, and preserving the covariances from real data. Any phylum resulting in negative values was set to zero.

Once the relative abundance of the simulated phyla was generated, the microbiome effect (***m***) for each animal was simulated as follows: 50 phyla were randomly selected out of the 86 phyla simulated before. Then, an effect (β_j) was sampled from a normal distribution N ~ (0, 1) and assigned to each of the selected 50 phyla. The m_i was then simulated as follow:

$$m_i = \sum_{\boldsymbol{j}} \beta_j \times F_{ij}$$

Where F_{ij} was the relative abundance of phylum j in animal i for the 50 selected phyla. The resulting $\{m_i\}$ were scaled to have a variance of σ_m^2.

Phenotypes were finally simulated in 4 different scenarios assigning a residual variance to obtain a heritability and a microbiability of 0.30. Phenotypes were respectively simulated under 4 different scenarios: (1) an additive genetic model, (2) microbiome model, (3) genetic and microbiome additive effects model and (4) same as 3 plus an interaction effect:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{u} + \mathbf{e}, \quad \textbf{(1)}$$

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{m} + \mathbf{e}, \quad \textbf{(2)}$$

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{u} + \mathbf{m} + \mathbf{e} \quad (3)$$

and

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{u} + \mathbf{m} + \mathbf{u} \times \mathbf{m} + \mathbf{e} \quad (4)$$

where $\boldsymbol{\mu}$ is the population mean, $\mathbf{u}$ is the genomic effect, $\mathbf{m}$ is the microbiome effect, $\mathbf{u} \times \mathbf{m}$ is a genomic-microbiome interaction effect and $\mathbf{e}$ is the residual error. The prior distribution of fixed effect is assumed to be a uniform distribution (-9999, 9999) and for σ_u^2 and σ_e^2 are assumed to be distributed as a scaled inverse chi-square $\{\sigma_u^2 \sim \upsilon_u s_u^2 \chi_{v_u}^{-1}$, $\sigma_e^2 \sim \upsilon_e s_e^2 \chi_{v_e}^{-1}\}$ where $\{\upsilon_u, \upsilon_e\}$ and $\{s_u, s_e\}$ are known degrees of freedom and scale parameters, respectively; with $\{\upsilon_u, \upsilon_e\} = 5$ and $\{s_u, s_e\}$ = 0.5 times the variance of the phenotypes (From default parameters of the BGLR package).

Real data

Methane measurement and sampling. Four hundred and thirty-seven lactating Holstein cows (primiparous or second lactation) from 14 commercial farms in 4 Northern Spanish regions (Cantabria, País Vasco, Navarra and Gerona) were included in this study. Methane emissions (ppm) were measured as described by Rey et al. (2019) using a non-dispersive infrared methane detector (The Guardian® NG infrared gas monitor from Edinburg Sensors; measure range 0-1%), installed in the feed bin of the automatic milking system (AMS). Methane concentration in breath samples was measured individually for each cow, during the milking time at each cow visit to the AMS for 2-3 weeks periods. Eructation peaks recorded were averaged to obtain a single record per cow. During this sample period herds underwent a test-day recording as part of the official test day recording scheme. All traits related to milk yield and composition were also available.

During the ruminal content sampling, cows were placed in individual stalls. A custom-made mechanical device was used to raise the snout of the animal. Samples of ruminal content (approximately 100 ml) were extracted from each cow by introducing a stomach tube (18 mm diameter and 160 mm long) orally through the esophagus connected to a mechanical pumping unit (Vacubrand ME 2SI, Wertheim, Germany) with a 1000 ml Erlenmeyer trapped in-between. Samples were then stored in a sterilized container. Hose and all material in contact with the samples were systematically washed between cows. Samples

were filtered through 4 layers of sterile cheesecloth, in order to remove the solid fraction and the filtered fraction was frozen in liquid nitrogen (N2) vapors immediately after. Then frozen samples were transported to the laboratory in liquid N2 containers and stored at −80 °C until analysis.

Metagenomic analysis. The samples were thawed, and then homogenized with a blender. The DNA extraction was performed using 250 µl from the homogenized samples with the commercial "DNeasy Power Soil Kit" (QIAGEN, Valencia, CA, USA). The quality control of the DNA and the protocol used for its sequencing using MinION device are more detailed described in (Saborío-Montero et al., 2020). After quality control (QS > 7 and length > 150bp), the remaining sequences were analyzed using the SqueezeMeta pipeline, with taxonomic resolution reaching genus level or family level in case of unclassified genus (1240 taxa). Count zero values were imputed for missing taxa to allow computing logarithms. The imputation was done using a Bayesian Multiplicative Replacement procedure with the geometric Bayesian multiplicative method (GBM) from the *cmultRepl* function of the *zCompositions* R package (Palarea-Albaladejo et al., 2019).

***Genotyping and analysis*.** Cows were genotyped using the EURO12K SNP chip from Illumina, and then imputed to 54609 SNPs (Bovine 50k SNP chip, Illumina, San Diego, California, USA) with BEAGLE software (http://faculty.washington.edu/browning/beagle/beagle.html), using the Spanish reference population provided by CONAFE (Spanish Friesian Associations Confederation), as described by (Jiménez-Montero et al., 2013). Monomorphic SNPs and those with MAF<0.05 were filtered out from the analysis, resulting in 42,372 SNP. Real data was analysed using the same models implemented in the simulation.

Ordination methods and distance matrices to build the Ks

Due to the compositional nature of metagenomic data (Gloor et al., 2017), a centred log ratio transformation (CLR) method (Aitchison, 1986) was implemented using the unweighted option of the *CLR* function from the *easyCODA* R package (Greenacre, 2018), as follows:

$$\mathbf{x_{clr}} = [\log(x_1/A(x)), \log(x_2/A(x)) \ldots log\ (x_D/A(x))],$$

with $A(x) = \sqrt[D]{x_1 * x_2 * \ldots * x_D}$.

Here, $\mathbf{x} = [x_1, x_2, \ldots, x_D]$ is a vector of counted features in a sample, and *A(x)* is the geometric mean of **x**.

A **K** matrix was built as $\mathbf{K}=1/p\mathbf{X}\mathbf{X}^{\mathbf{T}}$. Briefly, let **X** be the scaled matrix of centered log-ratio transformations of a $n \times p$ matrix **B** of equivalent dimentions, where n is the number of animals and p is the number of phyla. Each element of the **B** matrix $\{B_{ij}\}$ is the relative abundance of phylum j in animal i.

Additionally, other 11 methods to build the **K** matrices were implemented. First, five ordination methods were used to build **K**: Metric Multidimensional Scaling (MDS), Detrended correspondence analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy analysis (RDA) and Constrained correspondence analysis (CCA). Then, other six **K** were built using Euclidean, Bray-Curtis, Canberra, Jaccard, Mahalanobis and Aitchison metrics. A different **K** matrix was generated from each method. For that purpose, the **X** matrix mentioned before was independently replaced into the $(1/p)\mathbf{X}\mathbf{X}^{\mathbf{T}}$ approach with the elements in **X** $\{X_{ij}\}$differing according to the method to build the **K** matrix (Table 1).

The ordination matrices (MDS, DCA, NMDS, CCA and RDA) were built through the *ordinate* function of the *phyloseq* package in R. The Euclidean, Bray-Curtis, Canberra, Jaccard and Mahalanobis matrices were built using the *vegdist* function of *vegan* package in R. The Aitchison matrix was built with the *dist* function of the *coda.base* package also in R.

Variance component analysis and effects estimation

Genomic BLUP (GBLUP). A GBLUP model was implemented within a Bayesian framework using BGLR package (Pérez and De Los Campos, 2014), in order to estimate the proportion of the methane production variance attributable to the additive genetic variance, as:

$$\mathbf{y} = \mathbf{1}'\mu + \mathbf{Zu} + \mathbf{e}$$

where: **y** = methane production pre-corrected by fixed effects of parity, days in milk and herd-robot, μ = population mean , **1** = vector of ones of *n*x*1* dimensions, **u** = genetic effect, **Z** the corresponding incidence matrix for the genetic effect, and **e** = residual error, with $\mathbf{u} = \{u_i\} \sim N(0, \mathbf{G}\sigma_u^2)$ and $\mathbf{e} \sim N(0, \sigma_e^2)$. The prior distributions were the same as described in the simulation. The genomic relationship matrix (**G**), was constructed following method 2 of VanRaden (2008).

Microbiomic BLUP (MBLUP). Analogous as in the previous model, the proportion of the variance of methane production attributable to the microbiota variance was estimated using the following model:

$$\mathbf{y} = \mathbf{1}'\mu + \mathbf{Wm} + \mathbf{e}$$

Where: **m** = microbiota effect, **W** the corresponding incidence matrices for the microbiota effect, with $\mathbf{m} = \{m_i\} \sim N(0, \mathbf{K}\sigma_m^2)$ and **K** is the relationship matrix between cow rumen microbiotas. The other terms are as described above.

Hologenomic BLUP of additive effects (HBLUP). This approach assumes independent effects of genotype and microbiome. Mixed models were implemented using the following independent effect model in linear notation:

$$\mathbf{y} = \mathbf{1}'\mu + \mathbf{Zu} + \mathbf{Wm} + \mathbf{e}$$

With terms as described above.

Hologenomic BLUP with host interaction (H$_i$BLUP). Finally, a model accounting for the interaction between the genetic and the microbiota effects was tested:

$$\mathbf{y} = \mathbf{1}'\mu + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Tu}x\mathbf{m} + \mathbf{e}$$

Where: **y**, **μ, Zu**, **Wm** and **e** are the same as in the previous model and **u***x***m** stands for the interaction between genetic background of the host and her microbiome, **T** represent the corresponding incidence matrix, with **u***x***m** = $\{u_i \; x \; m_i\} \sim N(0, \mathbf{G\#K}\sigma^2_{uxm})$ where # stands for the Hadamard product.

All models were implemented in a Bayesian framework using the BGLR package in R (Pérez and De Los Campos, 2014). The mean and standard error of the posterior distribution for the parameters of interest were obtained. In the case of simulated data, the average, and its standard error (SEM) from 25 replicates is provided.

Holobiability estimation

The proportion of the phenotypic variance explained by the additive genetic variance plus the microbiome variance was defined as "holobiability" (ho^2), which is the addition of the heritability $\sigma^2_u/(\sigma^2_u + \sigma^2_m + \sigma^2_e)$ and the microbiability $\sigma^2_m/(\sigma^2_u + \sigma^2_m + \sigma^2_e)$ in an additive effects model without interaction: $ho^2 = (\sigma^2_u + \sigma^2_m)/(\sigma^2_u + \sigma^2_m + \sigma^2_e)$. It may also include the variance from the interaction effect. In biological terms, the holobiability would be the proportion of the phenotypic variance attributable to the host-microbiome holobiont effects as follow: $ho^2 = (\sigma^2_u + \sigma^2_m + \sigma^2_{u\times m})/(\sigma^2_u + \sigma^2_m + \sigma^2_{u\times m} + \sigma^2_e)$.

***Models comparison*.** The DIC is a hierarchical modelling generalization of the Akaike Information Criterion (AIC) and the Bayesian information criterion (BIC), which consists of two components, a term that measures goodness-of-fit and a penalty term for increasing model complexity, in which models with smaller DIC should be preferred to models with larger DIC, because this point to a better fit and a lower degree of model complexity (Spiegelhalter et al., 2002); however, some authors (Sorensen and Gianola, 2002) consider DIC as a preliminary metric for screening alternative models. DIC can be calculated as follow:

$$DIC = \overline{D} + p_D$$

The first term ($\overline{D}$), is a Bayesian measure of model fit, defined as the posterior expectation of the deviance, while the second component (p_D) measures the complexity of the model by the effective number of parameters, also defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters (Berg et al., 2004). The log-likelihood evaluated at the posterior mean, as well as the posterior mean of the log-likelihood, the effective number of parameters and the DIC, for each model and method used, were reported to evaluate and compare model's plausibility.

All the models previously described in the simulation were also evaluated using the real data.

RESULTS AND DISCUSSION

Simulated data

Microbiota Relationship Matrices (simulation). The diagonal elements in the K matrices represent the alpha-diversity and the off-diagonal elements, the beta-diversity. Correlations between diagonal elements (Figure 1A) and off-diagonal elements (Figure 1B) in the K matrices studied are shown.

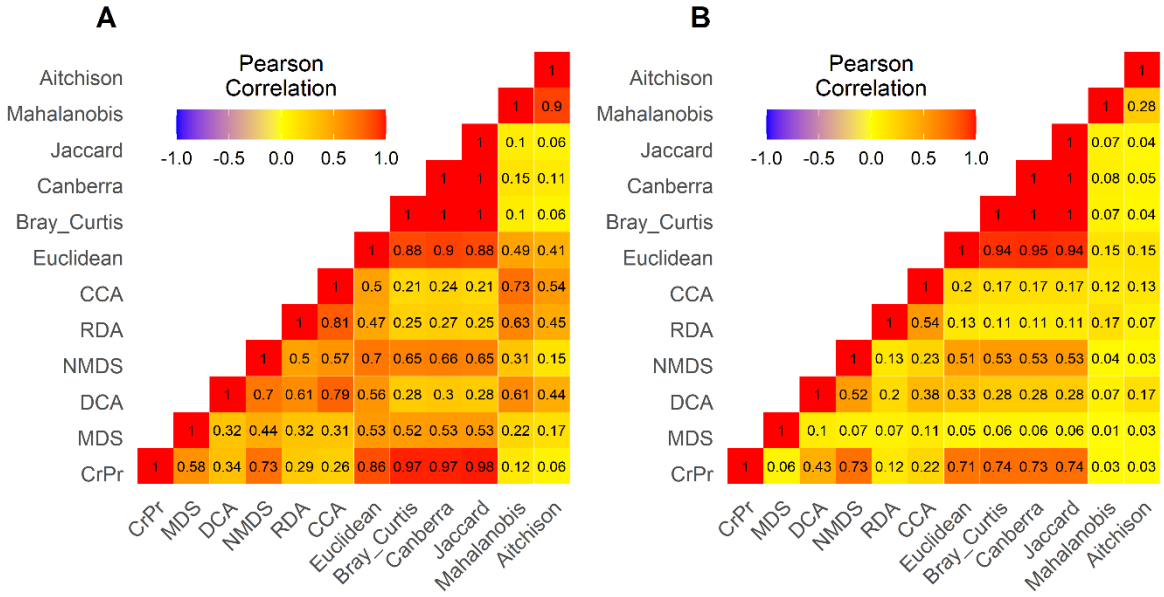

Figure 1. Pearson correlation between diagonal (A) and off-diagonal (B) elements of the centred log ratios of 1000 x 1000 simulated microbiome distance, dissimilarity, or index matrices according to ordination method*. The strength of the correlation is also represented with the intensity of the colour from -1 (blue) to 1 (red).
*$(1/p)\mathbf{X}\mathbf{X}^{\mathbf{T}}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Euclidean), Bray-Curtis dissimilarity (Bray_Curtis), Canberra distance (Canberra), Jaccard index (Jaccard), Mahalanobis distance (Mahalanobis) and Aitchison distance (Aitchison).

The correlation showed that all matrices except Bray-Curtis distance, Canberra distance and Jaccard index had different degrees of similitude between diagonal elements, ranging from null to high associations. The Jaccard index had a correlation of one with diagonal and out-diagonal elements of the Bray-Curtis matrix. However, the values on the diagonal are different between Ks. The Canberra distance also presented a correlation close to one (0.997) with diagonal and out-diagonal elements of the Bray-Curtis and Jaccard index matrices, which was driven by the CLR transformation of the data. Excluding that exceptions, the highest correlation ($\rho \geq 0.97$) between diagonal elements was obtained between the Bray-

Curtis, Canberra and Jaccard with the CrPr method. The lowest Pearson correlation ($\rho = 0.06$) between diagonal elements was obtained between CrPr, Bray-Curtis and Jaccard matrices with the Aitchison matrix. These matrices showed remarkable difference on how they calculate the distance between microbiota (Table 1).

In general terms, out-diagonal elements had lower correlations between methods than diagonal elements. There were correlations coefficients ranging from low to high between off-diagonal elements of Ks. The highest correlation ($\rho = 0.95$) was observed between Canberra distance with Euclidean distance methods (excluding correlations between Bray-Curtis distance and Jaccard index, which were equal to one). The lowest correlation ($\rho = 0.03$) between out-diagonal elements of the metagenome matrices was between CrPr, MDS and NMDS with the Aitchison method.

Heatmap graphic representations of the simulated microbiota relationship matrices according to method used to build them is depicted in Figure 2.

Table 1. Method and its linked distance, dissimilarity, or index used to build microbiota relationship matrices (K) via cross-product ($\mathbf{(1/p)XX^T}$) approach[1].

Method	Distance, dissimilarity, or Index d(x_j, x_k)	Analysis applied to CLR of RA	Elements in X matrix before $\mathbf{(1/p)XX^T}$
CrPr	$\mathbf{(1/p)XX^T}$	None	CLR of RA
MDS	$\boldsymbol{BrCr_{jk}} = \frac{\sum_{i=1}^{n}\lvert x_{ij} - x_{ik}\rvert}{\sum_{i=1}^{n} x_{ij} + \sum_{i=1}^{n} x_{ik}}$	PCoA	Vectors
DCA	$\boldsymbol{BrCr_{jk}} = \frac{\sum_{i=1}^{n}\lvert x_{ij} - x_{ik}\rvert}{\sum_{i=1}^{n} x_{ij} + \sum_{i=1}^{n} x_{ik}}$	CA, DT	Projections
NMDS	$\boldsymbol{BrCr_{jk}} = \frac{\sum_{i=1}^{n}\lvert x_{ij} - x_{ik}\rvert}{\sum_{i=1}^{n} x_{ij} + \sum_{i=1}^{n} x_{ik}}$	MR	Points
CCA	$x^2_{\mathrm{j,k}} = \sqrt{\sum_{j=1}^{p} \frac{(x_{ik} - x_{ij})^2}{a_{i+}}}$	CA	Projections
RDA	$x^2_{\mathrm{j,k}} = \sqrt{\sum_{j=1}^{p} \frac{(x_{ik} - x_{ij})^2}{a_{i+}}}$	PCA	Components
Euclidean	$\boldsymbol{Eucl_{jk}} = \sqrt{\sum_{i=1}^{n} (x_{ij} - x_{ik})^2}$	Euclidean	Distance of CLR of RA
Bray-Curtis	$\boldsymbol{BrCr_{jk}} = \frac{\sum_{i=1}^{n}\lvert x_{ij} - x_{ik}\rvert}{\sum_{i=1}^{n} x_{ij} + \sum_{i=1}^{n} x_{ik}}$	Bray-Curtis	Dissimilarity of CLR of RA
Canberra	$\boldsymbol{Canb_{jk}} = \frac{1}{NZ}\sum_{i=1}^{n} \frac{\lvert x_{ij} - x_{ik}\rvert}{\lvert x_{ij}\rvert + \lvert x_{ik}\rvert}$	Canberra	Distance of CLR of RA
Jaccard	$\boldsymbol{Jacc_{jk}} = 2\boldsymbol{BrCr}/(1 + \boldsymbol{BrCr})$	Jaccard	Index of CLR of RA
Mahalanobis	$\boldsymbol{Maha_{jk}} = \sqrt{\sum_{i=1}^{n}\sum_{l=1}^{n} (x_{ij} - x_{ik})^n\, \delta^{-1} (x_{lj} - x_{lk})}$	Mahalanobis	Distance of CLR of RA
Aitchison	$\boldsymbol{Aitc_{jk}} = \sqrt{\sum_{i=1}^{n} \left(log\left(\frac{x_{ij}}{A(x_j)}\right) - log\left(\frac{x_{ik}}{A(x_k)}\right)\right)^2}$	Aitchison	Distance of CLR of RA

[1] $\mathbf{(1/p)XX^T}$ = let **X** be the scaled matrix of centered log-ratio transformations of a $n \times p$ matrix **B** of equivalent dimentions, where n is the number of animals and p is the number of phyla. Each element of the **B** matrix $\{B_{ij}\}$ is the relative abundance of phylum j in animal i. CLR = Centred Log-Ratio transformation. RA= Relative Abundance. PCoA= Principal Coordinate Analysis. DT= Detrending. CA = Correspondence Analysis. MR= Monotone regression $\boldsymbol{a_{i+}}$ = total abundance of a given taxa in all the samples. NZ= number of non-zero entries. δ = Covariance matrix. A = geometric mean.

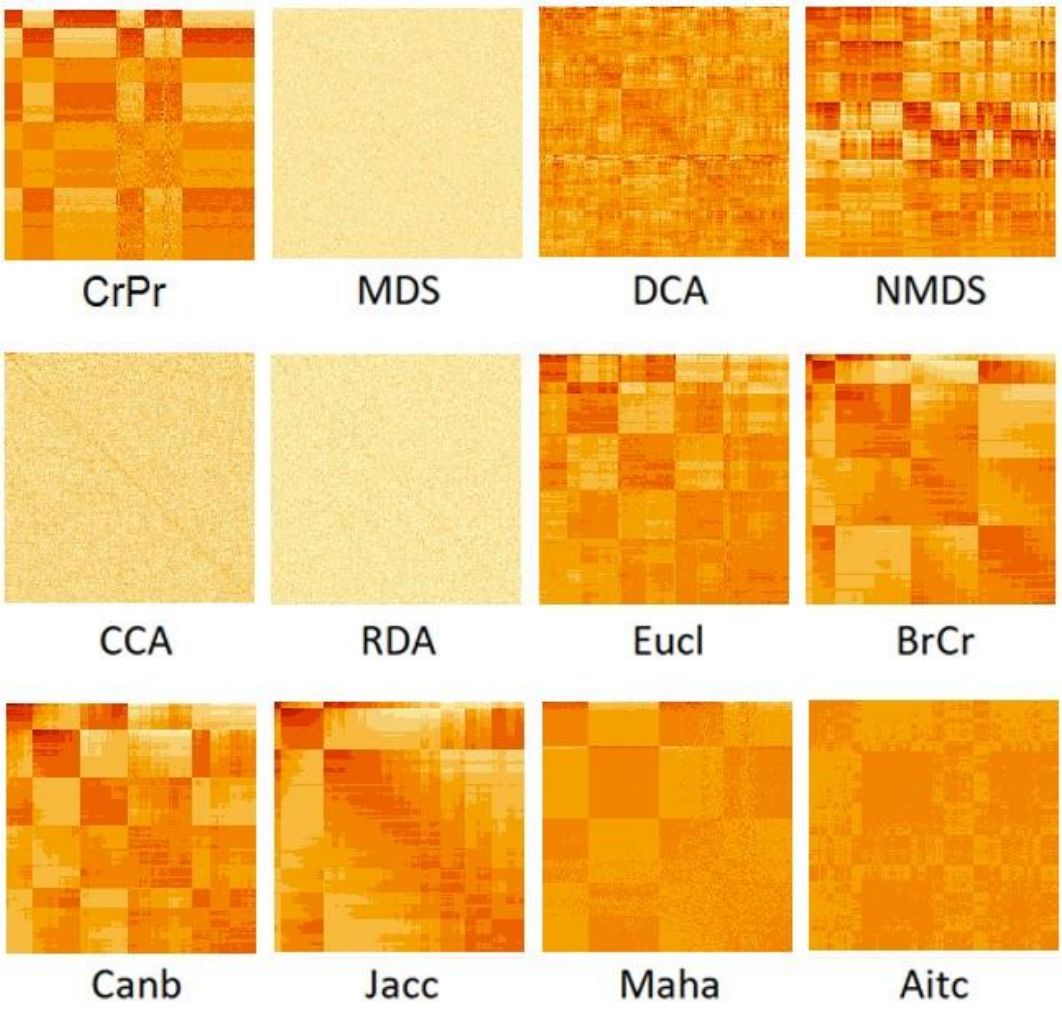

Figure 2. Simulated microbiota relationship matrices from rumen content (*n*=1000) according to ordination method*. Darker colours represent higher values and differ depending on the distance, dissimilarity or index used to build the K.
*$(1/p)\mathbf{X}\mathbf{X}^{\mathbf{T}}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Constrained Correspondence Analysis (CCA), Redundancy Analysis (RDA), Euclidean distance (Eucl), Bray-Curtis dissimilarity (BrCr), Canberra distance (Canb), Jaccard index (Jacc), Mahalanobis distance (Maha) and Aitchison distance (Aitc).

Figure 2 shows clear differences between Ks for the beta-diversity, in most methods represented as darker zones of the out-diagonal elements, compared to others showing flat homogeneous out-diagonal elements (MDS, RDA, CCA).

Heritability estimates. Mean heritability (±SE) from the GBLUP model was estimated at 0.34 (±0.01). Including the metagenome information resulted in a most accurate estimation of heritability, GBLUP might be capturing microbiome effect, and therefore, slightly overestimating heritability. The HBLUP provided the most accurate heritabilities: 0.30 (±0.01) in RDA, 0.29 (±0.01) in MDS and 0.29 (±0.01) in CCA methods (Figure 3A).

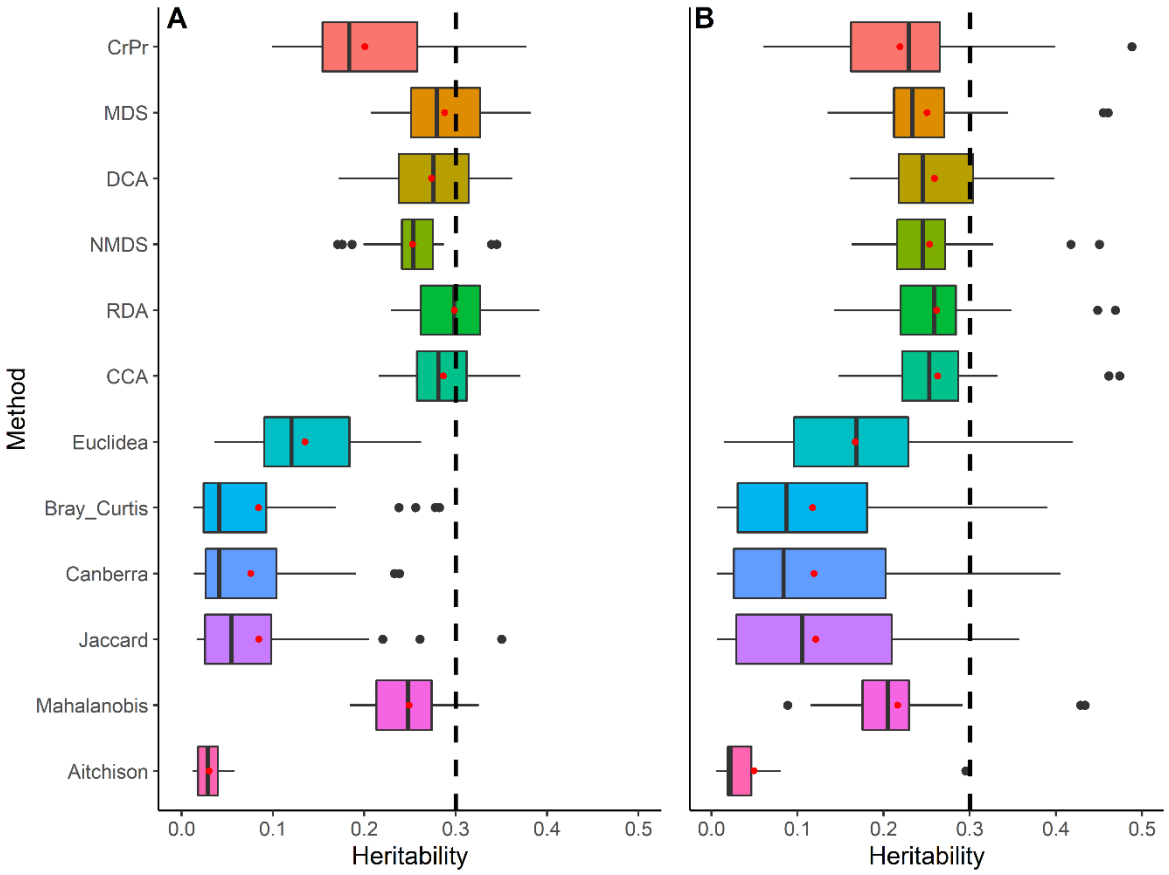

Figure 3. Heritability for methane production according to method to build the microbiota relationship matrix (**K**) for the ruminal microbiota using microbiome distance, dissimilarity or index matrices* from A) genetic and microbiome effects (HBLUP) and from B) genetic, microbiome and interaction effects model (H_iBLUP) using simulated data for 1000 cows and 25 replicates.

*$(1/p)\mathbf{XX}^{\mathbf{T}}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Euclidean), Bray-Curtis dissimilarity (Bray_Curtis), Canberra distance (Canberra), Jaccard index (Jaccard), Mahalanobis distance (Mahalanobis) and Aitchison distance (Aitchison).

The estimated heritability was slightly lower than the simulated (0.30) by matrices MDS, DCA, NMDS, RDA and CCA (Figure 3A). The other models underestimated the simulated value with larger differences. In the H_iBLUP model, simulated heritability (0.30) was underestimated by all methods following a similar pattern as in the HBLUP model (Figure 3B). In the H_iBLUP model, the most accurate estimation of heritability was performed by the CCA method (0.26 ± 0.02) followed by the RDA (0.26 ±0.02) and DCA (0.26 ± 0.01) matrices (Figure 3B). The Aitchison matrix resulted in the most biased estimated heritability (Fig. 3). The estimated heritability was slightly lower than the simulated (0.30) by method of MDS, DCA, NMDS, RDA and CCA (Figure 3A). The other models underestimated the simulated value with larger differences. In the H_iBLUP model, simulated heritability (0.30) was underestimated by all methods following a similar pattern as in the HBLUP model (Figure 3B).

Microbiability estimates. The microbiability was simulated to a value of 0.30. The model that ignored the genetic effect (MBLUP) resulted in closer estimates to the simulated value when the following matrices were used: RDA (0.29 ± 0.00), MDS (0.27 ± 0.00) and CCA (0.34 ± 0.01) (Figure 4A). In general, incorporating the genomic effect (HBLUP) resulted in similar or more accurate microbiability estimates: RDA (0.30 ± 0.00), DCA (0.30 ± 0.01), NMDS (0.29 ± 0.02), MDS (0.29 ± 0.01) and CCA (0.31 ± 0.01) (Figure 4B). The other matrices (Euclidean, Bray-Curtis, Canberra, Jaccard, Aitchison) showed a relevant biased estimate for the microbiability. Adding the interaction effect in the model (H_iBLUP) slightly reduced this bias but did not resulted in accurate estimates. Besides, it increased the SE in the rest of the matrices (Figure 4C).

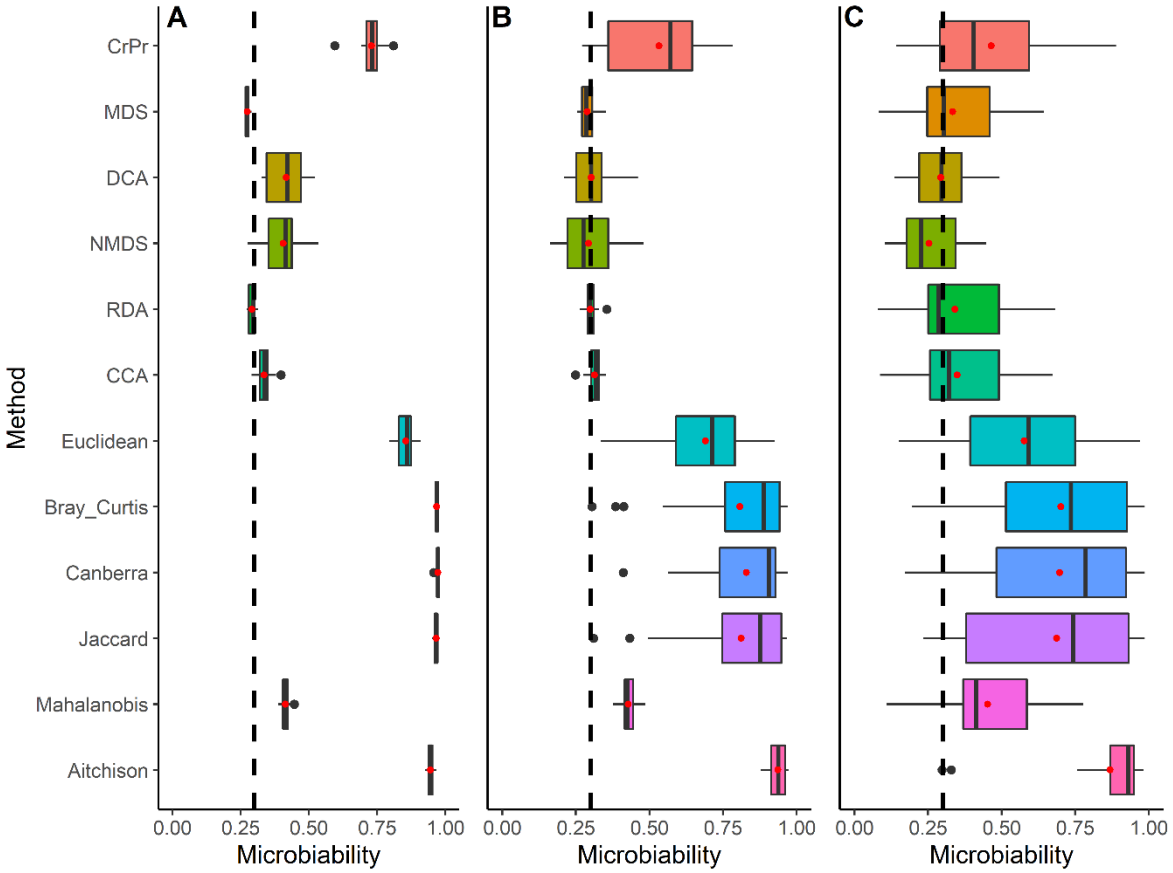

Figure 4. Microbiability for methane production according to method to build the microbiota relationship matrix (K) for the ruminal microbiota using microbiome distance, dissimilarity or index matrices*. A) from the microbiome effect model (MBLUP), B) genetic and microbiome additive effects model (HBLUP) and C) genetic and microbiome interaction effect model (H_iBLUP) using simulated data for 1000 cows and 25 replicates.
*$(1/p)\mathbf{XX^T}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Euclidean), Bray-Curtis dissimilarity (Bray_Curtis), Canberra distance (Canberra), Jaccard index (Jaccard), Mahalanobis distance (Mahalanobis) and Aitchison distance (Aitchison)

Correlation between GEBV and TBV. The average correlation between genomic estimated breeding values (GEBV) and true breeding values (TBV) from the GBLUP model was 0.62 ± 0.00. The correlation increased when the microbiome effect was included in the model (HBLUP). The largest correlation was for the CrPr method (0.99 ± 0.00), followed by the CCA (0.98 ± 0.00) and Euclidean (0.97 ± 0.00) methods. The lowest correlation was obtained for the NMDS method (0.71 ± 0.05). In the H_iBLUP model, the method with the highest correlation was CrPr (0.99 ± 0.00) followed by Euclidean (0.97 ± 0.00) and Canberra

(0.95 ± 0.01). The correlation increased considerably when the microbiome effect was included into the model, compared to the GBLUP model. Interestingly, the correlation between GEBV and TBV slightly decreased in some of the methods with the inclusion of the interaction effect (H_iBLUP), compared to HBLUP (Table 2).

Table 2. Heritability, microbiability and correlation between estimated and true breeding or microbiome values for methane production according to method to build the microbiota relationship matrix (K) for the ruminal microbiota, from a microbiome effect model (MBLUP), genetic and microbiome additive effects model (HBLUP) and for an interaction effect model (H_iBLUP), using simulated data for 1000 cows and 25 replicates[1].

Variance Components	Simulated	CrPr	MDS	DCA	NMDS	CCA	RDA	Eucl	BrCr	Canb	Jacc	Maha	Aitc
$\mathbf{y = 1'\mu + Wm + e}$, (MBLUP)													
Microbiability	0.3	0.73	0.27	0.42	0.41	0.34	**0.29**	0.85	0.97	0.97	0.97	0.41	0.95
SEM Microbiability	---	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Correlation EMV vs TMV	---	**0.98**	0.73	0.84	0.28	0.97	0.91	0.94	0.86	0.84	0.86	0.91	0.94
SEM Corr. EMV vs TMV	---	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\mathbf{y = 1'\mu + Zu + Wm + e}$, (HBLUP)													
Heritability	0.3	0.20	0.29	0.27	0.25	0.29	**0.30**	0.14	0.08	0.08	0.09	0.25	0.03
SEM Heritability	---	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.00
Microbiability	0.3	0.53	0.29	0.30	0.29	0.31	**0.30**	0.69	0.81	0.83	0.81	0.43	0.94
SEM Microbiability	---	0.03	0.01	0.01	0.02	0.01	0.00	0.03	0.04	0.03	0.04	0.01	0.01
Correlation GEBV vs TBV	---	**0.72**	0.67	0.69	0.68	0.71	0.70	0.71	0.71	0.71	0.71	0.70	0.71
SEM Corr. GEBV vs TBV	---	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Correlation EMV vs TMV	---	**0.99**	0.79	0.86	0.71	0.98	0.92	0.97	0.94	0.94	0.94	0.92	0.94
SEM Corr. EMV vs TMV	---	0.00	0.01	0.01	0.05	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.01
$\mathbf{y = 1'\mu + Zu + Wm + Tuxm + e}$, ($\mathbf{H_i}$BLUP)													
Heritability	0.3	0.22	0.25	**0.26**	0.25	**0.26**	**0.26**	0.17	0.12	0.12	0.12	0.22	0.05
SEM Heritability	---	0.02	0.02	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Microbiability	0.3	0.46	0.33	**0.29**	0.25	0.35	0.34	0.58	0.70	0.70	0.69	0.45	0.87
SEM Microbiability	---	0.05	0.03	0.02	0.02	0.03	0.03	0.05	0.05	0.05	0.05	0.04	0.04
Correlation GEBV vs TBV	---	**0.73**	0.69	0.70	0.69	**0.73**	0.72	0.73	0.72	0.72	0.72	0.72	**0.73**
SEM Corr. GEBV vs TBV	---	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Correlation EMV vs TMV	---	**0.99**	0.78	0.86	0.76	0.91	0.91	0.97	0.95	0.95	0.94	0.91	0.89
SEM Corr. EMV vs TMV	---	0.00	0.03	0.02	0.05	0.01	0.02	0.00	0.01	0.01	0.01	0.02	0.01

[1]Method of $(1/p)\mathbf{XX^T}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Eucl), Bray-Curtis dissimilarity (BrCr), Canberra distance (Canb), Jaccard index (Jacc), Mahalanobis distance (Maha) and Aitchison distance (Aitc). Values closest to the simulated value or better estimates are depicted in bold.

Correlation between EMV and TMV. The correlation between the estimated microbiome value (EMV) and the true microbiome value (TMV) were in general larger than 0.90. The larger values were estimated with CrPr (0.98 ± 0.00) and CCA (0.97 ± 0.00). In the HBLUP model, the CrPr method had the highest correlation (0.72 ± 0.01), followed by CCA (0.71 ± 0.01). In the H_iBLUP model the CrPr (0.99 ± 0.00) and Euclidean (0.97 ± 0.00) methods showed the higher values of correlation between EMV and TMV. The lowest correlation obtained for these parameters was for the NMDS (0.28 ± 0.00) method Table 2.

A graphic representation of the correlation between EMV and TMV from the microbiome effect model is shown in Figure 5, where different accuracies of the association patterns are depicted according to the microbiome matrices.

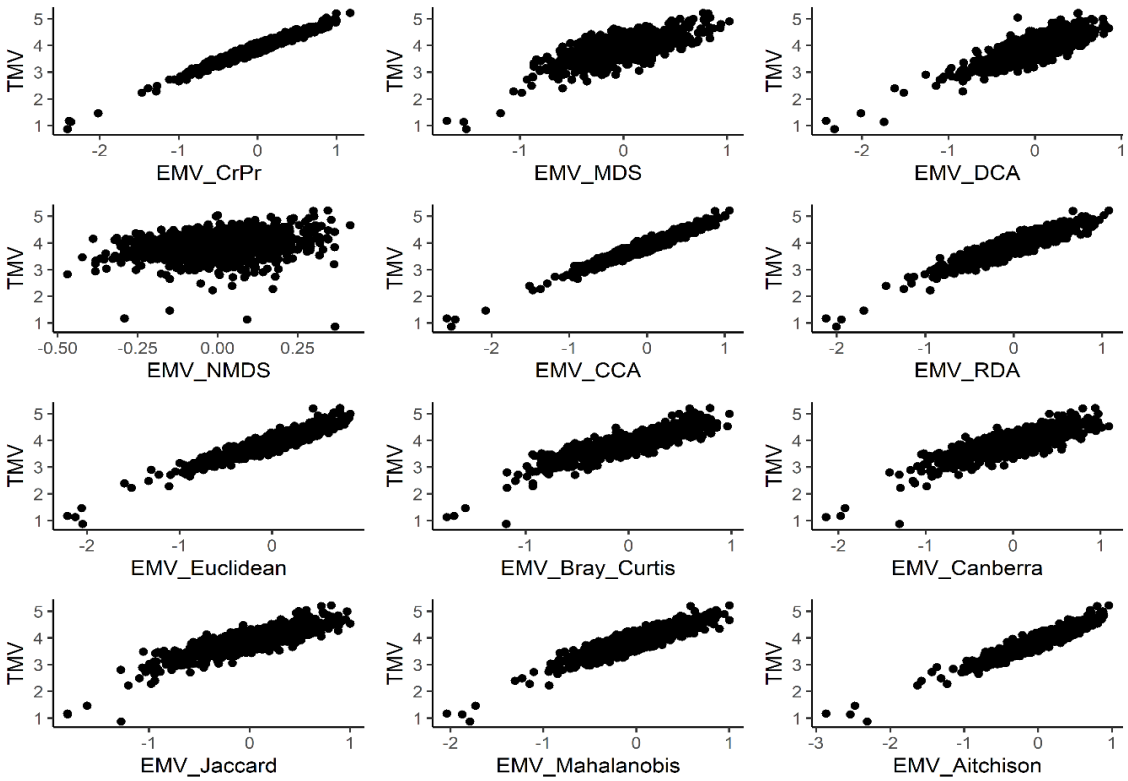

Figure 5. Correlation between estimated microbiome values (EMV) and true microbiome values (TMV) according to method to build the K using microbiome distance, dissimilarity or index matrices*from the microbiome effect model (MBLUP) using simulated data for 1000 cows and 25 replicates.

*$(1/p)\mathbf{XX}^{\mathbf{T}}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Euclidean), Bray-Curtis dissimilarity (Bray_Curtis), Canberra distance (Canberra), Jaccard index (Jaccard), Mahalanobis distance (Mahalanobis) and Aitchison distance (Aitchison).

Heritability, microbiability and correlation between estimated and true breeding or microbiome values for the simulated data are shown in Table 2. Results are summarized according to method to build the K for the ruminal microbiota from a MBLUP, HBLUP and for a H_iBLUP model, using simulated data for 1000 cows and 25 replicates. Simulations enable to get insight into expected performance of methods to be applied to real data. It also has been an effective way for the evaluation and development of new breeding strategies (Faux et al., 2016). Based in the results obtained from our simulated simplified reality, it is expected that MDS, RDA and CCA ordination methods will perform better in real data for variance components estimation.

Dairy data

Microbiome composition. The relative abundance of the 86 phyla and 1240 genus are depicted in Figure 6. Showing some variability for the microbiome composition between animals.

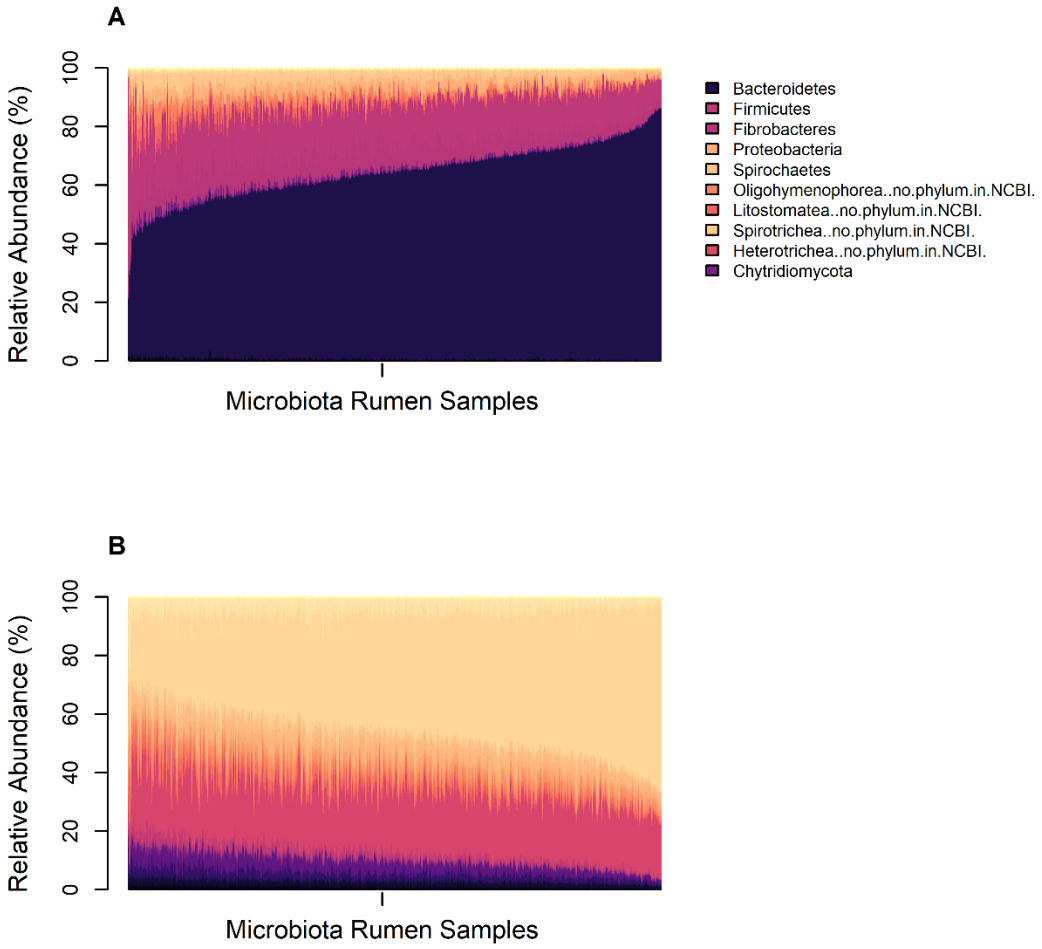

Figure 6. Relative abundance (%) of phyla sorted incrementally by the abundance of Bacteroidetes and coloured by the top 10 most frequent taxa (A) and genera sorted incrementally by the abundance of Prevotellaceae (B) according to microbiota rumen sample for 437 Holstein cows from 14 herds at 4 Northern Regions of Spain (País Vasco, Cantabria, Navarra, and Girona).

The variability between animals for microbiome composition implies a potential useful tool to improve traits of interest within animal breeding programs, its utility depends on a consistent and at least moderate microbiability across populations. Some specific phylum and genus were more frequent than others within an animal, but in general following a similar pattern of ranking of abundances according to phylum and genus. Most studies inform of higher proportions of Bacteroidetes and Firmicutes with consistent and conserved abundance rank structure of microbiome across geographical locations, breeds and diets

(Wallace et al., 2019). Consistently with our results, other studies (Deusch et al., 2017; Vaidya et al., 2018), also informed of the Prevotellaceae family dominating the phylum of Bacteroidetes and the overall bacterial community composition in rumen (Figure 6B).

Microbiota Relationship Matrices. In general terms, out-diagonal elements had lower correlations between methods than diagonal elements. There were correlations coefficients ranging from null to high between out-diagonal elements of Ks, the highest correlation (ρ = 0.78) excluding correlations between Bray-Curtis dissimilarity and Jaccard index was between Canberra distance with Bray-Curtis dissimilarity and Jaccard Index methods. The lowest correlations were between out-diagonal elements of the K matrices from CCA, RDA and Mahalanobis distance methods with the rest of methods (Figure 7). There are considerable differences between Ks, therefore, different performances in mixed models are expected.

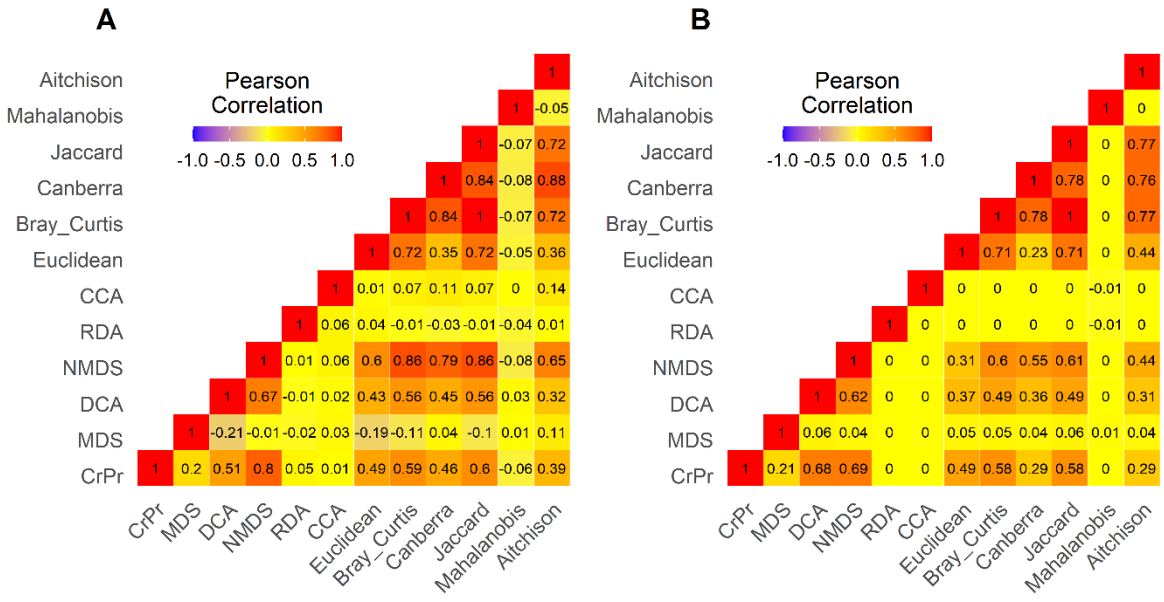

Figure 7. Pearson correlation between diagonal (A) and off-diagonal (B) elements of the centred log ratios of 437 x 437 real microbiome distance, dissimilarity, or index matrices according to ordination method*. The strength of the correlation is also represented with the intensity of the colour from -1 (blue) to 1 (red).

*$(1/p)\mathbf{X}\mathbf{X}^{\mathbf{T}}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Euclidean), Bray-Curtis dissimilarity (Bray_Curtis), Canberra distance (Canberra), Jaccard index (Jaccard), Mahalanobis distance (Mahalanobis) and Aitchison distance (Aitchison).

There were differences between the microbial matrices according to ordination method, as well as on the distance dissimilarity or index used. A heatmap graphic representation of the obtained K according to method used is depicted in Figure 8.

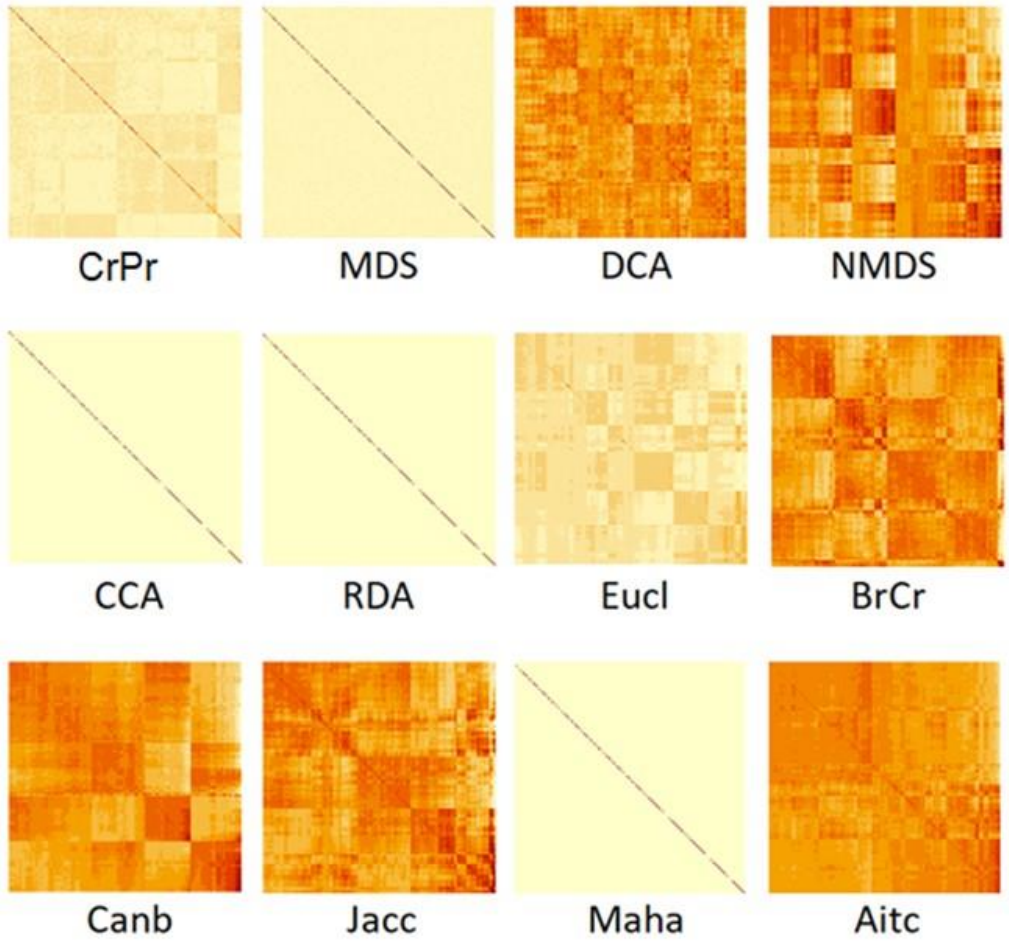

Figure 8. Microbiota relationship matrices of centred log-ratios of the microbial composition of rumen content for 437 Holstein cows according to ordination method*. Darker colours represent higher values and differ depending on the distance, dissimilarity or index used to build the K.

*$(1/p)\mathbf{X}\mathbf{X}^{\mathbf{T}}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Constrained Correspondence Analysis (CCA), Redundancy Analysis (RDA), Euclidean distance (Eucl), Bray-Curtis dissimilarity (BrCr), Canberra distance (Canb), Jaccard index (Jacc), Mahalanobis distance (Maha) and Aitchison distance (Aitc).

The Figure 8. illustrates that most Ks highlight the differences between alpha and beta-diversity, clearly evidencing the diagonal elements from the off-diagonal elements. There were also clear differences between Ks for the beta-diversity, in some methods (DCA, NMDS, Bray-Curtis, Canberra, Jaccard, Aitchison) represented as darker zones of the off-diagonal elements, compared to others showing faint slight dark zones (CrPr, Euclidean) or even absent asymmetric zones among off-diagonal elements of the Ks, depicting a flat homogeneous off-diagonal platform (MDS, RDA, CCA, Mahalanobis).

Variance components estimations. Genetic (σ_u^2), microbiome (σ_m^2), genetic x microbiome interaction (σ_{uxm}^2) and residual (σ_e^2) variances for methane emission were estimated when appropriate from an GBLUP, MBLUP, HBLUP, and H_iBLUP using real data (Figure 9).

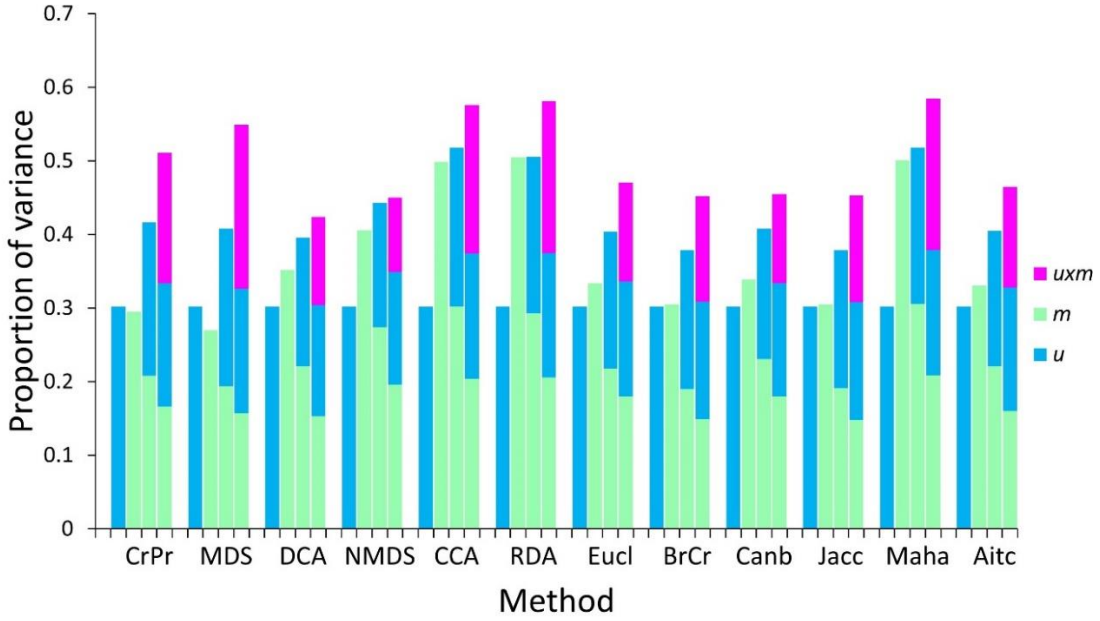

Figure 9. Proportion of phenotypic variance explained by the additive genetic variance (σ_u^2), microbiome variance (σ_m^2) and interaction variance (σ_{uxm}^2) according to a GBLUP, MBLUP, HBLUP, and H_iBLUP model, depending on the distance, dissimilarity or index used to build the K*.

*$(1/p)\mathbf{XX^T}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Constrained Correspondence Analysis (CCA), Redundancy Analysis (RDA), Euclidean distance (Eucl), Bray-Curtis dissimilarity (BrCr), Canberra distance (Canb), Jaccard index (Jacc), Mahalanobis distance (Maha) and Aitchison distance (Aitc).

Small variations were obtained between σ_u^2 across all the ordination methods, while for σ_m^2 and σ_e^2 there were higher differences between estimates according to K used. Results of h^2 estimates decreased with the inclusion of the microbiome effect and decreased even more with the inclusion of the microbiome × genetic effect into the models and were close across methods within a model. Results of m^2 estimates also decreased with the inclusion of the genetic effect and decreased even more with the inclusion of the microbiome × genetic effect into the models.

The estimated mean for methane h^2 was 0.30 for GBLUP, and it ranged from 0.15 (DCA, NMDS, Canberra) to 0.22 (CCA) in all the models that included the microbiome. The m^2 estimates ranged from 0.27 (MDS) to 0.51 (RDA) in the MBLUP, from 0.19 (MDS, Bray-Curtis, Jaccard) to 0.31 (Mahalanobis) in the HBLUP and from 0.15 (DCA, Bray-Curtis, Jaccard) to 0.21 (RDA, Mahalanobis) in the H_iBLUP. Microbiability estimates were close to heritability estimates but they showed larger differences between K matrices used.

Correlations between the posterior mean GEBVs and the methane emission phenotype were high and consistent across models that included the microbiome effect. These correlations ranged from 0.86 to 0.88 in the HBLUP model and from 0.84 to 0.86 in the H_iBLUP model, these values were slightly lower than the value obtained for the benchmark GBLUP model (0.90). The correlation between the posterior means of the EMV and the phenotype ranged from 0.18 (NMDS) to 1.00 (CCA, RDA, Mahalanobis) in the MBLUP model, from 0.15 (NMDS) to 0.99 (CCA, RDA, Mahalanobis) in the HBLUP model and from 0.14 (NMDS) to 0.99 (CCA, RDA, Mahalanobis) in the H_iBLUP model (Table 3).

The correlation between the estimated genetic × microbiome interaction value (EGMV) and methane production in the methods of: CrPr (0.97), MDS (0.99), CCA (0.99), RDA (0.99) and Mahalanobis (0.99) were remarkably higher than the correlation between GEBV with methane production obtained in the GBLUP (0.90). The correlation between EGMV and methane production for MDS (0.99) was higher than the correlation between EMV and methane production obtained in the MBLUP model (0.88) or the HBLUP model (0.88). For CCA, RDA and Mahalanobis correlations between phenotype and EMV or EGMV were remarkably similar for all models (Table 3).

The best methods, in terms of correlation between EMV or EGMV with methane concentration in the H_iBLUP model, were: CrPr (0.97), MDS (0.99), CCA (0.99), RDA (0.99) and Mahalanobis (0.99). These results were remarkably higher than those obtained from the GBLUP (0.90). The largest proportion of phenotypic variance was explained by the same methods mentioned before, with a clear advantage of the holobiability parameter over the sum of microbiability and heritability. The methods with lowest correlation were NMDS (0.66) and Aitchison (0.73) (Table 3).

Table 3. Variance components, heritability, microbiability, holobiability and correlations between GEBV and phenotype, between EMV and phenotype and between EGMV and phenotype for methane production according to ordination method for the ruminal microbiota relationship matrix, from a GBLUP model, microbiome effect model (MBLUP), genetic and microbiome additive effects model (HBLUP) and for an interaction effect model (H_iBLUP), using real data for 437 cows[1].

Variance Components	CrPr	MDS	DCA	NMDS	CCA	RDA	Eucl	BrCr	Canb	Jacc	Maha	Aitc
$\mathbf{y = 1'\mu + Wm + e}$, **(MBLUP)**												
Microbiability (m^2)	0.30	0.27	0.35	0.41	0.50	**0.51**	0.33	0.31	0.34	0.31	0.50	0.33
Correlation GEBV with methane	0.86	0.88	0.20	0.18	**1.00**	**1.00**	0.62	0.43	0.34	0.46	**1.00**	0.58
$\mathbf{y = 1'\mu + Zu + Wm + e}$, (**HBLUP**)												
Heritability (h^2)	0.21	0.21	0.18	0.17	**0.22**	0.21	0.19	0.19	0.18	0.19	0.21	0.18
Microbiability (m^2)	0.21	0.19	0.22	0.27	0.30	0.29	0.22	0.19	0.23	0.19	**0.31**	0.22
Holobiability (ho^2)	0.42	0.40	0.40	0.44	**0.52**	0.50	0.41	0.38	0.41	0.38	**0.52**	0.40
Correlation GEBV with methane	0.86	**0.88**	0.86	0.87	0.87	0.87	0.86	0.86	0.87	0.86	0.87	0.86
Correlation EMV with methane	0.81	0.87	0.19	0.15	**0.99**	**0.99**	0.49	0.34	0.29	0.37	**0.99**	0.46
$\mathbf{y = 1'\mu + Zu + Wm + Tuxm + e}$, **($H_i$BLUP)**												
Heritability (h^2)	**0.17**	**0.17**	0.15	0.15	**0.17**	**0.17**	0.16	0.16	0.15	0.16	**0.17**	**0.17**
Microbiability (m^2)	0.17	0.16	0.15	0.20	0.20	**0.21**	0.18	0.15	0.18	0.15	**0.21**	0.16
Holobiability (ho^2)	0.51	0.55	0.42	0.45	0.58	0.58	0.47	0.45	0.46	0.45	**0.59**	0.47
Correlation GEBV with methane	0.85	**0.86**	0.84	0.85	0.85	0.85	0.84	0.84	0.85	0.84	0.85	0.85
Correlation EMV with methane	0.78	0.88	0.18	0.14	**0.99**	**0.99**	0.45	0.31	0.27	0.33	**0.99**	0.38
Correlation EGMV with methane	0.97	**0.99**	0.81	0.66	**0.99**	**0.99**	0.82	0.80	0.75	0.82	**0.99**	0.73

[1]Method of $(1/p)\mathbf{XX^T}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Eucl), Bray-Curtis dissimilarity (BrCr), Canberra distance (Canb), Jaccard index (Jacc), Mahalanobis distance (Maha) and Aitchison distance (Aitc). Larger values between ordination methods are depicted in bold.

Models comparison. Different criteria, that could assist on selecting the best model, were obtained from the analysis: log-likelihood evaluated at posterior mean, the posterior mean of the Log-Likelihood, estimated effective number of parameters, as well as the deviance information criterion (DIC) were obtained.

DIC is considered particularly useful for Bayesian model selection where the posterior distribution of the models has been obtained by Markov Chain Monte Carlo (MCMC) simulation. When prior information is negligible, DIC results in an equivalent approximation to Akaike's criterion (Spiegelhalter et al., 2002), but the DIC uses the posterior expectation of the log likelihood as a measure of model fit (Sorensen, 2004).The DIC can be used to decide adequacy of a model; the difference between DIC of two given models should suffice to make a good choice. Sorensen (2004) indicates that it is difficult to affirm what is an important difference in DIC and suggests that 10 definitively exclude the model with the highest DIC, differences from 5 to 10 are still substantial, but to choose a model that differs only by a value below 5 in DIC could be misleading.

Results from those criteria were obtained for the GBLUP, MBLUP, HBLUP and H_iBLUP models (Table 4).

Table 4. Information criteria estimated from a microbiome effect model (MBLUP), genetic and microbiome additive effects model (HBLUP) and for an interaction effect model (H_iBLUP) according to method of ordination for the microbiota relationship matrix for real data from 437 cows, a GBLUP model is included as reference [1].

Variance Components	CrPr	MDS	DCA	NMDS	CCA	RDA	Eucl	BrCr	Canb	Jacc	Maha	Aitc	GBLUP
$\mathbf{y = 1'\mu + Wm + e}$, **(MBLUP)**													
LLPM	-3109	-3117	-3223	-3225	-2973	-2969	-3177	-3199	-3210	-3196	-2972	-3185	-3106
PMLL	-3164	-3171	-3227	-3228	-3073	-3070	-3202	-3214	-3219	-3212	-3073	-3207	-3164
pD	110	107	8	6	200	201	49	28	18	32	201	43	115
DIC	**6440**	6450	6462	6464	**6347**	**6341**	6454	6456	6457	6456	**6349**	6458	6443
$\mathbf{y = 1'\mu + Zu + Wm + e}$, **(HBLUP)**													
LLPM	-3042	-3043	-3130	-3132	-2969	-2972	-3100	-3115	-3122	-3113	-2972	-3106	-3106
PMLL	-3123	-3125	-3176	-3178	-3072	-3074	-3158	-3166	-3171	-3166	-3074	-3162	-3164
pD	161	164	92	91	206	204	115	103	97	104	203	111	115
DIC	**6407**	**6414**	6445	6447	**6350**	**6354**	**6431**	**6437**	**6439**	**6437**	**6352**	**6436**	6443
$\mathbf{y = 1'\mu + Zu + Wm + Tuxm + }\boldsymbol{e}$, **($H_i$BLUP)**													
LLPM	-2978	-2949	-3087	-3094	-2927	-2923	-3047	-3058	-3076	-3055	-2920	-3055	-3106
PMLL	-3080	-3059	-3152	-3156	-3043	-3040	-3125	-3131	-3143	-3130	-3038	-3130	-3164
pD	204	220	128	123	232	234	156	147	135	150	235	150	115
DIC	**6365**	**6340**	**6432**	**6435**	**6318**	**6314**	**6407**	**6411**	**6422**	**6410**	**6311**	**6410**	6443

[1]Method of $(1/p)\mathbf{XX^T}$ (CrPr), Multidimensional Scaling (MDS), Detrended Correspondence Analysis (DCA), Non-Metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), Constrained Correspondence Analysis (CCA), Euclidean distance (Eucl), Bray-Curtis dissimilarity (BrCr), Canberra distance (Canb), Jaccard index (Jacc), Mahalanobis distance (Maha) and Aitchison distance (Aitc). LLPM = log-likelihood evaluated at posterior mean. PMLL = posterior mean of the Log-Likelihood. pD = estimated effective number of parameters. DIC = deviance information criteria. DIC lower than that of GBLUP are depicted in bold.

Between the three models that included the microbiome effect, the log-likelihood evaluated at posterior mean, the posterior mean of the Log-Likelihood and the estimated effective number of parameters values increased as the model included more factors, in all building methods of K. The DIC decreased with the complexity of the model. Those results indicate better fit of models that included the microbiome effect. According to the DIC, CrPr (6440), CCA (6347), RDA (6341) and Mahalanobis (6349) fitted better than the GBLUP (DIC = 6443) in the MBLUP model; all other MBLUP models were outperformed by the GBLUP. However, all HBLUP and H_iBLUP showed better fit than the GBLUP model (except HBLUP with DCA and NMDS matrices).

In the H_iBLUP model, all methods fitted better than the GBLUP model, with the lowest DIC obtained for the Mahalanobis method (DIC=6311), and differences larger than 5 in DIC regarding GBLUP. The best models sorted according to the differences between DIC from the GBLUP model were Mahalanobis (132), RDA (129), CCA (125), MDS (103), CrPr (78), Euclidean (36), Jaccard (33), Aitchison (33), Canberra (21), DCA (11), NMDS (8). The Mahalanobis K matrix used in the H_iBLUP model showed to be the most plausible from real data. Furthermore, the highest holobiability and the highest correlation between the EGMV and methane production were obtained with this method.

Based on the obtained results, there are enough differences between DIC to consider that the H_iBLUP models including genomic information, microbiota information and its interaction using the Mahalanobis, RDA, CCA and MDS method should be chosen rather than the GBLUP and over the MBLUP and other metagenomic matrices.

The RDA, CCA and MDS methods within the H_iBLUP model performed better than most other methods in the simulation and were by far, together with Mahalanobis method, the most plausible methods of building the metagenomic matrices for the real data, as determined by the DIC. These methods yielded outstanding results compared to classical approaches of variance component and other population parameters estimation (*i.e.* GBLUP), allowing for improvements of accuracy in estimated variance components of complex traits influenced by microbiome.

CONCLUSIONS

This study incorporated microbiome relationship matrices into the models commonly used in genetic evaluation and variance component estimation. Several metrics were used to calculate distances between rumen microbiotas. The MDS, CCA and RDA matrices achieved unbiased estimation of variance components in simulated data. The genomic breeding values were accurately inferred when a microbiome effect and its interaction was accounted for. Similar results were obtained with real data.

The deviance information criterion (DIC) was substantially lower for the H_iBLUP model, providing enough evidence to recommend this model with K matrices built with Mahalanobis, CCA, RDA or MDS methods. Accounting for the genome × microbiome interaction improved variance component estimation and it may have yielded more accurate performance predictions of methane emissions.

Holobiont modelling might also be extended to other relevant traits in dairy cattle. If holobiability is consistent across studies and its value surpasses heritability for the trait of interest, it might be a better estimator to be included into genetic evaluations, which a priori would increase the selection accuracy compared to the classical usage of heritability.

ACKNOWLEDGEMENTS

This research was financed by RTA2015-00022-CO3 (METALGEN) project from the Spanish national plan of research, development, and innovation 2013-2020. Alejandro Saborío-Montero acknowledges the scholarship from Universidad de Costa Rica for his doctorate studies which partially conducted to the progress of this study.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest regarding the topics approached within this study.

REFERENCES

Aitchison, J., 1986. The statistical analysis of compositional data. Chapman and Hall.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., Kachman, S.D., Moriyama, E.N., Walter, J., Peterson, D.A., Pomp, D., 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. 107, 18933–18938. https://doi.org/10.1073/pnas.1007028107

Berg, A., Meyer, R., Yu, J., 2004. Deviance information criterion for comparing stochastic volatility models. J. Bus. Econ. Stat. 22, 107–120. https://doi.org/10.1198/073500103288619430

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., Clark, A.G., 2015. Host genetic variation impacts microbiome composition across human body sites. Genome Biol. 16, 191. https://doi.org/10.1186/s13059-015-0759-1

Buitenhuis, B., Lassen, J., Noel, S.J., Plichta, D.R., Sørensen, P., Difford, G.F., Poulsen, N.A., 2019. Impact of the rumen microbiome on milk fatty acid composition of Holstein cattle. Genet. Sel. Evol. 51, 1–8. https://doi.org/10.1186/s12711-019-0464-8

Camarinha-silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S., 2017. Host genome influence on gut microbial composition and microbial prediction of complex traits in pigs.

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S., Bennewitz, J., 2017. Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. Genetics 206, 1637–1644. https://doi.org/10.1534/genetics.117.200782

Ciuffreda, L., Rodríguez-Pérez, H., Flores, C., 2021. Nanopore sequencing and its application to the study of microbial communities. Comput. Struct. Biotechnol. J. 19, 1497–1511. https://doi.org/10.1016/j.csbj.2021.02.020

Deusch, S., Camarinha-Silva, A., Conrad, J., Beifuss, U., Rodehutscord, M., Seifert, J., 2017. A structural and functional elucidation of the rumen microbiome influenced by

various diets and microenvironments. Front. Microbiol. 8, 1–21. https://doi.org/10.3389/fmicb.2017.01605

Difford, G.F., Plichta, D.R., Løvendahl, P., Lassen, J., Noel, S.J., Højberg, O., Wright, A.-D.G., Zhu, Z., Kristensen, L., Nielsen, H.B., Guldbrandtsen, B., Sahana, G., 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLOS Genet. 14, e1007580. https://doi.org/10.1371/journal.pgen.1007580

Eckard, R.J., Grainger, C., de Klein, C.A.M., 2010. Options for the abatement of methane and nitrous oxide from ruminant production: A review. Livest. Sci. 130, 47–56. https://doi.org/10.1016/j.livsci.2010.02.010

Faux, A., Gorjanc, G., Gaynor, R.C., Battagin, M., Edwards, S.M., Wilson, D.L., Hearne, S.J., Gonen, S., Hickey, J.M., 2016. AlphaSim: Software for Breeding Program Simulation. Plant Genome 9, 1–14. https://doi.org/10.3835/plantgenome2016.02.0013

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 8, 2224. https://doi.org/10.3389/fmicb.2017.02224

Gonzalez-Recio, O., Zubiria, I., García-Rodríguez, A., Hurtado, A., Atxaerandio, R., 2017. Short communication: Signs of host genetic regulation in the microbiome composition in 2 dairy breeds: Holstein and Brown Swiss. J. Dairy Sci. 101, 1–8. https://doi.org/10.3168/jds.2017-13179

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

Hadfield, J.D., Nakagawa, S., 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23, 494–508. https://doi.org/10.1111/j.1420-9101.2009.01915.x

Jarquín, Diego, Crossa, José, Lacaze, Xavier, Philippe, ·, Cheyron, D., Daucourt, J., Lorgeou, Josiane, Piraux, F., Laurent, ·, Paulino Pérez, G.·, Calus, Mario, Burgueño, J., De Los Campos, G., Jarquín, D, Pérez, · P, De Los Campos, · G, Crossa, J, Burgueño, · J, Lacaze, X, Cheyron, P. Du, Daucourt, · J, Lorgeou, J, Piraux, · F, Guerreiro, L., Pérez, P., Calus, M, 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127, 595–

607. https://doi.org/10.1007/s00122-013-2243-1

Jiménez-Montero, J., Gianola, D., Weigel, K., Alenda, R., González-Recio, O., 2013. Assets of imputation to ultra-high density for productive and functional traits. J. Dairy Sci. 96, 6047–6058. https://doi.org/10.3168/jds.2013-6793

Khanal, P., Maltecca, C., Schwab, C., Fix, J., Bergamaschi, M., Tiezzi, F., 2020a. Modeling host-microbiome interactions for the prediction of meat quality and carcass composition traits in swine. Genet. Sel. Evol. 52, 41. https://doi.org/10.1186/s12711-020-00561-7

Khanal, P., Maltecca, C., Schwab, C., Fix, J., Tiezzi, F., 2020b. Microbiability of meat quality and carcass composition traits in swine. J. Anim. Breed. Genet. 00, 1:14. https://doi.org/10.1111/jbg.12504

Lu, H., Giordano, F., Ning, Z., 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics, Proteomics Bioinforma. https://doi.org/10.1016/j.gpb.2016.05.004

Maltecca, C., Lu, D., Schillebeeckx, C., McNulty, N.P., Schwab, C., Shull, C., Tiezzi, F., 2019. Predicting Growth and Carcass Traits in Swine Using Microbiome Data and Machine Learning Algorithms. Sci. Rep. 9. https://doi.org/10.1038/s41598-019-43031-x

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W.M., Wang, X., Kachman, S.D., Auwerx, J., Williams, R.W., Benson, A.K., Peterson, D.A., Ciobanu, D.C., 2012. Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS One 7. https://doi.org/10.1371/journal.pone.0039191

Myhre, G., Shindell, D., Bréon, F., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., Zhang, H., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P., 2013. Anthropogenic and natural radiative forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I 659–740.

Palarea-Albaladejo, J., Antoni, J., Maintainer, M.-F., 2019. Package “zCompositions”

Treatment of zeros, left-censored and missing values in compositional data sets. https://doi.org/10.1016/j.chemolab.2015.02.019

Pérez, P., De Los Campos, G., 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198, 483–495. https://doi.org/10.1534/genetics.114.164442

Rey, J., Atxaerandio, R., Ruiz, R., Ugarte, E., González-Recio, O., Garcia-Rodriguez, A., Goiri, I., 2019. Comparison Between Non-Invasive Methane Measurement Techniques in Cattle. Animals 9, 1–9. https://doi.org/10.3390/ani9080563

Roehe, R., Dewhurst, R.J., Duthie, C.A., Rooke, J.A., McKain, N., Ross, D.W., Hyslop, J.J., Waterhouse, A., Freeman, T.C., Watson, M., Wallace, R.J., 2016. Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. PLoS Genet. 12, 1–20. https://doi.org/10.1371/journal.pgen.1005846

Ross, E.M., Moate, P.J., Marett, L.C., Cocks, B.G., Hayes, B.J., 2013. Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle. PLoS One 8, 1–8. https://doi.org/10.1371/journal.pone.0073056

Saborío-Montero, A., 2018. Variance components estimation of complex traits including microbiota information. Master thesis. Universitat Politècnica de València. https://doi.org/http://hdl.handle.net/10251/110370

Saborío-Montero, A., Gutiérrez-Rivas, M., García-Rodríguez, A., Atxaerandio, R., Goiri, I., López de Maturana, E., Jiménez-Montero, J.A., Alenda, R., González-Recio, O., 2020. Structural equation models to disentangle the biological relationship between microbiota and complex traits: Methane production in dairy cattle as a case of study. J. Anim. Breed. Genet. 137, 36–48. https://doi.org/10.1111/jbg.12444

Santos, A., van Aerle, R., Barrientos, L., Martinez-Urtaza, J., 2020. Computational methods for 16S metabarcoding studies using Nanopore sequencing data. Comput. Struct. Biotechnol. J. https://doi.org/10.1016/j.csbj.2020.01.005

Sorensen, D., 2004. An introductory overwiew of model comparison and related topics. Course notes on model choice, model assessment and related topics from a likelihood and a Bayesian perspective 1–47.

Sorensen, D., Gianola, D., 2002. Likelihood of Bayesian, and MCMC Methods in Quantitative Genetics., 1st ed, Crop Science. Springer-Verlag New York, Inc, New York. https://doi.org/10.2135/cropsci2003.1574

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B (Statistical Methodol. 64, 583–639. https://doi.org/10.1111/1467-9868.00353

Tapio, I., Snelling, T.J., Strozzi, F., Wallace, R.J., 2017. The ruminal microbiome associated with methane emissions from ruminant livestock. J. Anim. Sci. Biotechnol. 8, 1–11. https://doi.org/10.1186/s40104-017-0141-0

Vaidya, J.D., van den Bogert, B., Edwards, J.E., Boekhorst, J., van Gastelen, S., Saccenti, E., Plugge, C.M., Smidt, H., 2018. The Effect of DNA Extraction Methods on Observed Microbial Communities from Fibrous and Liquid Rumen Fractions of Dairy Cows. Front. Microbiol. 9, 92. https://doi.org/10.3389/fmicb.2018.00092

VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wallace, R.J., 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., Cappelli, F.P., Cabezas-Garcia, E.H., Vilkki, J., Pinares-Patino, C., Fliegerová, K.O., Mrázek, J., Sechovcová, H., Kopečný, J., Bonin, A., Boyer, F., Taberlet, P., Kokou, F., Halperin, E., Williams, J.L., Shingfield, K.J., Mizrahi, I., 2019. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci. Adv. 5, 1–12. https://doi.org/10.1126/sciadv.aav8391

Weishaar, R., Wellmann, R., Camarinha-Silva, A., Rodehutscord, M., Bennewitz, J., 2020. Selecting the hologenome to breed for an improved feed efficiency in pigs—A novel selection index. J. Anim. Breed. Genet. 137, 14–22. https://doi.org/10.1111/jbg.12447

Zhao, L., Wang, G., Siegel, P., He, C., Wang, H., Zhao, W., Zhai, Z., Tian, F., Zhao, J., Zhang, H., Sun, Z., Chen, W., Zhang, Y., Meng, H., 2013. Quantitative Genetic Background of the Host Influences Gut Microbiomes in Chickens. Sci. Rep. 3, 1163. https://doi.org/10.1038/srep01163

Zoetendal, E.G., Akkermans, A.D.L., Vliet, W.M.A., Visser, J.A.G.M. De, Vos, W.M. De, Zoetendal, E.G., Akkermans, A.D.L., Akkermans-van, W.M., Visser, J.A.G.M. De, Vos, W.M. De, Host, T., Affects, G., Zoetendal, E.G., Akkermans, A.D.L., Vliet, W.M.A., 2001. Microbial Ecology in Health and Disease The Host Genotype Affects the Bacterial Community in the Human Gastronintestinal Tract. Microb. Ecol. Health Dis. 13, 129–134.

CHAPTER 6

A DIMENSIONAL REDUCTION APPROACH TO MODULATE THE CORE RUMINAL MICROBIOME ASSOCIATED TO METHANE EMISSIONS VIA SELECTIVE BREEDING.

Alejandro Saborío-Montero[1,2], Adrián López-García[1], Mónica Gutiérrez-Rivas[1], Raquel Atxaerandio[3], Idoia Goiri[3], Aser García-Rodriguez[3], José A. Jiménez-Montero[4], Carmen González[1], Javier Tamames[5], Fernando Puente-Sanchez[5], Luis Varona[6], Magdalena Serrano[1], Cristina Ovilo[1], Oscar González-Recio[1,7]

[1]Departamento de mejora genética animal. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Crta. de la Coruña km 7.5, 28040 Madrid, Spain; [2]Escuela de Zootecnia y Centro de Investigación en Nutrición Animal, Universidad de Costa Rica, 11501, San José, Costa Rica; [3]Department of Animal Production, NEIKER—Tecnalia, Granja Modelo de Arkaute, Apdo. 46, 01080 Vitoria-Gasteiz, Spain; [4]Spanish Holstein Association (CONAFE). Ctra. de Andalucía km 23600 Valdemoro, 28340 Madrid, Spain; [5]Department of Systems Biology, Spanish Center for Biotechnology, CSIC, 28049, Madrid, Spain; [6]Facultad de Veterinaria, Universidad de Zaragoza, 50013, Zaragoza, Spain; [7]Departamento de Producción Agraria. Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas. Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain.

Published in the Journal of Dairy Science
DOI: https://doi.org/10.3168/jds.2020-20005

ABSTRACT

The rumen is a complex microbial system of substantial importance in terms of greenhouse gas emissions and feed efficiency. This study proposes combining metagenomic and host genomic data for selective breeding of the cow hologenome towards reduced methane emissions. We analyzed nanopore long reads from the rumen metagenome of 437 Holstein cows from 14 commercial herds in 4 northern regions in Spain. After filtering, data were treated as compositional. The large complexity of the rumen microbiota was aggregated, through principal component analysis (PCA), into few principal components (PCs) that were used as proxies of the core metagenome. The PCA allowed condensing the huge and fuzzy taxonomical and functional information from the metagenome into few PCs. Bivariate animal models were applied using these PCs and methane production as phenotypes. The variability condensed in these PCs is controlled by the cow genome, with heritability estimates for the first PC of ~0.30 at all taxonomic levels, with a large probability (>83%) of the posterior distribution being > 0.20 and with the 95% highest posterior density interval (95%HPD) not containing zero. Most genetic correlation estimates between PC1 and methane were large (≥0.70), with most of the posterior distribution (>82%) being >0.50 and with its 95%HPD not containing zero. Enteric methane production were positively associated with relative abundance (RA) of eukaryotes (protozoa and fungi) through the first component of the PCA at Phylum, Class, Order, Family and Genus. Nanopore long reads allowed the characterization of the core rumen metagenome using whole metagenome sequencing, and the purposed aggregated variables could be used in animal breeding programs to reduce methane emissions in future generations.

Key words: Genetic correlation, heritability, methane, microbiome, principal component analysis.

INTRODUCTION

The rumen contains a plethora of anaerobic microorganisms of all known phylogenetic domains (Pace, 1997); Bacteria, eukaryotes and archaea conform this complex microbiota ecosystem (Deusch et al., 2017). The archaea superkingdom includes methanogenic microorganisms, responsible for the methane emissions emitted by the ruminants (Knapp et al., 2014). Reducing methane production from domesticated ruminants poses a large potential to reduce greenhouse gas from the agriculture industry (Negussie et al., 2017; Clauss et al., 2020), which is the main greenhouse gas contributor from livestock (Negussie et al., 2017). Methane production is also considered as a loss of energy, as it is not absorbed by the animal to increase the desire output of the productive system (Johnson and Johnson, 1995). Reduction of enteric methane from livestock has become an important area of research (Negussie et al., 2017), due to its environmental and economic implications.

Several microorganisms have been associated to complex traits of interest in dairy cattle (Schären et al., 2018). Previous studies focused on single taxa relationships rather than accounting for the whole microbiome simultaneously. Moderate to high heritability estimates (0.20-0.60) have been reported for single taxa (e.g., genus) previously (Wallace et al., 2019). The effect of a single taxa over methane emissions has also been studied (Saborío-Montero et al., 2020). Complex traits are commonly affected by global changes in the microbial composition, and rarely by a single type of microorganism (Martínez-Álvaro et al., 2020; Malmuthuge and Guan, 2017).

Some external forces (e.g. diet, diseases and medical treatments, stress) can determine the growth of some specific microorganisms, which can at the same time condition the abundance of others they are related with, accelerating or slowing down their multiplication (Bach et al., 2019).

In recent years, efforts have been made to characterize the rumen microbiome and their functionality, with the aim of implementing nutrition and selective breeding strategies to modulate it. The ruminal microbiota composition is partially controlled by the host genotype, and both affect important traits in livestock that are related to efficiency and sustainability, including methane production (Roehe et al., 2016; Gonzalez-Recio et al., 2017). Animal breeding aims to modulate the ruminal microbiota through selection and to

achieve a more efficient microbial composition that reduces the use of natural resources and generates less methane emissions without impairing health and productivity (López-Paredes et al., 2020; González-Recio et al., 2020). Selective breeding against methane emissions may use the effect that the host genetics exerts on the microbial composition, which can explain up to 40% of the variability between individuals for the relative abundance (RA) of some bacterial genera associated to methane emissions (Wallace et al., 2019; Saborío-Montero et al., 2020). However, selecting by a single genus is ineffective for methane emissions mitigation because the effect of the microbiota on this complex trait is commonly due to a set of microbes, rather than to a single type of them. Also, increasing the relative abundance of a given microorganism decreases the relative abundance of other taxa when analyzing metagenomic data. Interactions in the microbial systems are cumbersome, and the greater or lesser abundance of one type of microorganism can affect the composition of the rest of the microbiota (Newbold et al., 2015).

Data structured as proportions or with a constant or irrelevant sum, are designated as compositional data, sequence reads from the metagenome are compositional (Gloor et al., 2017). Therefore, it is necessary to develop an analytic strategy that considers the broad composition of the microbiota to modulate it as a holobiont organism entity, instead of acting only on a few genera.

The objective of this study was to aggregate the rumen microbiota complexity into few principal components and to estimate their heritability and genetic correlation with methane traits.

MATERIALS AND METHODS

Ethical statement

This study was conducted in accordance with Spanish Royal Decree 53/2013 for the protection of animals used for experimental and other scientific purposes and was approved by the Basque Institute for Agricultural Research and Development Ethics Committee (Neiker-OEBA-2017-004) on March 28, 2017.

Data

A total of 437 Holstein cows from 14 commercial herds in 4 northern regions in Spain (Cantabria, País Vasco, Navarra and Gerona) were included in this study. Methane concentration measurements were obtained for all cows using a non-disperse infrared methane detector (The Guardian NG infrared gas monitor, Edinburgh Sensors, Scotland) as described by Rey et al. (2019). Briefly, methane concentration was measured individually from breath at each cow visit (3-7 daily visits) to the automated milking system, during 2-3 week periods. Eructation peaks recorded were averaged to obtain a single record per cow. Cows were classified in different groups according to lactation parameters (parity and stage of lactation).

Genotyping

Animals were genotyped using the EuroG 10K and EuroG LD 12k (Illumina, San Diego, California, USA). Low density genotypes were imputed to the Bovine 50k SNP chip (Illumina, San Diego, California, USA) containing 54,609 SNPs using BEAGLE software (Browning et al., 2018) and 3,669 animals from the Spanish Holstein reference population provided by the Spanish Friesian Associations Confederation (Jiménez-Montero et al., 2013). The SNPs with MAF<0.05 were filtered out from the analysis resulting in 42,372 SNPs left for the analyses.

Ruminal sampling technique

Samples of the rumen content were extracted from each animal. Cows were placed in individual stalls during the process, and a mechanical device was used to raise the snout of the animal. Approximately 100 ml of content was extracted from each cow by introducing orally a stomach tube (18 mm diameter and 160 mm long) through the esophagus connected to a mechanical pumping unit (Vacubrand ME 2SI, Wertheim, Germany) with a 1000 ml Erlenmeyer trapped in-between. Samples were then stored in a sterilized container. Hose and all material in contact with the samples were thoroughly washed between cows. Samples were filtered through 4 layers of sterile cheesecloth, in order to remove the solid fraction and the filtered fraction was frozen in liquid nitrogen vapors immediately after. Frozen samples

were transported to the laboratory in liquid nitrogen containers and stored at −80 °C until analysis.

DNA extraction and sequencing

The samples were thawed, and then homogenized in a blender. The DNA extraction was performed using 250 µl from the homogenized samples with the commercial "DNeasy Power Soil" kit (QIAGEN, Valencia, CA, USA). The genomic DNA concentrations and their purity were measured using the Qubit fluorometer (ThermoFisher Scientific, 150 Waltham, MA, USA) and a Nanodrop ND-1000 UV/Vis spectrophotometer (Nanodrop Technologies Inc., DE, USA) with ratios 260/280 and 260/230 around 1.8 and 2.0, respectively. One µg of DNA from each sample was used as initial material for sequencing, following the ligation sequencing kit (SQK-LSK109) protocol from Oxford Nanopore Technologies (ONT), in a MinION sequencer. Twelve samples were multiplexed in each run with the 1D Native barcoding genomic DNA (EXP-NBD104 or EXP-NBD114) ONT kit. The barcoded samples (700 ng of DNA in total) were pooled in a 1.5 ml Eppendorf DNA LoBind tube to perform adapter ligation for sequencing using a R9.4.1 flow cell.

Bioinformatics

Base calling was performed with the *guppy* 4.2.2 software provided by ONT. After quality control (QS>7 and length>150bp), the remaining sequences were analyzed using the *SqueezeMeta* 1.1.0 pipeline (Tamames and Puente-Sánchez, 2019). Briefly, it runs a *blastx* search of the reads against *GenBank nr*, COGs and KEGG databases to taxonomically and functionally annotate putative ORFs. Taxa are annotated using a last common ancestor algorithm for finding the consensus taxon for each read. KEGG annotation was performed with *SqueezeMeta* long reads protocol (*SQM_longreads.pl*), which uses the *fun3* algorithm to assign functions. Functions were annotated using the best hit above a minimum score threshold in the *LCA* (last common ancestor) algorithm (60, 55, 50, 46, 42 and 40% for genus, family, order, class, phylum and superkingdom ranks, respectively). Hits below these thresholds were considered as unclassified at the respective taxonomical level. For instance, a protein will not be assigned to genus if it has no hits above 60% identity. This imposes rigorous taxonomy classification, but accurate assignation. The *euk* option, was used to

improve eukaryotic annotation. Non-microbial sequences (animalia, plantae and virus clades) were removed. Sequences unmapped to the family level (unclassified) were also removed. The results from the *SqueezeMeta* pipeline (taxonomy and functionality) were pruned through a prevalence filter in order to reduce sparsity and manage sequencing errors. Genera were arranged based on their superkingdom taxonomical classification (Archaea, Bacteria or Eukaryota). KEGG results were organized according to their participation in methanogenesis: KEGGs within the orthology pathway ko00680 (Methane metabolism) were classified as "CH4-KEGGs" (n=85), and the rest were considered as "Other" (n=6,559).

Association between microbial functionality with methane production

Here, cows were grouped based on their level of CH_4 (ppm) measurements according to quartile-based ranks, taking 4 levels: LOW, L-MID, H-MID and HIGH. The ALL-KEGGs group, which consisted of total classified KEGGs (n=6,644) within this study, were included in the differential abundance of KEGGs between samples regarding the different methane emissions levels was addressed through linear regression using Limma (Ritchie et al., 2015). Differential abundance threshold was set to $| \log_2 FC | \geq 0.5$. The Benjamini-Hochberg procedure was used to control for the false discovery rate at an adjusted significance threshold of $\alpha < 0.05$.

An additional classification of KEEGs was then included based on the results of the differential abundance analysis. This is, those KEGGs that resulted to be differentially abundant were classified as "KEGGs-DA" (n=279). These KEGGs were subsequently classified as differentially abundant in the high methane emitting group "KEGGs-DA-High" or in the low methane emitting group "KEGGs-DA-Low".

Compositional data

In order to deal with the compositional nature of metagenomic data, a centered log ratio transformation (CLR) method (Aitchison, 1986) was implemented as follows using the unweighted option of the *CLR* function from the *easyCODA* R package (Greenacre, 2019):

$$\mathbf{x_{clr}} = [log\ (x_1/G(x)), log\ (x_2/G(x)) \ldots\ log\ (x_D/G(x))],$$

with $G(x) = \sqrt[D]{x_1 * x_2 * \ldots * x_D}$.

Here, $\mathbf{x} = [x_1, x_2, \ldots, x_D]$ is a vector of counted features (taxa or KEGGs) in a sample, and *G(x)* is the geometric mean of **x**. Count zero values in the initial data frame were imputed in order to allow computing logarithms. The imputation was done using a Bayesian Multiplicative Replacement procedure. This procedure was performed with the geometric Bayesian multiplicative method (GBM) from the *cmultRepl* function of the *zCompositions* R package (Palarea-Albaladejo et al., 2019).

Principal component analysis of taxa and KEGGs

Principal component analysis was used to reduce the dimensionality and aggregate the metagenome variance into few variables. This was performed for the taxonomy and functional data sets, independently. Data matrices **K**s with *m* rows (*m* =437, the number of individuals in the data set) and *n* columns (*n*= the number of taxa or KEGGs according to taxonomy or KEGG subset, respectively) were created. The PCAs were performed on the CLR transformed **K**s using the centered and scaled option of *prcomp* function from the *stats* R package (R Core Team, 2020). The number of PCs from the PCA kept for further analysis was decided based on:

1. The number of PCs with eigenvalues higher than value 1,
2. The variance explained by the PCs (Threshold ≥ 1%),
3. The interpretation of the PCs (subjacent variables with biological meaning).

Thus, the first five PCs were kept as descriptive synthetic variables that explain the variability either of the rumen microbial diversity or KEGGs subsets. Pearson correlations between the RA of variables (Taxa or KEGGs) and PCs were extracted with the *cor* option of the *get_pca_var* function from the *factoextra* R package(Kassambara and Mundt, 2020). Analyses of variance (ANOVA) were performed to find differences between mean correlations (centroids) for the groups of taxonomy (Archaea, Bacteria, Eukaryote) or functionality (KEGGs) subsets. All the ANOVA were performed with the *aov* function of the *stats* package in R (R Core Team, 2020). The value of $P < 0.05$ was set as the threshold to consider that there were differences between groups from the ANOVA. Then, a *TukeyHSD* function from the *stats* package in R (R Core Team, 2020) was used as a *post hoc* analysis to determine differences between groups.

Heritability and genetic correlation

Principal components resulted from the above analyses were used as phenotypic variables describing the microbiota composition (taxonomical or functional). The genetic effect of the host over the PCs and methane emissions was simultaneously estimated using bivariate animal models on methane and each of the PC considered. Covariances components between methane and each PC were estimated under a Bayesian framework. Inferences were obtained through an MCMC approach with a modified version of the TM package (Legarra et al., 2011) that uses the genomic relationship matrix (VanRaden, 2008; Yang et al., 2010) instead of the pedigree numerator relationship matrix. A set of bivariate animal models were used to analyze the relationship between each of the principal components extracted from the PCA and methane production from cows. All models were implemented under a Bayesian framework assuming a joint multivariate normal distribution for principal components (PC) and CH_4. The observed data were modeled as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_{(h)}h} + \mathbf{Z_{(u)}u} + \mathbf{e},$$

Where $\mathbf{y} = \{y_{i,PC_k}, y_{i,CH_4}\}$ was the observed PC_k and CH_4 concentration for the *i*th individual (cow); **b** is a vector including effects of population mean, parity (2 levels: first and second calving) and days in milk (3 levels: <70d, 71 to 150d, >151d), assuming prior distributions for **b** as uniform (-9999, 9999); $\mathbf{h} = \{h_{i,PC_k}, h_{i,CH_4}\}$ were the herd-batch effect (24 levels) assumed distributed as a multivariate normal distribution with null mean vector and (co)variance matrix $\mathbf{H_0} \otimes \mathbf{I}$, with

$$\mathbf{H_0} = \begin{bmatrix} \sigma^2_{h_{PC_k}} & \sigma^2_{h_{PC_k,CH_4}} \\ \sigma^2_{h_{CH_4,PC_k}} & \sigma^2_{h_{CH_4}} \end{bmatrix}$$

where $\sigma^2_{h_{PC_k}}$ was the herd-batch variance on PC_k, $\sigma^2_{h_{CH_4}}$ was the herd-batch variance on CH_4 and $\sigma^2_{h_{PC_k,CH_4}}$ was the herd-batch covariance between PC_k and CH_4. The **I** was an identity matrix of corresponding order. Then, $\mathbf{u} = \{u_{i,PC_k}, u_{i,CH_4}\}$ were the genetic effect

assumed to be distributed as a multivariate normal, with a null mean and a (co)variance matrix $\mathbf{K_0} \otimes \mathbf{G}$, with **G** being the genomic relationship matrix between individuals (VanRaden, 2008) and

$$\mathbf{K_0} = \begin{bmatrix} \sigma^2_{u_{PC_k}} & \sigma^2_{u_{PC_k,CH_4}} \\ \sigma^2_{u_{CH_4,PC_k}} & \sigma^2_{u_{CH_4}} \end{bmatrix}$$

where $\sigma^2_{u_{PC_k}}$ was the additive genetic variance for the PC_k of a given taxa, $\sigma^2_{u_{CH_4}}$ was the additive genetic variance for CH_4, and $\sigma^2_{u_{PC_k,CH_4}}$ was the additive genetic covariance between PC_k and CH_4. Finally, $\mathbf{e} = \{e_{i,PC_k}, e_{i,CH_4}\}$ were the residuals assumed distributed as a multivariate normal distribution with zero mean and (co)variance matrix $\mathbf{R_0} \otimes \mathbf{I}$, where

$$\mathbf{R_0} = \begin{bmatrix} \sigma^2_{e_{PC_k}} & \sigma^2_{e_{PC_k,CH_4}} \\ \sigma^2_{e_{CH_4,PC_k}} & \sigma^2_{e_{CH_4}} \end{bmatrix}$$

and **I** was the identity matrix of corresponding order. The **X** and **Z** were incidence matrices of appropriate order. The $\otimes$ symbol stands for the Kronecker product.

The heritability for each trait (PC_k or CH_4) was calculated as follows:

$$h^2 = \frac{\sigma^2_u}{\sigma^2_u + \sigma^2_{\mathrm{h}} + \sigma^2_e}$$

where, σ^2_{u} is the additive genetic variance of the analyzed trait, σ^2_{h} is the herd-batch variance, and σ^2_e is the residual variance.

The genetic correlations between PC_k and CH_4 were computed as:

$$Corr_{u_{PC_k,CH_4}} = \frac{\sigma_{u_{PC_k,CH_4}}}{\sqrt{\sigma^2_{u_{PC_k}} \times \sigma^2_{u_{CH_4}}}}$$

where $\sigma_{u_{PC_k,CH_4}}$ is the additive genetic covariance between PC_k and CH_4, $\sigma^2_{u_{PC_k}}$ is the additive genetic variance for the PC_k and finally, $\sigma^2_{u_{CH_4}}$ is the additive genetic variance for CH_4.

Posterior analysis of genetic parameters estimated through MCMC

Probability intervals of the posterior distribution of heritability and genetic correlations were estimated as the highest posterior density intervals from quantile 0.025 to 0.975 (95%HPD) of the distribution. After preliminary runs, visual examination of trace plots, and additional diagnostic assessments, the length of the chain was set to 300,000 iterations with a burn-in of 100,000 iterations; a thin period of 10 was taken in order to reduce the autocorrelation between samples. A total of 20,000 samples were kept to infer the posterior distributions of the unknown parameters. The sample size adjusted for autocorrelation and convergence diagnosis were performed using the *coda* R package (Plummer et al., 2019).

RESULTS

Microbial composition taxonomy

The microbial dataset after filtering included a total of 6,318,344 reads, with a mean number of reads of 14,458 per sample, classified in 967 known genera (722 bacteria, 13 archaea and 232 eukaryotes), and 273 that only reached family rank (*i.e. Unclassified* denomination). Overall, 503 families, 277 orders, 158 classes and 86 different phyla (37 bacterial phyla, 3 archaeal phyla and 46 eukaryotic clades) were classified. The genera taxonomy classification is available in the REDIA repository (http://rdm.inia.es/dataset/rumen-microbial-taxonomy-in-dairy-cattle).

Predominant microorganisms in this core rumen subcomposition were from Bacteroidetes, Firmicutes and Fibrobacteres phyla, representing an average RA of 63%, 16% and 5%, respectively. Bacteroidetes fraction was majorly composed by *Prevotella*, and was the main representative in the total community (19.4% average RA), along with other Prevotellaceae members. The Firmicutes group included a large number of genera. The order

of Clostridiales dominated in terms of RA, with Lachnospiraceae and Ruminococcaceae families being the most representative ones. The remaining phyla (34) from the Bacteria superkingdom represented 7.6% averaged RA of the core metagenome. Eukaryotes represented a total average RA of 8.2% of the filtered dataset. Predominant eukaryotic clades were those included in the SAR supergroup (Stramenopiles-Alveolata-Rhizaria), accounting for 6% of total average RA, followed by Fungi (1.3% of total average RA). Alveolata clade was the most abundant among the eukaryotes, with a high representation of unclassified Ophryoscolecidae, *Stentor* and *Paramecium*. Archaea representation in the core subcomposition (0.24% of total average RA) consisted of methanogenic organisms, mostly Methanomicrobia and Methanobacteria members, and Thermoplasmata, which belong to the methylotrophic-methanogenic acidophilic organisms. Yet, many reads could not be assigned to a known genus. The relative abundance per animal of the most relevant taxonomic groups is depicted in Figure 1.

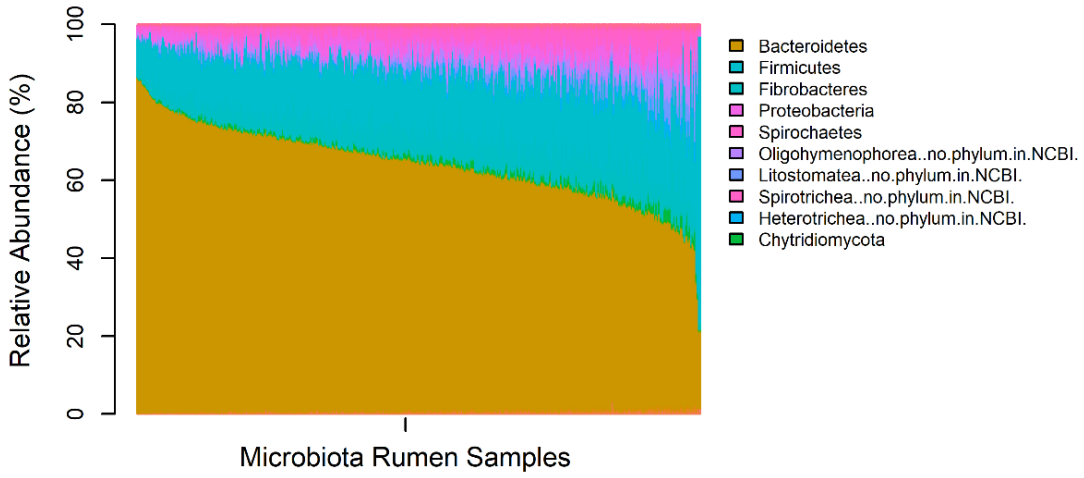

Figure 1. Relative abundance (%) of phyla colored by the top 10 most frequent Phylum and sorted by Bacteroidetes abundance according to microbiota rumen sample for 437 Holstein cows from 14 herds at 4 Northern Regions of Spain (País Vasco, Cantabria, Navarra, and Girona).

Functionality

The KEGG table was composed by 30,242,331 reads from 437 samples, classified in 6,644 KEGGs. Most of the rumen metagenome functions were in pathways that represent the metabolism of carbohydrate, amino acid and other biological compounds, as well as of energy metabolism proteins. Pathways related to pathogenic activity were also present, in agreement with the RA of several genera that included some known pathogenic species.

A total of 85 CH4-KEGGs were recovered from the KEGG orthology pathway ko00680 (Methane metabolism). After differential abundance analysis of ALL-KEGGs, a total of 279 KEGGs were classified as KEGGs-DA, with 182 KEGGs-DA-High and 97 KEGGs-DA-Low.

Dimensionality reduction of microbial composition

The rumen microbiome complexity was aggregated in few explanatory variables by applying principal component analysis (PCA) to the CLR data at the different taxonomical levels (phylum, class, order, family and genera) as well as for the KEGG classification (functional). We kept the information aggregated in the first five PC, based on eigenvalues >1, a percentage of variance explained from the original variables ≥ 1%, and biological meaning of the PCs (*i.e.* Clustering of known taxonomical groups). The proportion of the microbiome variance captured by these PC at the different levels was summarized in Table 1.

Table 1. Number of recovered taxonomic features from the rumen content of 437 Holstein cows and proportion of the variance explained by the first five principal components according to PCA by taxonomic classification.

	n	PC1	PC2	PC3	PC4	PC5
Phylum	86	0.310	0.027	0.027	0.018	0.018
Class	158	0.231	0.022	0.021	0.015	0.013
Order	277	0.172	0.023	0.018	0.016	0.012
Family	503	0.120	0.024	0.016	0.013	0.011
Genus	1240	0.076	0.021	0.014	0.012	0.011

The first five PC explained up to 40% of the phylum variance, but only 13% at genus level (Table 1). This is likely explained by the lower number of phyla (n=86), compared to other taxonomic levels, such as class, order, family or genus. Further, a lower accuracy at classifying more specific clades (*e.g.* genera) might create noisy variability, impairing the performance of PCA.

We found an evident clustering in PC1 by eukaryote proportion at all taxonomic levels (Figure 2). Animals located at the right-hand side of the PC1(positive coordinates) had higher proportion of eukaryotes, while lower proportion of eukaryotes were found in animals placed at the left-hand side of the PC1 (negative coordinates). Not evident clustering by herd, region or methane emissions were found from PCA biplots (results not shown).

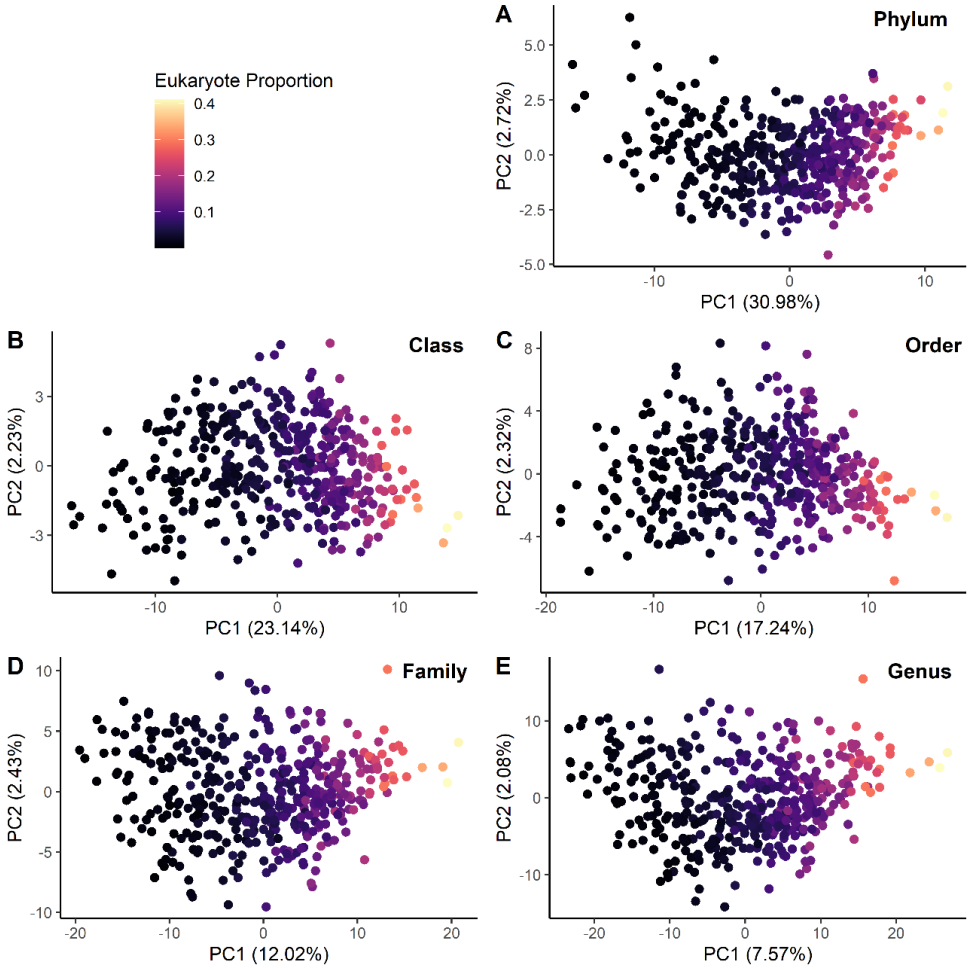

Figure 2. Principal component analysis depicting each animal coordinates for PC1 and PC2 at phylum (A), class (B), order (C), family (D) and genus (E) taxonomic level, colored according to the proportion of eukaryotes in the sample of rumen content for 437 Holstein cows.

The correlations between PC1 and the RA of microbial taxa clustered together within each superkingdom, eukaryotes with positive sign and bacteria with negative sign (Figure 3). The HSD Tukey test showed significant differences ($P < 0.05$) for these correlations between bacteria and eukaryote at all taxonomic levels and between archaea and eukaryotes at the genus level (Figure 3E). Here, some archaea showed null or negative correlations, whereas only two of them showed positive correlations. The latter were genus of *Methanobrevibacter spp.* (0.29) and *Methanosphaera spp.* (0.23).

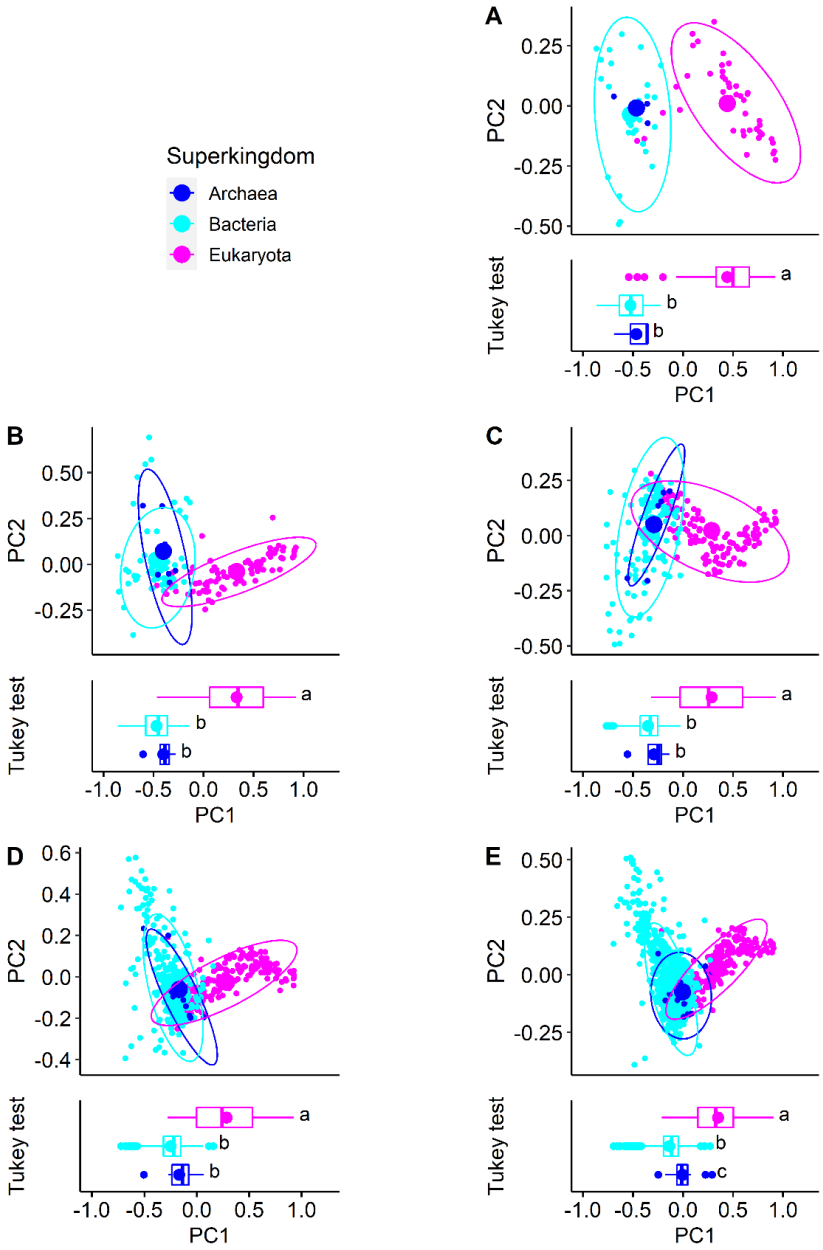

Figure 3. Correlations between relative abundance of taxa and the first two principal components (PC1 and PC2) from PCA according to taxonomic level for Phylum (A), Class (B), Order (C), Family (D) and Genus (E) clustered by superkingdom (Archaea•, Bacteria• and Eukaryota•). Centroids represent the mean correlation for each superkingdom and are represented by larger dots. Boxplot of correlations within a superkingdom are depicted below each biplot; points inside the boxes are the mean correlation in each superkingdom, points outside the boxes are outlier correlations. Different letters represent correlation differences between superkingdom ($P < 0.05$) according to HSD Tukey test.

The dimensionality of microbial functions was also reduced using the same strategy. In this case, three analyses were performed 1) with all 6,644 KEGGs, 2) with KEGGs specifically involved in the methanogenesis pathway (CH4-KEGGs, n=85) and 3) with KEGGs statistically associated to methane from the differential abundance analysis (KEGGs-DA, n=279). The first five PCs explained 31% of variance from ALL-KEGGs, 23% from CH4-KEGGs and 53% from KEGGs-DA (Table 2).

Table 2. Number of recovered KEGGs from the rumen content of 437 Holstein cows and proportion of the variance explained by the first five principal components according to PCA by KEGGs subset

	n	PC1	PC2	PC3	PC4	PC5
All-KEGGs	6,644	0.243	0.025	0.018	0.015	0.011
CH4-KEGGs	85	0.097	0.054	0.030	0.027	0.026
KEGGs-DA	279	0.479	0.015	0.013	0.010	0.009

Clustering was observed for cows with larger RA of KEGGs-DA (either low or high) (Figure 4). Cows with higher proportion of CH4-KEGGs also clustered on the positive coordinates of PC1, but not as clearly as for KEGGs-DA. Cows with higher proportion of KEGGs-DA-High were placed on negative values of PC1 (Figure 4B).

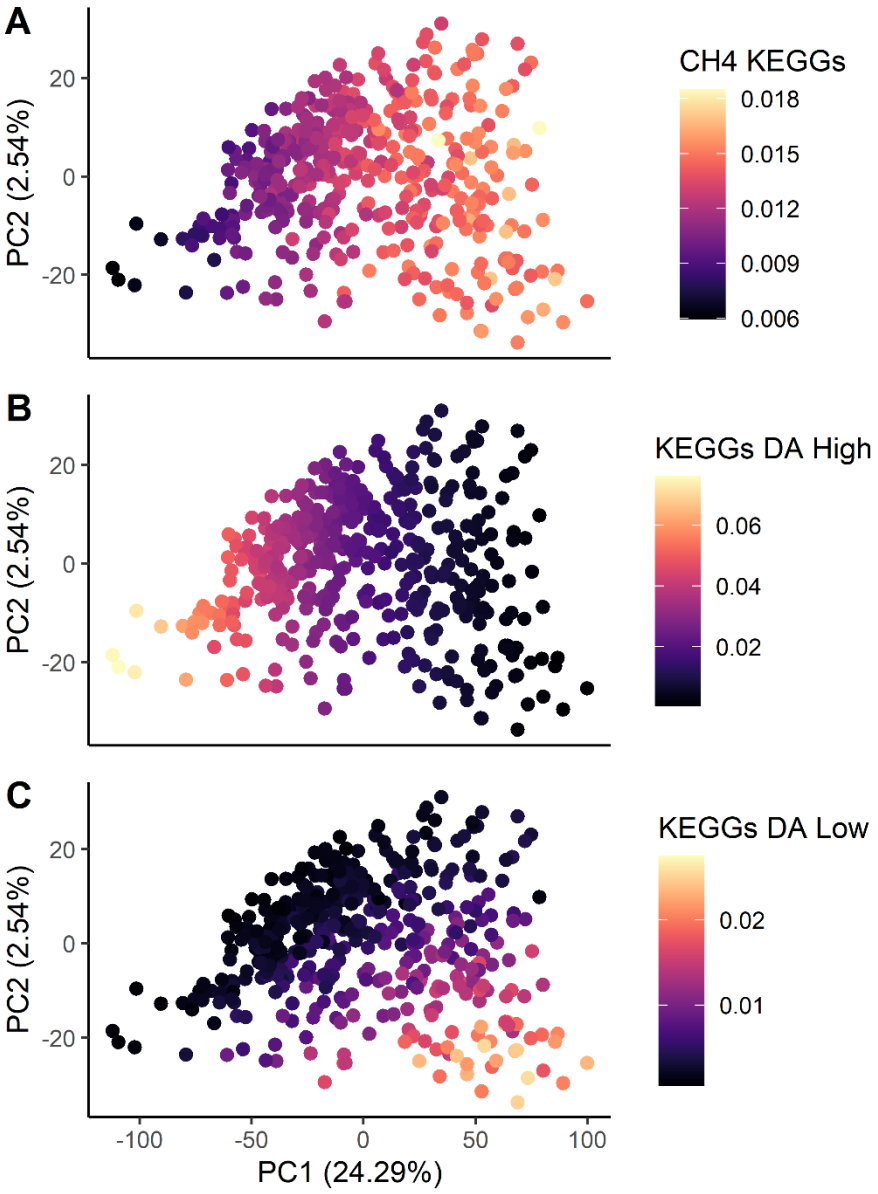

Figure 4. Principal component analysis depicting each animal coordinates for PC1 and PC2 from ALL-KEGGs PCA, colored according to the proportion in the sample of rumen content of methane specific KEGGs (A), KEGGs differentially abundant in the High (B) or Low (C) methane emitting animals for 437 Holstein cows.

Likewise, KEGGs-DA-Low (n=97) were positively correlated with PC1, and KEGGs-DA-High (n=182) negatively associated (Figure 5B), with mean correlations in KEGGs subsets statistically different (P< 0.05) between each other.

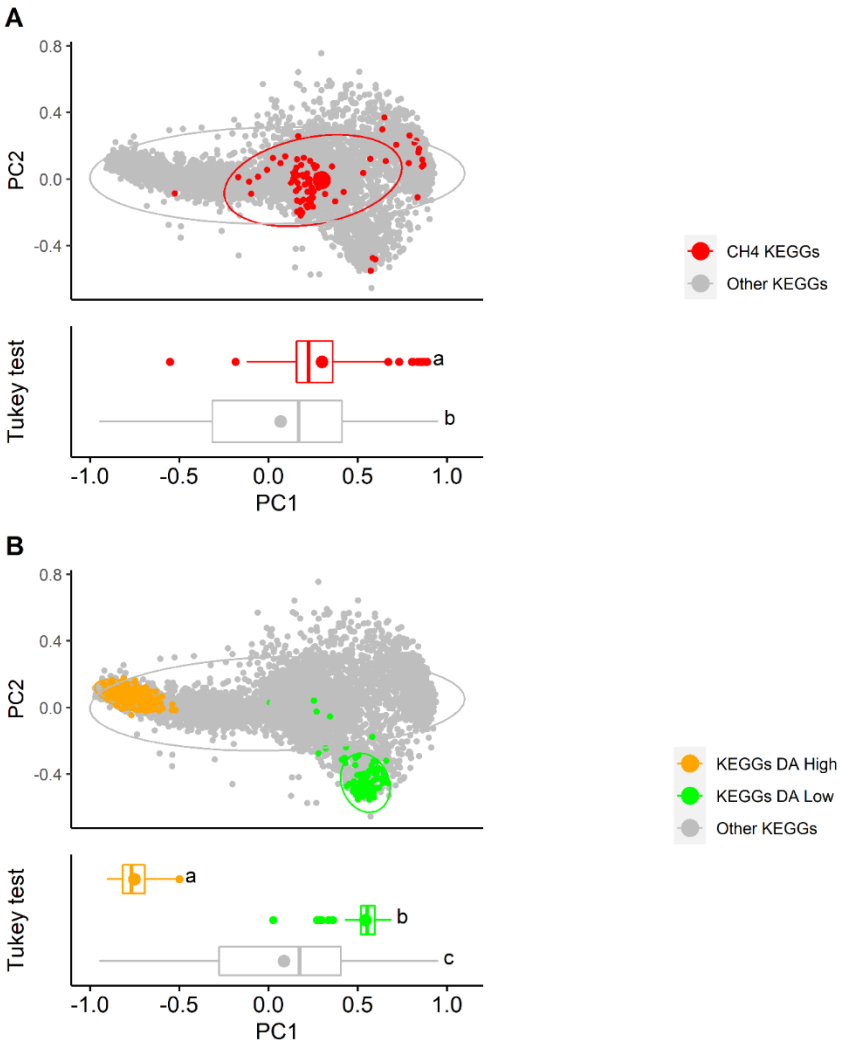

Figure 5. Correlations between relative abundance of KEGGs and the first two principal components (PC1 and PC2) from PCA from ALL-KEGGs for CH4-KEGGs• and other KEGGs• (A), and for differentially abundant KEGGs in the high (KEGGs-DA-High•) and low (KEGGs-DA-Low•) methane emitting cows (B). Boxplot of correlations within a subset according to PC1 are depicted below each biplot; points inside the boxes are the mean correlations in each KEGGs subset, points outside the boxes are outlier correlations. Different letters between KEGGs subset represent significant correlation differences ($P < 0.05$) according to HSD Tukey test.

The PCA using KEGGs-DA evidenced the clear clustering of low and high emitters (Figure 6) with extreme correlations differing ($P < 0.05$) between groups (Figure 7). The PC1 coordinates of each taxonomic level or KEGGs subset regarding every animal were kept for further analysis.

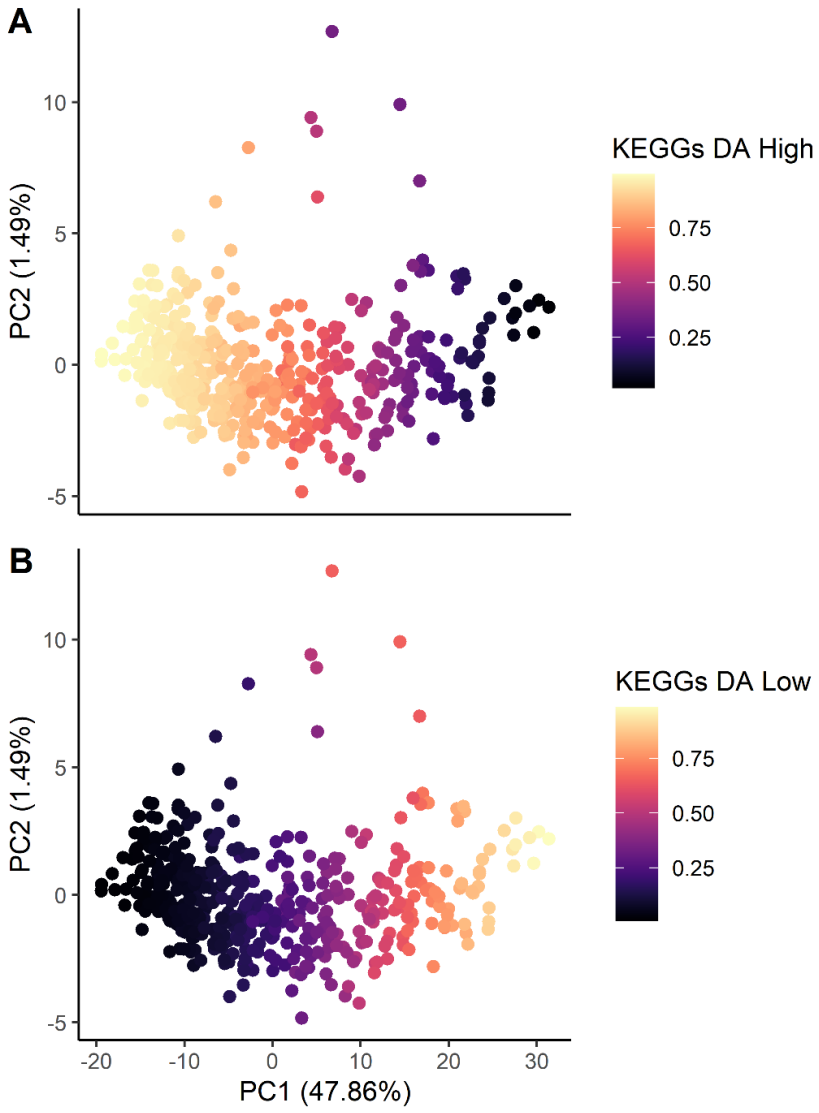

Figure 6. Principal component analysis depicting each animal coordinates for PC1 and PC2 from differentially abundant KEGGs PCA, colored according to the proportion in the sample of rumen content of KEGGs differentially abundant in the High (A) or Low (B) methane emitting animals for 437 Holstein cows.

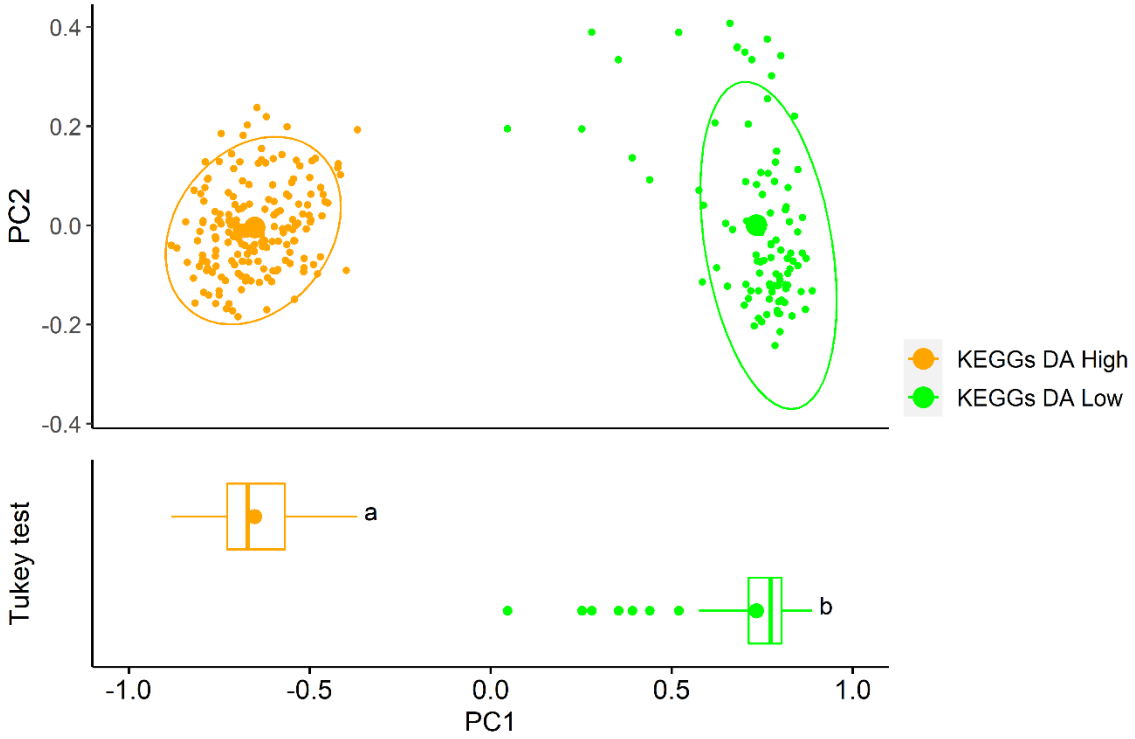

Figure 7. Correlations between relative abundance of KEGGs and the first two principal components (PC1 and PC2) from PCA for differentially abundant KEGGs in the high (KEGGs-DA-High•) and low (KEGGs-DA-Low•) methane emitting cows. Boxplot of correlations within a group according to PC1 are depicted below each biplot; points inside the boxes are the mean correlations in each KEGGs group, points outside the boxes are outlier correlations. Different letters between KEGGs groups represent significant correlation differences ($P < 0.05$) according to HSD Tukey test.

Host genomic control over the core microbiome

The posterior mean for the heritability of methane concentration was estimated at 0.16, with a 95% highest posterior density interval (95%HPD) of 0.02 to 0.35. The effective sample size of the Gibbs-sampling ranged between 162 and 940. The heritability estimates for the PCs were consistent across taxonomy level and ranged between 0.30 for the first PC to 0.11 for the 5th PC at phylum level (Figure 8).

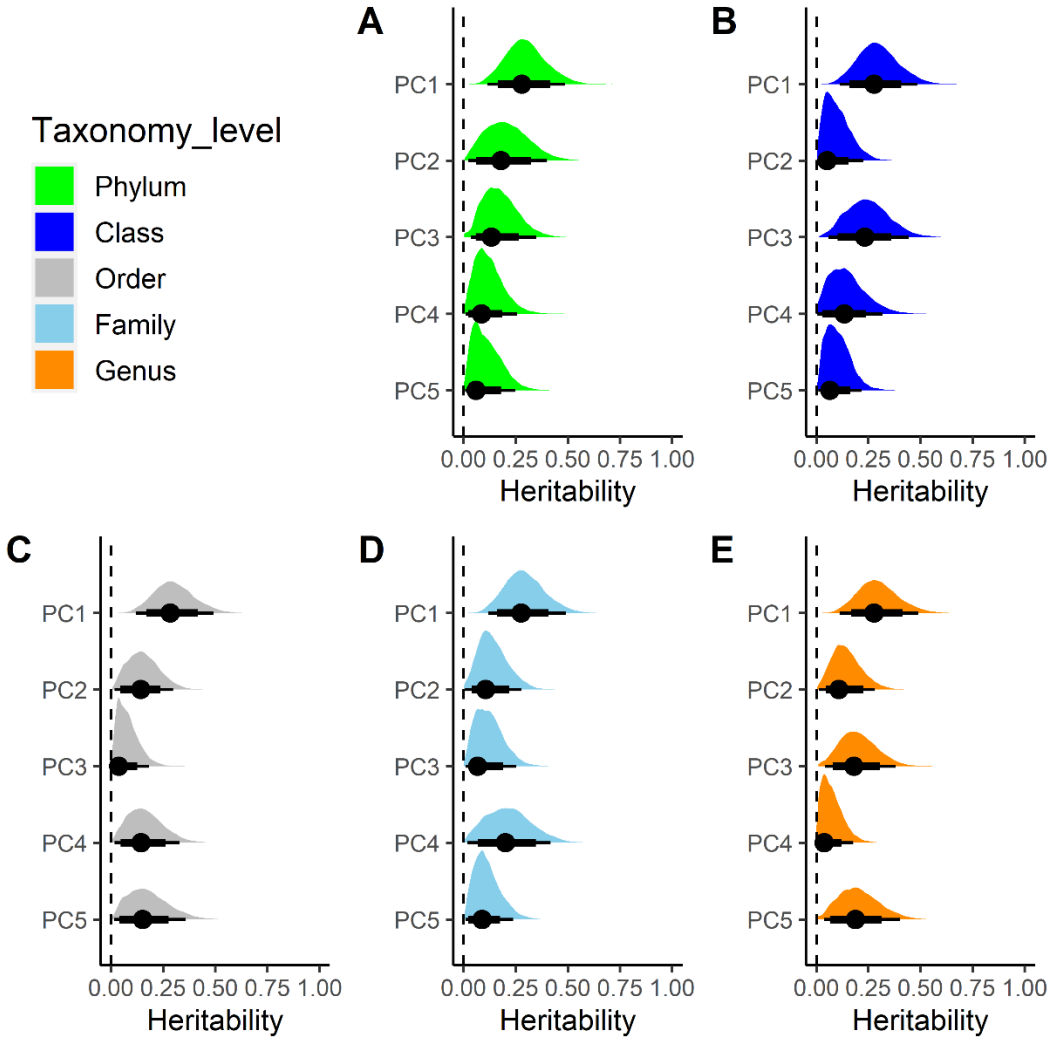

Figure 8. Posterior distributions of heritabilities for the first five principal components from bivariate animal models with methane emissions at different taxonomic levels (Phylum•, Class•, Order•, Family• and Genus•). Highest posterior density interval at 0.80 (thick) and 0.95 (thin) are displayed around the heritability mean at the base of the distributions.

Table 3 shows the heritability estimates for the first aggregated variable from PCA at different taxonomic levels and selection of KEGG function. In general, PC1 heritability had larger values at all taxonomic levels (means and medians close to 0.30, with at least 83% of the posterior distribution higher than 0.20) and were more consistent than those estimated for the remaining PCs (Figure 8). Similarly, heritabilities for the PC1s that aggregate functional information from KEGGs were estimated at 0.30, 0.11 and 0.34 for All-KEGGs, CH4-KEGGs and KEGGs-DA, respectively.

Table 3. Posterior mean heritability estimates and its highest posterior density intervals (95%HPD) for the metagenome first principal component, and its genetic correlation with methane emissions according to taxonomy level or KEGGs subset.

	h^2 PC1	Genetic correlation
Taxonomy level		
Phylum	0.298 (0.124 to 0.501)	0.737 (0.055 to 0.998)
Class	0.296 (0.122 to 0.498)	0.732 (0.039 to 0.998)
Order	0.302 (0.128 to 0.504)	0.760 (0.119 to 0.999)
Family	0.293 (0.123 to 0.498)	0.761 (0.112 to 0.999)
Genus	0.296 (0.124 to 0.505)	0.737 (0.059 to 0.998)
KEGGs subset		
All-KEGGs	0.295 (0.126 to 0.490)	-0.725(-0.999 to -0.061)
CH4-KEGGs	0.107(0.013 to 0.264)	0.460(-0.941 to 0.996)
KEGGs-DA	0.339(0.145 to 0.566)	-0.699(-0.998 to -0.020)

Strong genetic correlations were estimated between the microbiota aggregated variables (PC1) and methane emission both for taxonomy and functionality, with absolute values for taxonomy ≥0.70, with at least 82% of the posterior distribution higher than 0.5 (Table 3). The PCs aggregating KEGG functions showed negative genetic correlation with methane emissions. The sign of this genetic correlation depended on the side of the PC which KEGGs-DA-High fell on. Genes associated to larger methane fell on the negative sign of the PCs. Genetic correlations between PCs and methane emissions from PC2 to PC5 were close to zero with a large uncertainty on the estimates. An exception was the aggregated variable for KEGGs directly involved in the methanogenesis pathway. This set of KEGGs showed

positive but weaker genetic correlation with methane, and large 95%HPD overlapping zero. This suggests that standalone information from methanogenesis pathway is not sufficient to describe the complex processes in the rumen that lead to methane production and other pathways that fed with substrates to these genes must be considered. Residual correlations between PC1 and methane emissions were small at any taxonomic level and were not statistically different from zero. Similar results were obtained for KEGG PCs.

DISCUSSION

Cows are obligate dependent on their rumen microbiota to live. Indispensable functions such as rumen pH homeostasis, forage fiber fermentation or volatile fatty acids production are possible thanks to microorganisms inside their digestive tract. The methanogenesis regulate the rumen homeostasis during feed digestion, although it is also a loss of energy for the animal (Johnson and Johnson, 1995). Changes in ruminal microbiota composition can improve health, productive performance (Wallace et al., 2019), feed efficiency (Delgado et al., 2018) and methane emissions (Difford et al., 2018). However, rumen microbiome is a complex ecosystem of interrelated microorganisms, and there is not a straightforward strategy to modulate this microbial system (Wallace et al., 2019; Martínez-Álvaro et al., 2020; Saborío-Montero et al., 2020). This study evaluated the rumen microbiome composition using Nanopore long reads. Bacteroidetes, Firmicutes and Fibrobacteres were the most abundant phyla in the rumen metagenome in this study. Both Bacteroidetes and Firmicutes are common phyla in multiple environments, including animal digestive tracts. Bacteroidetes fraction is majorly composed by *Prevotella*, which covers a group of anaerobic gram-negative saccharolytic bacteria (Shah and Collins, 1990), whose large abundance in the digestive microbiota has been previously reported in both ruminant (Pitta et al., 2010; Li et al., 2020, Lee et al., 2012, Lopes et al., 2015) and monogastric species (Crespo-Piazuelo et al., 2018; Han et al., 2018). A wide representation of polysaccharide fermenters is represented in the rumen communities (Seshadri et al., 2018). Fibrobacteres comprises a small group of cellulose-degrading bacteria usually present in ruminant digestive system (Ransom-Jones et al., 2012). Our results were similar to previous findings using

different sequencing strategies (Wallace et al., 2019; Martínez-Álvaro et al., 2020). Previous studies also showed that nanopore sequencing is comparable to Illumina sequencing at the genus level using amplicons, while accurate taxonomic assignment at species level from Nanopore sequencing provides some benefits (Heikema et al., 2020). Longer reads improve classification at genus level, yielding better overall taxonomic classification due to the higher information content per read (Brandt et al., 2020). Longer reads also enable multiple genes presence within the same read, which might be particularly useful for functional assessments (Brandt et al., 2020). Nanopore sequencing offers an alternative strategy for metagenomics studies over amplicon-based approaches providing both taxonomical and functional information simultaneously, and for microbes from all superkingdom.

The results of this study highlight the complexity of the rumen microbiome, and the difficulty to disentangle their association with methane production. There remains a need to reduce the dimensionality of the problem for investigating applied genetic and nutritional solutions. It is important to point out that metagenomic data are compositional, and changes in the RA of some microbes will inevitably change the RA of some others. Modulating their relative abundance through nutrition or genetics is cumbersome, with many microbes and functions depending on each other. Acting on many variables simultaneously complicates the design of diets and breeding programs. We tackle these limitations by aggregating the metagenome information in few variables using PCA. The first PC separated Bacteria from Eukaryota composition. As shown in the previous sections, a larger abundance of Eukaryota were associated to larger methane emissions (Guyader et al., 2014), and some Bacteria, such as Proteobacteria indicated lower methane emission (Tapio et al., 2017; Granja-Salcedo et al., 2019; Saborío-Montero et al., 2020). Therefore, a positive correlation between methane emissions and PC1 at all taxonomic levels was expected, as most eukaryotes fell on positive PC1 values, whereas bacteria fell on the negative side. Those PCs explained a relevant amount of the rumen microbial composition (7 to 31%). The RA of Archaea genera most positively correlated with PC1 was *Methanobrevibacter spp.*, which is a well described hydrogenotrophic methanogen from the Methanobacteriaceae family; followed by *Methanosphaera spp.*, a methanogen from the same family that uses hydrogen to reduce methanol to methane (Bonin and Boone, 2006). In a previous study Martínez-Álvaro et al.

(2020), found that both *Methanobrevibacter spp.* and *Methanosphaera spp.* genera grouped together into the cluster of rumen microbiota that most variance explained of methane emissions in beef cattle. These results are consistent with the biological association found between methane and PC1. The KEGGs-DA clearly clustered the KEGGs-DA-High, from those in the KEGGs-DA-Low. Aggregated variables using KEGGs improved the interpretation of the genetic correlation with methane.

Previous results showed a causal relationship between certain genera and methane emissions (Saborío-Montero et al., 2020). We also evaluated the causality of PC variables on methane emissions, although they did not show a causal effect different from zero (results not shown). PC1 is a compendium of several taxa, which might generate noise or have opposite effects on methane, impairing the possibility of inferring a causal effect. However, and indirect response on methane production is expected if associated microorganisms are also modulated.

The host organism also exerts some control on the microbiome composition. For instance, eating behavior, rumen size and morphology, or its physiology may favor certain type of microorganisms. A novel strategy was proposed in this study: principal components from a PCA on the metagenome composition were used in bivariate analyses to estimate variance components and calculate heritability and genetic correlations with methane. The results showed that the core microbiome complexity can be captured by these heritable (~0.30) variables which are also genetically correlated with methane emissions at all taxonomic and functional levels. These results are supported by previous finding suggesting that the core rumen microbiome is heritable (Roehe et al., 2016; Gonzalez-Recio et al., 2017; Difford et al., 2018; Saborío-Montero et al., 2020). The principal components capturing partial information from the rumen microbiome composition could be incorporated into a breeding program. The positive genetic correlations between methane emissions and PC1 support the biological hypothesis that larger RA of eukaryotic microorganisms increase methane production. This association is explained due to the role of eukaryote as host for endo-symbiotic archaea (Hackstein and de Graaf, 2013) as well as a direct pathway of eukaryote genes linked to methane production (Liu et al., 2015). Highest posterior density for genetic correlation estimates were large, explained by the limited size of the data set, nonetheless the

95%HPD did not contain zero. These results highlight that the (co)variation between PCs and methane emissions is primarily mediated by the additive genetic relationship between those traits. The first principal component could be used in cattle breeding programs as representative of the core rumen metagenome to modulate the rumen metagenome towards larger efficiency and sustainability, although reproducibility across data sets needs to be confirmed. The PCs that aggregated taxonomic information at phylum level, and KEGGs-DA appeared to have an easier biological interpretation and explain a larger proportion of the microbiome variability. However, other taxonomical levels such as genus could also be considered. Selection in dairy cattle including the first principal component in the selection objective could reduce methane emissions in future generations.

CONCLUSIONS

Results in this study confirm the importance of Eukaryotes in the rumen microbiome and their role in the methanogenesis. Dimensionality reduction was implemented via principal component analysis, which allowed us to develop synthetic variables that aggregate the whole microbiome diversity (both at taxonomic and functional levels) in few PC aggregated variables. This approach simplifies the complexity and enable using these variables as phenotypes. The heritability estimates for these PC were relatively large and would allow selective breeding. Further, large genetic correlation with methane were estimated, which encourage us to pursue further studies using these aggregate variables as phenotypes in breeding programs. This strategy could modulate the rumen core metagenome and reduce methane production through correlated genetic response. A large enough reference population of cows with microbiome and genotype information had to be created for genomic selection implementation. However, it is necessary to evaluate possible collateral effects that might adversely affect the animal metabolism. Results in this study stimulate new opportunities for mitigating greenhouse-gas emissions from livestock, through direct modulation of the microbiota composition via animal breeding programs.

ACKNOWLEDGEMENTS

This research was financed by RTA2015-00022-C03-02 (METALGEN) project from the national plan of research, development and innovation 2013-2020 and the Department of Economic Development and Competitiveness (Madrid, Spain). We thank the regional Holstein Associations and farmers collaborating in the project. Computational support from the High-Performance Computing Centre in Galicia (Spain) is acknowledged. A.S.M. was granted a scholarship from Universidad de Costa Rica for course doctoral studies.

REFERENCES

Aitchison, J. 1986. The Statistical Analysis of Compositional Data. Chapman and Hall.

Bach, A., A. López-García, O. González-Recio, G. Elcoso, F. Fàbregas, F. Chaucheyras-Durand, and M. Castex. 2019. Changes in the rumen and colon microbiota and effects of live yeast dietary supplementation during the transition from the dry period to lactation of dairy cows. J. Dairy Sci. Sci 102:6180–6198. doi:10.3168/jds.2018-16105.

Bonin, A.S., and D.R. Boone. 2006. The Order Methanobacteriales. Springer.

Brandt, C., E. Bongcam-Rudloff, and B. Müller. 2020. Abundance Tracking by Long-Read Nanopore Sequencing of Complex Microbial Communities in Samples from 20 Different Biogas/Wastewater Plants. Appl. Sci. 10:7518. doi:10.3390/app10217518.

Browning, B.L., Y. Zhou, and S.R. Browning. 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am. J. Hum. Genet. 103:338–348. doi:10.1016/j.ajhg.2018.07.015.

Clauss, M., M.T. Dittmann, C. Vendl, K.B. Hagen, S. Frei, S. Ortmann, D.W.H. Müller, S. Hammer, A.J. Munn, A. Schwarm, and M. Kreuzer. 2020. Review: Comparative methane production in mammalian herbivores. Pages S113–S123 in Animal. Cambridge University Press.

Crespo-Piazuelo, D., J. Estellé, M. Revilla, L. Criado-Mesas, Y. Ramayo-Caldas, C. Óvilo, A.I. Fernández, M. Ballester, and J.M. Folch. 2018. Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. Sci. Rep. 8:1–12. doi:10.1038/s41598-018-30932-6.

Delgado, B., A. Bach, I. Guasch, C. González, G. Elcoso, J.E. Pryce, and O. Gonzalez-Recio. 2018. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. Sci. Rep. 9:1–13. doi:10.1038/s41598-018-36673-w.

Deusch, S., A. Camarinha-Silva, J. Conrad, U. Beifuss, M. Rodehutscord, and J. Seifert. 2017. A structural and functional elucidation of the rumen microbiome influenced by various diets and microenvironments. Front. Microbiol. 8:1–21. doi:10.3389/fmicb.2017.01605.

Difford, G.F., D.R. Plichta, P. Løvendahl, J. Lassen, S.J. Noel, O. Højberg, A.-D.G. Wright, Z. Zhu, L. Kristensen, H.B. Nielsen, B. Guldbrandtsen, and G. Sahana. 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLOS Genet. 14:e1007580. doi:10.1371/journal.pgen.1007580.

Gloor, G.B., J.M. Macklaim, V. Pawlowsky-Glahn, and J.J. Egozcue. 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 8:2224. doi:10.3389/fmicb.2017.02224.

González-Recio, O., J. López-Paredes, L. Ouatahar, N. Charfeddine, E. Ugarte, R. Alenda,

and J.A. Jiménez-Montero. 2020. Mitigation of greenhouse gases in dairy cattle via genetic selection: 2. Incorporating methane emissions into the breeding goal. J. Dairy Sci. 103:7210–7221. doi:10.3168/jds.2019-17598.

Gonzalez-Recio, O., I. Zubiria, A. García-Rodríguez, A. Hurtado, and R. Atxaerandio. 2017. Short communication: Signs of host genetic regulation in the microbiome composition in 2 dairy breeds: Holstein and Brown Swiss.. J. Dairy Sci. 101:1–8. doi:10.3168/jds.2017-13179.

Granja-Salcedo, Y.T., R.M. Fernandes, R.C. de Araujo, L.T. Kishi, T.T. Berchielli, F.D. de Resende, A. Berndt, and G.R. Siqueira. 2019. Long-Term Encapsulated Nitrate Supplementation Modulates Rumen Microbial Diversity and Rumen Fermentation to Reduce Methane Emission in Grazing Steers. Front. Microbiol. 10:1–12. doi:10.3389/fmicb.2019.00614.

Greenacre, M. 2019. Package "easyCODA". Compositional Data Analysis in Practice 35.

Guyader, J., M. Eugène, P. Nozière, D.P. Morgavi, M. Doreau, and C. Martin. 2014. Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. Animal 8:1816–1825. doi:10.1017/S1751731114001852.

Hackstein, J.H.P., and R.M. de Graaf. 2013. Anaerobic Ciliates and Their Methanogenic Endosymbionts. J.H.P. Hackstein, ed. Springer Berlin, Berlin.

Han, G.G., J.Y. Lee, G.D. Jin, J. Park, Y.H. Choi, S.K. Kang, B.J. Chae, E.B. Kim, and Y.J. Choi. 2018. Tracing of the fecal microbiota of commercial pigs at five growth stages from birth to shipment. Sci. Rep. 8:1–9. doi:10.1038/s41598-018-24508-7.

Heikema, A.P., D. Horst-Kreft, S.A. Boers, R. Jansen, S.D. Hiltemann, W. de Koning, R. Kraaij, M.A.J. de Ridder, C.B. van Houten, L.J. Bont, A.P. Stubbs, and J.P. Hays. 2020. Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota. Genes (Basel). 11:1105. doi:10.3390/genes11091105.

Jiménez-Montero, J., D. Gianola, K. Weigel, R. Alenda, and O. González-Recio. 2013. Assets of imputation to ultra-high density for productive and functional traits. J. Dairy Sci. 96:6047–6058. doi:10.3168/jds.2013-6793.

Johnson, K.A., and D.E. Johnson. 1995. Methane emissions from cattle. J. Anim. Sci. 73:2483–2492.

Kassambara, A., and F. Mundt. 2020. Package "factoextra", Extract and Visualize the Results of Multivariate Data Analyses.

Knapp, J.R., G.L. Laur, P.A. Vadas, W.P. Weiss, and J.M. Tricarico. 2014. Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. J. Dairy Sci. 97:3231–3261. doi:10.3168/JDS.2013-7234.

Lee, H.J., J.Y. Jung, Y.K. Oh, S.S. Lee, E.L. Madsen, and C.O. Jeon. 2012. Comparative survey of rumen microbial communities and metabolites across one caprine and three bovine groups, using bar-coded pyrosequencing and 1H nuclear magnetic resonance spectroscopy. Appl. Environ. Microbiol. 78:5983–5993. doi:10.1128/AEM.00104-12.

Legarra, A., L. Varona, and E. Lopez de Maturana. 2011. Threshold Model 1–33.

Li, J., H. Zhong, Y. Ramayo-Caldas, N. Terrapon, V. Lombard, G. Potocki-Veronese, J. Estelí, M. Popova, Z. Yang, H. Zhang, F. Li, S. Tang, F. Yang, W. Chen, B. Chen, J. Li, J. Guo, C. Martin, E. Maguin, X. Xu, H. Yang, J. Wang, L. Madsen, K. Kristiansen, B. Henrissat, S.D. Ehrlich, and D.P. Morgavi. 2020. A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading environment. Gigascience 9:1–15. doi:10.1093/gigascience/giaa057.

Liu, J., H. Chen, Q. Zhu, Y. Shen, X. Wang, M. Wang, and C. Peng. 2015. A novel pathway of direct methane production and emission by eukaryotes including plants, animals and fungi: An overview. Atmos. Environ. 115,:26–35. doi:10.1016/j.atmosenv.2015.05.019.

Lopes, L.D., A.O. de Souza Lima, R.G. Taketani, P. Darias, L.R.F. da Silva, E.M. Romagnoli, H. Louvandini, A.L. Abdalla, and R. Mendes. 2015. Exploring the sheep rumen microbiome for carbohydrate-active enzymes. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 108:15–30. doi:10.1007/s10482-015-0459-6.

López-Paredes, J., I. Goiri, R. Atxaerandio, A. García-Rodríguez, E. Ugarte, J.A. Jiménez-Montero, R. Alenda, and O. González-Recio. 2020. Mitigation of greenhouse gases in dairy cattle via genetic selection: 1. Genetic parameters of direct methane using noninvasive methods and proxies of methane. J. Dairy Sci. 103:7199–7209. doi:10.3168/jds.2019-17597.

Malmuthuge, N., and L.L. Guan. 2017. Understanding host-microbial interactions in rumen: searching the best opportunity for microbiota manipulation.. J. Anim. Sci. Biotechnol. 8:8. doi:10.1186/s40104-016-0135-3.

Martínez-Álvaro, M., M.D. Auffret, R.D. Stewart, R.J. Dewhurst, C.-A. Duthie, J.A. Rooke, R.J. Wallace, B. Shih, T.C. Freeman, M. Watson, and R. Roehe. 2020. Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. Front. Microbiol. 11:659. doi:10.3389/fmicb.2020.00659.

Negussie, E., Y. de Haas, F. Dehareng, R.J. Dewhurst, J. Dijkstra, N. Gengler, D.P. Morgavi, H. Soyeurt, S. van Gastelen, T. Yan, and F. Biscarini. 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.. J. Dairy Sci. 100:2433–2453. doi:10.3168/jds.2016-12030.

Newbold, C.J., G. De La Fuente, A. Belanche, E. Ramos-Morales, and N.R. McEwan. 2015. The Role of Ciliate Protozoa in the Rumen. Front. Microbiol. 6:1–14. doi:10.3389/fmicb.2015.01313.

Pace, N.R. 1997. A Molecular View of Microbial Diversity and the Biosphere. Science (80-.). 276:734–740.

Palarea-Albaladejo, J., J. Antoni, and M.-F. Maintainer. 2019. Package “zCompositions” Treatment of zeros, left-censored and missing values in compositional data sets 36.

doi:10.1016/j.chemolab.2015.02.019.

Pitta, D.W., W.E. Pinchak, S.E. Dowd, J. Osterstock, V. Gontcharova, E. Youn, K. Dorton, I. Yoon, B.R. Min, J.D. Fulford, T.A. Wickersham, and D.P. Malinowski. 2010. Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets. Microb. Ecol. 59:511–522. doi:10.1007/s00248-009-9609-6.

Plummer, M., N. Best, K. Cowles, K. Vines, D. Sarkar, D. Bates, A. Russell, and A. Magnusson. 2019. Package "coda." CRAN Repository.

R Core Team. 2020. Stats Package | R: A Lenguage and Environment for Statistical Computing. R Fundation for Statistical Computing, Vienna, Austria. Accessed August 3, 2020. https://www.rdocumentation.org/packages/stats/versions/3.6.2.

Ransom-Jones, E., D.L. Jones, A.J. McCarthy, and J.E. McDonald. 2012. The Fibrobacteres: An Important Phylum of Cellulose-Degrading Bacteria. Microb. Ecol. 63:267–281. doi:10.1007/s00248-011-9998-1.

Rey, J., R. Atxaerandio, R. Ruiz, E. Ugarte, O. González-Recio, A. Garcia-Rodriguez, and I. Goiri. 2019. Comparison Between Non-Invasive Methane Measurement Techniques in Cattle. Animals 9:1–9. doi:10.3390/ani9080563.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43. doi:10.1093/nar/gkv007.

Roehe, R., R.J. Dewhurst, C.A. Duthie, J.A. Rooke, N. McKain, D.W. Ross, J.J. Hyslop, A. Waterhouse, T.C. Freeman, M. Watson, and R.J. Wallace. 2016. Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. PLoS Genet. 12:1–20. doi:10.1371/journal.pgen.1005846.

Saborío-Montero, A., M. Gutiérrez-Rivas, A. García-Rodríguez, R. Atxaerandio, I. Goiri, E. López de Maturana, J.A. Jiménez-Montero, R. Alenda, and O. González-Recio. 2020. Structural equation models to disentangle the biological relationship between microbiota and complex traits: Methane production in dairy cattle as a case of study. J. Anim. Breed. Genet. 137. doi:10.1111/jbg.12444.

Schären, M., J. Frahm, S. Kersten, U. Meyer, J. Hummel, G. Breves, and S. Dänicke. 2018. Interrelations between the rumen microbiota and production, behavioral, rumen fermentation, metabolic, and immunological attributes of dairy cows. J. Dairy Sci. 101:4615–4637. doi:10.3168/jds.2017-13736.

Seshadri, R., S.C. Leahy, G.T. Attwood, K.H. Teh, S.C. Lambie, A.L. Cookson, E.A. Eloe-Fadrosh, G.A. Pavlopoulos, M. Hadjithomas, N.J. Varghese, D. Paez-Espino, R. Perry, G. Henderson, C.J. Creevey, N. Terrapon, P. Lapebie, E. Drula, V. Lombard, E. Rubin, N.C. Kyrpides, B. Henrissat, T. Woyke, N.N. Ivanova, W.J. Kelly, N. Palevic, P.H. Janssen, R.S. Ronimus, S. Noel, P. Soni, K. Reilly, T. Atherly, C. Ziemer, A.D. Wright, S. Ishaq, M. Cotta, S. Thompson, K. Crosley, N. McKain, J.J. Wallace, H.J.

Flint, J.C. Martin, R.J. Forster, R.J. Gruninger, T. McAllister, R. Gilbert, D.J. Ouwerkerk, A.J. Klieve, R. Al Jassim, S. Denman, C. McSweeney, C. Rosewarne, S. Koike, Y. Kobayashi, M. Mitsumori, T. Shinkai, S. Cravero, and M. Cerón Cucchi. 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat. Biotechnol. 36:359–367. doi:10.1038/nbt.4110.

Shah, H.N., and D.M. Collins. 1990. Prevotella, a new genus to include Bacteroides melaninogenicus and related species formerly classified in the genus Bacteroides. Int. J. Syst. Bacteriol. 40:205–208. doi:10.1099/00207713-40-2-205.

Tamames, J., and F. Puente-Sánchez. 2019. SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front. Microbiol. 9:1–10. doi:10.3389/fmicb.2018.03349.

Tapio, I., T.J. Snelling, F. Strozzi, and R.J. Wallace. 2017. The ruminal microbiome associated with methane emissions from ruminant livestock. J. Anim. Sci. Biotechnol. 8:1–11. doi:10.1186/s40104-017-0141-0.

VanRaden, P.M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

Wallace, R.J., G. Sasson, P.C. Garnsworthy, I. Tapio, E. Gregson, P. Bani, P. Huhtanen, A.R. Bayat, F. Strozzi, F. Biscarini, T.J. Snelling, N. Saunders, S.L. Potterton, J. Craigon, A. Minuti, E. Trevisi, M.L. Callegari, F.P. Cappelli, E.H. Cabezas-Garcia, J. Vilkki, C. Pinares-Patino, K.O. Fliegerová, J. Mrázek, H. Sechovcová, J. Kopečný, A. Bonin, F. Boyer, P. Taberlet, F. Kokou, E. Halperin, J.L. Williams, K.J. Shingfield, and I. Mizrahi. 2019. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci. Adv. 5:1–12. doi:10.1126/sciadv.aav8391.

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569. doi:10.1038/ng.608.

CHAPTER 7

GENERAL DISCUSSION

Alejandro Saborío-Montero[1,2]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain; [2]Universidad de Costa Rica, Escuela de Zootecnia, 11501 San Pedro de Montes de Oca, San José, Costa Rica.

Recapitulation

This thesis analyzed the host genetic control over the ruminal microbiota composition and their relationships with methane emissions in dairy cattle. For that, different methodological approaches were performed. A risk factor analysis for methane emissions was carried out aiming to determine risk factors and its impact over methane emissions. Productive and conformation traits, methane emissions GEBVs, and microbiota composition were evaluated through a probit threshold model. This analysis allowed to disentangle the modulation of methane emissions through the microbiota composition and to compare the host genetic effect with other phenotypic risk factors. Variance component estimations were performed including microbiota and methane emissions through different innovative approaches (single microbiota taxa within structural equation models (SEM); whole microbiota through microbiota relationship matrices (K); and aggregated microbiota via principal component analysis (PCA)). All variance component approaches targeted to estimate the partial genetic control of the microbiota composition and methane emissions simultaneously. The SEM approach also aimed to infer causality from microbiota over methane emissions from the hypothesized model. Several ordination methods to build the Ks were also analyzed to compare the accuracy performance of the different K under the variance component estimation. A new term, "Holobiability", was proposed to define the proportion of the phenotypic variance explained by the holobiont effects. From the PCA approach, a proposal for combining the core rumen microbiota into few aggregated variables (PCs), was postulated as an innovative way to include microbiota into animal breeding programs. Information of genotypes, phenotypes, and rumen microbiota composition from 437 Holstein cows belonging to herds of 14 commercial farms at four regions of the Northern Spain (País Vasco, Cantabria, Navarra, and Gerona) was used along this thesis.

Dealing with compositionality of microbiome

Compositional data are parts of some whole which only carry relative information (Lovell et al., 2015). Compositionality in datasets can lead to spurious correlations, distorting the results of the study. The centered log-ratio transformation (CLR) has been suggested as an appropriate way to deal with compositionality of datasets (Gloor et al., 2017). In this thesis, CLR was performed to the microbiome dataset in the Risk Factor, K and PCA

analyses. Despite of this difference in the microbiota data analysed (transformed and not-transformed microbiota), all the approaches yielded similar conclusions of high prevalence of eukaryote being preponderant in methane emissions rising. Similar results of CLR over microbiota having a small effect over main conclusions of the study had been reported before by Martínez-Álvaro et al. (2020), who obtained similar correlation structure between variables, and similar composition of larger clusters between transformed and not transformed data from their network analysis. These observed similarities do not change the compositionality nature of microbiota, which should be controlled through procedures analogous to CLR, to avoid artefactual or spurious correlations between variables.

Association of microbiota with methane emissions

Until now, evidence of methane emissions in dairy cows is strictly dependent on methane production by methanogenic archaea (Hook et al., 2010), though, a pathway linking direct production of methane to eukaryotes in the presence of oxygen (not the case in rumen) has been proposed (Liu et al., 2015). The archaea-methane association and the presence of methanogenic archaea in every ruminant, in addition to their relationships with other microorganisms, generates complex microbiota ecological networks. There was a positive association between the relative abundance of eukaryote and methane emissions, regardless of the statistical approach considered. This association has been reported previously (Guyader et al., 2014). The biological phenomenon behind this associations might rely on the endosymbiotic relationships between archaeas and eukaryotes (Hackstein and de Graaf, 2013). Some protozoa (belonging to the eukaryotes superkingdom) provide a high H_2 availability to archaeas (Belanche et al., 2014), through mitochondria like structures rich in H_2 called hydrogenosomes, responsible for H_2 supplying to surrounded archaea (Hackstein and de Graaf, 2013), and with protozoal size being a key factor influencing the number of methanogens per protozoal cell (Belanche et al., 2014).

The association between the RA of archaea and methane emissions was not as evident as that from eukaryotes; However, there was a positive association between some well-known methanogenic archaeas and methane emissions. For instance, the *Methanobrevibacter spp.* genus was positively associated to methane emissions in the SEM analysis, with a positive rate of change in methane emissions per unit change in the relative abundance of this

genus. The *Methanobrevibacter spp.* genus also showed the highest probability, of all archaea genera, of classifying cows into the upper quartile for methane concentration and methane intensity, per unit change in its relative abundance in the risk factors analysis. The *Methanobrevibacter spp.* genus has been previously associated to high methane emissions in ruminants. Wallace et al. (2015) reported abundances of 2.5 times larger in high compared to low methane emitting animals. Another study by Zhou et al. (2009) analyzed feed efficiency, showing prevalence of *Methanobrevibacter spp.* 2.26 times higher in inefficient animals and concluded that methanogenic ecology in the rumen may play relevant roles in the differences for methane gas production between cattle with different feed efficiencies. That study also found 1.92 times higher prevalence of *Methanosphaera spp*. in inefficient animals. Concordantly, the *Methanobrevibacter spp*. and *Methanosphaera spp*. were the two archaeal genera that showed the highest positive correlations (0.29 and 0.23, respectively) with PC1 in the PCA analysis of this thesis, with PC1 being genetically correlated (0.737, 95%HPD = 0.059 to 0.998) with methane concentration. Those two genus had been grouped together, in a previous study (Martínez-Álvaro et al., 2020), into the cluster of rumen microbiota that most variance explained of methane emissions in beef cattle. The archaea relationship with methane emissions were biologically more plausible when the associations were analysed at a more specific taxonomic level (*e.g.*, genus). Hence, point out that to using more specific taxonomic levels may better explain this association.

Most bacteria showed negative associations with methane emissions along this thesis. For instance, *Prevotella spp*., *Succinimonas spp*. and *Succiniclasticum spp*. genera showed negative genetic correlations with methane concentration in the SEM study. Those genera are from the Bacteroidetes, Proteobacteria and Firmicutes phyla, respectively. Concordantly with the SEM study, an increment of the relative abundance of three mentioned phyla, would have the lowest probability of classifying cows into the upper quartile of methane concentration, in the risk factors study. Larger relative abundance of Bacteroidetes and *Prevotella spp*. had been previously related to lower abundance of *Methanobrevibacter spp*. and better feed efficiency in dairy cattle (Delgado et al., 2018). The association between better feed efficiency and lower methane emissions lies on the basis of energy loss through methane emissions (Tapio et al., 2017; Mizrahi and Jami, 2018). A possible explanation for the negative association between bacteria and methane emissions is that bacteria relative

abundance is inversely dependent of the relative abundance of eukaryotes and archaea, given the compositional nature of microbiome (Gloor et al., 2017), thus relating larger proportion of bacteria to smaller proportion of eukaryote and archaea. Perhaps, rather than the mentioned mathematical issue, it could be related to the ecological associations between protozoa and bacteria, in which protozoa predate on bacteria (Newbold et al., 2015), thereby, resulting in larger bacterial presence as protozoa prevalence diminish, with subsequent reduction on methane emissions. Other plausible explanation could be a direct effect from bacteria competing with methanogens for H_2 uptake, generating a methane emissions reduction. This latter hypothesis has been studied for bacteria genera such as *Selenomonas spp*. using H_2 for fumarate and nitrite reduction, or *Blautia spp*. using H_2 for acetogenesis (Greening et al., 2019).

The genus was the most specific taxonomic level assigned for microbiota within this thesis, which may pose a limitation of the study. More specific taxonomic levels informed better about biological relationships, as showed by the association between methane emissions and archaea. Lower taxonomic levels than genus (*i.e.* species) might have informed of associations not evidenced at the available classifications. Despite of this limitation, longer reads obtained through nanopore sequencing improves taxonomical classification at genus level (Brandt et al., 2020), compared to amplicon sequencing; which has previously showed misclassification at genus or lower taxonomic levels (Poretsky et al., 2014). Longer reads from nanopore sequencing allows better overall classification, due to the higher information content per read; Therefore, more reliable taxonomy assignation than amplicon sequencing (Brandt et al., 2020). Other advantage of nanopore over amplicon sequencing is its capacity of taxonomic assignation for microbiome from all superkingdom. However, reading species level is challenging even with long reads, as reference databases are yet incomplete.

Host genetic control over the methane emissions

The probit threshold models (Gianola, 1982), showed that the holobiont effect exerts some influence on methane emissions. The probability of a cow of being classified in the upper quartile for methane concentration (ppm CH_4) and methane intensity (ppm CH_4 /kg milk) was mainly affected by conformation traits, as well as by the microbiota. Results from the risk factor analysis showed that larger scores for conformation traits related to structure

and capacity (body depth, chest width, BCS, loin fortress, overall structure and capacity value, body weight and stature) resulted in increased probability of classifying cows into the upper quartile for methane concentration. Conformation traits are heritable (Rupp and Boichard, 1999; Miglior et al., 2017) and some of them are genetically correlated with methane emissions (López-Paredes et al., 2020). In addition to that, the rumen size has been previously related to enteric methane emissions (Goopy et al., 2014; Kamke et al., 2016), with larger rumen size having greater mean retention time of digesta, resulting in larger methane emissions (Barnett et al., 2012). Hence, we can assume that larger animals have bigger rumen, and therefore larger mean retention time of digesta, associated to higher methane emissions. From this assumption, it is straightforward to presume that rumen size is a heritable trait transmitted from one generation to another and, likely, genetically correlated with methane emissions. The mean heritability estimate for methane concentration in the SEM approach was 0.12 (ranging from 0.09 to 0.18 depending on the bivariate model). When the microbiota was included as whole through K, it ranged from 0.15 to 0.17 depending on the distance metric used. Similar values were obtained in the PCA approximation: 0.16 (95%HPD = 0.02 to 0.35). The accuracy between heritability estimates from SEM, K and PCA statistical approaches supports the variance component estimation between methods. The heritability estimates for methane emissions in this thesis are close to that reported for this trait from other studies (Difford et al., 2018; van Engelen et al., 2018; Lassen and Difford, 2020; Zhang et al., 2020). The estimated heritability indicates that there is an opportunity to select for lower enteric methane emissions to tackle this trait through animal breeding programs.

Host genetic control over the microbiota composition

Heritability estimates for microbiota single genus in the SEM study showed a mean heritability value for the genus analyzed of 0.25, with a wide range (0.08 to 0.48) depending on the genus studied. Similar heritabilities estimates have been previously reported for single taxa (Wallace et al., 2019). A close heritability to that obtained from single genus in the SEM study, was obtained in the PCA study for the aggregated microbiota into PC1 (0.30, 95% HPD = 0.12 to 0.50). Similar results (0.29 ±0.11) were obtained for the proportion of variance explained by host genetic markers, for a PC1 limited to selected archaeal ruminal OTUs, in

a previous study using an amplicon sequencing based approach (Zhang et al., 2020). There was consistency between SEM and PCA statistical approaches, reaching similar results for the heritability estimation of microbiota. This moderate heritability for microbiota indicates that there would be a quite fast response to selection for this phenotype. Genetic correlations between microbiota relative abundance and methane concentration were found through the SEM variance component estimation analyses. Methane concentration showed moderate positive genetic correlations (> 0.50) with some eukaryote genera (*i.e. Stentor spp., Pseudocohnilembus spp., Paramecium spp, Oxytricha spp., Stylonychia spp. and Neocallimastix spp.*), as well as moderate but negative genetic correlations (< −0.40) with some bacteria genera (*i.e. Prevotella spp. and Succinimonas spp.*). Analogous results were obtained from the variance components estimation using PC1, with large genetic correlations (> 0.70) between PC1 and methane concentration, in which PC1 values increased along the relative abundance of eukaryote. The microbiota dimensionality reduction through PCA approach has advantages over the single genus SEM approach. The PCA aggregates the core microbiota into a single vector explaining most of the variance from the original variables (PC1). The analysis of the core microbiota, as an aggregated variable in a single vector, allows to include it as a phenotype in an animal breeding program context. This thesis provides information on the possible inclusion of microbiota composition as a proxy for methane emissions and other correlated traits of interest through genomic selection in animal breeding programs.

Causality of microbiome over methane emissions

The SEM approach was performed to infer causality from single microbiota genus over methane emissions. This approach has been used before to infer causality from associated traits in animal production (Wu et al., 2007; Rosa et al., 2011; De Los Campos et al., 2014). However, this is an innovative approach in metagenomics, as a methodology to deal with the "driver-passenger dilemma" to infer causality rather than association. In addition to the positive association found between eukaryotes and methane emissions discussed before, the SEM analyses established a causal relationship from single taxa from all eukaryotes analyzed (*Stentor spp., Pseudocohnilembus spp., Paramecium spp., Oxytricha spp., Stylonychia spp., Neocallimastix spp., Tetrahymena spp.* and *Ichthyphthirius spp.*) over

the methane concentration. Larger relative abundance of eukaryotes boosted methane concentration. The study of the role of microbiome as cause or effect has been previously tackled in ruminants (Morgavi et al., 2012; O'Hara et al., 2020), but as far as we know, not under a SEM approach. The case of microbiota relative abundances, as the causative factor affecting methane emissions variation in ruminants, fits neatly into a recursive rather than into a simultaneous association. The biological support for the latter affirmation lies on the evidence of methanogenic archaea being imperative for enteric methane production in ruminants, as well as in the physiological phenomenon of methane emissions itself. The release of methane from the rumen prevents from a negative feedback of methane production, thereby supporting the hypothesis of the mentioned recursive association. The findings of the SEM analyses contributed to the elucidation of microbiome acting as a cause rather than as an effect, over the methane emissions phenotype. Limitations from the SEM study were that all genera were analysed one at a time, thus informing of their single associations. The partial genetic control exerted by the host over each genus was also estimated, through the variance component estimation. However, it lacked to inform of the genetic modulation of methane emissions from the whole microbiota. The microbiota relationship matrices (Ks) approach analysis was performed to estimate the whole microbiota effect over methane emissions.

Single taxa vs whole microbiome effect over methane emissions

The whole microbiota was integrated into Ks through different methods. Similarly, as for the genomic relationship matrix (G), the K matrices had $n \times n$ dimensions, where n was the number of animals used to build G and K. The conformability between G and K allowed to generate a square interaction matrix (G#K) through the Hadamard ("#") product between them. All matrices had identical dimensions, which allowed to include them together into mixed models. The K approach to integrate whole microbiota has been approached previously (Ross et al., 2013; Camarinha-Silva et al., 2017; Khanal et al., 2020); However, the evaluation of performance accuracy between Ks has not been reported before in a variance component estimation framework, as far as known. The accuracy comparison from the variance component analyses, was possible due to the conformability of the G, K and G#K matrices, for the GBLUP, MBLUP, HBLUP and H_iBLUP. The Holobiability from the H_iBLUP model was larger than the combination of the heritability and the proportion of the

phenotypic variance explained by the microbiome effect. This result suggests that the G and K interaction accounts for a certain proportion of the phenotypic variance, that variated according to the K included into the model. The H_iBLUP showed to be the most plausible model according to the deviance information criterion (DIC), suggesting that holobiont modeling could outperform the classical approaches used so far, but more research should be done before affirming this. A clear limitation from the K analyses is that microbiome information remains dimensionally unmanageable to estimate the host genetic control over the whole microbiota. Its matrix shape precludes the assignment of genetic parameters estimates. Another limitation linked to the K analyses is the impossibility of allocate a single microbiome value for each animal, which might allow its inclusion as phenotype into breeding programs. The PCA analyses tackled both limitations by conferring a vectorial shape to the microbiota, allowing the estimation of host genetic control, as well as a single microbiota value for each animal.

Microbiome dimensionality reduction

The whole microbiome was aggregated, through PCA, into principal components (PCs). The PCs contain the maximized variance from the original variables (Macciotta et al., 2010). The PCs were vectors of coordinates, containing aggregated microbiome variables, with a single value for each animal. The vectorial shape of the PCs allowed to consider them as phenotypes. The PCs were then included into bivariate models jointly with methane emissions. The inclusion of the whole microbiota information as PCs into variance component estimation is a novelty in this thesis. It allowed the estimation of the host genetic control over the core microbiota, which reached moderate values (≈ 0.30) for PC1 at all taxonomic levels. It also allowed to estimate the genetic correlation of the core microbiota with methane concentration, yielding large positive values (> 0.70). The PC1 separated eukaryotes from bacteria and archaea, where animals with larger proportions of eukaryote had larger values of PC1, supporting the direct effect of eukaryotes on methane emissions from previously obtained in the risk factors and SEM analyses. The PCs inclusion into bivariate animal models accomplished to estimate the host genetic effect over methane emissions and the core microbiota simultaneously, and finally the genetic correlation between the core microbiota and methane emissions.

Implications

The inclusion of microbiome into animal breeding programs through aggregated variables (PCs) could accelerate the reduction of methane emissions from livestock. The larger heritability of the core microbiome compared with that from methane emissions and the high genetic correlation among both traits is encouraging. The response to selection is dependent on heritability, with faster response to selection for traits with larger heritabilities. The findings from this thesis indicate that the estimated heritability of the core microbiome is near double of that from methane emissions and highly genetically correlated. However, both methane emissions and microbiome are difficult to measure traits. Therefore, more cost-effective methodologies, to sample microbiota from animals under commercial conditions are to be studied. There are evidences of bolus and saliva samples being good predictors of ruminal microbiota (Tapio et al., 2016), similar sampling approaches as the mentioned before could enable large-scale phenotyping. If aggregated variables (PCs) of proxies for rumen microbiota behave similarly as in this thesis, microbiome inclusion into breeding programs could become feasible. The dimensionality reduction of microbiota through a robust methodology (PCA) is considered a relevant contribution of this thesis, for the plausible inclusion of microbiota information into animal breeding programs. This thesis also contributed to the understanding of association between genotype, microbiota, and methane emissions. It also provides information and methodology approaches for the possible inclusion of microbiota composition as a proxy for methane emissions and other correlated traits of interest through genomic selection in animal breeding programs.

REFERENCES

Barnett, M.C., J.P. Goopy, J.R. McFarlane, I.R. Godwin, J. V. Nolan, and R.S. Hegarty. 2012. Triiodothyronine influences digesta kinetics and methane yield in sheep. Anim. Prod. Sci. 52:572. doi:10.1071/AN11303.

Belanche, A., G. De La Fuente, and C.J. Newbold. 2014. Study of methanogen communities associated with different rumen protozoal populations. doi:10.1111/1574-6941.12423.

Brandt, C., E. Bongcam-Rudloff, and B. Müller. 2020. Abundance Tracking by Long-Read Nanopore Sequencing of Complex Microbial Communities in Samples from 20 Different Biogas/Wastewater Plants. Appl. Sci. 10:7518. doi:10.3390/app10217518.

Camarinha-Silva, A., M. Maushammer, R. Wellmann, M. Vital, S. Preuss, and J. Bennewitz. 2017. Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. Genetics 206:1637–1644. doi:10.1534/genetics.117.200782.

Delgado, B., A. Bach, I. Guasch, C. González, G. Elcoso, J.E. Pryce, and O. Gonzalez-Recio. 2018. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. Sci. Rep. 9:1–13. doi:10.1038/s41598-018-36673-w.

Difford, G.F., D.R. Plichta, P. Løvendahl, J. Lassen, S.J. Noel, O. Højberg, A.-D.G. Wright, Z. Zhu, L. Kristensen, H.B. Nielsen, B. Guldbrandtsen, and G. Sahana. 2018. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLOS Genet. 14:e1007580. doi:10.1371/journal.pgen.1007580.

van Engelen, S., H. Bovenhuis, P.P.J. van der Tol, and M.H.P.W. Visker. 2018. Genetic background of methane emission by Dutch Holstein Friesian cows measured with infrared sensors in automatic milking systems. J. Dairy Sci. 101:2226–2234. doi:10.3168/jds.2017-13441.

Gianola, D. 1982. Theory and Analysis of Threshold Characters. J. Anim. Sci. 54:1079. doi:10.2527/jas1982.5451079x.

Gloor, G.B., J.M. Macklaim, V. Pawlowsky-Glahn, and J.J. Egozcue. 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 8:2224. doi:10.3389/fmicb.2017.02224.

Goopy, J.P., A. Donaldson, R. Hegarty, P.E. Vercoe, F. Haynes, M. Barnett, and V. Hutton Oddy. 2014. Low-methane yield sheep have smaller rumens and shorter rumen retention time. Br. J. Nutr. 111:578–585. doi:10.1017/S0007114513002936.

Greening, C., R. Geier, C. Wang, L.C. Woods, S.E. Morales, M.J. McDonald, R. Rushton-Green, X.C. Morgan, S. Koike, S.C. Leahy, W.J. Kelly, I. Cann, G.T. Attwood, G.M. Cook, and R.I. Mackie. 2019. Diverse hydrogen production and consumption pathways influence methane production in ruminants. ISME J. 13:2617–2632. doi:10.1038/s41396-019-0464-2.

Guyader, J., M. Eugène, P. Nozière, D.P. Morgavi, M. Doreau, and C. Martin. 2014.

Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. Animal 8:1816–1825. doi:10.1017/S1751731114001852.

Hackstein, J.H.P., and R.M. de Graaf. 2013. Anaerobic Ciliates and Their Methanogenic Endosymbionts. J.H.P. Hackstein, ed. Springer Berlin, Berlin.

Hook, S.E., A.-D.G. Wright, and B.W. Mcbride. 2010. Methanogens: Methane Producers of the Rumen and Mitigation Strategies. Archaea 2010:11. doi:10.1155/2010/945785.

Kamke, J., S. Kittelmann, P. Soni, Y. Li, M. Tavendale, S. Ganesh, P.H. Janssen, W. Shi, J. Froula, E.M. Rubin, and G.T. Attwood. 2016. Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. Microbiome 4. doi:10.1186/s40168-016-0201-2.

Khanal, P., C. Maltecca, C. Schwab, J. Fix, and F. Tiezzi. 2020. Microbiability of meat quality and carcass composition traits in swine. J. Anim. Breed. Genet. 00:1:14. doi:10.1111/jbg.12504.

Lassen, J., and G.F. Difford. 2020. Review: Genetic and genomic selection as a methane mitigation strategy in dairy cattle. Animal 1–11. doi:10.1017/S1751731120001561.

Liu, J., H. Chen, Q. Zhu, Y. Shen, X. Wang, M. Wang, and C. Peng. 2015. A novel pathway of direct methane production and emission by eukaryotes including plants, animals and fungi: An overview. Atmos. Environ. 115,:26–35. doi:10.1016/j.atmosenv.2015.05.019.

López-Paredes, J., I. Goiri, R. Atxaerandio, A. García-Rodríguez, E. Ugarte, J.A. Jiménez-Montero, R. Alenda, and O. González-Recio. 2020. Mitigation of greenhouse gases in dairy cattle via genetic selection: 1. Genetic parameters of direct methane using noninvasive methods and proxies of methane. J. Dairy Sci. 103:7199–7209. doi:10.3168/jds.2019-17597.

De Los Campos, G., D. Gianola, P. Boettcher, and P. Moroni. 2014. A structural equation model for describing relationships between somatic cell score and milk yield in dairy goats. J. Anim. Sci. 84:2934–2941. doi:10.2527/jas.2006-016.

Lovell, D., V. Pawlowsky-Glahn, J.J. Egozcue, S. Marguerat, and J. Bähler. 2015. Proportionality: A Valid Alternative to Correlation for Relative Data. PLOS Comput. Biol. 11(3):e1004075, 1–12. doi:10.1371/journal.pcbi.1004075.

Macciotta, N.P.P., G. Gaspa, R. Steri, E.L. Nicolazzi, C. Dimauro, C. Pieramati, and A. Cappio-Borlino. 2010. Using eigenvalues as variance priors in the prediction of genomic breeding values by principal component analysis. J. Dairy Sci. 93:2765–2774. doi:10.3168/jds.2009-3029.

Martínez-Álvaro, M., M.D. Auffret, R.D. Stewart, R.J. Dewhurst, C.-A. Duthie, J.A. Rooke, R.J. Wallace, B. Shih, T.C. Freeman, M. Watson, and R. Roehe. 2020. Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. Front. Microbiol. 11:659. doi:10.3389/fmicb.2020.00659.

Miglior, F., A. Fleming, F. Malchiodi, L.F. Brito, P. Martin, and C.F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100:10251–10271. doi:10.3168/jds.2017-12968.

Mizrahi, I., and E. Jami. 2018. Review: The compositional variation of the rumen microbiome and its effect on host performance and methane emission. Animal 12:220–232. doi:10.1017/S1751731118001957.

Morgavi, D.P., C. Martin, J.P. Jouany, and M.J. Ranilla. 2012. Rumen protozoa and methanogenesis: Not a simple cause-effect relationship. Br. J. Nutr. 107:388–397. doi:10.1017/S0007114511002935.

Newbold, C.J., G. De La Fuente, A. Belanche, E. Ramos-Morales, and N.R. McEwan. 2015. The Role of Ciliate Protozoa in the Rumen. Front. Microbiol. 6:1–14. doi:10.3389/fmicb.2015.01313.

O'Hara, E., A.L.A. Neves, Y. Song, and L.L. Guan. 2020. The Role of the Gut Microbiome in Cattle Production and Health: Driver or Passenger?. Annu. Rev. Anim. Biosci. 8:199–220. doi:10.1146/annurev-animal-021419-083952.

Poretsky, R., L.M. Rodriguez-R, C. Luo, D. Tsementzi, and K.T. Konstantinidis. 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9. doi:10.1371/journal.pone.0093827.

Rosa, G.J., B.D. Valente, G. de los Campos, X.-L. Wu, D. Gianola, and M.A. Silva. 2011. Inferring causal phenotype networks using structural equation models. Genet. Sel. Evol. 43:6. doi:10.1186/1297-9686-43-6.

Ross, E.M., P.J. Moate, L.C. Marett, B.G. Cocks, and B.J. Hayes. 2013. Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle. PLoS One 8:1–8. doi:10.1371/journal.pone.0073056.

Rupp, R., and D. Boichard. 1999. Genetic Parameters for Clinical Mastitis, Somatic Cell Score, Production, Udder Type Traits, and Milking Ease in First Lactation Holsteins.

Tapio, I., K.J. Shingfield, A. Bonin, D. Fischer, A.R. Bayat, J. Vilkki, P. Taberlet, T.J. Snelling, and R.J. Wallace. 2016. Oral Samples as Non-Invasive Proxies for Assessing the Composition of the Rumen Microbial Community. PLoS One 1–15. doi:10.1371/journal.pone.0151220.

Tapio, I., T.J. Snelling, F. Strozzi, and R.J. Wallace. 2017. The ruminal microbiome associated with methane emissions from ruminant livestock. J. Anim. Sci. Biotechnol. 8:1–11. doi:10.1186/s40104-017-0141-0.

Wallace, R.J., J.A. Rooke, N. McKain, C.A. Duthie, J.J. Hyslop, D.W. Ross, A. Waterhouse, M. Watson, and R. Roehe. 2015. The rumen microbial metagenome associated with high methane production in cattle. BMC Genomics 16:1–14. doi:10.1186/s12864-015-2032-0.

Wallace, R.J., G. Sasson, P.C. Garnsworthy, I. Tapio, E. Gregson, P. Bani, P. Huhtanen, A.R. Bayat, F. Strozzi, F. Biscarini, T.J. Snelling, N. Saunders, S.L. Potterton, J.

Craigon, A. Minuti, E. Trevisi, M.L. Callegari, F.P. Cappelli, E.H. Cabezas-Garcia, J. Vilkki, C. Pinares-Patino, K.O. Fliegerová, J. Mrázek, H. Sechovcová, J. Kopečný, A. Bonin, F. Boyer, P. Taberlet, F. Kokou, E. Halperin, J.L. Williams, K.J. Shingfield, and I. Mizrahi. 2019. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci. Adv. 5:1–12. doi:10.1126/sciadv.aav8391.

Wu, X., B. Heringstad, Y. Chang, G.D.L. Campos, and D. Gianola. 2007. Inferring Relationships Between Somatic Cell Score and Milk Yield Using Simultaneous and Recursive Models. J. Dairy Sci. 90:3508–3521. doi:10.3168/jds.2006-762.

Zhang, Q., G. Difford, G. Sahana, P. Løvendahl, J. Lassen, M.S. Lund, B. Guldbrandtsen, and L. Janss. 2020. Bayesian modeling reveals host genetics associated with rumen microbiota jointly influence methane emission in dairy cows. ISME J. 14:2019–2033. doi:10.1038/s41396-020-0663-x.

Zhou, M., E. Hernandez-Sanabria, and L.G. Le. 2009. Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. Appl. Environ. Microbiol. 75:6524–6533. doi:10.1128/AEM.02815-08.

CHAPTER 8

CONCLUSIONS

Alejandro Saborío-Montero[1,2]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain; [2]Universidad de Costa Rica, Escuela de Zootecnia, 11501 San Pedro de Montes de Oca, San José, Costa Rica.

- Results in this thesis confirm the importance of eukaryotes in the rumen microbiome and their role in the methanogenesis.

- According to the probit threshold models, increments in milk yield as well as in structure and capacity scores were risk factors for being classified in the upper quartile of methane concentration (ppm). Rumen eukaryotes and some hydrogenotrophic archaea were the main phenotypic risk factors in microbiota for both traits.

- This study applied SEM as a tool to integrate genomic, metagenomic and phenotypic information in order to jointly analyze plausible biological relationships. Ciliate protozoa (7 genus) showed moderate heritabilities and consistent positive genetic correlation to CH_4 in both statistical model approaches (non-recursive and recursive).

- Models considering a causal relationship of the microbiota over methane emissions are statistically more plausible.

- Methane evaluated in this thesis showed enough heritability as to be included in dairy cattle breeding programs, in order to obtain more efficient animals while diminishing their environmental footprint.

- SEM could be used to also include metagenomic information into genetic evaluations analysis accounting for the recursive relationship between traits, and potentially increasing the reliability.

- Microbiome relationship matrices can be incorporated into the models used in genetic evaluation and variance component estimation in order to account for variance explained by the microbiota. The MDS, CCA and RDA matrices seemed preferable for unbiased estimation of variance components.

- Larger GEBVs accuracy can be obtained by including a microbiome effect into the model.

- Accounting for the genome × microbiome interaction in the H_iBLUP model improves variance component estimation and might yield more accurate performance predictions of methane emissions.

- Considering the holobiont organism and estimating the holobiability in the models must be considered in the future due to its potential implications. If the holobiability is consistent across studies and its value surpasses heritability value for the trait of interest, it might be a better estimator to be included into the selection index, which a priori would increase the response to selection.

- Dimensionality reduction can be implemented through PCA in order to include the microbiota composition as a trait under selection, both at taxonomic and functional levels. This approach simplifies the complexity and enable using these variables as phenotypes.

- The heritability estimates for these PC were relatively large and genetically correlated to methane concentration. This strategy could modulate the rumen core metagenome and reduce methane production through correlated genetic response.

- A large enough reference population of cows with microbiome and genotype information needs to be created for genomic selection implementation. However, it is necessary to evaluate possible collateral effects that might adversely affect the animal metabolism.

- Results in this thesis stimulate new opportunities for mitigating greenhouse-gas emissions from livestock, through direct modulation of the microbiota composition via animal breeding programs.