UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE VETERINARIA

TESIS DOCTORAL

Uso de datos de microbioma para explicar la expresión de caracteres productivos en especies domésticas

Use of microbiome data to explain the expression of productive traits in domestic species

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Adrián López García

Directores

Cristina Óvilo Martín
Óscar González Recio

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE VETERINARIA

TESIS DOCTORAL

USO DE DATOS DE MICROBIOMA PARA EXPLICAR LA EXPRESIÓN DE CARACTERES PRODUCTIVOS EN ESPECIES DOMÉSTICAS

USE OF MICROBIOME DATA TO EXPLAIN THE EXPRESSION OF PRODUCTIVE TRAITS IN DOMESTIC SPECIES

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Adrián López García

DIRECTORES

Cristina Óvilo Martín

Óscar González Recio

Madrid, 2022

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE VETERINARIA

TESIS DOCTORAL

USO DE DATOS DE MICROBIOMA PARA EXPLICAR LA EXPRESIÓN DE CARACTERES PRODUCTIVOS EN ESPECIES DOMÉSTICAS

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Adrián López García

DIRECTORES

Cristina Óvilo Martín

Óscar González Recio

El trabajo experimental que ha dado lugar a esta tesis ha sido realizado en el Departamento de Mejora Genética Animal del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), y ha sido financiado mediante los siguientes proyectos de investigación:

- Ayudas para la contratación de personal investigador predoctoral en formación (**FPI-SGIT2016-06**) **del Ministerio de Economía y Competitividad**.
- Proyecto **TREASURE** (grant 634476) **de la Unión Europea** (H2020).
- Proyecto **S2013/ABI-2913** (MEDGAN-CM) **de la Comunidad de Madrid y cofinanciado por el FEDER**.
- Proyecto **RTA2015-00022-C03-02** (METALGEN) **del Plan Nacional de Investigación, Desarrollo e Innovación 2013-2020 y el Ministerio de Economía y Competitividad**.

AGRADECIMIENTOS

Esta tesis ha sido posible gracias a mucha gente que me ha prestado su conocimiento, su experiencia, su ayuda y su apoyo.

En primer lugar, agradezco a mis directores de tesis, Cristina Óvilo Martín y Óscar González Recio, por darme la oportunidad de llevar a cabo este proyecto y enseñarme prácticamente todo lo que sé sobre la investigación científica. Gracias por vuestro trabajo como directores, por vuestra confianza, vuestra paciencia y vuestra disponibilidad. Gracias por un ambiente de trabajo inmejorable, por haber hecho de esta etapa algo más fácil y llevadero.

A Juan Pablo Gutiérrez García, por tu fundamental papel como tutor y por ser mi nexo con la Universidad Complutense de Madrid.

Al grupo de NEIKER, Aser García, Raquel Atxaerandio e Idoia Goiri, por hacer posible la parte de esta tesis del proyecto METALGEN, además de por aportar vuestro conocimiento y experiencia y por vuestra buena disposición para ayudar en cualquier momento.

A Rita, Yolanda y Fabián, por hacer posible la parte de la tesis del proyecto MEDGAN, por vuestra cercanía y amabilidad, y por vuestra buena voluntad para ayudarme con las dificultades y dudas que he tenido durante estos cuatro años.

A Juan García Casco y María Muñoz, por darme la oportunidad de entrar en el mundo de la ciencia y por seguir ahí hasta este momento. Gracias por la confianza que pusisteis en mí, y por la experiencia que me llevé después de trabajar con vosotros. Gracias también a los demás compañeros de Zafra, Miguel, Carmen y José, por crear un buenísimo ambiente de trabajo mientras estuve allí.

A Alejandro, Silvia, Eli, Adrián, Mónica, María Saura, Anita, Consuelo, Marta y todo el grupo del INIA, los que seguís ahí y los que no, por vuestra amistad y apoyo, y por "obligarme" a tomarme la pausa del café.

A Jordi Estellé, Manuel Revilla y al grupo del INRAE-GABI, por permitirme hacer una estancia muy enriquecedora y acogerme con gran hospitalidad.

Un especial agradecimiento a mis padres, Aniria y Carlos, por habérmelo dado todo en la vida y hacer de mí la persona que soy ahora; a mi hermano Andrés, por estar siempre ahí; y a toda mi familia y amigos, por confiar siempre en mí y darme vuestro apoyo incondicional.

Y, por último, a Elena, mi pareja, por estar a mi lado durante tantos años, por cuidar de mí y animarme a seguir adelante. Gracias por formar parte de mi vida.

ACKNOWLEDGEMENTS

This thesis has been possible thanks to many people who has lent me their knowledge, experience, help and support.

First of all, I would like to thank my thesis supervisors, Cristina Óvilo Martín and Óscar González Recio, for giving me the opportunity to carry out this project and for teaching me almost everything I know about scientific research. Thank you for your labour as supervisors, for your trust, your patience and your accessibility. Thank you for and excellent working environment, for making this stage easier and more bearable.

To Juan Pablo Gutiérrez García, for your vital role as tutor and for being my nexus with the Complutense University of Madrid.

To the NEIKER team, Aser García, Raquel Atxaerandio and Idoia Goiri, for making possible the part of the thesis from the METALGEN project, as well as for providing your knowledge and experience, and for your good will to help at any time.

To Rita, Yolanda and Fabián, for making possible the part of the thesis from the MEDGAN project, for your closeness and kindness, and for your willingness to help me with the multiple difficulties and doubts during these four years.

To Juan García Casco and María Muñoz, for giving me the opportunity to initiate in the world of science, and for staying close until now. Thank you for trusting me and for the background I acquired working with you. Thanks also to the rest of workmates in Zafra, Miguel, Carmen and José, for building a fantastic working atmosphere.

To Alejandro, Silvia, Eli, Adrián, Mónica, María Saura, Anita, Consuelo, Marta and all the INIA group, to those who are still there and those who are not, for your friendship and support, and for "forcing" me to take the coffee breaks.

To Jordi Estellé, Manuel Revilla and all the staff at INRAE-GABI, for allowing me to do a very fulfilling residence and welcome me with great hospitality.

Special thanks to my parents, Aniria and Carlos, for giving me everything in life and making me the person I am today; to my brother, Andrés, for always being there; and to all my family and friends, for always trusting me and giving me your unconditional support.

And finally, to Elena, my partner, for staying by my side for so many years, for taking care of me and encouraging me to keep going. Thank you for being part of my life.

Index

Index of figures

Index of tables

Abbreviation index

ADG:	average daily gain
AFOLU:	Agriculture, Forestry and Land Use
ALR:	additive log ratio
AS:	amplicon sequencing
ASV:	amplicon sequence variant
BOG:	best overlap graph
CAIT:	Climate Analysis Indicators Tool
CDS:	clade-specific markers
CLR:	centred log ratio
CO_2e/eq:	CO_2 equivalent
cPAL:	combinatorial probe-anchor ligation
DA:	differential abundance
DBG:	de Bruijn graph
DNB:	DNA nanoball
EDGAR:	Emissions Database for Global Atmospheric Research
ETEC:	enterotoxigenic *Escherichia coli*
emPCR:	emulsion PCR
FA:	fatty acid
FAO:	Food and Agriculture Organization
FC:	fold-change
GCP:	Global Carbon Project
GG:	GreenGenes
GH:	growth hormone
GHG:	greenhouse gas
GLEAM:	Global Livestock Environmental Assessment Model
GM:	gut microbiota
HO/HF:	high-oleic/high-fibre
IEA:	International Energy Agency
iHMP:	Integrative Human Microbiome Project
ILR:	isometric log ratio
IMF:	intramuscular fat
IP2G:	InterPro2GO
IPCC:	Intergovernmental Panel on Climate Change
ITS:	internal transcribed spacers
KEGG:	Kyoto Encyclopaedia of Genes and Genomes
LCA:	lowest common ancestor

LSU: large sub-unit
M5nr: MD5 non-redundant
MICCA: MICrobial Community Analysis
MUFA: monounsaturated fatty acids
Mya: million years ago
NGS: next-generation sequencing
OLC: overlap layout consensus
ONT: Oxford Nanopore Technologies
OTU: operational taxonomic unit
PCA: principal component analysis
PCR: polymerase chain reaction
PERMANOVA: permutational analysis of variance
PUFA: polyunsaturated fatty acids
PYY: peptide YY
QIIME: Quantitative Insights Into Microbial Ecology
RA: relative abundance
RCA: rolling circle amplification
RDP: Ribosomal Database Project
RM: rumen microbiota
rRNA: ribosomal RNA
SBL: sequence by ligation
SBS: sequence by synthesis
SFA: saturated fatty acids
SMS: single molecule sequencing
SMRT: single molecule real-time
SNP: single nucleotide polymorphism
SOLiD: sequencing by oligonucleotide ligation and detection
SSCP: single-stranded conformational polymorphism
SSU: small sub-unit
STR: short tandem repeats
TGS: third-generation sequencing
UNFCCC: United Nations Framework Convention on Climate Change
VFA: volatile fatty acids
WMS: whole metagenome sequencing
ZMW: zero-mode waveguides

Resumen

El descubrimiento de comunidades microbianas asociadas simbióticamente a organismos eucariotas ha llevado a un cambio de paradigma en la definición de individuo biológico, que ahora se ve como una combinación codependiente del hospedador y su microbioma, u holobionte. Por tanto, el estudio de los microbiomas se ha convertido en algo fundamental para comprender la biología de los organismos vivos complejos. De hecho, se ha observado que las comunidades microbianas poseen un papel crucial en la salud, supervivencia, desarrollo y metabolismo del hospedador. Los recientes avances en secuenciación genética han supuesto un importante impulso para la investigación en microbiología, al permitir la obtención de bases de datos de secuenciación masiva que abarcan una gran parte de la diversidad presente dentro de los microbiomas. La era del *next-generation sequencing* ha aportado nuevos conocimientos sobre el efecto de las comunidades microbianas sobre el fenotipo del hospedador, con especial relevancia del microbioma intestinal. Para la industria ganadera este hecho ha dado lugar a importantes avances en la comprensión de los mecanismos biológicos que influyen en productividad, sostenibilidad y bienestar animal, lo que podría ser útil para afrontar los desafíos existentes en este sector.

La ganadería española está mayoritariamente representada por dos especies de gran importancia económica, el cerdo doméstico, en especial la raza ibérica, y el vacuno. El cerdo ibérico es una raza rústica autóctona de la Península Ibérica con una alta tendencia a la acumulación de grasa y criada tradicionalmente en un sistema semiextensivo denominado montanera, lo que confiere unas propiedades organolépticas especiales a sus productos cárnicos, muy demandados por su calidad. Sin embargo, su baja prolificidad y productividad han favorecido la intensificación y la práctica del cruzamiento con Duroc, lo que ha supuesto una pérdida de calidad y un efecto negativo en la composición de sus productos. En cuanto al vacuno, es la especie ganadera más importante a nivel mundial, siendo útil para la obtención tanto de carne como de leche. Sin embargo, su proceso digestivo genera una cantidad importante de metano, lo que contribuye de forma relevante al calentamiento global.

En esta tesis se han abordado los retos más importantes de estos sistemas de producción ganadera desde la perspectiva del microbioma. Conocer el papel de la microbiota intestinal en el metabolismo del cerdo ibérico es fundamental para determinar el verdadero alcance de las consecuencias del cruzamiento con Duroc y la intensificación sobre el metabolismo y la calidad de los productos cárnicos. Respecto al vacuno, si bien los programas de selección genética se erigen actualmente como la estrategia más eficiente para reducir las emisiones de metano, la microbiota ruminal juega un

importante papel en la producción de metano y, por lo tanto, la modulación del microbioma podría ayudar a mitigar su contribución al calentamiento global. Para lograr estos objetivos, se ha estudiado la composición de la microbiota intestinal en ambas especies a través de diferentes metodologías de secuenciación y análisis. Se ha utilizado secuenciación de amplicones (16S rRNA) para explorar la fracción bacteriana de la microbiota fecal de cerdos ibéricos y Duroc, mientras que la composición de la microbiota ruminal se ha explorado mediante secuenciación de metagenoma completo. En ambos casos se ha realizado un análisis composicional de los datos.

La tesis se compone de tres experimentos. En el primero se evaluaron las diferencias de rendimiento de dos herramientas de procesamiento de secuencias e identificación de OTUs, Mothur y QIIME, y dos bases de datos de referencia de rRNA diferentes, GreenGenes y SILVA, con el fin de determinar el protocolo más fiable para agrupar y clasificar datos de secuenciación de amplicones. Para este análisis se utilizaron secuencias de 16S (región V4) procedentes del rumen, de una cohorte de 18 vacas lecheras. Los resultados obtenidos indican que es preferible el uso de SILVA sobre GreenGenes como base de datos de referencia para la clasificación de OTUs, mientras que las diferencias entre las dos herramientas bioinformáticas probadas se limitan a OTUs de baja abundancia, acentuándose más con el uso de GreenGenes como base de datos. En el segundo experimento se caracterizó la microbiota fecal de cerdos con diferente genotipo (Duroc e Ibérico) y alimentados con dietas con diferente fuente de energía (carbohidratos versus ácido oleico) con el fin de explorar los efectos, tanto del fondo genético como de la suplementación con oleico, en la composición de la microbiota intestinal. Se secuenciaron las regiones hipervariables V3-V4 del gen rRNA 16S, utilizando el protocolo *Illumina MiSeq paired-end*, en 48 cerdos machos castrados de raza pura, 19 Duroc y 29 ibéricos, alimentados durante la fase de crecimiento con una dieta control o con una dieta enriquecida con aceite de girasol. Después de tener en cuenta los resultados del primer experimento, el procesamiento de las secuencias se realizó usando la herramienta QIIME y la base de datos de referencia SILVA. Los resultados han revelado la existencia de diferencias en la composición de la microbiota fecal entre razas y, en menor medida, entre dietas. Se observa una menor riqueza de OTUs en los cerdos Duroc, que poseen más abundancia de *Prevotella* y *Actinobacteria*, mientras que los cerdos ibéricos poseen una microbiota más diversa, con menos presencia de taxa típicos del período de lactancia, lo que indica una tasa de transición microbiana más rápida. Por otro lado, los animales alimentados con una dieta alta en oleico poseen una mayor abundancia de varios OTUs de *Prevotella*, debido al mayor contenido de fibra en esta dieta. Se han encontrado interacciones significativas entre la raza y la dieta para la abundancia de varios OTUs, principalmente de los géneros *Bifidobacterium*, *Lactobacillus* y *Corynebacterium*, lo que podría indicar una respuesta diferencial a la dieta entre ambas razas. Este estudio ha aportado

pruebas de la influencia de la genética y la dieta en la composición de la microbiota intestinal del cerdo. Por último, en el tercer experimento se ha caracterizado la microbiota ruminal del vacuno lechero, tanto taxonómica como funcionalmente, con el fin de desentrañar el papel de microbios y genes microbianos concretos en la producción de metano. Se dispuso de un total de 437 vacas lecheras de 14 granjas comerciales del norte de España para tomar medidas individuales de concentración de metano, durante las visitas de las vacas al sistema de ordeño automático. Se realizó secuenciación completa del metagenoma del contenido ruminal utilizando el dispositivo MinION de *Oxford Nanopore Technologies*. Los resultados sugieren una asociación entre la producción de metano y la composición general de la microbiota del rumen, encontrándose 36 géneros microbianos y 279 KEGGs diferencialmente abundantes entre animales con diferentes niveles de producción de metano. Los géneros asociados a altas emisiones de metano son eucariotas, en su mayoría ciliados y hongos, mientras que los asociados a bajas emisiones son grupos bacterianos, principalmente fermentadores de azúcares solubles. Los genes que participan en la metanogénesis no han resultado más abundantes en los animales de altas emisiones, pero un análisis más profundo ha revelado que genes implicados en la respiración anaeróbica basada en nitratos son más abundantes en los animales de bajas emisiones. Este estudio ofrece una descripción integral de las poblaciones microbianas del rumen utilizando secuenciación de lectura larga y una gran cohorte de animales, destacando el papel de los eucariotas en la producción de metano a través de arqueas metanogénicas endosimbiontes, así como la contribución de microbios capaces de utilizar nitrato para reducir la metanogénesis a través de inhibición competitiva.

La investigación realizada en esta tesis evidencia la relación entre el microbioma intestinal y las características del hospedador, en el contexto de la producción animal, mediante el uso de diferentes protocolos de secuenciación y clasificación. Los resultados obtenidos ponen de manifiesto la alta complejidad de los microbiomas, enfatizando así la necesidad de contabilizar la mayor cantidad de biodiversidad posible al realizar estudios de microbiomas, y arrojan luz sobre la relación entre grupos específicos de microorganismos intestinales y rasgos fenotípicos del hospedador de alta importancia económica, abriendo una nueva puerta para afrontar los retos actuales en la industria ganadera.

Abstract

The discovery of microbial communities symbiotically associated with eukaryotic organisms has led to a paradigm shift in the definition of the biological individual, which is now seen as a co-dependent combination of the host and its microbiome, or holobiont. Thus, the study of microbiomes has become essential to understand the biology of complex living organisms. Indeed, current research points to a crucial role of microbial communities in host health, survivability, development and metabolism. Recent advances in DNA sequencing have entailed a significant boost to microbial research, allowing the generation of massive sequencing databases encompassing a large proportion of the diversity inside microbiomes. The era of next-generation sequencing has brought new knowledge about the role of microbial communities, with special significance for gut microbiomes, in host phenotype. For livestock industry, this has led to important advances in the understanding of biological mechanisms influencing animal welfare, productivity and sustainability, which could be useful to face existing challenges in animal production.

Spanish livestock is majorly represented by two species of great economic importance, the domestic pig, especially the Iberian breed, and taurine cattle. Iberian pig is a rustic breed native to the Iberian Peninsula with a high trend to fat accumulation and traditionally reared in a semi-extensive system called *montanera*, which confers special organoleptic properties to its products, highly demanded for their quality. However, its low prolificacy and productivity have encouraged both intensification and the practice of crossbreeding with Duroc, thereby involving a loss in quality and a negative effect on the composition of their products. As for cattle, it is the most important livestock species worldwide, being useful for obtaining both beef and milk. However, their digestive process comes with an important amount of greenhouse gas emissions in the form of methane, thus relevantly contributing to global warming.

In this thesis, the most important challenges in these livestock production systems have been addressed from a microbiome perspective. Knowing the role of gut microbiota in Iberian pig metabolism is essential to determine the true extent of the consequences of crossbreeding and intensification on metabolism and final products' quality. In regard to cattle, while genetic selection programmes currently stand as the most efficient strategy for reducing methane emissions, rumen microbiota plays a significant role in methane production, and thereby microbiome modulation could help to mitigate its contribution to global warming. In order to achieve these goals, gut microbiota composition has been studied in both species, through different sequencing and analysis methodologies. An amplicon sequencing approach (16S rRNA) has been used for exploring the bacterial fraction of the faecal

microbiota of Iberian and Duroc pigs, while rumen microbiota composition has been explored through a whole metagenome sequencing approach. In both cases, compositional analysis of microbial data has been performed.

Three experiments have been carried out. In the first one, differences in classification performance were evaluated for two different OTU-like post-processing pipelines used for amplicon sequencing data, Mothur and QIIME, and two different rRNA reference databases, GreenGenes and SILVA, in order to determine the most reliable protocol to cluster and classify amplicon sequencing data. Ruminal 16S-V4 sequences from a cohort of 18 dairy cows was used for this analysis. The results state that SILVA reference database is preferred over GreenGenes for classifying OTUs, while differences between pipelines were restricted to low-abundance OTUs, and were more accentuated when GreenGenes database was used. In the second experiment, faecal microbiota of pigs with different genotype (Duroc and Iberian) and fed diets with different energy source (carbohydrates versus oleic acid) was characterized in order to explore the effects of both genetic background and oleic supplementation on gut microbiota composition. V3-V4 hypervariable regions of 16S rRNA gene were sequenced, using Illumina MiSeq paired-end protocol, in 48 castrated male purebred pigs, 19 Duroc and 29 Iberian, fed either a control diet or a sunflower oil-enriched diet during growth stage. After taking into account the results from the first experiment, QIIME pipeline and SILVA reference database were used for sequence post-processing. Results exposed the existence of differences in faecal microbiota composition between breeds and, to a lesser extent, between diets. A lower OTU richness was observed in Duroc pigs, which possessed more abundance of *Prevotella* OTUs and *Actinobacteria*, while Iberian pigs possessed a more diverse microbiota with less presence of pre-weaning typical taxa, indicating a faster microbial transition rate. On the other hand, animals fed high-oleic diet possessed more abundance of several *Prevotella* OTUs, due to the higher content of fibre in this diet. Significant interactions between breed and diet were found for the abundance of several OTUs, mostly from *Bifidobacterium*, *Lactobacillus* and *Corynebacterium* genera, which might indicate a differential response to diet in both breeds. This study has brought some evidence to the influence of pig genetics and diet supply on gut microbiota composition. Lastly, in the third experiment, rumen microbiota of dairy cattle has been characterized, both taxonomically and functionally, in order to disentangle the role of specific microbes and genes in methane production. A total of 437 dairy cows from 14 commercial farms at northern Spain were used to take individual methane concentration measures during the cow's visits to the automatic milking system. Whole metagenome sequencing of ruminal content was performed using the MinION device from Oxford Nanopore Technologies. Results suggested an association between methane production and rumen

overall microbiota composition, and 36 microbial genera and 279 KEGGs were found as differentially abundant between animals with different methane production levels. Genera associated with high methane emissions were eukaryotes, mostly ciliates and fungi, while those associated with low emissions were bacterial clades, majorly with a role in soluble sugar fermentation. Genes participating in methanogenesis were not over-abundant in high methane emitters, but a deeper analysis revealed that genes related to nitrate-based anaerobic respiration pathways were more abundant in low methane emitters. This study offers a comprehensive overview of rumen microbial populations using long reads and a large cattle cohort, highlighting the role of eukaryotes in methane production through the presence of endosymbiotic methanogenic Archaea, as well as the contribution of microbes capable of utilize nitrate in reducing methanogenesis through competitive inhibition.

The research done in this thesis evinces the relationship between gut microbiome and host characteristics, in the context of animal production, using different sequencing and classification approaches. The obtained results provide evidence of the high complexity of microbiomes, thus emphasizing the necessity of accounting as much biodiversity as possible when performing microbiome studies, and shed light on the relationship between specific groups of gut microorganisms and host phenotypic traits of high economic importance, opening a new door to face current challenges in livestock industry.

Chapter 1: Background

1.1. First landmarks in microbial research

Microbiota and microbiome terms are usually employed indistinctly, although they were originally used for different purposes. Current predominant definition of microbiome was stated by **Whipps *et al.* (1988)** as "the characteristic microbial community occupying a reasonably well-defined habitat which has distinct physio-chemical properties", referring not only to the microorganisms, but also to their interactions and functions inside the ecosystem. As for the term microbiota, it refers to the set of microorganisms themselves, and it is an older concept which has been used at least since the 1960s, when germ-free animal models started to be incorporated in laboratory experiments **(Lane-Petter, 1962)**.

The first studies of microbiota date from the seventeenth century, when Robert Hooke and Antonie van Leeuwenhoek observed the first microorganisms **(Gest, 2004)**. Indeed, van Leeuwenhoek observations of *animalcules* living in his own teeth led to the first description of human microbiome. But it wasn't until the late 1800s that the concept of microbiota started to flourish. The manuscript *A Flora and Fauna within Living Animals*, published by the palaeontologist and parasitologist Joseph Leidy in 1853 on the gut parasites, is considered as the first scientific research in microbiota. Later, a series of experiments stated that microorganisms interact between them and with their host **(Pariente, 2019)**. The discoveries made can be accounted in three main research fields: (1) the involvement of microorganisms in chemical processes; (2) the relationship between pathogens and illness; and (3) the existence of microbes linked to healthy humans.

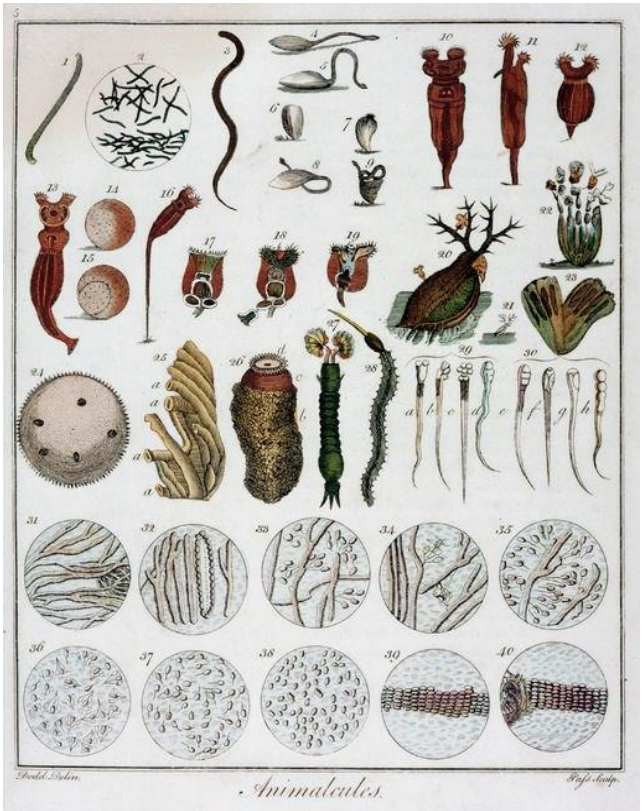

Figure 1: *Animalcules of Antonie van Leeuwenhoek (unknown artist)*

The participation of microbes in chemical processes was firstly stated by Charles Cagniard-Latour, Friedrich T. Kützing and Theodor Schwann, who identified yeasts as living organisms and suggested their role in alcoholic fermentation through conversion of sugar in alcohol and carbon dioxide, in the absence of oxygen **(Barnett, 1998)**. But it was in 1860 when two parallel publications from J. H. van

den Broek and Louis Pasteur associated alcoholic fermentation with yeast growth, also stating that the conversion process involves an intake of matter and thereby considering alcoholic fermentation as a physiological process. Pasteur was also the coiner of the terms aerobic and anaerobic, after describing the differences in fermentation under conditions of oxygen presence and absence **(Barnett, 2000)**. Other two important findings were made in the microbial ecology field by Sergei Winogradsky and Lorenz Hiltner. Winogradsky uncovered the interactions between aerobic and anaerobic microorganisms in the nitrogen fixation process, revealing the existence of ecological communities with different niches occupied by different species **(Winogradsky, 1890)**. Besides, Hiltner referred to Winogradsky's research to coin the term "rhizosphere" in 1904, arguing that plant root environment favours the development of bacterial communities which in turn impact plant nutrition **(Hartmann *et al.*, 2008)**.

Agostino Bassi was the first one to prove the link between a microorganism and a disease. In 1835, he identified a vegetable parasite – currently known as the fungus *Beauveria bassiana* – as the cause of the muscardine disease in silkworms **(Porter, 1973)**. Some years later, the surgeon John Goodsir discovered a gastrointestinal bacteria, *Sarcina ventriculi*, in one of his patients, and suggested the association between the bacteria and the patient's condition, providing one of the first descriptions of gut microorganisms **(Farré-Maduell & Casals-Pascual, 2019)**. But it was Louis Pasteur who formally established the pathogen-disease relationship with his research on *pébrine* and *flacherie* silkworm diseases **(Pasteur, 1870)**, that led to the consolidation of the Germ Theory of Disease, the first great contribution to understand the host-microbe interactions and the currently accepted scientific theory for infectious disease functioning **(Pariente, 2019)**. Pasteur also established that microbes are responsible of the putrefaction, which, joined to his experiments on product spoilage in silk and beverage industries, led him to introduce and patent the heating-sterilization process known as pasteurization. These advances, together with the isolation of the bacteria responsible for cholera by Pasteur and for anthrax and tuberculosis by Robert Koch, culminated in the publication of the Koch's postulates in 1890, that were key to establish a causative relationship between the presence of a microbe and a disease **(Cohen, 2017)**.

Lastly, a series of events evidenced the presence of bacteria in healthy humans' gut and faeces, as well as their link to health. The most important contributors in this scientific research field were Theodor Escherich and Henry Tissier. Escherich discovered in 1885 the so-called *Bacterium coli commune* (later named as *Escherichia coli*) and its role in food digestion, and also presented a comprehensive microbiological study of the faeces of children, which makes up one of the first characterizations of gut microbiota transition with age. Likewise, Tissier applied the anaerobic culture technique to babies'

stool, observing in 1899 the *Bacillus bifidus communis* (i.e., *Bifidobacterium*). Later research led Tissier to conclude that some gut bacteria (*Enterococcus*, *B. acidiparalactici* and others) were beneficial for babies affected by diarrhoea and he used them as the first probiotics, initially in babies and later in adult patients **(Farré-Maduell & Casals-Pascual, 2019)**.

Figure 2: *From left to right: Sergei Winogradsky, Robert Koch, Louis Pasteur and Theodor Escherich.*

Following these findings, pure culture experiments were carried out during the 20th century, which led to important advances in the comprehension of microbiota. Hungate's anaerobe culture tube allowed the culture of *Clostridium cellobioparum* from bovine rumen in 1944. In subsequent years, this method was enhanced and adapted for large-scale culturing or for methanogen isolation **(Hungate & Macy, 1973)**. On the other hand, the development of germ-free animals in 1960s **(Lane-Petter, 1962)** and the first faecal microbiota transplantations (FMT) (1958) allowed to conduct the first *in vivo* experiments – although FMT is documented to be used in Chinese medicine from the 4th century to treat poisoning and diarrhoea **(Valiquette & Laupland, 2013)**. In fact, germ-free animals have been, and still are, an essential instrument for studying new associations between microbiota and host phenotype.

But the greatest boost to microbiota research has been the development of DNA sequencing techniques, which led to the uncovering of the enormous biodiversity and complexity of microbial ecosystems and to the popularization of the holobiont concept, referring to the integrated and co-dependent organism composed by microorganisms and their eukaryotic host **(Baedke *et al.*, 2020)**.

1.2. DNA sequencing landmarks

Sequencing technology has experienced a rapid evolution over the last 50 years, from the first-generation methods developed at late 1970s to the long-read sequencing currently under development.

1.2.1. First-generation sequencing

With the first rudimentary sequencing techniques, Carl Woese and George Fox carried out multiple phylogenetic studies using ribosomal RNA sequences (16S-18S). These experiments allowed them to define the actual classification of the three domains of life, *Bacteria*, *Archaea* and *Eukaryota*, in 1977 **(Woese & Fox, 1977; Fox *et al.*, 1980)**. However, this classification faced a strong rejection by the scientific community until some years later.

That same year, two new sequencing methods came to light. The Maxam-Gilbert method used a direct approach of DNA radioactive marking and cleavage **(Maxam & Gilbert, 1977)**, becoming rapidly popular, as it did not require fragment cloning, but also rapidly abandoned with the emergence of new technologies. The second technique was developed by Frederik Sanger (**Figure 3**) and consisted in a primer-extension approach adding dideoxynucleotides **(Sanger *et al.*, 1977)**, which in fact involved sequencing fragmented clones later assembled to recover the original sequence.

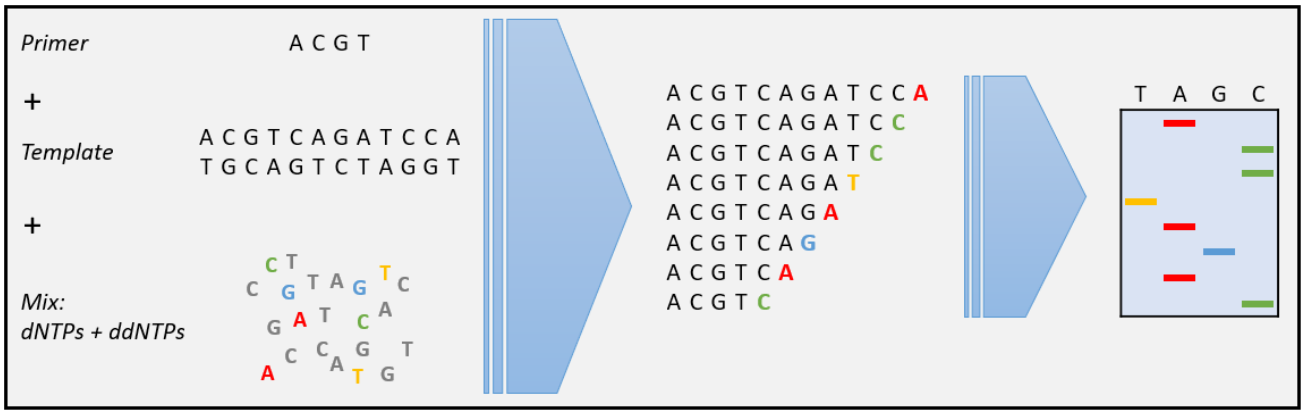

Figure 3: *Sanger's sequencing technique by primer extension with dideoxynucleotides.*

The Sanger method would become predominant over the next two decades, being gradually enhanced as new techniques such as PCR or capillary electrophoresis were developed. Sanger sequencing promoted the discovery of important applications like the perfectioning of 16S rRNA sequencing technique **(Lane *et al.*, 1985)**, the detection of SNPs (*single nucleotide polymorphisms*) or the analysis of SSCP (*single-stranded conformational polymorphisms*) and STR (*short tandem repeats*).

1.2.2. Second-generation sequencing (NGS)

Parallel to the modernization and automation of Sanger sequencing, new methods were developed during the 1990s-2000s. These new technologies, which are currently known as "Next Generation Sequencing (NGS) Technologies" or "High Throughput Sequencing Technologies", allowed the massive parallel production of high-quality cost-efficient DNA sequences, resulting in a more sensitive, more precise, faster and cheaper way to sequence genomes compared to first-generation techniques. NGS technologies emerged alongside bioinformatic advances that facilitated the processing and analysis of the increasing amount of data.

Two main technical advantages place NGS above classic Sanger sequencing: (1) massive parallel sequencing and (2) *in vitro* creation of DNA libraries.

1) **Massive parallel sequencing**. All NGS techniques are based on fixation of DNA fragments to a surface, fragment amplification through PCR and simultaneous sequencing and detection of DNA. The main reason for NGS to be able to sequence large amounts of DNA in less time and with higher resolution is the process following the amplification, which produces the sequencing data. While Sanger method produces fragments that are then separated and individually size-detected to produce the final sequence, NGS pass through a single process of nucleotide addition, detection of incorporated nucleotides and washing to remove unwanted signals, this stepwise process being simultaneously performed in a huge amount of spots **(Mardis, 2013)**.

2) **Creation of DNA libraries**. Template preparation in first-generation Sanger sequencing was usually made through *in vivo* clonal amplification by means of bacterial mechanisms. On the contrary, NGS uses *in vitro* clonal amplification of fragments without employing bacterial vectors, allowing to process larger genomes and metagenomes. This can be considered a minor advantage of NGS, since modern Sanger also works with *in vitro* libraries, either with specific or non-specific primers (targeted or untargeted, respectively).

Although all NGS methods rely on the same principles, multiple technologies emerged, differing among them both in the preparation of the template libraries and in the amplification methodology (**Table 1**).

NGS Platform	Company	Template preparation	Chemistry
MiSeq/HiSeq	Illumina	Bridge Amplification	Reversible Dye Terminator
454	Roche	emPCR	Pyrosequencing
Ion Torrent	Life Technologies	emPCR	Proton detection
SOLiD	Life Technologies	emPCR	Oligo Chained Ligation
Revolocity	Complete Genomics	RCA	Oligo Unchained Ligation

Table 1: *A brief summary of the most important NGS platforms and the technologies they use. Data withdrawn from* ***Buermans & den Dunnen (2014)****.*

Clonal amplification methods

As mentioned, NGS technologies use clonal amplification from single DNA molecules to prepare templates. The most common methods are explained hereunder:

A. ***Emulsion PCR (emPCR)***. This amplification method is based on embedding DNA templates in water-in-oil emulsions containing silica beads and all the PCR reagents (**Figure 4a**). The beads carry complementary adapter sequences, thus acting as a solid attach surface. Each emulsion droplet, ideally enclosing a single bead attached to a unique template, acts as a PCR micro-reactor, cloning each template and improving the signal detection **(Nakano *et al.*, 2003; Kanagal-Shamanna, 2016)**.

B. ***Bridge amplification***. In this method, the surface is heavily coated by adapters and their complementaries. The single-stranded templates attach to the surface at one end and hybridise their free end to complementary adapters in the surface, which act as primers for amplification (**Figure 4b**). The double-stranded DNA bridges are then denatured and, after several cycles, clusters of the same DNA fragment (*polonies*) are created **(Ansorge, 2009)**. This amplification method is only used by Illumina platforms.

C. ***Rolling circle amplification (RCA)***. This method emulates the natural circular DNA replication process used by bacteria or viruses. A DNA fragment is circularized by enzymatic ligation and a linear amplification from a single primer produces a long single-stranded DNA concatemer, which will be visualized through fluorescent detection **(Ali *et al.*, 2014)**. RCA is commonly used in immunoassays or for rapid detection of pathogens, but not in microbiota studies. It is also the amplification method used by Complete Genomics' whole genome sequencing systems, which modify the RCA products to DNA Nanoballs (DNBs).

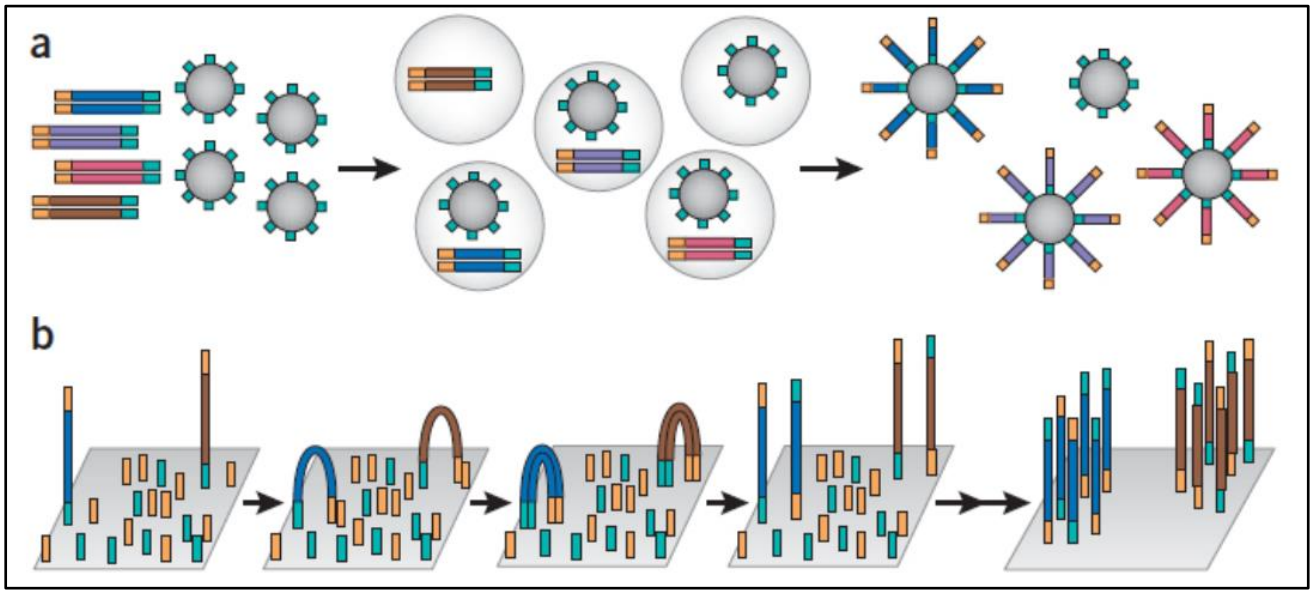

Figure 4: *Scheme of two NGS clonal amplification methods: emPCR (a) and bridge amplification (b). Figure from* ***Leong et al. (2014)****.*

Sequencing chemistry methods

According to the sequencing and detection chemistry, two approaches are predominant: (1) sequence by synthesis (SBS) and (2) sequence by ligation (SBL). The main difference between both is that SBS adds one base at a time, using DNA polymerase to elongate the template's complementary strand, while SBL adds multiple bases through pre-designed probes and anneals them using a DNA ligase. The methods used by the principal platforms are described hereafter:

A. ***Reversible dye termination***. In this SBS method, nucleotides used by the DNA polymerase are labelled with a fluorescent molecule and blocked in 3', which prevents additional base incorporation. Thus, each cycle includes four steps: (1) addition of one nucleotide per template, (2) washing of unattached nucleotides, (3) imaging of fluorescence and (4) cleavage of fluorescent group and deblocking of 3' ends **(Mardis, 2013)**. It is used in Illumina platforms.

B. ***Proton detection***. This technology is used by the Ion Torrent system, and it is an SBS method based on the detection of pH changes occurring when a nucleotide is incorporated into the DNA. After clonal amplification by emPCR, enriched beads carrying amplified templates are each deposited into a well of an Ion chip. The chip is then cyclically flooded with a solution with one of the four nucleotides. Whenever a nucleotide is incorporated into the template, protons are released, altering the pH of the solution in the well. The Ion chip detects this change and reports it as a voltage variation **(Mardis, 2013)**.

C. ***Pyrosequencing***. An SBS approach used by Roche 454 systems whose process begins similarly to Ion Torrent sequencing, as enriched beads are deposited in a plate to separate the reactions. However, it then uses chemoluminescent detection of released pyrophosphate during the elongation of the fragment's complementary strand, through a luciferase-mediated reaction **(Buermans & den Dunnen, 2014)**. This technology is currently obsolete, as other platforms are more cost-efficient.

D. ***Oligonucleotide ligation***. This method is based on SBL approach, using oligo probes in each elongation cycle. Two variations of this method are used by different platforms.

a) ***Chained ligation***. Used by SOLiD sequencing. Probes are pre-designed octamers consisting of – from 3' to 5'– 2 probe-specific bases, 6 degenerated bases and a fluorescent label. The pair of probe-specific bases is one of the 16 possible combinations of DNA bases, and each combination is associated with one of four different fluorescent labels (**Figure 5A**). The probe anneals to the former one (or the primer) if these two initial bases are complementary to the template. After annealing through the ligase, fluorescence is recorded and the three last degenerated bases are cleaved, leaving space for the next probe. This extension product is denatured from the template and a new elongation cycle starts with the primer n-1, to a total of 7 cycles **(Voelkerding *et al.*, 2009)**.

b) ***Unchained ligation***. Complete Genomics' cPAL chemistry (combinatorial Probe-Anchor Ligation) uses an unchained ligation system in which a set of nonameric probes degenerated in all but one position, and labelled with fluorescence according to the non-degenerated base (**Figure 5B**), are ligated to an anchor complementary to the adaptors in nanoballs. Each cycle ligates one random probe and washes away the anchor-probe complex, so there is no chaining of probes **(Blankenship, 2017)**.

In summary, NGS technologies have prevailed over the last thirty years, and some of these platforms are still used for multiple applications. In microbiome research, Illumina platforms have become the gold standard, as they are capable to generate up to 25 million high-quality reads from 25 to 300 base pairs length with MiSeq system – or 1.5 billion from 35 to 100 bp with HiSeq system – at a relatively low cost, with the possibility of multiplexing up to 96 samples.

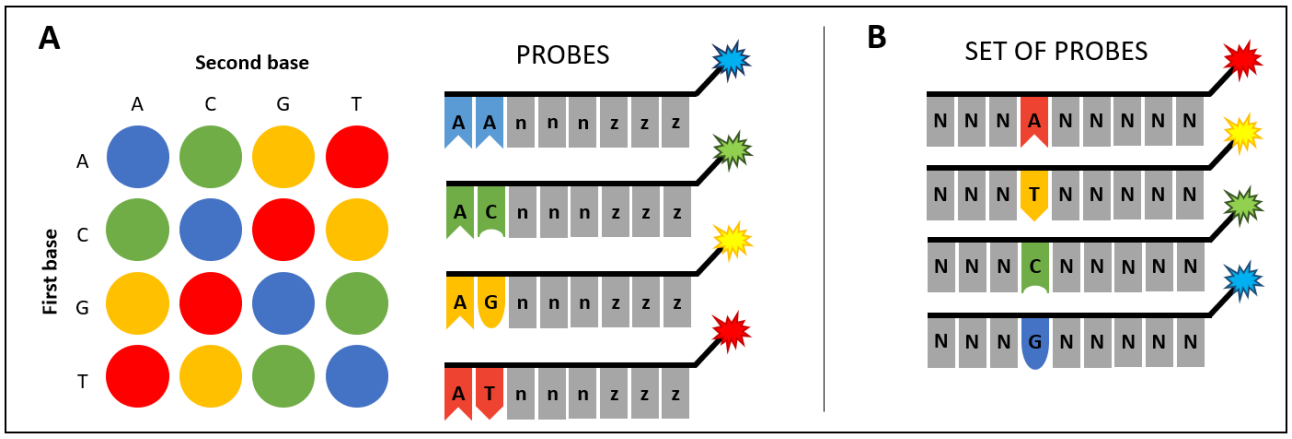

Figure 5: *Probes used in SBL technologies.* ***A****: chained ligation probes (Life Technologies' SOLiD platform);* ***B****: unchained ligation probes (Complete Genomics' cPAL approach).*

1.2.3. Third-generation sequencing

Third-generation sequencing (TGS) or "Long-Read Sequencing" technologies are the current research field seeking to improve output from NGS techniques. There is not a consensus about the exact time at which NGS transitioned to TGS, as the early 2010s were a time of continuous technology improvement. Some authors consider the development of Ion Torrent technology as a turning point between NGS and TGS, as it allowed to speed up the sequencing runs by getting rid of the imaging processes to register the progress, although still possessing the characteristics of second-generation platforms **(Schadt *et al.*, 2010)**. Another relevant change that opened the doors to third generation was the introduction of technologies capable of working with single DNA templates, being the Helicos sequencing platform the first one developed **(Thompson & Steinmann, 2010)**.

Summarizing, TGS technologies have five advantages over NGS: (1) no need for PCR, as these systems can work with single DNA molecules, avoiding the bias generated by template clonal amplification; (2) generation of longer reads, which might facilitate assembly of genomes and reduce problems with repeated regions; (3) sequencing speed boost, as no pre-processing is required and the lesser DNA fragmentation allows real-time analysis; (4) higher precision, allowing rare variant detection; and (5) low cost, a human genome can now be whole-sequenced for less than $1000.

There are three main platforms considered as TGS, i.e., functioning without template amplification by PCR and outputting long reads (**Table 2**).

A. ***Helicos Single Molecule Sequencing (SMS)***. This technology was developed by the now bankrupted Helicos Biosciences. As mentioned before, it is considered as the bridge between NGS and TGS, as it was the first system working with single templates. It also uses SBS by

reversible dye termination, with some differences with regard to Illumina platforms. While in Illumina all four nucleotides are present in each extension cycle, as each nucleotide is labelled with a different fluorescent group, Helicos system presents only one type of nucleotide each cycle, and only template fragments having the complementary bases throw fluorescence **(Thompson & Steinmann, 2010)**. However, it is currently out-to-date due to its low speed, short read length and high error rate due to noise, coupled with the overcoming of better technologies.

B. ***PacBio Single Molecule Real-Time (SMRT)***. This system uses a metallic film containing thousands of nanostructures light-focusing zero-mode waveguides (ZMW), capable of detecting fluorescence to the single-molecule level at the incorporation of each labelled nucleotide by the DNA polymerase at the bottom of each ZMW. The templates, so-called SMRTbells, are circular single-stranded DNA created by ligation of hairpin adapters to the ends of double-stranded DNA molecules generated after the fragmentation of the original DNA **(Ardui *et al.*, 2018)**. This technology produces reads 100 times longer than NGS platforms – average of 8-15 kb, but can reach 70 kb – also being time-effective and highly resolutive.

C. ***ONT nanopore sequencing***. In this sequencing method, long fragments of single-stranded DNA directly pass through a protein nanopore attached to an electrically resistant polymer membrane. Ionic current changes occur when a nucleotide passes through the pore, and sensors detect and register these voltage changes **(Kono & Arakawa, 2019)**. This method presents important advantages: (1) high speed, thanks to both real-time sequencing and parallelization, as each flow cell contains thousands of nanopores; (2) high device miniaturization, as no imaging equipment is required (MinION device); (3) sequencing of very long reads with no theoretical length limit (reads of 882 kb length have been reported **(Jain *et al.*, 2018)**), which, in turn, makes possible library preparation without excessive fragmentation nor size selection ("whale watching" protocol); and (4) sequencing is not limited to the four conventional nucleotides (A, T, G and C), as it detects ion current changes and not fluorescence, making it ideal for detecting DNA

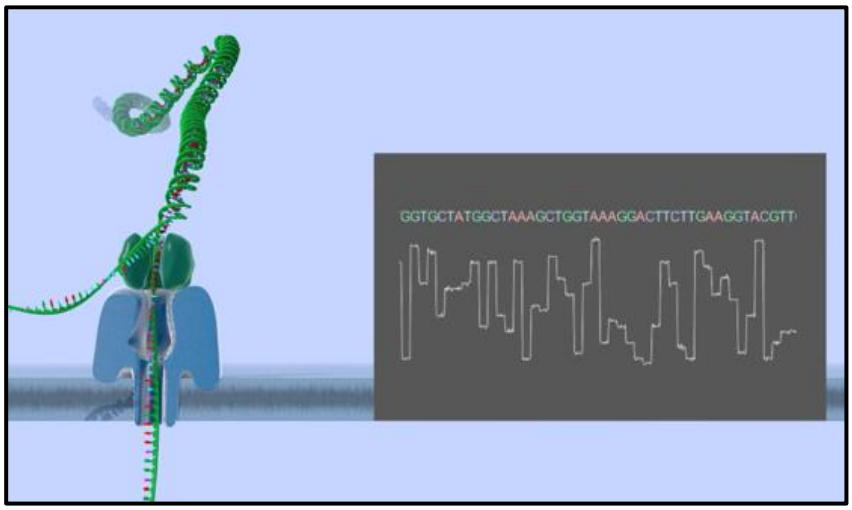

Figure 6: *Oxford Nanopore sequencing system (https://nanoporetech.com/)*

modifications or for RNA sequencing. This successful technology marketed by Oxford Nanopore Technologies (ONT) is broadly used for multiple applications, as the portability of its platforms and the easiness of library preparation allows low-cost and high-speed *in situ* sequencing.

TGS technology	Company	Chemistry
Helicos SMS	Helicos Biosciences	Reversible Dye Terminator
PacBio SMRT	Pacific Biosciences	ZMW Fluorescent Detection
ONT nanopore sequencing	Oxford Nanopore Technologies	Protein nanopores

Table 2: *A brief summary of the most important TGS technologies.*

TGS technologies have marked a new milestone in sequencing performance and resolution, opening new ways to explore genetic information at multiple levels, such as whole genome sequencing, transcriptomics, metagenomics or epigenomics. These technologies have driven the creation of huge global consortiums trying to sequence all the metagenomes, such as the TerraGenome project **(Vogel *et al.*, 2009)** for sequencing soil metagenome, the Integrative Human Microbiome Project (iHMP) **(The Integrative HMP Research Network Consortium, 2014)** for human metagenome, or the Earth Microbiome Project (EMP) **(Thompson *et al.*, 2017)** for characterizing all microbial life on Earth.

However, TGS still have an important drawback, as all SMS technologies possess an increased sequencing error rate **(Laehnemann *et al.*, 2016)**. While sequencing error rates from Illumina platforms go from 0.087% to 0.613% **(Stoler & Nekrutenko, 2021)**, raw sequence error rate in SMRT goes from 13 to 15%, although randomness of error and high-fold coverage due to repetitive sequencing allow to find consensus sequences with an accuracy higher to 99% **(Ardui *et al.*, 2018)**. On the other hand, Nanopore sequencing has a similar error rate (5-20%), also including indels, which makes this technology less suitable for single variant detection, unless high coverage is achieved, or for allele distinction **(Jain *et al.*, 2018; Ebler *et al.*, 2019)**. Future research might solve this problem by improving raw sequencing and/or consensus assembly.

1.3. DNA sequencing in microbiota studies

Microbial research has benefitted a lot from sequencing advances. As stated, first-generation sequencing opened the door for unveiling the basic composition of microbial environments, such as the first microorganism classifications using 16S rRNA genes. But the real boost occurred in the NGS era. During this period, microbial research experimented an impressive progress and a high number of microbiota characterization studies were carried out (**Figure 7**), not only in host-associated communities, but also in food and environmental samples.

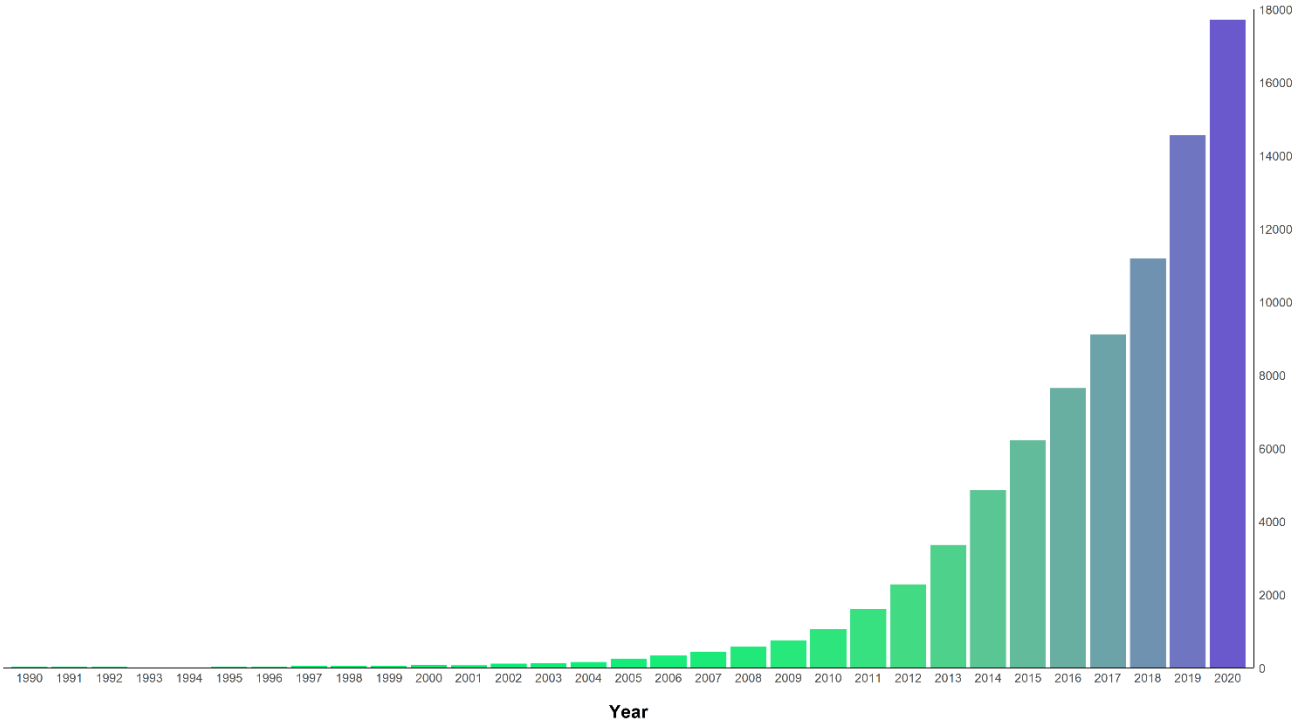

Figure 7: *Number of published papers related to microbiota, from 1990 to 2020. Source: PubMed.gov (https://pubmed.ncbi.nlm.nih.gov/), accessed 2021-10-07.*

First-generation sequencing limited the information from microbiomes, as only cultured microbes could be quantified. But this new era allowed the exploration of new microbial diversity, as well as a more precise classification of formerly known ones. NGS era established the gold standard protocols for exploring and characterizing microbiota composition. Basically, two approaches have been popularized: amplicon sequencing and whole metagenome sequencing.

1.3.1. Amplicon sequencing

Amplicon sequencing (AS) has been broadly used since the era of classical sequencing to study the phylogenetic relationships between different species. This method requires the cloning/amplification and sequencing of specific genes fulfilling a series of conditions: (1) being highly conserved in a huge

group of organisms, to allow the construction of universal primers; (2) possessing highly variable regions between individual species to make identification possible; and (3) being in a large number of copies, for an easier isolation.

The first gene used for this purpose was the 16S rRNA, which is highly conserved in both *Bacteria* and *Archaea* domains. This gene encodes the 16S ribosomal RNA, a component of the small subunit of the prokaryotic ribosome (30S subunit), which has nine hypervariable regions (V1 to V9) separated between them by highly conserved regions (**Figure 8**), making it ideal for studying prokaryotic phylogenetics **(Woese & Fox, 1977; Fox *et al.*, 1980; Neefs *et al.*, 1993)**. Later, its utilization in microbiota studies became popular as a response to the economic and time expensiveness of massive sequencing of complete genomes during the early-NGS era, as amplifying and sequencing only one broadly-represented and highly-variable gene was cheaper and faster, both for sequencing and computational processing, outputting sufficiently accurate results.

However, AS has some disadvantages. First of all, 16S rRNA does not cover all the microbiota biodiversity, as eukaryotic organisms do not possess this gene and, although it is present in bacteria and archaea, universal primers completely encompassing both domains do not exist. For this reason, *Archaea* and *Bacteria* domains must be characterized using different primers, while protozoan or fungi communities must be studied using alternative genes, such as 18S rRNA in protozoa or ITS regions in fungi. The need for multiple amplifications to account the total microbial variability involves an additional bias due to PCR **(Bonk *et al.*, 2018)** and hampers the joint

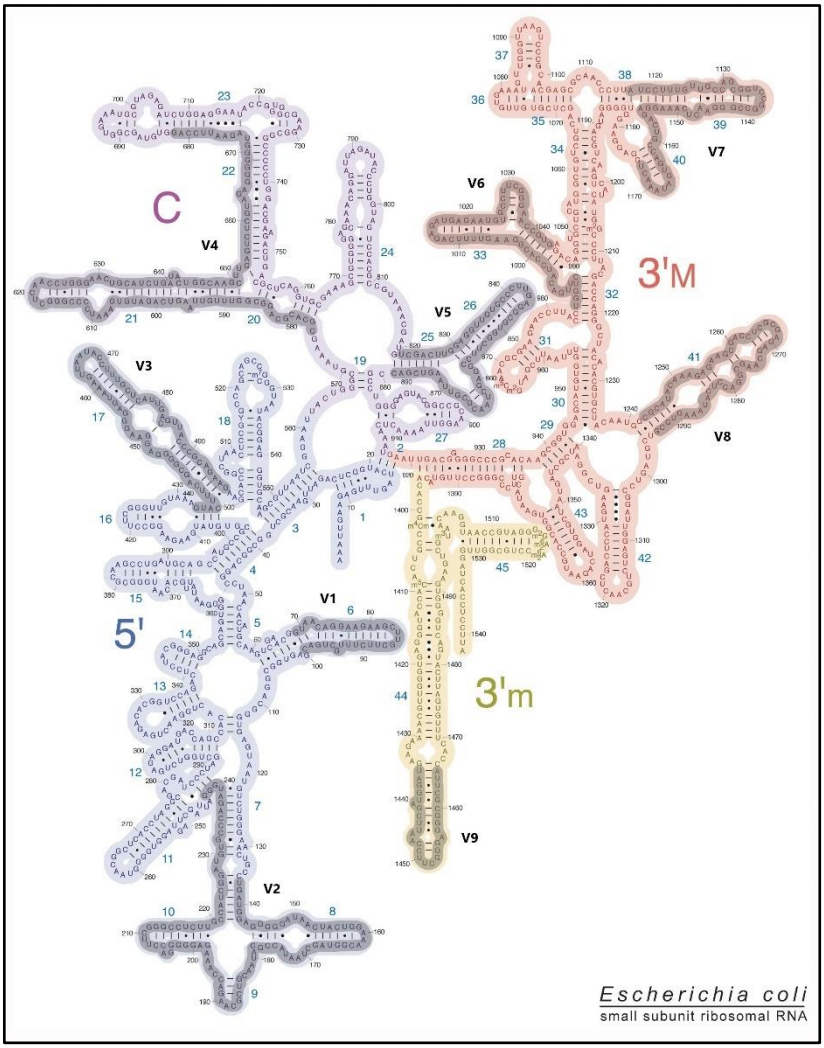

Figure 8: *Secondary structure of 16S rRNA of Escherichia coli. Hypervariable regions are marked in grey. Adapted from http://rna.ucsc.edu/rnacenter.*

analysis of all domains. On the other hand, amplicon analysis can reveal microorganism abundance, but not gene functionality, as only one gene is being considered.

Other genes used for amplicon sequencing

As mentioned above, the exploration of eukaryotic biodiversity requires the amplification of other highly-conserved genes different to 16S.

A. 18S rRNA is the homologue version of 16S in eukaryotes, encoding the RNA component of the eukaryotic ribosome small subunit (40S). Its structure is similar to the 16S rRNA gene, containing eight hypervariable regions (V1-V5 and V7-V9), as the region corresponding to V6 is more conserved among eukaryotes **(Neefs *et al.,* 1993)**. This gene is used for characterization of protozoan communities in microbiota studies, being the regions V2, V4 and V9 the ones with the highest taxonomic resolution **(Hadziavdic *et al.,* 2014)**.

B. Although hypervariable regions of SSU (small sub-unit) rRNA genes are suitable to identify organisms for all the domains, fungal organisms possess less hypervariable regions in the 18S gene, so internal transcribed spacers (ITS) were proposed as an alternative. These are spacer DNA sequences located between SSU and LSU (large sub-unit) rRNA genes, ITS1 between 18S and 5.8S genes and ITS2 between 5.8S and 28S genes. These regions have been proposed as DNA barcodes for fungi due to its higher potential for species discrimination and their better yield in PCR amplification compared to 18S hypervariable regions **(Schoch *et al.,* 2012)**.

An overview of the amplicon sequencing post-processing pipelines

AS data have the advantage of being relatively easy to process. Quality-curated reads can be directly mapped against a reference database, as no assembly is required, and mapping results only need to be quantified to obtain relative abundance values. Still, a surprising amount of different software have been developed in order to optimize each processing step (i.e., data curation, alignment/clustering, and taxonomic classification):

Data curation. Quality control is a key step in every data analysis procedure. In AS context, it consists of an initial standard trimming step, in which adapters, primers and other unwanted artifacts are removed from sequences, and a denoising step, at which sequencing and PCR errors are dealt with.

1) ***Trimming***. Adapters and low-quality fragments are chopped from the sequences using specialized tools such as Cutadapt **(Martin, 2011)**, Trimmommatic **(Bolger *et al.,* 2014)** or PRINSEQ **(Schmieder & Edwards, 2011)**.

2) ***Denoising***. Sequencing errors are present in every platform, although different technologies differ in their accuracy **(Laehnemann *et al.*, 2016)**. Getting rid of ambiguous sequences is crucial, as mistakes in clustering and/or taxonomy assignment may lead to incorrect estimates of microbial composition. Multiple sequencing error correction tools have been developed – compendium available at **Laehnemann *et al.* (2016)** – and some pipelines implement their own correction strategy. PCR errors are common as well, and usually lead to the formation of artifact sequences known as chimeras. This occurs by an incorrect joining of two or more sequences, due to either incompletely extended primers acting as original primers in the last PCR cycles, or polymerase template-switching **(Kanagawa, 2003)**. It is important to locate and filter these contaminant sequences, as they might be incorrectly interpreted as new species and contribute to overestimate microbial diversity. The first programs performed multiple alignment against a reference database for detecting chimeras, but recent methods are reference-free and use error modelling to infer taxonomically-accurate putative sequences (**Table 3**).

Software	Approach	Reference
Bellerophon	Reference-based	**Huber *et al.* (2004)**
ChimeraChecker	Reference-based	**Nilsson *et al.* (2010)**
ChimeraSlayer	Reference-based	**Haas *et al.* (2011)**
Perseus	Reference-based	**Quince *et al.* (2011)**
UCHIME	Reference-based	**Edgar *et al.* (2011)**
MED	Sequence inference	**Eren *et al.* (2015)**
Deblur	Sequence inference	**Amir *et al.* (2017)**
DADA2	Sequence inference	**Callahan *et al.* (2016)**
UNOISE2	Sequence inference	**Edgar (2016)**

Table 3: *Summary of the most used chimera-detection tools.*

Sequence alignment and clustering. This process performs automatic classification of sequences according to their similarity, either among them or with a reference database. Clustering methods can be classified according to two criteria: (1) the clustering unit; and (2) the picking strategy.

1) ***Clustering units***. Traditionally, sequence clustering is done into operational taxonomic units (OTUs), which are groups of similar sequences above a pre-defined threshold (usually, 97% similarity). A broad variety of clustering algorithms have been developed for OTU clustering, either using greedy heuristics or hierarchical clustering, or even model-based methods. A list of the most used ones is presented in **Table 4**. But the recent development of new and powerful algorithms has redefined the clustering paradigm, tending to create less arbitrary groups called

amplicon sequence variants (ASVs), which can differ in only one base pair **(Callahan *et al.*, 2017)**. Although ASVs provide more taxonomic resolution, they are more sensitive to data quality. Thus, the utilization of OTUs is still a valid approach for studies not targeting high-level resolution.

2) ***Picking strategy***. According to the dependence on an external database, different clustering methods can be distinguished: (1) *de novo* approach matches sequences within the dataset itself; (2) closed-reference approach uses an external reference database to cluster the reads, ignoring those without a hit in the database; and (3) open-reference approach combines the other, by first matching known sequences to a database and then performing *de novo* clustering within the set of remaining reads **(Bhat *et al.*, 2019)**. Generally speaking, database-based approaches are considered less optimal, as sequence similarity is calculated as read-vs-database but not read-vs-read, so a 97% OTU might contain reads with 95% similarity between them. Additionally, clustering results are dependent on the database completeness, i.e., the prior knowledge of the microbes in the studied environment. Open-reference clustering fixes the problem of unknown reads, also being faster than *de novo* approach, but mixes reference OTUs (i.e., clusters of < 97% similarity) with *de novo* OTUs (≥ 97% similarity) **(Westcott & Schloss, 2015)**. On the other hand, *de novo* OTU picking has been reported to lack OTU stability, being sensitive to the input order of the sequences **(He *et al.*, 2015)**, but still outperforms the other approaches. Picking strategy performance is also influenced by both OTU clustering (**Table 4**) and sequence alignment algorithms (**Table 5**), as well as by the chosen reference database.

Algorithm	Approach	Reference method	Pipelines	Reference
Cd-hit	Greedy heuristic	*de novo*	QIIME, Mothur, MG-RAST	**Li & Godzik (2006)**
UCLUST	Greedy heuristic	open, closed, *de novo*	QIIME	**Edgar (2010)**
VSEARCH	Greedy heuristic	open, closed, *de novo*	QIIME, Mothur	**Rognes *et al.* (2016)**
OTUCLUST	Greedy heuristic	*de novo*	MICCA	**Albanese *et al.* (2015)**
OptiClust	Greedy heuristic	*de novo*	Mothur	**Westcott & Schloss (2017)**
ESPRIT	Hierarchical	*de novo*	N/A	**Sun *et al.* (2009)**
NN / AN / FN	Hierarchical	closed, *de novo*	Mothur	**Schloss *et al.* (2009)**
Swarm	Model-based	*de novo*	QIIME	**Mahé *et al.* (2014)**
CROP	Model-based	*de novo*	N/A	**Hao *et al.* (2011)**
Deblur	ASV-based	*de novo*	QIIME2	**Amir *et al.* (2017)**
DADA2	ASV-based	*de novo*	QIIME2	**Callahan *et al.* (2016)**

Table 4: *Summary of the most used OTU clustering algorithms. Adapted from* ***Bhat et al. (2019)***.

Algorithm	Approach	Pipelines	Reference
Needleman-Wunsch	Global alignment	Mothur	**Needleman & Wunsch (1970)**
Suffix tree	Local alignment	Mothur	**Weiner (1973)**
BLAST	Local alignment	QIIME, Mothur, MG-RAST	**Altschul *et al.* (1990)**
MUSCLE	Multiple alignment	QIIME, MICCA	**Edgar (2004)**
UBLAST	Local alignment	N/A	**Edgar (2010)**
USEARCH	Global alignment	Mothur	**Edgar (2010)**
PyNAST	Multiple alignment	QIIME, MICCA	**Caporaso *et al.* (2010a)**
Clustal Omega	Multiple alignment	N/A	**Madeira *et al.* (2019)**

Table 5: *Summary of the most used algorithms for sequence alignment.*

Taxonomic classification. The next step attempts to assign the taxonomy of each OTU/ASV, using alignment programs or other classifiers against a reference database. Usually, a consensus taxonomy is found for each OTU/ASV based on the classification of their individual sequences. BLAST tools and RDP naïve Bayesian classifier **(Wang *et al.*, 2007)** are the most frequently used tools for AS taxonomy assignment. But the choice of the reference database is determinant for the final results. The most used available databases are summarised below.

Available ribosomal RNA reference databases

As stated before, sequence alignment, chimera detection and taxonomic assignment require, in most cases, a reference database containing known templates. For AS, four reference databases have become popular, each of them containing sequences from different sources. The majority of bioinformatic pipelines for AS processing provide compatible versions for these reference databases.

A. **GreenGenes (DeSantis *et al.*, 2006)**. This database holds annotated full-length 16S rRNA sequences from *Bacteria* and *Archaea* organisms. For its construction, high-throughput chimera detection with Bellerophon was performed, and *de novo* tree inference was used for taxonomy assignment. The latest version (gg_13_8_99) contains 202,421 sequences. Unfortunately, it has not been updated since August 2013.

B. **SILVA (Quast *et al.*, 2013)**. This is a more complete database, containing information from both rRNA sub-units – SSU-16S / LSU-23S for *Bacteria* and *Archaea*, and SSU-18S / LSU-28S for *Eukaryota*. In its latest version (Release 138.1), from August 2020, SILVA includes 510,508 non-redundant SSU and 95,286 LSU sequences, and adopts the Genome Taxonomy Database

(GTDB) for taxonomy assignment **(Parks *et al.*, 2018)**, also including the EMBL-EBI/ENA taxonomy for every sequence.

C. **UNITE (Kõljalg *et al.*, 2013)**. UNITE database exclusively includes information from fungal ITS regions. The current version (8.3) has been last updated in November 2020, and contains 2,688,805 ITS sequences, classified as 120,183 fungal species (considering the UNITE fungal Species Hypotheses with DOIs at 1.5% threshold).

D. **RDP (Cole *et al.*, 2014)**. The Ribosomal Database Project is a complete toolkit that includes a Bayesian classifier and a rRNA taxonomy database among others. The taxonomy database contains around 3 million aligned and annotated bacterial and archaea 16S sequences and 60-65 thousand fungal 28S sequences. It has been recently updated (training set 18, April 2020) to include 800 new genera and 4,000 new species.

Available pipeline toolkits

Due to the difficulty in implementing all the necessary steps for amplicon sequence processing, experts have developed several pipelines including the most adequate tools for each step, adding functionalities for a more efficient and ergonomic output. Although their final results are similar, differences in abundance and/or taxonomy assignment might be detected, as default tools and quality curation methods that each one employs are different.

A. **Mothur (Schloss *et al.*, 2009)**: This toolkit is an open-source single program in C++ language, developed as an expansion of DOTUR, an OTU-assignment software using nearest, average and furthest neighbour algorithms, and SONS, for pairwise alignment. It has progressively re-implemented the code from other specialized tools to become a complete pipeline, useful for processing amplicon data from Sanger, PacBio, IonTorrent and Illumina platforms. The main advantage of Mothur over other pipelines is its plasticity, allowing the user to change a wide variety of options in order to optimize the processing performance, depending on the type of inputted data.

B. **QIIME (Caporaso *et al.*, 2010b)**. Quantitative Insights Into Microbial Ecology is another popular open-source pipeline that connects multiple external bioinformatics protocols and tools, including Mothur applications, under a python-based interface. It is also capable of processing data from different platforms, offering a more user-friendly result visualization at the cost of less availability of customization options. Its default output format is a .biom table, which includes

OTU relative abundances and taxonomy. QIIME recently stopped being supported as it has been replaced by QIIME2 pipeline.

C. **QIIME2 (Bolyen *et al.*, 2019)**. This pipeline replaced original QIIME in 2019, offering a new way to curate and cluster rRNA sequences, based on ASVs classification, through DADA2 and Deblur approaches. QIIME2 output is less typical, as it generates .qza files called Artifacts which can be visualized in "QIIME2 View" application through transformation to .qzv, or be converted into typical .biom files or text tables and exported to other tools.

D. **MICCA (Albanese *et al.*, 2015)**. MICrobial Community Analysis is another open-source pipeline written in python and C, and also includes external software for AS processing. It operates similarly to QIIME, but includes less command options, being less extended in use than the other tools.

1.3.2. Whole metagenome sequencing

Cost reduction and speed improvement in sequencing technologies during the late-NGS and TGS eras fostered the appearance of whole metagenome sequencing (WMS). This approach consists in sequencing all the genetic material present in a microbiome sample, in order to obtain as much information as possible of microbial genomes. WMS technique goes on in three steps **(Quince *et al.*, 2017)**:

1) **DNA shotgun fragmentation**, generating random sequenceable fragments. The length of the fragments will be decisive for the accuracy of the results, as longer reads are expected to be better aligned or assembled.

2) **Sequencing of fragments**, using either NGS or TGS platforms. The most used platforms for WMS are Illumina MiSeq or Hiseq, PacBio SMRT and ONT nanopore sequencing. Although Illumina platforms are considered second-generation, they have been established as the gold standard for metagenome sequencing, due to its high accuracy. However, Illumina is currently being outperformed by TGS technologies, as their continuous refinement allows to obtain longer fragments with dwindling error rates.

3) **Sequence processing**. Two approaches are used when analysing WMS reads. Mapping-based analysis directly aligns them to a pangenomes database, then calculating the abundance of each species according to the number of hits over each species' gene. On the other hand, assembly-based analysis joins overlapping DNA fragments into contigs, in order to rebuild the original complete microbial genomes **(Quince *et al.*, 2017)**.

WMS technologies have important advantages over AS. They cover all the taxa variability without bias from the use of primers or from gene selection, thereby facilitating the evaluation of overall microbial composition and the study of the interactions among microbes. Additionally, whole genome information can be used to explore gene functionality and find active metabolic paths in the studied community. This approach has been successfully used to register new microbial species, contributing to expand knowledge about microbial diversity in multiple hosts and environments **(Almeida *et al.*, 2019; Pasolli *et al.*, 2019; Stewart *et al.*, 2019; Peña-Ocaña *et al.*, 2021)**.

However, this method generates a huge amount of data, and due to random DNA fragmentation, post-processing of raw sequences becomes complicated. Thus, WMS is slower and more expensive than AS, and more important, it requires more computational resources and specialized software to properly analyse the data. An additional constraint is related to the experimental design, as a mapping-based approach is better to find known organisms in a faster way, while contig assembly is mandatory for *de novo* characterization, at the cost of a higher memory consumption. Moreover, as stated before, TGS techniques still carry an important drawback related to sequencing error, therefore impeding the correct characterization of total biodiversity. Hence, even though WMS has opened a new window to microbial exploration, successfully contributing to deepen in the knowledge of microbiomes' diversity, it is crucial to address its current limitations.

An overview of assembly strategies

Computational resources have been a key factor for the evolution of metagenomic techniques. The constantly increasing production of sequencing data during NGS and TGS eras required more and more processing power and storing capacity, but more importantly, specialized software tools optimizing the data processing.

Assembly of overlapping reads is a key step in metagenomics data treatment, as final results rely on the alignment accuracy, but also because it is a high time and memory consuming process. As a consequence, optimizing the assemblers has been a major issue during the last decades. Four major assembly strategies have been developed **(Miller *et al.*, 2010; Ghurye *et al.*, 2016)**:

A. **First-generation assemblers**. The first assemblers were based on greedy algorithms, which merge reads with maximum overlapping into contigs, iterating until no more contigs or reads can be assembled. Several examples of greedy assemblers are available, such as Phrap (http://www.phrap.org/phredphrap/phrap.html), TIGR **(Sutton *et al.*, 1995)** or SSAKE **(Warren *et al.*, 2007)**.

B. **Second-generation assemblers**. These programs used the Overlap Layout Consensus (OLC) three-step approach. It first calculates pairwise overlaps between all reads and builds a graph with reads as nodes and overlaps as edges. Then, the layout stage tries to simplify the overlap graph to find a path corresponding to the real sequence. Finally, the consensus stage uses the layout information to generate a multiple alignment of the reads. There are several assemblers using this algorithm, such as Celera **(Myers *et al.*, 2000)**, ARACHNE **(Batzoglou *et al.*, 2002)**, Edena **(Hernandez *et al.*, 2008)**, for short reads, or Omega **(Haider *et al.*, 2014)**.

C. **Third-generation assemblers**. De Bruijn graphs (DBG) are based on the utilization of fixed k-length fragments, known as *k-mers*, which are generated from fragmentation of the reads. A graph structure is generated, using the *k-1-mers* prefixes and suffixes as nodes and the unions between *k-1-mers* as edges. The different reads are aligned based on the *k-mers* they share, and the final assembly corresponds to the path visiting each edge once, i.e., an Eulerian path. The most used DBG programs are SOAPdenovo2 **(Luo *et al.*, 2012)**, MEGAHIT **(Li *et al.*, 2015)** and metaSPAdes **(Nurk *et al.*, 2017)**.

D. **Fourth-generation assemblers**. The assembly of long reads from TGS technologies (PacBio and ONT) poses an additional challenge due to their high sequencing error rates. The necessity of new software adapted to long and noisy sequences has led to the creation of hybrid assemblers using modified versions of OLC algorithms joined to either greedy or DBG approaches.

a) **Canu**. This complex program uses an optimized Celera Assembler (CA) best overlap graph (BOG) algorithm. CABOG is a hybrid approach that identifies reads as anchored to others (i.e., likely to overlap) if they share enough *k-mers*, iteratively making pairwise alignments between anchored reads to determine overlaps. A multigraph is then constructed, with each pair of nodes representing each read's start and end, connected by undirected edges, and directed edges representing the best overlaps (i.e., overlaps covering most of the sequence). This greedy process is used to form unitigs (high-confidence contig seeds) which are then combined into contigs and scaffolds using the corresponding part of the Celera Assembler OLC pipeline **(Miller *et al.*, 2008)**. The difference between original BOG and the optimized version used by Canu is that, due to the high error rates of single-molecule sequencing, the new approach runs multiple rounds of read and overlap error correction before constructing the graph, thus including only overlaps below the calculated global error rate threshold **(Koren *et al.*, 2017)**.

b) **Flye**. As mentioned, an approach combining OLC and DBG can also be used for long reads, and assemblers such as ABruijn **(Lin *et al.*, 2016)** or HINGE **(Kamath *et al.*, 2017)**

have been developed under this strategy. Flye software goes one step further, trying to address the problem of repeats by using repeat graphs. These graphs are built from concatenates of arbitrary and error-prone read assemblies known as disjointigs. The repeat graphs are then solved by aligning the reads against it, first resolving the bridged repeats (i.e., those completely contained by at least one read) and then the unbridged repeats to output more accurate contigs **(Kolmogorov *et al.*, 2019)**.

An overview of non-assembly strategies

Even with the constant improvement of bioinformatic resources, assembly strategies can generate computational speed and memory problems, as well as certain tendency to create chimeric contigs or additional bias due to contig binning methods. Assembly-free approaches can be useful to explore core microbiota of short read sequencing data, if pangenomic databases are sufficiently complete (**Table 6**). Most of these direct-mapping methods are based on BLAST methods followed by a relatively simple algorithm known as lowest common ancestor algorithm (LCA), which assigns each read to the bottom taxonomic rank common to all hits over a specific identity score **(Huson *et al.*, 2007)**, MEGAN tool being the most popular. Other programs, such as MetaPhlAn, use clade-specific markers (CDS), i.e., sequences which are highly conserved within clades but not similar to other sequences outside the clade, before applying BLAST **(Segata *et al.*, 2012)**. This ensures a speed boost in taxonomic profiling, while keeping good classification resolution. Another strategy is the one carried out by MG-RAST pipeline, which uses MD5-based non-redundant protein database (M5nr), an integrative indexed database for similarity search with BLAST, including several taxonomy and function databases. Functional profiling is also possible through BLAST approaches, as multiple gene databases are available, though these approaches are always based on entire sequence datasets.

Mapping tool	Read profiling	Reference
MEGAN6	Taxonomic (LCA) Functional (KEGG, IP2G, SEED, eggNOG)	**Huson *et al.* (2016)**
SqueezeMeta	Taxonomic (LCA) Functional (KEGG, COG, Pfam)	**Tamames & Puente-Sánchez (2019)**
MG-RAST	Taxonomic (M5nr) Functional (M5nr)	**Keegan *et al.* (2016)**
MetaPhlAn 3	Taxonomic (CDS)	**Beghini *et al.* (2021)**
HUMAnN 3	Functional (UniRef90)	**Beghini *et al.* (2021)**
Kaiju	Taxonomic (LCA)	**Menzel *et al.* (2016)**

Table 6: *Summary of the most used metagenome read-mapping tools for taxonomic and functional profiling.*

Available metagenomic reference databases

As it happens with amplicon sequencing, reference database is determinant for metagenomic profiling, using either assembly or direct mapping strategies. In this case, databases are supposed to contain full genomes from every living organism, as sequencing is not bound to a unique marker gene, and these databases will normally be compendiums of other smaller ones. Additionally, computational effort both building the database and mapping the reads will be much higher. The most used reference databases in metagenomic taxonomic and functional assignment are summarised below.

A. **Taxonomy databases**. The most used database for taxonomic profiling is the one from the National Center for Biotechnology Information (NCBI) **(Federhen, 2012)**. This sequence repository includes the GenBank, ENA (EMBL) and DDBJ databases, with lineages and names from organisms pertaining to all living dominions. Although most metagenomic classification tools use this database as standard reference, custom databases are constantly built and updated for specific microbiomes. Examples are the Hungate 1000 database **(Seshadri *et al.*, 2018)**, including genomes and contigs from cow rumen microbiota, or the UHGG collection for human gut microbiota **(Almeida *et al.*, 2021)**. These databases are useful when profiling well-known microbiomes, reducing computational effort and removing possible misclassifications. The NCBI database relies on the sequences uploaded by users around the world, so it might contain incorrect taxonomic assignments or contaminated genomes (i.e., built-in genomes with bad-assigned contigs). NCBI RefSeq project **(Tatusova *et al.*, 2015)** offers an alternative to the classic database, as it curates GenBank sequences to exclude highly fragmented or contaminated genomes.

B. **Functional databases:** Gene function databases are more diverse, as gene functions can be understood at different cellular, biological and physiological levels. In addition, functional prediction can be addressed through multiple approaches **(Prakash & Taylor, 2012)**. The most common one is the classic homology-based approach, which compares predicted query amino acid sequences with a reference protein from collections such as NCBI or UniProt. Homology-based databases use different strategies to assign functions to genes. As examples, GO database uses the protein cell location, its participation in biological processes and its molecular function. COG **(Galperin *et al.*, 2018)** and NOG **(Huerta-Cepas *et al.*, 2019)** databases assign proteins to orthologous groups (i.e., sets of genes with common ancestors), Pfam **(Mistry *et al.*, 2021)** uses domains to assign proteins to families, and CARD **(Alcock *et al.*, 2020)** matches proteins to antibiotic resistance ontologies. Another approach is the one followed by KEGG **(Kanehisa**

***et al.*, 2019)** and SEED **(Overbeek *et al.*, 2014)** databases, which assign proteins to functional pathways or subsystems, respectively. The other function analysis approach is the motif/pattern-based approach, useful for partial protein sequences generated by short read sequencing. Some databases store a collection of specific sequence or structure motifs that proteins usually share to keep their normal function. InterPro database **(Camon *et al.*, 2005)** integrates other smaller databases relying on these common patterns, and annotates proteins to gene ontology terms (InterPro2GO).

1.3.3. Statistical analysis of microbiota data: current problematics

A recent issue which raised concern in the scientific community is the methodology used for post-processing and statistical analysis of microbiota data. Taxonomic and functional profiling of microbial communities, either through amplicon or through metagenomic sequencing, results in data matrices with some specific characteristics:

Compositional nature

Microbiota datasets are inevitably compositional. This is, the number of reads or contigs mapping to different features (taxa or genes) is limited by the total amount of reads that can be outputted by the sequencer. In other words, the abundance of each feature is relative to the number of reads from each sample. Compositional data are mathematically closured to the unit (or 100%), as each component is part of a wholesome, and are represented in an (n-1)-dimensional simplex space. Ignoring the data compositionality is dangerous, as it can involve statistical problems such as subcompositional incoherence (i.e., results change when working with a subset of the data), increment of false positives when testing differential abundance or large appearance of spurious correlations. Several solutions have been proposed for this problem, most of them based on using log-ratio transformations both to relativize values, reducing spurious correlation problems, and to relocate the sampling space in all the multivariate space (not only in the positive octant), making possible to apply classic statistical methods **(Greenacre, 2018)**. For applying log-ratios, data must be divided by different components and then calculate the logarithm of the quotient. While several techniques have been described, the most used are centred, additive and isometric log-ratios (CLR, ALR and ILR, respectively). CLR divides absolute data by the geometric mean of each feature, ALR uses the value of an arbitrary reference feature, which is also a part of the composition, as denominator, and ILR uses two subsets of parts from the composition and calculates the geometric mean of one respect to the other **(Greenacre, 2018)**. This

methodology has proven useful to mitigate the effects of false correlations and subcompositionality, and it is more and more utilized in microbiome research.

High-dimension multivariate data

Microbiota data contain thousands of different organisms, so the matrices usually include more variables than samples. This makes standard statistical methods unreliable, as intercorrelated variables will appear by chance due to high dimensionality problems. Multivariate and dimension reduction techniques might be used to mitigate the effect of undesired correlations, without losing too much variability.

High sparsity

Zero values are usually present in high number in multivariate compositions. They pose a problem for analysis, because statistical significance might be biased, and because data transformations are not possible when zeros exist. A common patch to avoid zeros when transforming data is replacing them by a very low value, by simple addition or by imputing them using Bayesian methods. However, zeros in microbiota datasets possess an uncertain true nature, which makes it difficult to differentiate true structural zeros (i.e., absence of a microbe in a specific group of animals) from sampling zeros (i.e., rare taxa not detected due to the sequencer limited sampling depth capacity) or from outlier zeros **(Kaul *et al.*, 2017)**. Currently, no standard procedure has been defined for resolving this problem. Although the zero replacement by addition or imputing are not optimal, since microbial absence cannot be correctly accounted, it might result appropriate enough for observing compositional differences between groups of samples.

In summary, the application of data transformations and multivariate statistical techniques addressing these problems is mandatory for a correct interpretation of the results. Microbiomes are complex environments with a high diversity of microbes, of which we can only observe a small window. Taking some correlations as true might lead to misinterpretations, since they can be consequence of compositionality.

1.4. Study of the microbiota as a way of understanding and modulating host metabolism and phenotype

The holobiont concept has been crucial to deepen in the characteristics of the microbiota-host association. This term was conceived by Lynn Margulis as a way to define the new organism originated by a symbiotic association between two or more individuals **(Margulis, 1990)**. All animals and plants are holobionts, as they establish symbiotic relationships with microorganisms, hosting different microbial populations in their tissues. The phytomicrobiome comprises multiple communities associated with the aerial parts, roots or inner tissues of the plant, biomes also known as phyllosphere, rhizosphere and endosphere, respectively **(Bhatt *et al.*, 2020)**. As for animals, they offer a great variety of niches for microbial communities, although gut and skin microbiota are the most important communities due to their microbial mass and major impact on host.

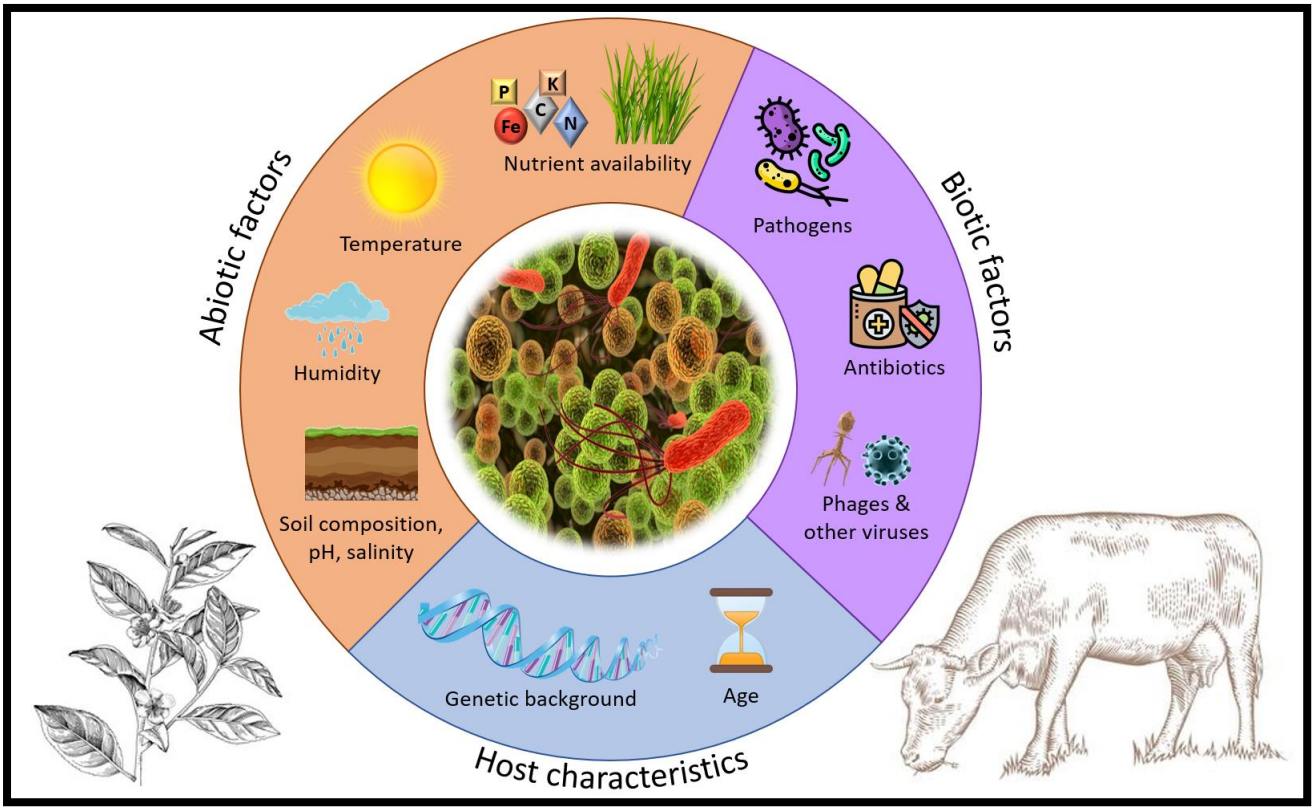

Figure 9: *Main factors influencing the structure of host-associated microbiomes.*

The composition of plant and animal-associated microbiomes is highly variable. Phytomicrobiome is more sensitive to abiotic stress than animal microbiome: solar radiation, temperature, humidity and nutrient availability, as well as soil composition, are important shapers of the phytomicrobiome, even causing a very active seasonal variation in the structure of the phyllosphere **(Copeland *et al.*, 2015)**,

while the influence of abiotic factors on animal microbiome mostly relies on thermal stress conditions and diet variations.

On the other hand, biotic stress caused by pathogen infection can lead to dysbiosis, i.e., a disruption to the microbiome normal homeostasis (eubiosis), either through the direct effects of the infection (competing with commensal microbes), the natural host immunity (host-mediated or exploited by the pathogens) or the utilization of antibiotics. Competition between commensal and pathogenic microorganisms can occur at several levels: some microbes such as *Clostridium difficile* are known to compete for the attachment to mucins at the intestinal mucus layer, as a way to successfully colonize the gut environment **(Cai *et al.*, 2020)**; other microbes such as *enterotoxigenic E. coli* (ETEC) have developed strategies to use alternate nutrient sources to increase their fitness and overcome commensal bacteria **(Kamada *et al.*, 2013)**. As for the host immunity, some pathogens like *Salmonella spp.* can exploit it through activation of the inflammatory pathway to reduce the eubiotic microbial population, promoting infection **(Behnsen *et al.*, 2015)**. Lastly, antibiotics are the most important biotic stress source responsible of dysbiosis, causing short-term (opportunistic infections or diarrhoeas) and long-term effects (allergic conditions) **(Konstantinidis *et al.*, 2020)**. Dysbiosis is related to multiple disorders and diseases, both in plants and animals **(Carding *et al.*, 2015; Liu *et al.*, 2020)**. Nevertheless, not only pathogens are biotic sources of microbiota shifting, neither every shifting result in malfunction of microbiota. Low-virulent viruses can eventually infect holobionts and function as mutualistic symbionts, even conferring resistance to other infections **(Barton *et al.*, 2007)**. Bacteriophages are also shifters of prokaryotic communities in a beneficial way by favouring high-richness populations. As they attack more frequently to the most abundant species, phages tend to drift the microbiome to a more ecologically diverse one **(Scanlan *et al.*, 2015)**, therefore enhancing the response capacity, i.e. resilience, of the microbiome ecosystem, and providing health benefits to the host **(Elmqvist *et al.*, 2003; Roslund *et al.*, 2020)**.

But the largest differences in microbiome between holobionts are normally attributed to host characteristics, which play a central role as microbiome modulators. Host age is determinant, since transitions in microbiota composition have been reported for different ages in humans and other animals **(Lu *et al.*, 2003; Zhao *et al.*, 2015; Xu *et al.*, 2019)**. Interestingly, these transitions are more related either to dietary changes occurring through different aging stages, such as the passage from lactation to weaning, or to health status degeneration inherent to aging processes (reviewed in **Clark & Walker (2018)**). Several studies have also reported a major influence of host genotype on microbial composition, due to different physical, metabolic or behavioural phenotypes, thereby involving a heritable component of the microbiota **(Crespo-Piazuelo *et al.*, 2019; López-Paredes *et al.*, 2020)**.

Significant variations in gut microbiota composition have been observed between different breeds in domestic animal species **(Bergamaschi *et al.*, 2020; Massacci *et al.*, 2020b; Yang *et al.*, 2020)**.

However, the direction of causality in microbiota-host associations is still unknown for several phenotypic traits, such as health measures, feed intake, or animal behaviour. Indeed, relationships between microbiota and host cannot be reduced to a unidirectional dependence, but bidirectional. It is widely known that host-associated microbiota has an important role in host development and survivability. Gut microbiota (GM) has been far more studied than other communities, because digestive tract possesses the highest abundance and diversity of endosymbiotic microbes in most animals **(White *et al.*, 1999; Sender *et al.*, 2016)**, and also because its contribution to host biology has a broader scope (**Table 7**), from nutritional functions to neurobiological influence through the recently discovered gut-brain axis **(Carabotti *et al.*, 2015)**.

Animal species	Influence on host phenotype	References
Insects	Nutrient acquisition Improved digestibility Improved fecundity Infection resistance	**Dillon & Charnley (2002)** **Dillon & Dillon (2004)** **Rosengaus *et al.* (2011)**
Fish	Nutrient acquisition Development Infection resistance	**Rawls *et al.* (2004)** **Romero *et al.* (2014)**
Birds	Nutrient acquisition Detoxification of plant materials Infection resistance	**Grond *et al.* (2018)**
Mammals	Nutrient acquisition Increased feed efficiency Development Improved fertility and fecundity Increased longevity Infection resistance GHG emissions Higher carcass yield Product quality	**Hanning & Diaz-Sanchez (2015)** **Dearing & Kohl (2017)** **Tapio *et al.* (2017)** **Khanal *et al.* (2021)**
Humans	Nutrient acquisition Increased longevity Infection resistance Gut-brain axis	**Chassard & Lacroix (2013)** **Carabotti *et al.* (2015)** **Badal *et al.* (2020)**

Table 7: *Examples of gut microbiota contributions to host biology.*

Medical sciences have used these discoveries to improve human health and longevity, while livestock industry tries to unravel the influence of microbiota on both animal health and productive traits. Livestock industry faces multiple challenges related to productivity, animal welfare and health and sustainability, which traditionally have been addressed through environment modulation (i.e., nutrition improvement, temperature controlling, health conditions and antimicrobials as last option) or classic genetic improvement breeding programs. The holobiont concept and the interactions between their components (i.e., microbiota and host) opens a new way to confront old and new challenges in animal production.

1.4.1. The case of the Iberian pig

Origins of the Iberian pig

The Iberian pig is a rustic racial group raised mostly in Spain, whose origins date back to the domestication of the Mediterranean wild boar (*Sus scrofa meridionalis*) in several areas of the Iberian Peninsula **(Toro *et al.*, 2000)**. Due to the genetic and physiological singularities of this breed, it is interesting to know its evolutionary background.

Domestic pigs come from a single Suidae species, the wild boar (*Sus scrofa*), which originated from the tropical forests of Southeast Asia and spread across Eurasia and North Africa during the Pliocene (1-2 Mya). During this expansion period, *S. scrofa* populations adapted to multiple environments in a relatively short period of time **(Frantz *et al.*, 2016)**. Up to date, 16 *S. scrofa* subspecies have been catalogued **(ITIS, 2021)**, as a result of adaptations to new habitats and foraging strategies.

Archaeological evidence supports the hypothesis of independent domestication events in East Asia and Near East, 12,000 to 8,000 years ago **(Ervynck *et al.*, 2001)**. On the other hand, the first genetic studies revealed that Asian and European domestic pigs are closer to regional wild boars than to each other, suggesting the existence of more independent domestication event hotspots, thereby contradicting the archaeological and ancient DNA registers **(Giuffra *et al.*, 2000)**. But later research pointed to introgression phenomena from regional wild boars to imported domestic pigs (introgressive capture) **(Larson & Fuller, 2014)**, which causes a higher genetic resemblance between them. The joining of both the archaeological and genetic evidences lead to conclude that long-term independent domestication occurred only in two regions, China and Eastern Anatolia, and that domestic pigs accompanied humans in their dispersion across Eurasia, then hybridizing with regional wild boars **(Frantz *et al.*, 2016)** (**Figure 10**). Moreover, human search for improved pig breeds led to hybridization between different domestic Asian and European populations during the 19th and 20th

centuries. These repeated admixture processes have generated a considerable amount of pig breeds. There are currently around 480 pig breeds according to Domestic Animal Diversity Information System (DAD-IS) **(FAO, 2021)**, most of them being local breeds with a high level of regional wild boar introgression. Only a few are improved breeds selected for high yield and fertility, which are majorly distributed through Europe and USA.

Figure 10: *Schematic overview of the domestication process of Sus scrofa. Black arrows represent the dispersion of ancestral S. scrofa from Southeast Asia to Eurasia and North Africa, 1 to 2 Mya. Red zones are the currently proposed hotspots for pig domestication events (**A:** Eastern Asia; **B:** Anatolia), which occurred 12,000 to 8,000 years ago. Red arrows represent the spreading of domesticated pigs accompanying humans.*

Iberian pig is a special case of local breed. While other local breeds have experimented crossbreeding with Asian lineages to improve their prolificacy and productivity, Iberian pig does not possess Asian introgression **(Alves *et al.*, 2003)**, being distant to other local European breeds **(Muñoz *et al.*, 2018)** and highly similar to regional wild boars **(Ribani *et al.*, 2019)**. For this reason, this breed keeps many adaptative traits from its Mediterranean ancestry. Iberian pig's morphology is brevilinear, with stretched and thin legs, a long and pointed snout and dark skin with thick hair. These adaptations make

it resistant to summer temperature rising and enable it to optimally profit the Mediterranean forest resources, spending long timeslots scouting for acorns, grass and small roots. Its physiological adaptations – low basal metabolism, high feed intake and increased fat deposition rate – make it extremely resilient to food scarcity periods, as they allow to store high amounts of energy as fat deposits **(Toro *et al.*, 2000; Forero Vizcaíno & Andrés, 2008)**. These adaptations have been described as thrifty genotype and related to obesity **(Neel, 1962)**.

The Iberian pig production system

Iberian pig traditional production is deeply linked to the Dehesa ecosystem, which is profited by human as an agroforestry production system with extensive livestock use as the main activity. Although mixing different farming species for maximizing territory exploitation is very common, Iberian pig is currently the most important one, both economically and in terms of total population. Dehesas are Mediterranean sparse forests composed by oak species, mostly *Quercus ilex* and *Q. suber*, coexisting with other tree species such as ashes (*Fraxinus angustifolia*), carobs (*Ceratonia siliqua*) or junipers (*Juniperus communis, J. sabina*), all accompanied by a scarce and low-diversity bush stratum (e.g., *Cistus ladanifer*, *Thymus spp*., *Phyllea latifolia*, *Arbutus unedo*). The prevailing climate of these forests, mostly present in central, southern and southwestern Iberian Peninsula, is Mediterranean with continental influences. It is characterized by cold winters and hot summers, as well as variable rainfall, mainly distributed during winter months and almost absent from May to September. Dehesa pastures are composed by an annual community with high species richness (e.g., *Corynephorus spp*., *Vulpia spp*., *Trifolium spp*., *Bromus spp*.) and different perennial populations, either dominated by *Poa bulbosa* and legume species (majadales) or associated with flooded areas (vallicares). Pasture abundance is dependent on climatology, presenting two growth peaks during spring and autumn and two minimum during winter and summer **(Forero Vizcaíno & Andrés, 2008)**.

The vegetation profile of the Dehesa makes it ideal for extensive exploitation of Iberian pigs. The arboreal stratum, dominated by *Quercus ilex*, contributes to improve soil quality and creates microclimatic conditions under their treetops, softening temperatures and retaining moisture **(Montoya, 1982)**. These factors are essential for growth of pasture, which ensures a productive and high-quality nutrient source for livestock – especially through the presence of legume species. Oaks also provide both foliage and acorns as nutrient supply, the latest being the most important dietary contribution for Iberian pigs during the extensive fattening period of their productive cycle, known as *montanera*.

The traditional Iberian pig production system is semi-extensive, consisting in a growing period with food restriction, to avoid premature fat deposition due to its thrifty genotype, and a *montanera* fattening period occurring in the Dehesa ecosystems, in which animals mostly feed on pastures and acorns **(Daza *et al.*, 2008)**. This feeding strategy has an important effect on final products: first, because acorn-based diet produces high quality fat, with a unique fatty acid profile richer in unsaturated acids, mainly oleic acid (C18:1n-9); and second, because food restriction during growing period will cause a compensatory growth during *montanera*, thus fat deposition will mainly occur when diet is richer in acorns. On the other hand, the genetic distinctiveness of Iberian pigs causes an increase in the amount of both external and intramuscular fat (IMF) deposition, compared to other pig breeds. All these factors confer a higher quality and better organoleptic properties to Iberian pig products compared to products from indoor commercial pigs **(Ventanas *et al.*, 2007)**.

Figure 11: *Iberian pigs in the Dehesa. Image by Cristina Óvilo.*

Crossbreeding and intensification

Due to the high quality of their final products, *montanera* Iberian pig rearing represents one of the most internationally successful examples of sustainable local breed production systems. Indeed, Spain ranks as the fourth country in swine meat exportations, with an important contribution of Iberian dry-cured products, mostly ham **(FAO, 2020)**. But the widespread demand for these products has not been always as it is now. During the 1960s, Iberian pig populations suffered an important downsizing process mostly due to the low interest in high-fat products and the emergence of new selected breeds, more efficient for meat production because of their faster muscle development, increased growth rate and higher fertility and prolificacy. Later changes in consumer demands, more focused on meat quality, allowed to recover the Iberian pig production, occupying a new high-price niche in the market. Unfortunately, the low productivity and prolificacy of the Iberian pig remained significant problems for farmers, and crossbreeding with Duroc pigs became a common practice.

Duroc breed was originated in the United States of America as a result of multiple crossbreeding events during the 18th and 19th centuries. Their African ancestors were transported from the Guinea coast to New England during the era of slave trade, and subjected to crossbreeding during the 1800s with different European breeds such as Berkshire or some Iberian pig's varieties. The actual Duroc is an improved commercial variety created during the 19th century, as a result of crossbreeding between New York's old Duroc and Jersey Red pig (originally, it was known as Duroc-Jersey pig). Duroc pigs are known by their dark coloured coat from orange to brown and their high productivity, as they have been strongly selected. They possess fast growing and high prolificacy, with litters of 10-15 piglets, and most of the selected lines have a high carcass yield, producing large amounts of meat **(Vaughan, 1950)**, making this breed ideal for intensive farming. Although different crossbreeding between Iberian and commercial pigs have been tested before, Duroc has been established as the preferred parental line in Iberian crossbreds, not only for its high performance, but also because of both its genetic closeness and its physical resemblance – i.e., the dark coat colour – to Iberian pig **(Rothschild & Ruvinsky, 2011)**.

This strategy has helped to solve some of the Iberian pig production problems, as crossbred animals have higher prolificacy, growth rate and carcass yield than Iberian purebreds. However, their litter size and maternal aptitude is lower in comparison to improved breeds. In order to preserve the breed, its genetics and the production system, the legislation demands the employment of pure Iberian sows for crossbreeding with Duroc boars (*Quality Norm for Iberian products*, BOE, 2014). This implies that reproductive ability is far below to that reached by modern pig genotypes. The reduced uterine capacity of the Iberian pig **(Gonzalez-Añover *et al.*, 2011)** limits the development of foetuses and induces differential growth among littermates leading to increased heterogeneity and high incidence of low birth-weight piglets (LBW). LBW offspring are a huge problem as their health is compromised at birth and they show altered developmental patterns, health status, and adult phenotype **(Quiniou *et al.*, 2002; Gondret *et al.*, 2006; Rehfeldt & Kuhn, 2006)**. Thus, improvement of prolificacy and piglet viability in early periods has major implications on the profitability and future development of Iberian pork production.

In summary, Duroc x Iberian crossbreeding is a strategy which offers farmers the opportunity to intensify and improve their production, while preserving the Iberian pig's characteristics. However, the loss in product quality respect to purebred Iberians, as well as the mentioned maternal aptitude problems, are the cause of the overall shifting to crossbred intensive production, which represents 75% of total Iberian pig production, while traditional purebred *montanera* is less than 10% **(MAPAMA, 2019)**. Solutions have been searched through diet optimization, as high-oleic acid supply could

improve IMF composition, or breeding programs, mostly in intensive rearing systems, which have larger room for improvement.

State-of-art of microbiome studies in pigs

Swine microbiome has been broadly studied from different perspectives. Although the vast majority of available studies are based on 16S sequencing of V3 or V4 hypervariable regions, thereby focusing on bacterial sub-populations, a notable amount of knowledge on pig microbiota has been generated, both in taxonomic composition and metabolic functioning.

Most of the microbial niches contribute somehow to host health. Both oral and vaginal microbiomes are known to house pathogens responsible of important epidemics. Murase *et al.* reported saliva as the most probable source of infection from *Streptococcus suis* and other streptococcal species **(Murase *et al.*, 2019)**. However, the composition and normal functioning of vaginal microbiota in pigs is still a matter of research. Characterization studies have been carried out, stating that vaginal microbiota is pretty similar to faecal microbiota due to anatomical disposition of rectum and genitalia **(Lorenzen *et al.*, 2015)**. But the major evidence of pathogen presence in vaginal region has been so far observed only by culture techniques **(Bara *et al.*, 1993)**, and so plenty of microbial diversity in both eubiotic and dysbiotic conditions is still unknown. Instead, vaginal microbiome in pigs is better known for its participation in neonatal development, as some studies have found relationship between gilt vaginal dysbiosis and piglet gut dysfunction, even affecting brain development **(Jašarević *et al.*, 2015)**.

However, these niches have been scarcely explored, as the majority of the current research focuses on the gut microbiome. The composition of pig gut microbiota is well known at each gut portion, with large presence of *Firmicutes*, *Bacteroidetes* and *Proteobacteria* **(Ramayo-Caldas *et al.*, 2016; Xiao *et al.*, 2016)**. Small intestine possesses a less stable community and a lower microbial richness than colon regions, as reported by several authors **(Yang *et al.*, 2016b; Kelly *et al.*, 2017)**. Studies in distal colon and faeces reveal *Prevotella spp.* as a common and a very abundant genus in those regions **(Ramayo-Caldas *et al.*, 2016; Crespo-Piazuelo *et al.*, 2019)**. Other studies have also delved into the age-related changes in microbial composition, mostly in colon region. According to current knowledge, pre-weaning stage is characterized by a predominance of *Proteobacteria* from *Escherichia-Shigella* group – such as *E. coli* –, accompanied by some *Bacteroidetes* and *Firmicutes* genera, such as *Bacteroides*, *Streptococcus*, *Clostridium* or *Lactobacillus*, the latest playing a crucial role in disease prevention **(Mach *et al.*, 2015; Fouhse *et al.*, 2016)**. *Proteobacteria* are gradually replaced by *Bacteroidetes* representatives, mostly *Prevotella*, with the transition into adulthood, while *Firmicutes* are known to increase with age **(Slifierz *et al.*, 2015; Zhao *et al.*, 2015; Ke *et al.*, 2019)**.

Maturation of GM also involves an increment in richness and evenness and lower abundance of facultative anaerobes (*Escherichia-Shigella* or *Streptococcus*) and opportunistic pathogens, due to increased competition (ETEC, *Helicobacter*, *Lactococcus*, *Campylobacter* or *Actinobacillus*) **(Slifierz *et al.*, 2015; Gresse *et al.*, 2019)**. All these age-related changes are highly linked to diet transition from milk to plant-based, thus weaning period plays a significant role in adult GM composition, and so in host health. For instance, Massacci *et al.* reported that late weaning might be beneficial for adult pigs, through increasing colon microbiota diversity and the abundance of *Faecalibacterium prausnitzii* **(Massacci *et al.*, 2020a)**, a bacteria known for improving gut health in humans **(Miquel *et al.*, 2013)**.

Gut microbiome has also been associated with multiple swine productive traits, such as growth, performance, feed efficiency or carcass composition **(Maltecca *et al.*, 2020)**. As bacterial metabolism is directly correlated to host nutrient supply, it is reasonable to deduce that microbiota will contribute to host development. Some bacteria have been found as significantly correlated to pig body weight, either in pre-weaning phase (*Streptococcus* and *Lactobacillus*) or in adult phase (*Turicibacter*, *Clostridium* or *Akkermansia*) **(Derrien *et al.*, 2017; Wang *et al.*, 2019)**. GM also affects adiposity, as several studies have associated an elevated *Firmicutes*/*Bacteroidetes* ratio to obesity phenotypes, while *Akkermansia* is known as inversely correlated to obesity **(Cani & de Vos, 2017; Yang *et al.*, 2018)**. As fat deposition is in turn related to organoleptic parameters, F/B ratio might as well modify meat quality. In fact, a significant proportion of variance in several carcass composition and meat quality traits can be explained by microbiota composition, including average daily gain, fat thickness, premium cut's weight, tenderness, redness and yellowness **(Khanal *et al.*, 2021)**.

So far, the state-of-art of gut microbiome studies in Iberian pig is still scarce. Crespo-Piazuelo *et al.* have already described the progression in GM through the different intestinal regions **(Crespo-Piazuelo *et al.*, 2018)**, with results matching former studies in other breeds, and have also analysed the association between the pig genome and the relative abundance of gut microbes **(Crespo-Piazuelo *et al.*, 2019)**, stating that pig genome has control on GM composition. However, research linking GM to phenotypic traits in Iberian pigs is limited, including non-published associations between GM composition and birth weight in crossbred animals **(Vázquez-Gómez *et al.*, 2019; Heras-Molina *et al.*, 2021)**. In addition, other authors have proposed to use GM composition as proxy for traceability of farming systems **(García-Casco *et al.*, 2019)**. Given that this breed possesses some peculiarities related to metabolism, development, fat deposition and fatty acid composition, it might be interesting to explore the relationships between gut microbiome, tissue metabolism and meat composition.

1.4.2. The case of dairy cattle

Origins and evolution of bovine cattle

Bovine cattle originated during the Neolithic era, about 10,000 years ago, through domestication of currently extinct wild aurochs (*Bos primigenius*). Current studies point to two independent domestication processes, one in the Fertile Crescent and another in Asia, which resulted in two different domestic taxa, taurine cattle (*B. taurus*) and indicine cattle (*B. indicus*), respectively. Current mtDNA studies suggest that farmers spread taurine cattle through Europe, where hybridization with regional aurochs occurred **(Upadhyay *et al.*, 2017; Pitt *et al.*, 2019)** (**Figure 12**). It must be pointed out that it is still not clear whether the domestic taxa form separate species or if they are aurochs' subspecies, since they have remarkable morphological and genetic differences, but interbreeding between them and with related species (e.g., *Bos grunniens* or *B. javanicus*) is possible, producing hybrids that can be fertile or sterile **(Zhang, 2000; Nijman *et al.*, 2003)**.

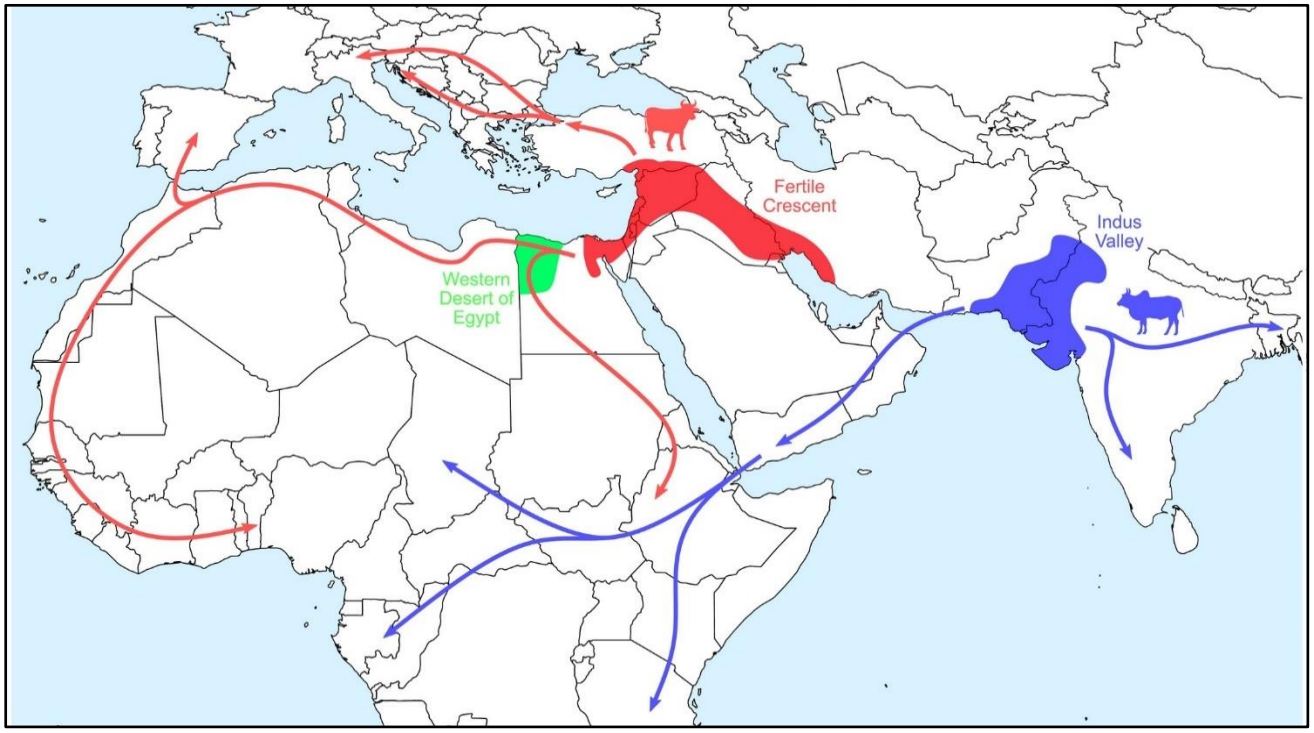

Figure 12: *Proposed domestication points and dispersion routes for taurine and indicine cattle. An additional domestication site has been proposed in Western Egypt, though current research points to the existence of only two independent events. Figure from* ***Pitt et al. (2019)***.

During the European expansion, *B. taurus* populations diverged under the influence of different human tribes, specializing in beef (Charolais, Limousin or Angus), dairy (Holstein or Jersey) or dual-purpose breeds (Simmental, Brown Swiss or Dexter), either large or medium-sized **(FAO, 2021)**. A number of

breeds adapted to semi-arid and tropical environments emerged after cattle introduction in multiple parts of the worlds. For instance, Spanish and Portuguese empires introduced cattle in Central and South America during the 16th century, originating breeds such as Creole or Senepol, while Australia received Shorthorns and Herefords from United Kingdom during the 1800s **(O'Neill *et al.*, 2010)**.

The upcoming arrival of new technologies in the 1950s, within the framework of the so-called Green Revolution, led to a mechanization and intensification of livestock production. This was accompanied by the emergence of quantitative genetics methods and the definition of selection indices, leading to the establishment of breeding programs, which significantly reduced the genetic diversity of taurine and indicine cattle by favouring high-productive strains **(O'Neill *et al.*, 2010)**. However, selection for high production has generated undesired responses in other traits **(Rauw *et al.*, 1998)**. In dairy cattle, increased milk production requires a higher mobilization of the body reserves, which causes a negative energy balance that has been related to metabolic disorders (hypocalcaemia, ketosis or ruminal acidosis) and a higher incidence of production diseases (mastitis or lameness) **(Mulligan & Doherty, 2008)**. Milk production has also been correlated with reproductive loss **(Lucy, 2001)**. In addition, massification and intensification promoted the onset of other problems, such as animal welfare issues or the overwhelming increment of livestock greenhouse gas (GHG) emissions.

Contribution of cattle to global warming

Ruminants are a suborder of Artiodactyl mammals characterized by possessing a four-chambered stomach specialized for plant digestion. This group includes multiple families such as tragulids, cervids, giraffids, antilocaprids and bovids **(Hernández Fernández & Vrba, 2005)**. The ruminant stomach can be anatomically divided in the forestomach complex and the true stomach. The forestomach includes the three first chambers, the rumen, the reticulum and the omasum, while the abomasum constitutes the glandular stomach. Each compartment fulfils different functions in food digestion. The rumen contains an important microbial community with a crucial role fermenting plant fibre. In the reticulum, partially digested material is separated: while undigested particles are regurgitated and chewed (rumination), thereby facilitating their reduction and later digestion, digested particles pass to the omasum, where volatile fatty acids, minerals and reticulum fluid are absorbed. The enzymatic digestion of dietary remnants and microbes is carried out in the abomasum, this process continuing in the small intestine **(Clauss & Hofmann, 2014)**.

Rumen microbiota (RM) plays an essential role in digestion, through a process known as enteric fermentation. This process consists in the digestion of plant complex carbohydrates by several microbial species, producing nutrients usable by the host and methane as a residual, which is expelled

by the host through eructation. The microbes participating in this process constitute an anaerobic trophic pyramid in which each level uses the metabolic residuals from the former as nutrient source. Thus, primary fermenters decompose complex fibres and generate nutrients usable by the host and by other secondary fermenters, which in turn produce methane precursors. Lastly, methanogenic archaea utilize these precursors to carry out their metabolism and output methane as a residual in a process known as methanogenesis **(Crable *et al.*, 2011; Mizrahi *et al.*, 2021)**. Methanogenesis is carried out by multiple archaeal groups, which use different precursors as metabolites to produce energy under anaerobic conditions. Three methanogenesis pathways have been described until now, hydrogenotrophic, methylotrophic and acetoclastic (**Figure 13**).

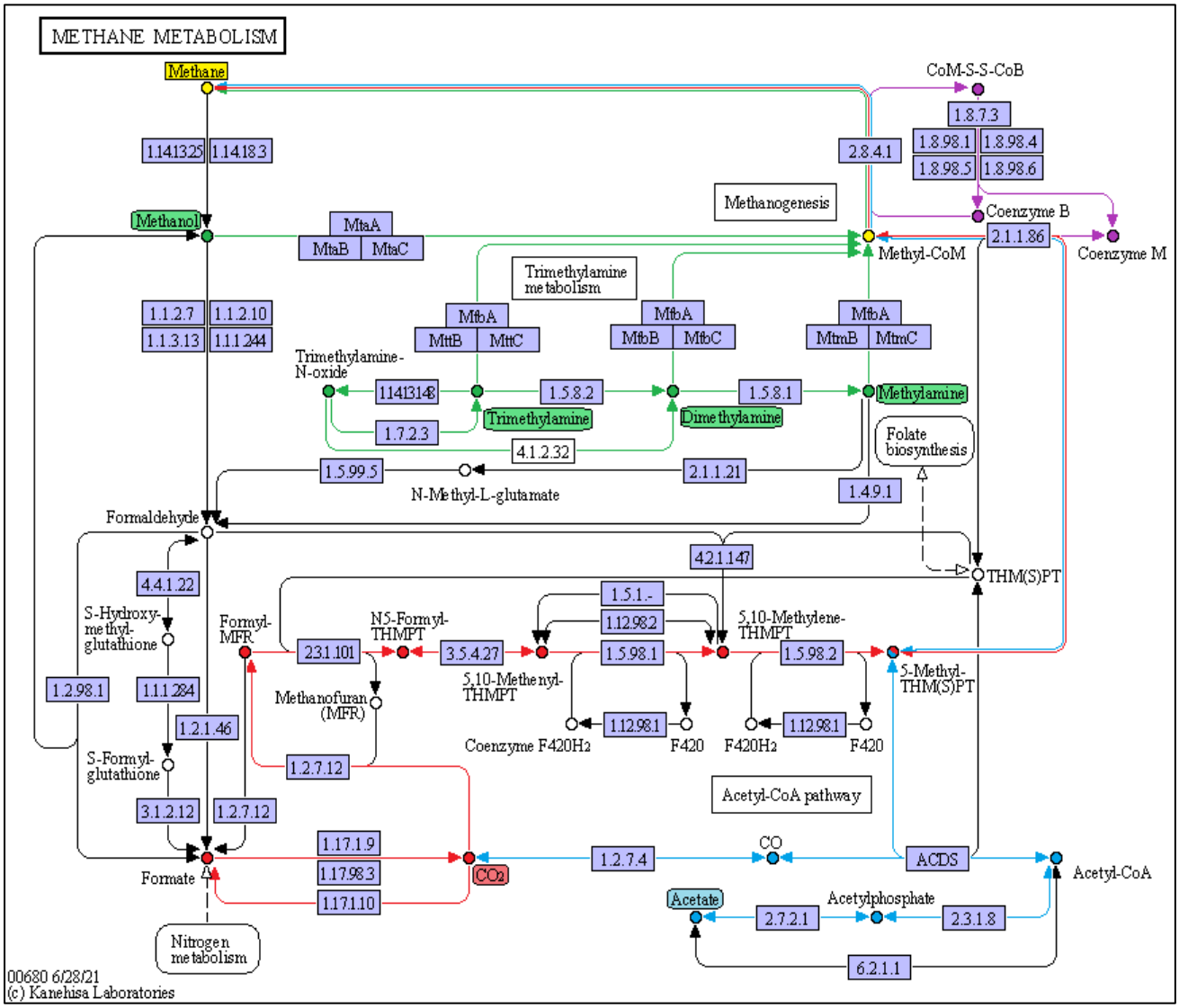

***Figure 13**: The three methanogenesis pathways: hydrogenotrophic (red), acetoclastic (blue) and methylotrophic (green). Adapted from KEGG methane metabolism pathway (https://www.genome.jp/pathway/ko00680) and **Galagan et al. (2002)**.*

The most common one is hydrogenotrophic methanogenesis, present in all the basal groups of methanogens and using CO_2 or formate as electron donors. Methylotrophic methanogenesis is less exergonic than hydrogenotrophic and is only present in *Methanosarcinales*, *Methanomassiliicoccales* and *Methanobacteriales*, which use methanol or methylamines as methyl donors to Coenzyme-M. Finally, acetoclastic methanogenesis is the least exergonic of the three mechanisms, reserved only to *Methanosarcinales*, and uses the activation of acetyl-CoA to acetate, whose cleavage produces CO and a methyl group **(Wolfe, 1993; Lang *et al.*, 2015)**.

The ongoing climate emergency has highlighted the need to urgently reduce GHG emissions from human activities. Latest data from IPCC report an increase in atmospheric GHG concentrations up to an average of 410 ppm for carbon dioxide (CO_2), 1866 ppb for methane (CH_4), and 332 ppb for nitrous oxide (N_2O) in 2019. This concentration increment has resulted in a global warming of the Earth's surface, whose temperature has risen 1.09°C in the 2011-2020 period compared to 1850-1900, being the 2003-2012 period critical for temperature rising (+0.19°C) **(IPCC, 2021)**. As stated by IPCC, *"Observed increases in well-mixed greenhouse gas (GHG) concentrations since around 1750 are unequivocally caused by human activities"* **(IPCC, 2021)**. Hence, global measurement and control of human GHG emissions has become important in order to mitigate climate change effects. Although a consensus on the correct assignment of GHG emissions to each human activity sector has not been still reached, an approximation is currently accepted (**Figure 14**).

Data from different countries have been aggregated by several projects such as the Global Carbon Project (GCP), the Climate Analysis Indicators Tool (CAIT) from the World Resources Institute, the Emissions Database for Global Atmospheric Research (EDGAR) or the data from the UNFCCC (United Nations Framework Convention on Climate Change). Other resources more focused on specific sectors can be found as well, like the databases from the IEA (International Energy Agency) for emissions generated by energy supply, the FAO (Food and Agriculture Organization) for emissions derived from agriculture and food production, or the GLEAM (Global Livestock Environmental Assessment Model) for livestock GHG emissions.

Although energy generation activities are the main source of GHG, representing a 70% of total human emissions, those derived from farming activities are not despicable. The most updated disaggregated data show a contribution from livestock direct emissions (i.e., CH_4 and N_2O from manure management and CH_4 from digestive processes) of 3.9 - 4.1 GT(CO_2eq)/year for 2017 and 2018 **(FAO, 2017; Poore & Nemecek, 2018)**, of which around 70% are generated by cattle (2.8 GT CO_2eq in 2017). Although part of cattle emissions ends as neutral due to carbon sequestration by pastures feeding the same

animals, intensification and overpopulation have outweighed this closed carbon cycle, resulting in a net increase of atmospheric GHG concentrations **(Garnett *et al.*, 2017)**. The relevance of cattle in this context can be mostly attributed to the digestion process characteristic of ruminants, which produces high amounts of methane (2.51 GT CO_2eq in 2017).

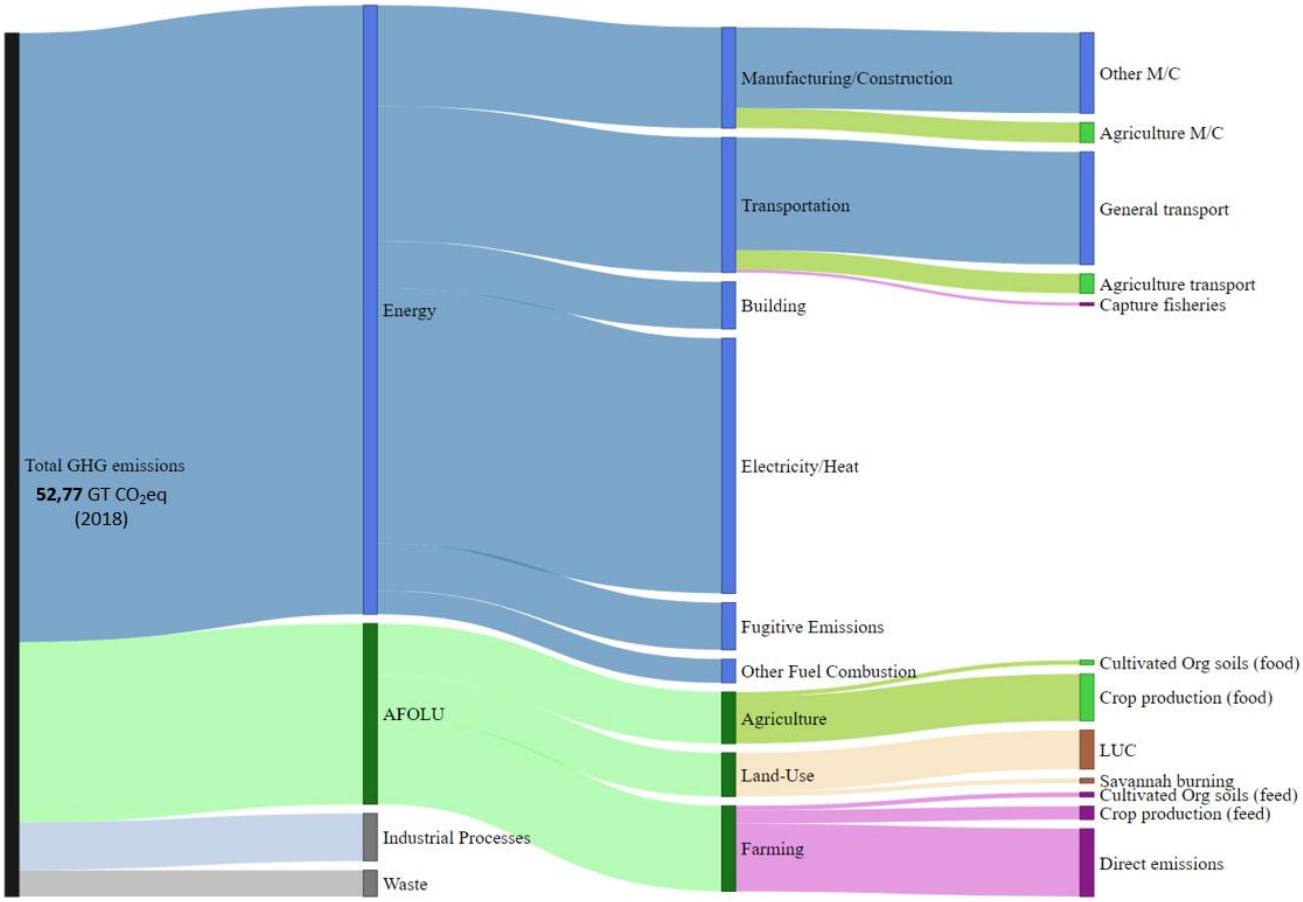

Figure 14: *Total GHG emissions (CO_2eq) from human activities during the year 2018. Data for major sectors (energy, industry & waste) adapted from Climate Watch database (https://www.climatewatchdata.org/). AFOLU data adapted from* ***Poore & Nemecek (2018)****.*

As said before, methanogen activity is closely related to the fermentation of plant carbohydrates by bacteria, which makes cattle emissions dependent on their diet, feed intake and feed efficiency **(de Haas *et al.*, 2011)**, in addition to the heritable component of RM **(Wallace *et al.*, 2019)**. Multiple approaches have been proposed for methane reduction, including de-intensification of livestock towards more sustainable production systems, diet modification or the inclusion of emissions as a phenotypic trait into modern breeding programmes.

So far, nutritional strategies have proven successful for GHG reduction, both by manipulation of nutritional balance or by the use of additives. Rendón-Huerta *et al.* revisited the current knowledge on nutrient modulation applied for this goal, thus concluding that lower crude protein and forage content

could decrease both CH_4 and N_2O emissions in dairy cattle **(Rendón-Huerta *et al.*, 2018)**, while also being effective measures for beef cattle **(van Gastelen *et al.*, 2019)**. On the other hand, several additives have been described as potent inhibitors of methanogenesis. For example, 3-nitrooxypropanol (3-NOP), which inhibits the methyl-CoM reductase, is reported to reduce up to 60% of methane production in cattle **(Haisan *et al.*, 2014; Van Wesemael *et al.*, 2019)**. Another promising additive is the red algae *Asparagopsis taxiformis*, which contains various anti-methanogenic compounds such as bromoform or dichloromethane, reducing up to 80% of methane emissions in beef cattle **(Roque *et al.*, 2021)**. Lastly, competitive inhibitors of methanogenesis, such as nitrate and sulphate, have also proven effective to reduce methane production through reducing the bioavailability of precursors for methanogens **(van Zijderveld *et al.*, 2010; Yang *et al.*, 2016a; Sela-Adler *et al.*, 2017)**.

Other approaches have focused on genetic selection, which currently stands as the most efficient strategy for reducing enteric methane, with predictive studies calculating a potential reduction of 20-25% in methane emissions in 10 to 20 years when applying modern selection indices **(González-Recio *et al.*, 2020; López-Paredes *et al.*, 2020; de Haas *et al.*, 2021)**. The heritability for methane concentration, production and/or intensity has been estimated in values from 0.11 to 0.30, using different regression models **(Lassen & Løvendahl, 2016; Pszczola *et al.*, 2017; Brito *et al.*, 2018; Breider *et al.*, 2019; López-Paredes *et al.*, 2020)**, and even 0.35 when using predicted methane production based on feed intake instead of direct measurement **(de Haas *et al.*, 2011)**. Thereby, many breeding programmes have begun to consider including methane production as a trait for selection, either from direct measurement or predicted through feed intake and diet composition **(IPCC, 2000)**. Other programmes use indirect selection through productive traits such as residual feed intake **(Basarab *et al.*, 2013)**, since methane production is linked to food consumption and digestibility. Furthermore, the existence of a relationship between rumen microbiota and host genetics plays a key role to consider the microbiome information in the selection indices. There is evidence of host genetic control on microbiota composition via physiological and metabolic characteristics **(Roehe *et al.*, 2016; González-Recio *et al.*, 2018; Abbas *et al.*, 2020)**. Also, genetic correlation between methane production and rumen microbiota composition has been reported. In this sense, some authors have proposed to add microbial information to selection indices, since mixing host genetic and metagenomic information to predict methane production could result in an improvement of selection response **(Saborío-Montero *et al.*, 2021)**.

Both the direct role of rumen *Archaea* and the existence of a host-microbiome interaction have placed the spotlight on RM for reducing enteric methane production from ruminants. RM modulation through

direct approaches or through genetic selection are arising as valid strategies to improve feed efficiency, therefore positively affecting to the animal's productivity, and to mitigate cattle impact on global warming.

State-of-art of microbiome studies in dairy cattle

Rumen microbiome has a major contribution to multiple aspects of cattle production, and therefore it has been broadly studied in recent years. The current knowledge about RM is fairly broad, and WMS techniques have been repeatedly applied to characterize its composition in different cattle breeds. *Firmicutes* and *Bacteroidetes* have been reported to predominate in RM composition, with high representation of fibre fermenters, as it occurs in other herbivore animals. While *Bacteroidetes* are mainly represented by *Prevotella*, *Firmicutes* group is more variable in genera and includes both primary fermenters (*Ruminococcus*, *Lachnospira*, *Butyrivibrio* or *Pseudobutyrivibrio*) and secondary fermenters (*Selenomonas* or *Acetitomaculum*). *Fibrobacter* is another cellulose-degrader highly prevalent in rumen populations. Other representative rumen bacterial phyla are *Proteobacteria* and *Actinobacteria* **(Mizrahi *et al.*, 2021)**. Methanogenic *Archaea* community is less abundant, although different abundance measures have been reported **(Wallace *et al.*, 2015; Martínez-Álvaro *et al.*, 2020; Xue *et al.*, 2020)**. In general terms, *Methanobrevibacter* from Methanobacteriales is the most abundant representative, with a relative abundance that can reach up to 70% of *Archaea*. In contrast to pig microbiota studies, early research through microscopic counting added to new WMS techniques allowed estimating the abundance of eukaryotic rumen populations. Protozoa are not present in newborn animals, being acquired through direct contact with saliva from other individuals **(Stewart *et al.*, 1988)**, but they are very diverse and usually reach high abundances in rumen, up to half of rumen biomass **(Hungate, 1966; Newbold *et al.*, 2015)**. However, as diversity has been mostly measured through morphological identification and amplicon sequencing, it might still be underestimated. Less information is available for fungi, though it is known that their abundance is far lower than protozoa **(Fouts *et al.*, 2012)**. New research is being carried out in this sense, as the elaboration of rumen databases through metagenomic assembly allows to identify new eukaryotic genomes, thus expanding the knowledge of their true diversity **(Stewart *et al.*, 2019; Watson, 2021; Xie *et al.*, 2021)**.

Microbiota modulation strategies for mitigating methane production have been investigated at different levels. The most direct strategies are mainly focused on methanogen destruction through the utilization of viruses (*Siphoviridae* phages) and vaccines, though this approach is still ineffective, as rumen siphoviruses have not been yet isolated and vaccine effects might be limited due to the extensive and unknown biodiversity encompassed inside methanogens **(McAllister & Newbold, 2008; Mizrahi *et***

***al.*, 2021)**. Other strategies try to reduce the bioavailability of methane precursors by targeting the abundance or activity of microbes up in the trophic chain. One largely studied example is protozoa defaunation, as these microorganisms are responsible of generating high amounts of hydrogen, thus providing substrate to methanogens **(Williams & Coleman, 1992)**. However, defaunation experiments generally show a decrease in organic matter digestibility due to the loss of fibrolytic activity from protozoa **(Newbold *et al.*, 2015)**. The replacement of methanogens by acetogenic bacteria as a way to boost alternative electron sinks has also been tested, but so far reported effects are weak and temporary **(Ungerfeld *et al.*, 2003; Fonty *et al.*, 2007)**. In brief, for the proposed approaches to be truly effective, a better understanding of the structure of the rumen microbiota and its interaction with the host is still needed.

Chapter 2: Justification and objectives

2.1. Study of gut microbiota in Iberian pigs through amplicon sequencing

As stated before, Iberian x Duroc pigs possess higher prolificacy, growth rate and carcass yield than purebred Iberians. However, crossbreeding has a negative impact on the quality and composition of their final products, directly, as meat becomes leaner, and indirectly, as crossbred pigs are usually reared in intensive systems and fat composition lacks the additional unsaturated fatty acids provided by acorns. Moreover, differential development of adipose tissue has been attributed to differences in gene expression **(Óvilo *et al.*, 2014)**. This practice might represent an important issue for the Iberian quality-based market, and research efforts have focused on improving IMF content and organoleptic properties of the meat from crossbred animals, from multiple approaches. For instance, diet optimization is an important research field. Reduction of dietary protein during fattening period has proven effective for increasing IMF content **(Tejeda *et al.*, 2020)**, while vitamin A restriction during early growth phases increases the pig's lipogenic potential **(Ayuso *et al.*, 2015)**. Oleic acid supplementation is far more extended, as modulation of oleic proportion can be used for simulating IMF fatty acid profile resulting from *montanera* finishing, as well as for improving meat quality. The inclusion of oleic-enriched ingredients in diet, i.e., a partial replacement of diet carbohydrates, saturated or polyunsaturated fat with monounsaturated fatty acids, has proven effective to replicate the *montanera* FA profile in Iberian and crossbred pigs **(Daza *et al.*, 2005, 2008)**, this enrichment affecting also muscle gene expression **(Benítez *et al.*, 2019)**. Handling optimization has also been used, mostly for reducing heat stress in farming facilities, which can negatively affect growth through deterioration of intestinal fermentation and reduction of feed intake **(Pardo *et al.*, 2020)**. Lastly, genetic breeding programs are starting to work, due to the crescent interest in pure Iberian production, although they are still scarce. So far, the main genetic selection programme is managed by AECERIBER, mostly targeting ADG and premium cuts' performance **(Nieto *et al.*, 2019)**.

Microbiota research has opened a new door for studying the effects of genetic background and rearing systems on Iberian pig physiology and meat quality. It has been proven that modification in gut microbiota through diet and other environmental factors has an impact on animal phenotype, affecting nutrient acquisition, metabolism, health, productivity and even appetite, thus altering final product's quality. In addition, GM composition has also been related to host genetic background, conferring it a certain degree of heritability. Disentangling this close and complex bond between host and GM is critical to understand the effects of crossbreeding or diet management on the pig phenotype, and might as well offer the possibility to directly modify GM composition to increase fat deposition, composition or feed intake, as it is being proposed for other species **(Dahiya *et al.*, 2017; Olmos Soto, 2021)**.

2.2. Study of rumen microbiota in cattle through whole metagenome sequencing

As the main microorganisms responsible of methane production are well known, methanogens are in the spotlight of microbial modulation experiments in the search for reducing cattle methane emissions. However, their true diversity and abundance are still unknown, and their participation in methane production might be covered by other communities such as ciliates and other protozoa, which are known for engulfing endosymbiotic methanogenic archaea **(Tapio *et al.*, 2017; Hackstein & de Graaf, 2018)**. In addition, other microbes participating in further metabolic pathways are also present in rumen microbiome, including predators, parasites, competitors or co-dependent microbes. In fact, this community possesses a huge diversity and biomass **(Weimer, 2015)**, making it one of the most studied microbiomes and a model case of complex ecological interactions between microbes. Hence, modulation of rumen microbiome becomes a complicated task, as it requires a deep knowledge of its main population structure, gene expression and interspecific relationships to precisely modify a desired parameter, such as methane production, without interfering in the normal functioning of the microbiota and the host.

WMS techniques have opened a new way to characterize microbial communities, both through taxonomy and functionality. Considering the rumen microbiome as a whole is important to shed light on the true extent of microbial interactions. In the context of methane production, it might be useful to unveil whether free methanogens or endosymbiotic methanogens are the main producers of methane in cattle, as well as to identify the main methane precursor sources.

2.3. Compositional analysis of microbiota

As stated before, classical statistical approaches are suboptimal to analyse microbiome data. Their high dimensionality forces to utilize multivariate methods, and their compositional nature involves the use of log-ratio transformations for an optimal treatment of data. High sparsity must be taken into account as well, since zeros might be an important source of bias, and also the application of logarithms is not possible without removing zeros.

Although a gold standard for microbiome analysis has not been established yet, multiple procedures and tools are currently available to patch these hindrances. CLR transformation has been established as a valid method to explore microbiota composition, while zero treatment is usually addressed through replacement by a constant value or by Bayesian imputation. The experiments carried out in this thesis have been performed following these criteria, in order to reduce the effect of spurious correlations and compositionality bias.

2.4. Objectives

This thesis tries to address these challenges from a holobiont perspective, in order to unravel the role of the gut microbiota through different analytical methods, also taking into account the problems of microbiota compositional nature.

Principal objectives of this thesis

1. To develop novel bioinformatic approaches to study the role of microbiota on several phenotypic traits of interest in livestock species, considering the compositional nature of metagenomic data.
2. To identify the main gastrointestinal microbes related to the current productive challenges concerning two main livestock species.
3. To characterize the microbiome-mediated physiological and metabolic mechanisms and pathways related to host relevant traits.

Specific objectives of this thesis

1. Methods:
    a. To identify the best amplicon sequencing post-processing bioinformatics protocol for microbiota studies in livestock species.
2. Iberian pig:
    a. To understand the differences in gut microbiota composition according to two main factors affecting Iberian pig production and quality: genetic background (Iberian versus Duroc genotypes) and diet composition (oleic acid and fibre content), as well as their interaction.
    b. To infer potential relationships between pig gut microbiota composition and the phenotypic and metabolic peculiarities of Iberian pigs.
3. Dairy cattle:
    a. To identify the main rumen microbes involved in methane production.
    b. To disentangle the microbial metabolic pathways driving the methane production.

Chapter 3: Comparison of Mothur and QIIME for the analysis of rumen microbiota composition based on 16S rRNA amplicon sequences

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 December 2018
doi: 10.3389/fmicb.2018.03010

Frontiers in Microbiology

Q1 Microbiology best quartile

SJR 2020 1.7

powered by scimagojr.com

Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences

*Adrian López-García[1], Carolina Pineda-Quiroga[2], Raquel Atxaerandio[2], Adrian Pérez[1,3], Itziar Hernández[1,3], Aser García-Rodríguez[2] and Oscar González-Recio[1,3]**

[1] Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid, Spain, [2] Departamento de Producción Animal, NEIKER-Tecnalia, Vitoria-Gasteiz, Spain, [3] Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Madrid, Spain

Background: Microbiome studies need to analyze massive sequencing data, which requires the use of sophisticated bioinformatics pipelines. Up to date, several tools are available, although the literature is scarce on studies that compare the performance of different bioinformatics pipelines on rumen microbiota when 16S rRNA amplicons are analyzed. The impact of the pipeline on the outcome of the results is also unknown, mainly in terms of the output from studies using these tools as an intermediate phenotype (pseudophenotypes). This study compares two commonly used software (Quantitative Insights Into Microbial Ecology) (QIIME) and mothur, and two microbial gene data bases (GreenGenes and SILVA) for 16S rRNA gene analysis, using metagenome read data collected from rumen content of a cohort of dairy cows.

Results: We compared the relative abundance (RA) of the identified OTUs at the genus level. Both tools presented a high degree of agreement at identifying the most abundant genera: *Bifidobacterium*, *Butyrivibrio*, *Methanobrevibacter*, *Prevotella*, and *Succiniclasticum* (RA > 1%), regardless the database. There were no statistical differences between mothur and QIIME ($P > 0.05$) at estimating the overall RA of the most abundant (RA > 10%) genera, either using SILVA or GreenGenes. However, differences were found at RA < 10% ($P < 0.05$) when using GreenGenes as database, with mothur assigning OTUs to a larger number of genera and in larger RA for these less frequent microorganisms. With this database mothur resulted in larger richness ($P < 0.05$), more favorable rarefaction curves and a larger analytic sensitivity. These differences caused significant and relevant differences between tools at identifying the dissimilarity of microbiotas between pairs of animals. However, these differences were attenuated, but not erased, when SILVA was used as the reference database.

Conclusion: The findings showed that the SILVA database seemed a preferred reference dataset for classifying OTUs from rumen microbiota. If this database was used,

OPEN ACCESS

Edited by:
Garret Suen,
University of Wisconsin-Madison,
United States

Reviewed by:
Irene Wagner-Doebler,
Helmholtz-Zentrum für Infektionsforschung,
Helmholtz-Gemeinschaft Deutscher Forschungszentren (HZ), Germany
Luciano Takeshi Kishi,
São Paulo State University, Brazil
Kimberly Anne Dill-McFarland,
University of British Columbia, Canada

***Correspondence:**
Oscar González-Recio
gonzalez.oscar@inia.es;
ogrecio@gmail.com

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *31 July 2018*
Accepted: *21 November 2018*
Published: *13 December 2018*

Citation:
López-García A, Pineda-Quiroga C, Atxaerandio R, Pérez A, Hernández I, García-Rodríguez A and González-Recio O (2018) Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. Front. Microbiol. 9:3010. doi: 10.3389/fmicb.2018.03010

both QIIME and mothur produced comparable richness and diversity, and also in the RA of most common rumen microbes. However, important differences were found for less common microorganisms which impacted on the beta diversity calculated between pipelines. This may have relevant implications at studying global rumen microbiota.

Keywords: mothur, QIIME, 16S rRNA, rumen microbiota, bovine

INTRODUCTION

Research on ruminal microbiota is becoming increasingly important in dairy cattle as the microbial communities and their genome expression are related to important traits as health condition (Zilber-Rosenberg and Rosenberg, 2008), feed enteric fermentation (Zhou et al., 2009, 2010), or methane emissions (Wallace et al., 2015; Kamke et al., 2016; Roehe et al., 2016). The differences in the microbiota composition have also been proposed as a predictor or proxy of the differences in complex traits and environmental phenotypes (Ross et al., 2013; Kamke et al., 2016). Improving these traits is relevant for farm profitability and sustainability (Basarab et al., 2013; Bell et al., 2013; González-Recio et al., 2014a). Further, there is increasing interest on inferring the host genetic influence on the microbiota composition (Goodrich et al., 2016; Roehe et al., 2016). Tools that accurately estimate the microbial composition are essential to associate microbiota to phenotype variability.

Advances in sequencing technologies allow for obtaining genomic information in a fast and affordable manner. Whole metagenome and rRNA amplicons sequencing provide useful information to characterize the microbial composition in a given environment. Metagenomic information from hypervariable regions in the 16S and 18S ribosomal RNA amplicons are so far preferred in microbiome research due to their lower cost and reasonable accuracy. The results of these kind of studies rely on computational tools that provide accurate characteristics from large data sets of DNA sequences from the community under investigation (Lindgreen et al., 2016). Several authors have reviewed the specifications of different bioinformatics tools to analyze 16S rRNA gene sequences (Lozupone et al., 2005; Nilakanta et al., 2014; Oulas et al., 2015). Among these tools, mothur (Schloss et al., 2009; Kozich et al., 2013) and Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso et al., 2010) are currently two of the most used suits of tools to analyze sequencing information from rRNA amplicons. However, comparisons between these tools on real data sets are scarce. For instance, other authors performed a benchmark study in order to investigate the performance of several tools in terms of microbial taxonomy and function (Lindgreen et al., 2016). These authors applied the methods on synthetic whole sequence metagenomes, which aim to represent the complexities encountered in a non-specific environment. In that study, QIIME resulting on a high specificity at determining the genus level but low sensitivity, whereas mothur was not tested. A recent study evaluated QIIME and mothur in fecal samples collected from preterm infants, showing slight differences in terms of the effective number of genera, richness and relative abundance (RA) detected (Plummer and Twin, 2015). Up to the best of our knowledge, the performance of these tools has not been yet evaluated in aligning rumen metagenome samples to public amplicons databases. Rumen microbiota poses the difficulty that most species have not been yet isolated, and therefore gene data bases may lack of many of the species in the rumen.

The aim of this study was to compare the rumen microbiota composition resulted from two different software: mothur and QIIME, when aligned against GreenGenes (GG) or SILVA databases. The null hypothesis is that the software and data base used to determine the ruminal microbial composition do not impact the results and conclusion from rumen microbiota studies.

RESULTS

This study used sequence data from the hypervariable region V4 of the 16S amplicon from ruminal content in 18 dairy cows. The libraries were generated by means of Nextera kit. The 250 bp paired-end sequencing reactions were performed on a MiSeq platform (Illumina, San Diego, CA, United States). The sequences were processed using the two softwares: QIIME package version 1.9.1 (Caporaso et al., 2010) and mothur version 1.39.5 (Kozich et al., 2013). The RA of 16S rRNA gene reads for each sample and bioinformatics tool was used to infer the taxonomical composition of the samples, taking into account the copy number of 16S genes calculated from each tool. Two reference panels were considered for this purpose: GreenGenes (GG) database (May 2013 version) and SILVA (release 132). The detailed pipeline from each software is shown in **Figure 1**.

After filtering and chimera removal, both tools used a similar number of sequences to cluster ($P > 0.05$), regardless the database used. In average, QIIME left 54,544 reads (SD = 9,041) per animal, whereas mothur worked with 53,790 reads per sample (SD = 7,709). However, mothur clustered these sequences in a larger number of OTUs regardless the database (**Figure 2**). Using QIIME with GG as reference database kept the lowest number of OTUs for classification.

Taxonomical Richness

The performance of each tool was evaluated by looking at the assignment of individual OTUs and the number of genera classified. The RA of genera in each sample was calculated after excluding those genera that appeared at RA $< 0.1\%$ across all samples.

Figure 3 shows the rarefaction curves from each tool. Mothur detected a larger number of OTUs (**Figures 3A,B**) and also of microbial taxa at the genus level (**Figures 3C,D**) ($P < 0.01$) in the samples than QIIME using both GG and SILVA databases.

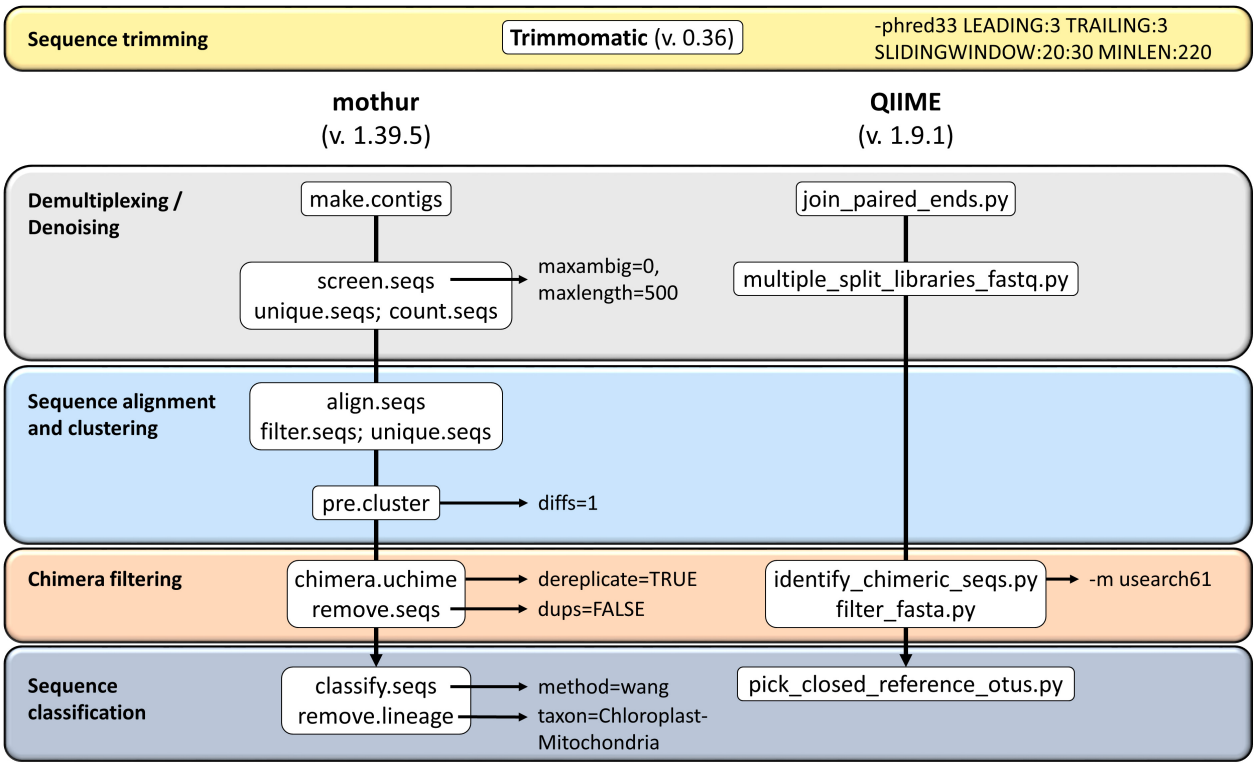

FIGURE 1 | Overview of the workflows used in this study on QIIME and mothur for 16S rRNA amplicons analysis.

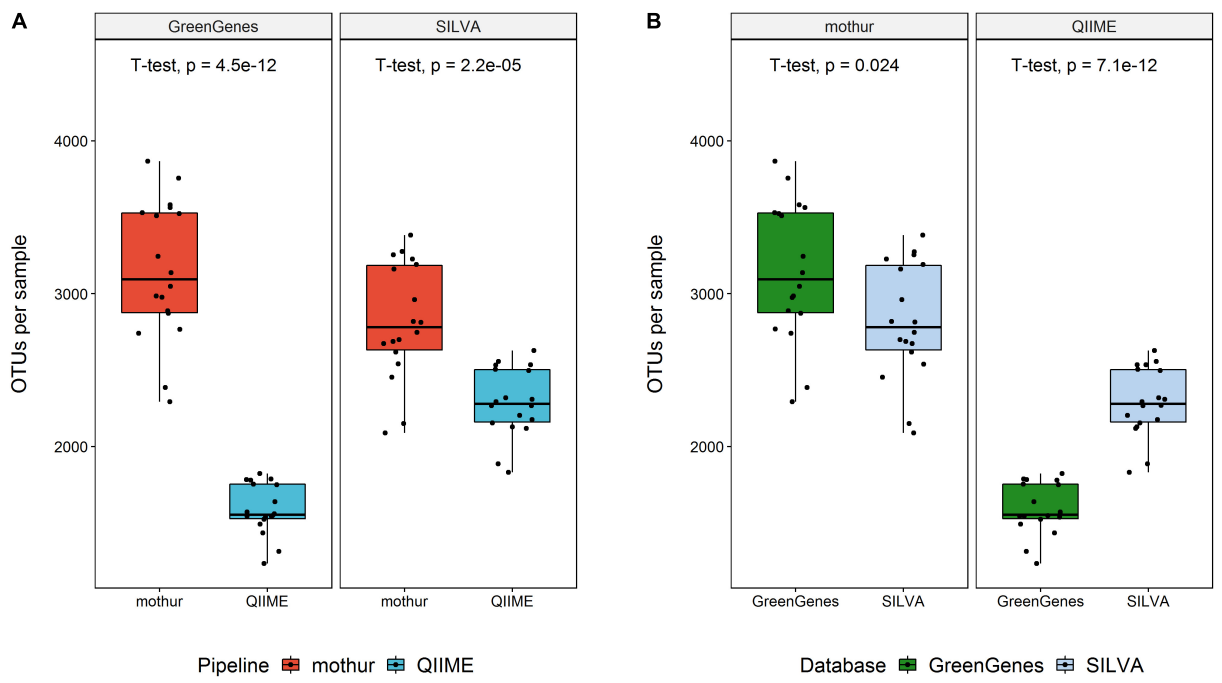

FIGURE 2 | Number of OTUs per sample left for taxonomic classification within reference database (above) or within software (below).

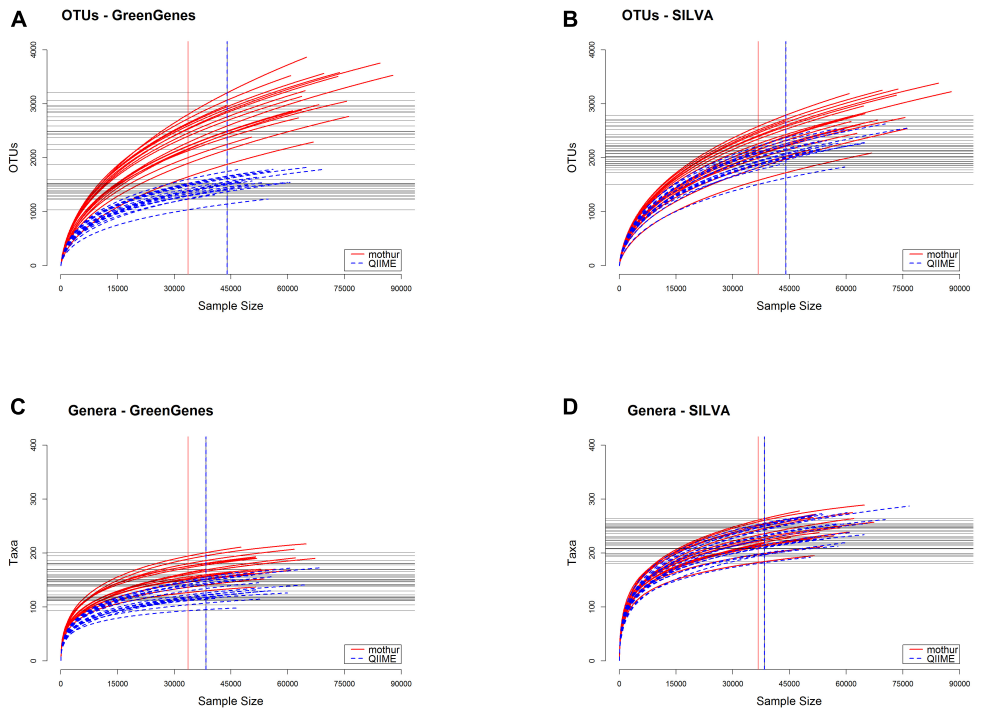

FIGURE 3 | Rarefaction curves on OTUs **(A,B)** or classified genera **(C,D)** for the samples analyzed with each tool prior to filtering by relative abundance <0.1%. **(A,C)** represent curves from GreenGenes as the reference database, whereas **(B,D)** are obtained from SILVA database.

Opposite, QIIME classified a larger number of genera after filtering by RA > 0.1% (**Table 1**). Most of the additional genera encountered by mothur were in very low abundance.

Classification

GreenGenes

On average, mothur clustered a significantly ($P < 0.001$) higher number of OTUs per sample than QIIME. In average per sample, QIIME could not assign 61% ($SD = 2.7$) of clustered OTUs to a known genus, considering known every genus not named as "unclassified," "uncultured," "ambiguous," "unidentified," "unknown," or null, whereas mothur could not assign a larger proportion (67%, $SD = 2.5$) of OTUs. QIIME was less restrictive at assigning OTU to genus level ($P < 0.001$), which might be related with the higher initial number of OTUs clustered by mothur, as we mentioned before. With this database, mothur identified a total of 29 different genera appearing in more than one sample, whereas QIIME assigned 24. Twenty three of these genera were common to both pipelines. The former aligned sequence data to six additional known genera, although most of them appeared in an average RA lower than 0.5%. Three out of these six genera had low representativeness, appearing in less than four out of 18 samples. On the other hand, the only QIIME-exclusive genus, *Bacillus*, had a low average RA and low representativeness, appearing only in three samples. **Table 1** shows the average RA of genera assigned by one or both tools, highlighting that reads that were assigned to a known genus by only one of the tools appeared in very low RA. Both tools were capable of assigning around 99% of reads to any known taxonomy rank belonging to either bacteria or archaea kingdoms.

A scatter plot of the RA estimated by each tool for each genus within sample are shown in **Figure 4**. A strong Pearson correlation (0.996; $P < 0.001$) was found between RA obtained from each tool. Although some small variability can be seen for some samples, there were not statistical differences in the overall RA between tools ($P > 0.05$). However, this disagreement was more evident for microorganisms at RA < 10%, for which significant differences were found between both tools ($P < 0.05$), and these differences were even higher at RA < 1% ($P < 0.01$), and the regression coefficient of RA from QIIME on RA from mothur differed from one, becoming even lower when subsetting the RA dataset (**Table 2**).

SILVA

Mothur also clustered a significantly higher number of OTUs in known taxa than QIIME ($P < 0.001$) when SILVA was the reference database even though mothur filtered out a larger number of reads, but the differences were more attenuated than using GG. These OTUs from mothur were nonetheless classified into a lower number of known taxa than using QIIME (**Table 1**). Both tools identified a total of 52 known genera. Mothur aligned sequence data to three additional exclusive known genera that appeared in more than 1 sample, and QIIME identified 13 genera that did not appeared in mothur. With SILVA as database, mothur could not assign an average of 36.1% per sample ($SD = 1.37$)

TABLE 1 | Total of genera (N) and its average relative abundance (standard deviation within brackets) assigned by each tool (only by QIIME, only by mothur or by both).

Reference database	Total genera assigned			Relative abundance		
	QIIME	Mothur	Mothur and QIIME	QIIME	Mothur	Mothur and QIIME
GreenGenes	1	6	23	0.19 (-)[1]	2.89 (9.67)	2.60 (8.30)
SILVA	13	3	52	0.28 (0.13)	1.90 (6.51)	1.79 (5.67)

Results from each reference data set are presented separately. [1] *No standard deviation calculated with n = 1.*

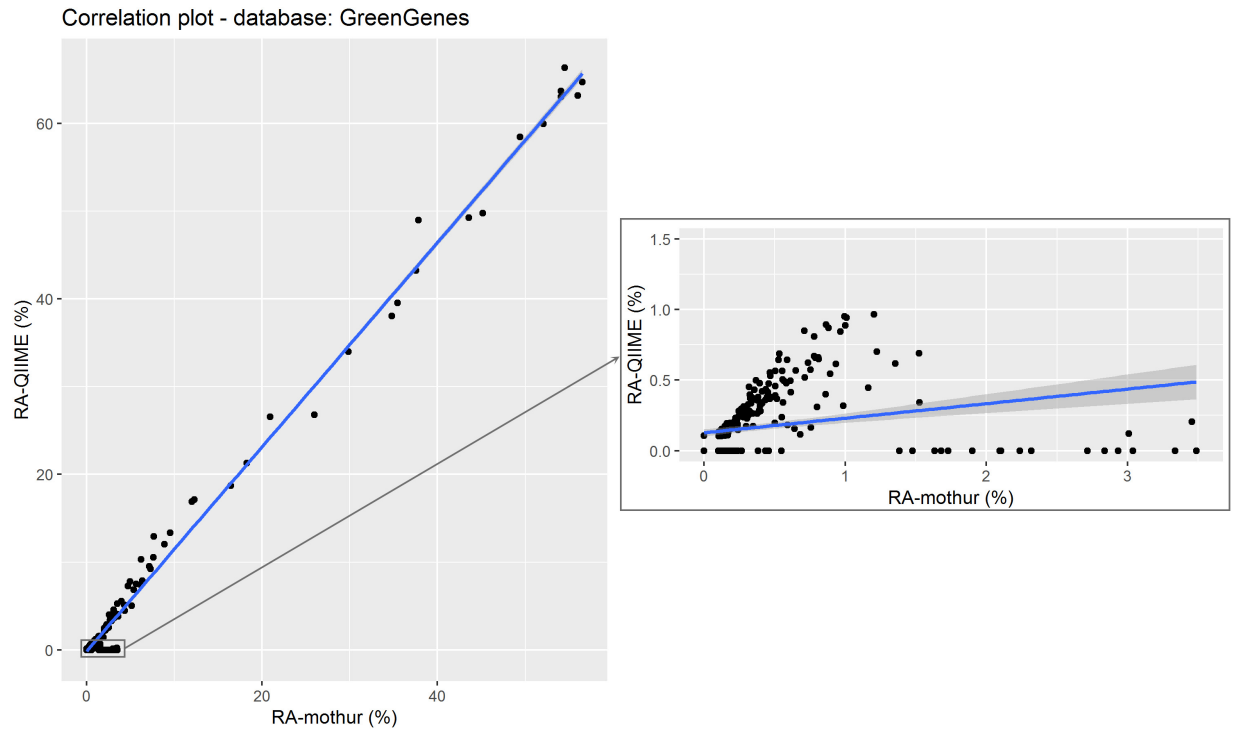

FIGURE 4 | Relative abundance of the different microorganisms (by genera) detected by QIIME and mothur within the 18 samples using GreenGenes as reference data set. The subset shows the correlation between data with QIIME RA < 1%. Points represent individual RA within each sample.

of clustered OTUs to a known genus, but with QIIME only 9.1% (*SD* = 1.36) of OTUs were not assigned to known genera. Thus, mothur appeared to be much more restrictive ($P < 0.001$) at assigning OTUs to genus level when SILVA was used as the reference database.

Figure 5 shows a scatter plot of the RA estimated by each tool. As in the previous case for GG, a strong correlation (0.996; $P < 0.001$) was found between RA obtained from each tool. However, the regression coefficient of RA from QIIME on RA from mothur deviated from 1 at RAs < 10%, although smaller differences were observed compared to GG (**Table 2**). This suggests that mothur detects larger RA of microbes that are present in lower proportion in the rumen.

In summary, both tools were able to classify microorganism from the following genus: *Prevotella*, *Butyrivibrio*, *Succiniclasticum*, *Methanobrevibacter*, *Treponema*, *Bifidobacterium*, *Pseudobutyrivibrio*, *Ruminococcus*, *Mogibacterium*, *Lachnospira*, *Acetobacter*, *Methanosphaera* and *Desulfovibrio*, regardless the database. In addition, other microbes were not identified to genus level, but as members of *Lachnospiraceae* and *Ruminococcaceae* families, regardless the database. The OTUs unable to be classified by QIIME at the genus level were from the *Paraprevotellaceae, Coriobacteriaceae, Prevotellaceae,* and *Succinivibrionaceae* families when GG was used as the reference database and from the *Christensenellaceae* family when SILVA was used as the reference dataset. The OTUs that were not assigned to a genus level by mothur belonged to *Enterobacteriaceae* and *Spirochaetaceae* families using GG, and to *Bacillaceae, Enterobacteriaceae, Erysipelotrichaceae, Family_XIII, Prevotellaceae,* and *Spirochaetaceae* families using SILVA. Also, members from *Bacteroidetes*, *Firmicutes,* and *Proteobacteria* phyla were not assigned to family level when using mothur, regardless the database.

The genera that were identified exclusively by either mothur or QIIME are shown in **Table 3**. This table includes the reference database they were detected with, and previous studies reporting these microbes in rumen microbiota. Other genera were classified by both tools, but not for both databases. Among those with

TABLE 2 | Regression (slope and regression coefficient estimates) of the relative abundance from QIIME on the relative abundance from mothur using GreenGenes or SILVA as reference data set.

	GreenGenes		SILVA	
Reference population	**Coefficient**	R^2	**Coefficient**	R^2
All	1.10**	0.90**	1.14**	0.98**
Relative abundance < 10%	0.50**	0.31**	0.86**	0.84**
Relative abundance < 1%	−0.03**	0.05**	0.12**	0.14**

**$P < 0.001$.

RA > 0.5% we found *YRC22* and *Clostridium* when GG was the reference database, and *Acetitomaculum*, *Saccharofermentans*, *Schwartzia*, *Candidatus_Saccharimonas* and some groups from families *Ruminococcaceae*, *Christensenellaceae*, *Rickenellaceae*, *Lachnospiraceae,* and *Prevotellaceae* when SILVA was used. Five taxa were identified for any combination of tool and database that have not been reported in rumen so far: *Eubacterium hallii group, Eubacterium_nodatum_group, Ruminococcaceae UCG-011, Ruminococcus gauvreanii group,* and *Prevotella P9*.

Diversity

Beta-diversity was calculated to investigate the dissimilarity between rumen microbiotas within tool. Results clearly clustered by software at taxonomical levels of genus, family and Phylum, regardless the reference database used (**Figure 6**). This figure also shows that the dissimilarities between samples were larger between than within software at lower taxonomic levels (genus and family), whereas distances at the phylum level were similar between and within software.

Computation requirements were not reported here as they greatly depend on the computational strategy applied in terms of parallelization, available number of cores, and the bioinformatician's creativity to design more efficient pipelines.

DISCUSSION

The results of this study support previous research showing differences between bioinformatics tools analyzing 16S rRNA amplicons. The number of OTUs and the taxonomic classification resulted thereof was compared between mothur and QIIME.

The results herein show that the final number of taxa and their overall RAs are not statistically different between tools using SILVA as reference database, but beta-diversity between samples clustered together by software used. Mothur worked with a larger number of OTUs, and these were classified into a larger number of genera than by QIIME when GG was the reference database. Plummer and Twin (2015) showed larger richness (total number of different genera) using QIIME compared to mothur, using the same reference database for alignment, although they used human preterm gut samples.

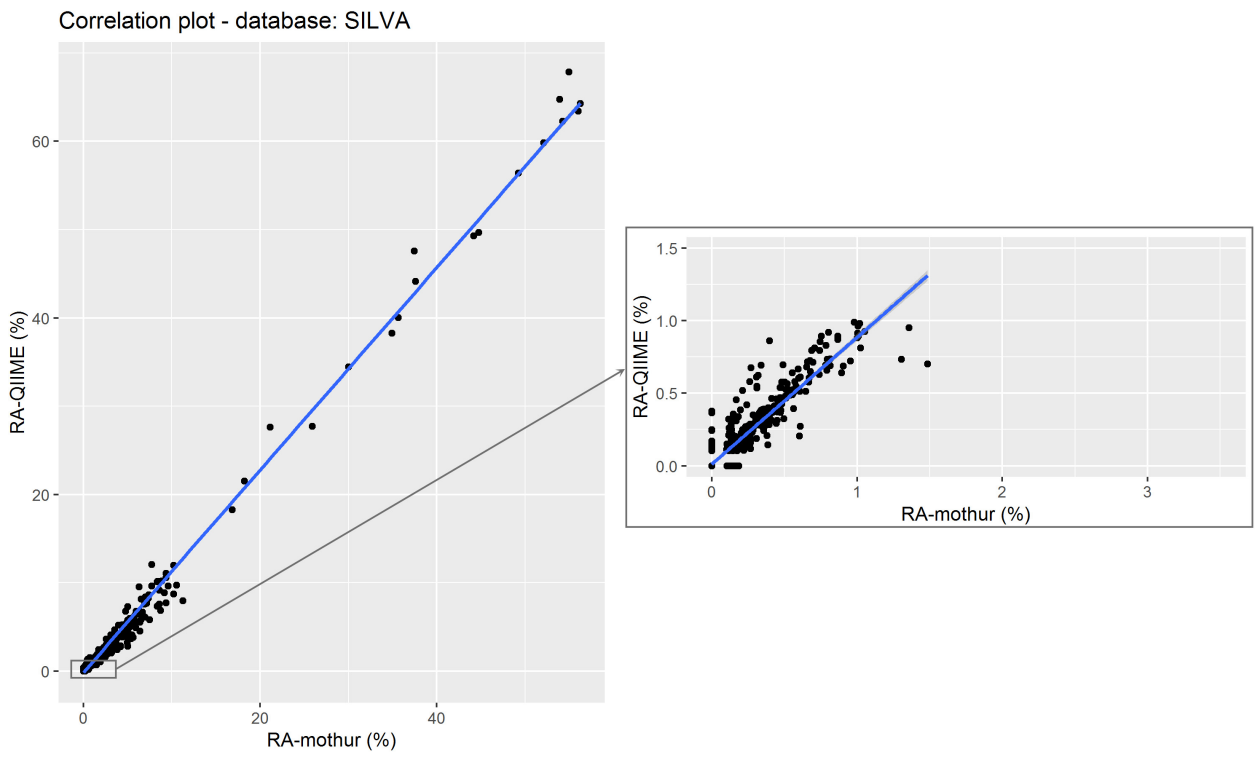

FIGURE 5 | Relative abundance of the different microorganisms (by genera) detected by QIIME and mothur within the 18 samples using SILVA as reference data set. The subset shows the correlation between data with QIIME RA < 1%. Points represent individual RA within each sample.

TABLE 3 | Genera identified exclusively by mothur or QIIME, their function or activity in the rumen (if known), the reference database it was identified from, and information source or reference.

Genus	Function/activity in rumen	Reference database	Tool	Previous source(s)
p-75-a5	Detected in ruminal liquid fraction	GreenGenes	mothur	Jewell et al., 2015
SHD-231	Detected in rumen. Reduced in diets containing linseed diets	GreenGenes	mothur	de Carvalho et al., 2017
Lachnospira bacterium FD2005	Detected in rumen	SILVA	mothur	Azevedo et al., 2015.
Papillibacter	Detected in rumen. Cellulose- degrading bacteria	SILVA	mothur	Zhang et al., 2014
Ruminococcus	Cellulolytic bacteria	SILVA	QIIME	Wallace et al., 2015.
Bacillus	Amylolytic bacterium	GreenGenes, SILVA	QIIME	Gallo et al., 2016
Eubacterium_ cellulosolvens_group	Fibrolytic (Sika deer) Detected in sheep rumen	SILVA	QIIME	Li et al., 2013 Azevedo et al., 2015
Eubacterium_coprosta noligenes_group	Detected in rumen	SILVA	QIIME	Tong et al., 2018 Popova et al., 2017
Eubacterium_ ruminantium_group	Present in rumen with an appropriate balance of degradable protein and carbohydrates	SILVA	QIIME	Abdelmegeid et al., 2018
Eubacterium_ ventriosum_group	Present in forestomach (Alpacas and Sheep)	SILVA	QIIME	Abdelmegeid et al., 2018
Lachnospiraceae_ NK4A136_group	Detected in rumen	SILVA	QIIME	Azevedo et al., 2015
Roseburia	Adherent bacteria community involved in plant degradation	SILVA	QIIME	Huws et al., 2016

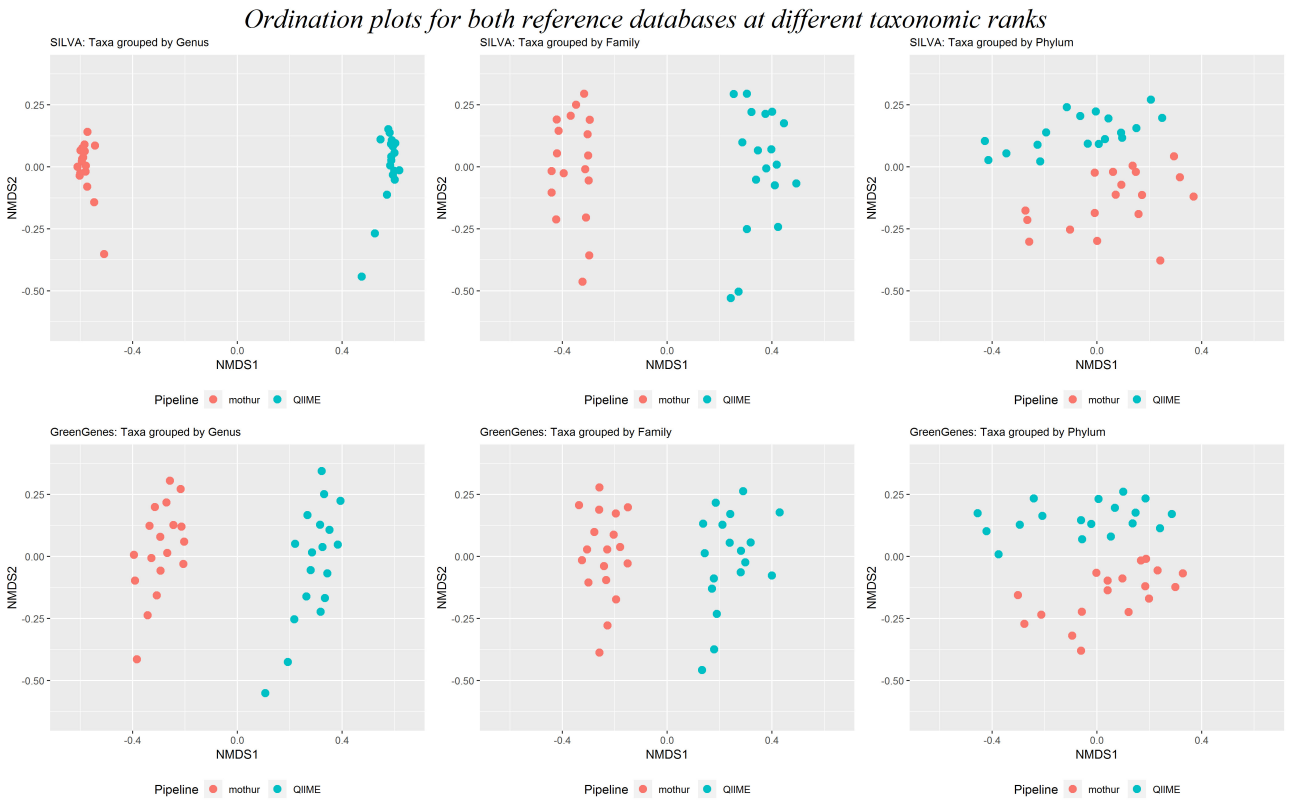

FIGURE 6 | Two first dimensions of a non-metric multidimensional scaling for samples analyzed with mothur (red dots) or QIIME (blue dots) using either SILVA (above) or GreenGenes (below) as the reference dataset at the genus, family, and phylum taxonomical levels.

They also showed statistical differences between tools on the comparison for diversity within samples. Our results show that using more updated SILVA releases mitigated these differences in terms of richness and diversity, suggesting that not only the implemented pipeline/program strongly influences diversity results, but database should also be considered in microbiota analysis. Therefore, it must be pointed out that comparison between pipelines in terms of number of OTUs must be done

within database, and in terms of number of OTUs remaining for classification. It is also worth to highlight that in the current study both pipelines utilize a reference database for chimera filtering as well as a differentiated OTU labeling, making the comparison for raw OTUs unfeasible. Analyzing OTUs instead of taxonomical levels might be of interest in some circumstances. The results at the taxa level showed differences between tools and databases, which may be extended at the more specific level of OTUs.

It must be pointed out that the objective of this study was not to determine what tool provides a more accurate picture of the true microbiota composition, since the latter is unknown in cultivated rumen samples. There is no gold standard microbiota with known composition as many of the microbes in the rumen cannot be isolated or have not been yet cultivated. Sequencing the 16S rRNA gene poses the limitation that closely related microbes can be indistinguishable as they harbor almost identical sequences at this amplicon, and the different tools handle these drawbacks differently. A favorable pipeline should maximize sensitivity with a minimum impair in specificity. According our results, we infer that mothur provides higher diversity than QIIME regardless the reference database. There were five taxonomy groups detected in our study that have not been reported in rumen microbiota before. The rumen microbiota is largely underrepresented in databases and most of them have not been cultured yet (Stewart et al., 2018). Therefore, we consider that these groups of microorganisms might be considered as new candidates, although it must be corroborated in future studies through deep sequencing analysis or culture isolation. If they are true positive, they may be potential candidates to create mock communities to challenge bioinformatics tools. Lindgreen et al. (2016) showed low sensitivity scores and an impaired prediction performance with QIIME using a benchmark metagenome. However, as noted by these authors, QIIME uses custom databases that only contain specific marker sequences such as 16S rRNA. Therefore, performance of QIIME in whole metagenome analyses cannot be extrapolated to 16S or 18S amplicons studies. In addition, mothur was not used in such a study.

There are two main differences between mothur and QIIME: the OTU clustering algorithm and the algorithm for taxonomic classification. The alignment and clustering processes differ between software, as well as the chimera detection. Mothur handles the taxonomic classification using a naïve Bayesian classifier using a pseudo-bootstrapping to generate a confidence score, which must be over 80% to assign a read to a given taxonomy (Wang et al., 2007). QIIME uses the usearch algorithm to find the closest match in a reference data base (Edgar, 2010), which has been reported problematic at identifying the closest reference because it is sensitive to the order of the reference sequences as they can be identical over the region being considered (Westcott and Schloss, 2015). Usearch shows a high level of sensitivity to detect reference sequences, however, the specificity of those matches was poor relative to the true best match. High error rates have been previously reported with GG, and could be substantially improved by randomizing the sequences (Westcott and Schloss, 2015). Further the poorly GreenGenes-aligned sequences artificially increases the distances between sequences, which may also impair the accuracy of the classification. QIIME uses a closed reference pick up strategy in a single step, which implies some difficulties at disentangling all dissimilarities with mothur. The way mothur is implemented here resembles a pseudo open-reference pick up, because there were a previous step of chimera filtering and a known reference database was used for classification. These arguments might explain the poorest performance of QIIME in our data set when GG was used as the reference data set.

This is a proof of principle analysis showing how the choice of bioinformatics pipeline and the reference data set can impact the analysis of 16S rRNA gene sequencing data from rumen microbiotas. Nonetheless, the bio-informatics tools could perform differently in samples from different sites as different body parts may host different taxonomic composition, making the algorithms more or less efficient at detecting the true composition.

In the light of the results obtained in this study we can conclude that the impact of the tool is relatively small in terms of richness as a more updated and comprehensive reference database is used. SILVA seems to be a preferred reference data set as a larger number of different genera were identified, and more consistent results were obtained between tools. SILVA is a more updated database, whereas GG has not been updated since 2013. However, differences were detected in terms of beta-diversity, and differences between pipelines were obtained for microbes in lower abundances, yet belonging to the core microbiome. In this sense, mothur showed larger sensitivity at detecting microorganisms that can potentially populate the cow rumen. This may be important, as differences in the RA of less frequent groups of microbes may be relevant. These differences affect the microbiota similarity between samples or individuals. In turn, this would affect the phenotypic variance of a complex trait explained by the microbiota using mixed models that accounted for the microbiota composition as a random effect with a covariance structure given by these similarities between samples. Performance of mixed model methodology under Best Linear Unbiased Prediction or any other Reproducing Kernel Hilbert Space scenario greatly depends on the structure of the covariance or kernel matrix used as reported in González-Recio et al. (2009, 2014b). Incorrect or improper microbiotas similarity matrices might bias the proportion of variance explained by the microbiota or genetic correlation estimates between host genome and its metagenome. Up to the best of our knowledge, there is no proof of concept to determine what tool provides a more suitable similarity matrix. The degree of aimed microbiota specificity may affect the choice of the pipeline. Mock communities that mimic the true composition of rumen microbiota are not yet available. This study also highlights the necessity to create benchmark samples with a known composition of cultivated ruminal microorganisms to evaluate different bioinformatics tools, as well as the convenience of including more rumen specific communities into the gene databases. In this sense, those mock samples could include the genera that have been detected by only one of the tools (**Table 3**). Moreover, It must be consider that the samples used in this study combined the four possible ruminal fractions and the RAs

in the samples might differ from the true composition in the rumen. Nonetheless, this is not expected to affect the comparison between pipelines.

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. An ethics committee was not necessary in this case because it was conducted on pre-existing data from a previous trial based on routine management practices in commercial farms. Data used in this study were described in Gonzalez-Recio et al. (2018). In brief, samples were obtained from ruminal content from 18 cows from 2 breeds (10 Holstein and 8 Brown Swiss) allocated in the Fraisoro Farm School (Zizurkil, Gipuzkoa, Spain). Ruminal samples were collected from each dairy cow using a stomach tube connected to a mechanical pumping unit. About 100 ml of each ruminal extraction were placed into a container and were frozen immediately after the extraction and then stored at -20 ± 5°C until analysis. Samples were gradually thawed overnight at refrigeration (5 ± 3°C) and squeezed through four layers of sterile cheesecloth to separate solid (solids with a particle size smaller than the diameter of the sampling tube) from liquid digesta phases. This latter phase was subsequently separated into planktonic organisms and bacteria associated with the liquid fraction. The solid phase was separated in associated and adherent fractions. Fractionation procedures were carried out following the methodology described in Yu and Foster (2005). The four fractions were lyophilized and combined to obtain a unique sample with the four fractions represented proportionally (on dry matter basis).

After composition, DNA extraction was performed using the commercial Power Soil DNA Isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA, United States) following manufacturer's instructions. The extracted DNA was subjected to paired-end Illumina sequencing of the V4 hypervariable region of the 16S rRNA. Universal bacterial 16S rRNA gene primers (515F: 5′-GTGCCAGCMGCCGCGGTAA-3′ and 806R: 5′-GGACTACHVHHHTWTCTAAT-3′ (Caporaso et al., 2011) were used to generate the bacterial amplicon libraries (expected amplicon size 250 bp). The libraries were generated by means of Nextera kit. The 250 bp paired-end sequencing reactions were performed on a MiSeq platform (Illumina, San Diego, CA, United States). Data are publicly available at http://www.ebi.ac.uk/ena/data/view/PRJEB26635.

Sequences were pre-processed using Trimmomatic tool (v 0.36) (Bolger et al., 2014). Sequences below 220 bp in length and average quality score below 30 on a window of 20 bases were discarded. In total, 3,261,168 reads were analyzed. The remaining sequence data were then processed using the two softwares: QIIME package version 1.9.1 (Caporaso et al., 2010) and mothur version 1.39.5 (Schloss et al., 2009; Kozich et al., 2013). In the case of QIIME, forward and backward reads were joined with join_paired_ends.py. Chimeras were identified and filtered using *usearch* method (Rognes et al., 2016). Finally, the tool was used to pick closed-reference OTUs from the GreenGenes database (May 2013 version) or SILVA database (Quast et al., 2013)[1] (release 132) and representative sequences with a 99% of similarity were kept. The pipeline for mothur also began by joining forward and backward reads. Chimeras and unique sequences were removed using UCHIME (Edgar et al., 2011). Sequences were then preclustered, and finally classified using the default method (naïve Bayesian classifier; Wang et al., 2007) on *classify.seqs()*, with the same cut-off for sequence identity and reference databases as above. OTUs were summarized at phylum, class, order, family, and genus. Phylogenetic groups with an abundance lower than 0.1% in all samples were excluded from the final analyses. The pipelines used can be found in a git-hub repository[2].

All statistical analyses were performed in R v3.5.1 (R Core Team, 2015). When Pearson correlation was calculated, the statistical significance was tested using the cor.test() command from the base package.

Filtering and Chimera Removal

Differences in the number of sequences left after chimera removal from each tool (mothur vs. QIIME) was computed using a least squared mean regression.

The linear model was:

$$\mathbf{y} = \mu + \mathbf{x}\beta + \mathbf{e}$$

where $\mathbf{y}$ was the vector of the number of reads left for each sample after filtering and chimera removal with either mothur or QIIME ($n = 2 \times 18$), μ is the intercept, $\mathbf{x}$ is the incidence vector assigning each record to the corresponding tool (mothur vs. QIIME), β is the coefficient estimate, and $\mathbf{e}$ is the vector of residuals assumed to be independently and identically normally distributed. The level of significance was set to $\alpha = 0.05$.

Richness and Relative Abundance

Differences between mothur vs. QIIME were computed using a simple generalized linear model. Sequence reads from each sample ($n = 18$) were analyzed with the mothur or QIIME pipelines, and using either SILVA or GG databases. The statistical analysis for the resulting richness and RAs were computed within database as follows:

$$\mathbf{y} = \mu + \mathbf{x}\beta + \mathbf{e}$$

where $\mathbf{y}$ was the vector of number of microbial taxa at the genus level (or their RA) assigned either with mothur or QIIME ($n = 2 \times 18$) using GG, μ is the intercept, $\mathbf{x}$ is the incidence vector assigning each record to the corresponding tool (mothur vs. QIIME), β is the coefficient estimate, and $\mathbf{e}$ is the vector of residuals assumed to be independently and identically normally distributed. The level of significance was set to $\alpha = 0.05$.

[1] http://www.arb-silva.de

[2] https://github.com/alopgar/16S-analysis/tree/master/16S-mothur-and-QIIME-pipeline-comparison

Further, the same statistical analysis was performed using the RAs obtained with SILVA as the reference data base.

Similarly, the number of unclassified reads from each tool within reference database were analyzed using the same model as above.

Dissimilarity Matrix and Principal Component Analyses

Non-metric multidimensional scaling (nMDS) was performed to explore the ruminal community structure, using the phyloseq package (v 1.24.2). The ordinate function was used to estimate dissimilarities using Bray–Curtis distances. Plot_ordination was used to plot these dissimilarities between mothur and QIIME pipelines with either SILVA or GreenGenes as the reference databases, grouping taxa by genus, family, and phylum levels.

AUTHOR CONTRIBUTIONS

AL-G, AP, and IH selected the steps in the pipelines for mothur and Qiime and analyzed the sequence files. AG-R, CP-Q, and RA executed the animal experiments, collected and analyzed the samples, discussed the results and helped to write the manuscript. AL-G and OG-R implemented the statistical analyses. OG-R designed the experiment and wrote the first draft of the manuscript. AL-G, CP-Q, RA, AG-R, and OG-R discuss the results. All authors read and approved the final manuscript.

FUNDING

Funding from grants INIA RTA2012-00065-C04, RTA2015-00022-C3-02, and LIFE SEEDCAPITAL-12 ENV/ES/590 is acknowledged.

ACKNOWLEDGMENTS

The authors wish to thank Iurancha González for assistance during sampling and Ana Hurtado, Beatriz Oporto, and Medelin Ocejo for DNA preparations, as well as the staff from Escuela Agraria de Fraisoro (Basque Department of Environment, Rural Development, Agriculture and Fishery).

REFERENCES

Abdelmegeid, M. K., Elolimy, A. A., Zhou, Z., Lopreiato, V., McCann, J. C., and Loor, J. J. (2018). Rumen-protected methionine during the peripartal period in dairy cows and its effects on abundance of major species of ruminal bacteria. *J. Anim. Sci. Biotechnol.* 9:17. doi: 10.1186/s40104-018-0230-8

Azevedo, A. C., Bento, C. B. P., Ruiz, J. C., Queiroz, M. V., and Mantovania, H. C. (2015). Distribution and genetic diversity of bacteriocin gene clusters in rumen microbial genomes. *Appl. Environ. Microbiol.* 20, 7290–7304. doi: 10.1128/AEM.01223-15

Basarab, J. A., Beauchemin, K. A., Baron, V. S., Ominski, K. H., Guan, L. L., Miller, S. P., et al. (2013). Reducing GHG emissions through genetic improvement for feed efficiency: effects on economically important traits and enteric methane production. *Animal* 7(Suppl. 2), 303–315. doi: 10.1017/S1751731113000888

Bell, M. J., Eckard, R. J., Haile-Mariam, M., and Pryce, J. E. (2013). The effect of changing cow production and fitness traits on net income and greenhouse gas emissions from Australian dairy systems. *J. Dairy Sci.* 96, 7918–7931. doi: 10.3168/jds.2012-6289

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl.), 4516–4522. doi: 10.1073/pnas.1000080107

de Carvalho, I. P. C., Fiorentini, G., de Souza Castagnino, P., de Jesus, R. B., Messana, J. D., Granja-Salcedo, Y. T., et al. (2017). Supplementation with lipid sources alters the ruminal fermentation and duodenal flow of fatty acids in grazing Nellore steers. *Anim. Feed Sci. Technol.* 227, 142–153. doi: 10.1016/j.anifeedsci.2017.02.017

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Gallo, A., Giuberti, G., Duval, S., Moschini, M., and Masoero, F. (2016). Short communication: the effect of an exogenous enzyme with amylolytic activity on gas production and *in vitro* rumen starch degradability of small and large particles of corn or barley meals. *J. Dairy Sci.* 99, 3602–3606. doi: 10.3168/jds.2015-9904

González-Recio, O., Gianola, D., Rosa, G. J., Weigel, K. A., and Kranis, A. (2009). Genome-assisted prediction of a quantitative trait measured in parents and progeny: application to food conversion rate in chickens. *Genet. Sel. Evol.* 41:3. doi: 10.1186/1297-9686-41-3

González-Recio, O., Pryce, J. E., Haile-Mariam, M., and Hayes, B. J. (2014a). Incorporating heifer feed efficiency in the Australian selection index using genomic selection. *J. Dairy Sci.* 97, 3883–3893. doi: 10.3168/jds.2013-7515

González-Recio, O., Rosa, G. J. M. M., and Gianola, D. (2014b). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest. Sci.* 166, 217–231. doi: 10.1016/j.livsci.2014.05.036

Gonzalez-Recio, O., Zubiria, I., García-Rodríguez, A., Hurtado, A., and Atxaerandio, R. (2018). Short communication: signs of host genetic regulation in the microbiome composition in two dairy breeds: Holstein and Brown Swiss. *J. Dairy Sci.* 101, 2285–2292. doi: 10.3168/jds.2017-13179

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., et al. (2016). Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 19, 731–743. doi: 10.1016/j.chom.2016.04.017

Huws, S. A., Edwards, J. E., Creevey, C. J., Rees Stevens, P., Lin, W., Girdwood, S. E., et al. (2016). Temporal dynamics of the metabolically active rumen bacteria colonizing fresh perennial ryegrass. *FEMS Microbiol. Ecol.* 92:fiv137. doi: 10.1093/femsec/fiv137

Jewell, K. A., McCormick, C., Odt, C. L., Weimer, P. J., and Suen, G. (2015). Ruminal bacterial community composition in dairy cows is dynamic over the course of two lactations and correlates with feed efficiency. *Appl. Environ. Microbiol.* 81, 4697–4710. doi: 10.1128/AEM.00720-15

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., et al. (2016). Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4:56. doi: 10.1186/s40168-016-0201-2

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

Li, Z. P., Liu, H. L., Li, G. Y., Bao, K., Wang, K. Y., Xu, C., et al. (2013). Molecular diversity of rumen bacterial communities from tannin-rich and fiber-rich forage fed domestic Sika deer (*Cervus nippon*) in China. *BMC Microbiol.* 13:151. doi: 10.1186/1471-2180-13-151

Lindgreen, S., Adair, K. L., and Gardner, P. P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Sci. Rep.* 6:19233. doi: 10.1038/srep19233

Lozupone, C., Knight, R., Firrell, S., Foulkes, M. A., Jablonski, K. A., Collman, R., et al. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Nilakanta, H., Drews, K. L., Firrell, S., Foulkes, M. A., and Jablonski, K. A. (2014). A review of software for analyzing molecular sequences. *BMC Res. Notes* 7:830. doi: 10.1186/1756-0500-7-830

Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., et al. (2015). Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform. Biol. Insights* 9, 75–88. doi: 10.4137/BBI.S12462

Plummer, E., and Twin, J. (2015). A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *J. Proteomics Bioinform.* 8, 283–291. doi: 10.4172/jpb.1000381

Popova, M., McGovern, E., McCabe, M. S., Martin, C., Doreau, M., Arbre, M., et al. (2017) The structural and functional capacity of ruminal and cecal microbiota in growing cattle was unaffected by dietary supplementation of linseed oil and nitrate. *Front. Microbiol.* 8:937. doi: 10.3389/fmicb.2017.00937

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Available at: https://www.r-project.org/

Roehe, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., McKain, N., Ross, D. W., et al. (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:e1005846. doi: 10.1371/journal.pgen.1005846

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584

Ross, E. M., Moate, P. J., Marett, L. C., Cocks, B. G., and Hayes, B. J. (2013). Metagenomic predictions: from microbiome to complex health and environmental phenotypes in humans and cattle. *PLoS One* 8:e73056. doi: 10.1371/journal.pone.0073056

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source. platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* 9:870. doi: 10.1038/s41467-018-03317-6

Tong, J., Zhang, H., Yang, D., Zhang, Y., Xiong, B., and Jiang, L. (2018) Illumina sequencing analysis of the ruminal microbiota in high-yield and low-yield lactating dairy cows. *PLoS One* 13:e0198225. doi: 10.1371/journal.pone.0198225

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C.-A., Hyslop, J. J., Ross, D. W., et al. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16:839. doi: 10.1186/s12864-015-2032-0

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Westcott, S. L., and Schloss, P. D. (2015). De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ* 3:e1487. doi: 10.7717/peerj.1487

Yu, Z., and Foster, R. (2005). *Methods in Gut Microbial Ecology for Ruminants*. Berlin: Springer.

Zhang, R., Zhu, W., Zhu, W., Liu, J., and Mao, S. (2014). Effect of dietary forage sources on rumen microbiota, rumen fermentation and biogenic amines in dairy cows. *J. Sci. Food Agric.* 94, 1886–1895. doi: 10.1002/jsfa.6508

Zhou, M., Hernandez-Sanabria, E., and Guan, L. L. (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

Zhou, M., Hernandez-Sanabria, E., and Guan, L. L. (2010). Characterization of variation in rumen methanogenic communities under different dietary and host feed efficiency conditions, as determined by PCR-denaturing gradient gel electrophoresis analysis. *Appl. Environ. Microbiol.* 76, 3776–3786. doi: 10.1128/AEM.00010-10

Zilber-Rosenberg, I., and Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* 32, 723–735. doi: 10.1111/j.1574-6976.2008.00123.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter 4: Influence of genetic background and dietary oleic acid on gut microbiota composition in Duroc and Iberian pigs

PLOS ONE

OPEN ACCESS

Citation: López-García A, Benítez R, Núñez Y, Gómez-Izquierdo E, de Mercado E, García-Casco JM, et al. (2021) Influence of genetic background and dietary oleic acid on gut microbiota composition in Duroc and Iberian pigs. PLoS ONE 16(5): e0251804. https://doi.org/10.1371/journal.pone.0251804

Editor: Martina Zappaterra, University of Bologna, ITALY

Received: January 12, 2021

Accepted: May 3, 2021

Published: May 20, 2021

Data Availability Statement: Raw microbial sequence data are available from the ENA repository (accesion number PRJEB42303) (https://www.ebi.ac.uk/ena/browser/view/PRJEB42303).

Funding: This research was funded by the Comunidad de Madrid (Spain) and European Structural and Investment Funds (project MEDGAN-CM S2013/ABI-2913), and the Spanish Ministry of Science and Innovation (grant

RESEARCH ARTICLE

Influence of genetic background and dietary oleic acid on gut microbiota composition in Duroc and Iberian pigs

Adrián López-García[1]*, Rita Benítez[1], Yolanda Núñez[1], Emilio Gómez-Izquierdo[2], Eduardo de Mercado[2], Juan M. García-Casco[1], Óscar González-Recio[1], Clemente López-Bote[3], Jordi Estellé[4], Cristina Óvilo[1]

1 Department of Animal Breeding, INIA, Madrid, Spain, **2** Pig Test Center ITACYL, Hontalbilla, Segovia, Spain, **3** Department of Animal Production, Veterinary Faculty, UCM, Madrid, Spain, **4** GABI, INRAE, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

* adrian.lopez@inia.es

Abstract

Background

Phenotypic variability for productive and meat quality traits has been largely studied in Iberian pigs, especially in genetic selection and nutritional experiments. Complex interactions among genetic background, diet composition and gut microbiota hinder the correct assessment of each factor's contribution on phenotypes. In order to disentangle these interactions, we evaluated changes in gut microbiota composition comparing 48 Iberian and Duroc pigs fed diets with different energy source (standard diet with carbohydrates vs sunflower oil-enriched diet with high oleic acid content).

Results

A higher richness was observed for Iberian pigs ($p < 0.05$) and compositional analysis was applied for beta-diversity, differential abundance and pairwise log-ratio analyses. We found significant differences in overall microbiota composition between breeds, and also between diets inside breeds, to a lesser extent. Differential abundance analysis revealed that Duroc animals have more proportion of Actinobacteria and *Prevotella*, while Iberian replace those microorganisms with other more variable taxa. According to dietary differences, high-oleic fed animals were richer in *Prevotella*. We also found microbial ratios capable of separating animals by breeds and diets, mostly related to Actinobacteria.

Conclusion

This study reveals that both genetic background and diet composition might have a relevant impact in gut microbiota composition. The application of compositional data analysis has facilitated the identification of microorganisms and ratios as possibly related to metabolic changes due to genetic background and, to a lower extent, to dietary changes. This may

PID2019-108695RB-C31). A.L.G. was funded by FPI-INIA grant with reference FPI-SGIT2016-06. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AD, Abundance decrease; **AI**, Abundance increase; **ATO/ENT**, *Atopobiaceae/Enterobacteriaceae*; **BW**, Body weight; **CLR**, Centred log-ratio; **COR/SPR**, *Corynebacteriaceae/Spirochaetaceae*; **DA**, Differential abundance; **FA**, Fatty acid; **FC**, Fold-change; **FDR**, False discovery rate; **IMF**, Intramuscular fat; **MUFA**, Monounsaturated fatty acids; **NR**, no response; **OA**, Over-abundant; **OTU**, Operational taxonomic unit; **PCA**, Principal Component Analysis; **PEP/STR**, *Peptostreptococcaceae/Streptococcaceae*; **PERMANOVA**, Permutational multivariate analysis of variance; **p-LR**, Pairwise log-ratio; **PUFA**, Polyunsaturated fatty acids; **RA**, Relative abundance; **RPART**, Recursive Partitioning and Regression Trees; **SCFA**, Sort-chained fatty acids; **SFA**, Saturated fatty acids; **sPLS-DA**, sparse Partial Least Square Discriminant Analysis; **UA**, Underabundant; **VIP**, Variable Influence of Projection.

lead to a relevant progress in the knowledge of interactions between pig genetics, environment and gut microbiota.

Introduction

The contribution of Spanish pig production to worldwide pork market is meaningful. According to FAOSTAT, Spain produces almost 4% of worldwide swine meat, being the fourth producer, and the second one inside EU, with a 19% of communal production, after Germany. One of the most eye-catching contribution of Spanish pig production to international markets comes from Iberian pig, a local breed whose high-quality final products are much appreciated. In fact, Spain ranks as the third country in lard and ham exportations (data from 2018) [1], with a crucial contribution of dry-cured Iberian ham. The high value of this product is due to the unique organoleptic properties conferred by Iberian pig genetic, phenotypic and production particularities. This rustic and lowly-prolific local breed has been traditionally raised under a semi-extensive production system in the "Dehesa" ecosystems, taking advantage of the grass and acorn resources available [2]. The combination of their natural high tendency to fat accumulation and the acorn and grass feeding in the late fattening phase ("Montanera"), confers their final products a unique fatty acid (FA) profile composition and a higher intramuscular fat (IMF) content than commercial breeds.

The interest in this high-quality market has led to the search and legal regulation of strategies for the improvement of its efficiency and productivity. In this sense, one of the most common practices nowadays is crossbreeding with Duroc, resulting in a negative impact on final product composition and quality [3]. Crossbreeding consequences have been largely studied through different points of view—such as phenotypic traits, meat quality, genetic markers or gene expression [4–6]. As the high quality of Iberian products has become the flag for the Iberian pig market to enter international markets, producers keep an interest in improving their products' organoleptic attributes through FA profile and IMF content optimization [7, 8], especially in the recent more intensive production systems. Several strategies such as genetic selection, handling optimization and diet design, are methods of modifying final product composition. Strategies such as dietary FA supplementation are common to enhance these organoleptic traits. Genetic, nutrition and nutrigenomic studies have shown the complexity of the molecular regulation of physiological processes associated to relevant phenotypic traits in Iberian pig genotypes subjected to different diets [4, 9–11].

Besides, with the advent of new mass-sequencing techniques, alternative ways have been opened for animal breeding and for in-depth studies of the molecular basis of phenotypic variability. Microbiota studies have hugely benefitted from these NGS techniques, as culture-free sequencing allows obtaining a lot more information from microbiome communities. Gut microbiota has become a key focus in animal breeding industry, as well as in human health, since gut microbiota and host are linked in a bidirectional way; first, diet composition and other environmental elements, as well as animal genetics, might modify microbial population structure [12]; second, microbiota has been reported to have an impact on animal physiology, on multiple animal productive traits and also on the final products' quality and composition [13, 14]. This intricate correlation network makes gut microbiota composition a key factor for understanding the relationships between genotype, phenotype and environment. As a consequence, it might result obvious that different genetic backgrounds can promote different microbiota population structures as it has been proven in other studies [15]. For this reason,

the two main breeds employed in Iberian pig production (Iberian and Duroc) might have differences in gut microbiota composition possibly affecting relevant phenotypic traits. Also, dietary modification or supplementation strategies such as FA addition might have an impact on gut microbiota composition, potentially contributing to differences in animal traits, with special relevance on final product composition and quality. Understanding all these relationships and their implications is crucial to comprehend their global effect on animal phenotype.

In this study we compared the gut microbiota composition in pigs from Iberian and Duroc breeds, fed two different isocaloric and isoproteic diets with different energy sources (oleic acid vs carbohydrates), in order to dissect the diet and genetic background effects on the gut microbiota composition, as well as to deepen in the metabolic differences between breeds and diet groups, potentially mediated by gut microbiota.

Methods

Experimental design and sampling

The experiment was carried out at the experimental facilities of the Instituto Tecnológico Agrario de Castilla y Leon (ITACYL) Pig Test Center (Hontalbilla, Segovia, Spain). A total of forty-eight castrated male purebred pigs, 19 Duroc (DU) and 29 Iberian (Torbiscal strain) [16] (IB), were raised in the same commercial farm (IBERPEX S.L., Guareña, Badajoz, Spain), weaned at 28 days old and transferred to the experimental facility one month after weaning. The animals were kept under identical management conditions, housed in batches of 4 pigs/pen (1 m^2 pig^{-1}), with a concrete floor and straw bedding. Temperature was controlled at a mean of 23.8 ˚C throughout the experiment. At 70 days old (±2 days) and with a mean body weight of 19.7 ± 3.6 kg, both breeds were distributed in two different experimental groups with a factorial 2x2 design, and fed a control commercial diet with carbohydrates as energy source (C) (9 DU and 13 IB) or a sunflower oil-enriched diet with high oleic acid content (6%) (O) (10 DU and 16 IB) (Initial weights per group: 17.3 ± 2.1 kg BW in DU-C; 17.3 ± 3.5 kg BW in DU-O; 20.5 ± 2.8 kg BW in IB-C; 22.0 ± 3.5 kg BW in IB-O, these differences not being statistically significant). Both diets were formulated according to FEDNA nutritional guidelines (2006) using the Brill Formulation software (Brill Co., Georgia, USA) to be isocaloric and isoproteic (3.3 kcal digestible energy and 15.6% crude protein) (S1 File) and were provided *ad libitum*. Diets were provided for 47 days, until animals reached 50.4 ± 7.8 kg BW (Final weights per group: 53.2 ± 6.4 kg BW in DU-C; 49.0 ± 10.8 kg BW in DU-O; 48.7 ± 7.4 kg BW in IB-C; 51.3 ± 6.8 kg BW in IB-O, these differences not being statistically significant). Fresh water was provided *ad libitum*, with two drinking troughs available in each pen.

Animals were slaughtered in the same experimental facilities at the end of treatment, at 117 days old (±2 days), and individual stool samples were then collected from rectum and rapidly frozen on liquid N_2 and preserved at -80˚C until its use for microbiome analysis. All experiments were performed in accordance with the regulations of the Spanish Policy for Protection of Animals employed in Research and other scientific purposes RD53/2013, which meet the European Union Directive 2010/63/EU on the protection of animals used in experimentation. The project was approved on March 20, 2015, by the Comunidad de Madrid animal welfare and protection committee (reference number PROEX-007/15).

DNA extraction and 16S sequencing

Microbiota DNA was extracted from the 48 stool samples using QIAamp PowerFecal® kit (QIAGEN, Hilden, Germany) according to manufacturer's standard protocol. Illumina MiSeq® paired-end sequencing protocol (Illumina, San Diego, CA, USA) was performed by an external service (FISABIO bioinformatics, Valencia, Spain) targeting 16S rRNA gene

V3-V4 amplicon. The employed primers were S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21 [17], which produce amplicons of 464 bp. Raw microbial sequence data have been uploaded to the ENA repository and are available at https://www.ebi.ac.uk/ena/browser/view/PRJEB42303.

Sequence processing pipeline

Raw reads were pre-processed using Prinseq-lite tool [18] and in-house software. After trimming low quality bases in each read, sequences shorter than 50 bp and with an average quality score below 30 on a window of 20 bases were discarded. After joining forward and reverse files, a total amount of 11,599,350 reads were processed using Quantitative Insights Into Microbial Ecology (QIIME) version 1.9.1 [19] into operational taxonomic units (OTUs), following the *de novo* OTU-picking approach, with a clustering threshold of 97% identity. Chimeras were filtered using USEARCH algorithm (v. 6.1) [20]. SILVA reference database (release 132) [21] was used for taxonomic assignment and chimera filtering. An OTU abundance threshold of 0.005% was established for final quality-filtering, as described in Bokulich *et al.* [22]. Processed OTU table was composed by 9,428,777 reads from 1,398 OTUs.

Finally, two additional processing steps were performed: a taxonomy filtering, excluding OTUs not assigned to Bacteria at kingdom level in the taxonomic classification, as used primers are designed for bacterial amplicons; and a sample pruning, removing those with sequencing depth lower than 30k total reads, which eliminated one sample for the IB-O group. The final OTU table was composed by 9,162,809 reads from 1,246 OTUs and 47 samples.

Microbial community analysis

All analyses were performed in R, using the following packages: ALDEx2 [23], DESeq2 [24], easyCODA [25], limma [26], mixOmics [27], phyloseq [28], vegan [29] and zCompositions [30]. Microbiota composition was compared between breed-diet animal pens. Three different approaches were then followed to analyse microbiome composition:

Alpha-diversity. For this approach, OTU table was rarefied to minimum sample depth (30,136 reads). Four alpha-diversity measures were calculated: observed richness, Chao1 index, Shannon index and Inverse-Simpson index, using the function estimate_richness from phyloseq. ANOVA for each alpha-diversity measure was carried out following this model:

$$\boldsymbol{\alpha}_{ijk} = \boldsymbol{\mu} + \boldsymbol{b}_i + \boldsymbol{d}_j + \boldsymbol{bd}_{ij} + \boldsymbol{e}_{ijk}$$

Being α the alpha-diversity index for each sample, b the breed effect (with $i = 2$ levels), d the diet effect (with $j = 2$ levels) and bd the interaction of both effects.

Data transformation: Compositional data analysis. Most of the statistical analysis were made taking into account the compositional nature of microbiome data [31]. For the compositional approach our non-rarefied dataset was processed following a number of steps as defined by [32]: (1) Zero-counts imputing by Geometric Bayesian Multiplicative replacement; (2) Data closure computed to the total counts per sample; (3) Weighted centred log-ratio (CLR) transformation. This process moves the data to an additive scale, which makes the performance of multivariate hypothesis testing easier. The analysis was repeated glomming the dataset to OTU, genus, family and phylum levels. R packages zCompositions and easyCODA were used for this transformation.

Beta-diversity. Beta-diversity was computed using a compositional Principal Component Analysis (i.e., PCA after CLR transformation of the data). For a better visualization of the behaviour of both breed and diet effects, a mixed "breed-diet" factor was considered as PCA grouping variable. The distance matrix was built using Aitchison distance measure, which is

the Euclidean distance after CLR transformation. Individual beta-diversities were calculated as the distance to centroid for groups from mixed "breed-diet" effect (using vegan::betadisper). Homogeneity of dispersions respect to the group centroids was measured through ANOVA-like permutation test (vegan::permutest) and post-hoc Tukey HSD test. For testing the significance of multivariate effects in our *a priori* groups, non-parametric permutational multivariate analysis of variance (PERMANOVA) [33, 34] was performed using adonis2 function from vegan package with this model:

$$\boldsymbol{D}_{ij} = \boldsymbol{\mu} + \boldsymbol{b}_i + \boldsymbol{d}_j + \boldsymbol{bd}_{ij} + \boldsymbol{e}_{ij}$$

The additive model applied was built sequentially adding breed (b_i; $i = [1,2]$) and diet (d_j; $j = [1,2]$) factors and their interactions (bd_{ij}). The Aitchison distance matrix (D) was used as dependent variable.

Differential abundance analysis

Differential abundance analysis was carried out using both DESeq2 and ALDEx2 R packages. Default DESeq2 normalization by estimation of size factors as the median ratio of counts [35] and negative binomial GLM fitting with Wald significance tests were performed. Differences between breeds and between diets were analysed in separate models correcting by the other factor. Interactions between factors were also evaluated with an additional model (Table 1). P-values were corrected for multiple testing through Benjamini-Hochberg method. Significantly differentially abundant (DA) OTUs were considered with a false discovery rate (FDR) cut-off of 0.05 and a fold-change (FC) higher than 1.5 or lower than -1.5 (i.e., $|\log2FC| > 0.59$).

Input for ALDEx2 analysis was prepared as the CLR-transformed posterior distribution of the data generated by 128 Monte Carlo samples from the Dirichlet distribution. CLR transformation was made through geometric mean of all features' abundance. Mono-factorial breed and diet contrasts were performed, using Welch's t-test. P-values were corrected through Benjamini-Hochberg method. OTUs were considered as significantly differentially abundant between breeds or diets when corrected p-value was lower than 0.05 and diff.btw (i.e., median difference in CLR values between factor groups) was higher than 1 or lower than -1. Visualization of coincident DA OTUs between contrasts and pipelines was carried out with VennDiagram package in R [36].

Pairwise log-ratios analysis

Closured data were also used for another approach, using pairwise log-ratios (p-LR) instead of centred log-ratios [25], as a way to evaluate individual relationships among OTUs and features glommed to genus, family and phylum levels. Recursive Partitioning and Regression Trees (RPART) were generated, including the microorganism pairwise log-ratios most suitable to split samples according to the breed and diet phenotypic criteria.

Pairwise log-ratios analysis is computationally expensive, as feature combinations produce matrices much larger than the original, i.e., a $m \times n$ matrix (*samples* × *features*) would be transformed in a $m \times n(n-1)/2$ matrix. For this reason, OTU-grouped dataset was size-reduced by

Table 1. DESeq2 differential abundance models for each contrast, with R notation.

Contrast	Model design
Breed contrast	*RA ratio ~ Diet + Breed*
Diet contrast	*RA ratio ~ Breed + Diet*
Interaction effects	*RA ratio ~ Breed + Diet + Breed: Diet*

https://doi.org/10.1371/journal.pone.0251804.t001

selecting a relevant feature subcomposition generated by sparse Partial Least Square Discriminant Analysis (sPLS-DA) (mixOmics package), using the 4-leveled mixed trait "breed-diet" as classification factor. Five-fold cross-validation with permutation was used to test the optimal component structure for the model. In order to generate the subcomposition, OTUs with a Variable Influence of Projection (VIP) value higher than 1 for each component were selected. As p-LR analysis keeps subcompositional coherence [25], data subsetting has no significant effect on the results.

In order to check which microbes were mainly responsible of separation between groups at each ratio, supplementary differential abundance analyses were made using ALDEx2 (as this pipeline uses the same methods as for RPART trees building) with phylum, family and genus-grouped database.

Microbiome-phenotype association

CLR data were also used to detect associations between microbiome and several phenotypic traits. IMF content was measured in *Biceps femoris* and *Longissimus dorsi*. FA profile was analysed in backfat outer layer (Back) and in ham outer fat layer (Ham). Total FA fractions (saturated, SFA; mono-unsaturated, MUFA; and poly-unsaturated, PUFA) as well as relevant FA contents (oleic, C18:1(9); linoleic, C18:2; palmitic, C16:0; and stearic, C18:0) were also measured, with all quantification methods being described in formerly published studies [11].

PERMANOVA controlling both breed and diet effects was performed for each phenotypic variable using vegan::adonis2 function, in order to elucidate linear or nonlinear relationships between samples distribution and each phenotypic trait:

$$\boldsymbol{D_{ij}} = \boldsymbol{\mu} + \boldsymbol{b_i} + \boldsymbol{d_j} + \boldsymbol{P} + \boldsymbol{e_{ij}}$$

The additive model applied was built sequentially adding breed (b_i; $i = [1,2]$) and diet (d_j; $j = [1,2]$) factors, and each phenotype (P) was treated as linear regression at the end. The Aitchison distance matrix (D) was used as dependent variable.

Additionally, individual OTU association was addressed using Limma linear regression. To avoid the influence of the known breed effects on phenotype which could lead to potential spurious associations, both microbiome and phenotypic datasets were split for carrying out independent DU and IB microbiome-phenotype analyses. To measure diet weight in correlations, two methods were carried out: (1) linear regression of each OTU on each phenotypic variable without diet correction; (2) linear regression of each OTU on the diet regression residuals of each phenotypic variable. P-values were adjusted using Benjamini-Hochberg method (adjusted significance threshold set to FDR = 0.05). Representation of associations between OTUs and phenotypes was addressed through interaction network in Cytoscape software (v. 3.8.1), setting Kamada-Kawai algorithm (Edge-weighted spring embedded layout) [37] for optimal visualization.

Results

Overall microbiota composition

The overall composition is summarised in Table 2. Our final OTU table was composed by 47 samples, as we pruned one sample with less than 30k reads. As mentioned in methods, we applied an additional filter removing those reads not assigned to Bacteria kingdom. After this step, 97.44% of total reads were kept, representing 1,246 OTUs. From this bacteria-classified fraction, 99.93% reads were assigned to phylum level. A total of 17 phyla were identified. Prevalence and abundance of phyla are represented in Fig 1.

Table 2. Overall microbiota composition.

	Kingdom	Phylum	Class	Order	Family	Genus
% Sequences [1]	97.44	99.93	99.91	99.91	97.30	86.70
Classified OTUs	1246	1245	1243	1243	1184	988
Identified taxa	1	17	23	33	48	148

Total OTU number, % of sequences assigned to a known taxon at different ranks and unique taxa found at each rank.

[1] Proportion of total reads classified to phylum and lower taxonomic levels are calculated relative to the total reads after filtering OTUs not assigned to Bacteria kingdom.

https://doi.org/10.1371/journal.pone.0251804.t002

The 988 annotated OTUs were classified to 148 different genera, covering 86.7% of total reads. 20 genera had an average relative abundance (**RA**) higher than 1%, being *Prevotella_9*, *Lactobacillus*, *Prevotellaceae NK3B31 group*, *Alloprevotella* and *Treponema_2* the most abundant ones (S2 File). Note that SILVA database subdivides some genera according to sequence clustering, appearing under multiple designations in our database.

On the other hand, 79 genera were present at an average abundance lower than 0.1% (S2 File). This fact might be taken into account when interpreting the results, as they might be poorly represented across the samples (i.e., low prevalence) and/or have a minimum number of reads per sample (i.e., low coverage). Nonetheless, all the analyses have been performed including these low-abundance OTUs.

Microbiota composition analysis

Two-way ANOVA for alpha-diversity measures in rarefied data showed no significant differences between diet groups (C vs O), but all of them (richness: $p = 0.021$; Chao1 index: $p = 0.018$; Shannon index: $p = 0.008$; and Inverse Simpson index: $p = 0.002$) were significantly affected by breed, these measures showing higher values in the IB group (Fig 2). No significant interaction between breed and diet was detected in the ANOVA. Initial average number of OTUs per sample was 1,046 for DU population and 1,097 for IB population. After rarefaction, these numbers were corrected to richness values, with an average of 880 in DU and 944 in IB.

Compositional PCA for beta-diversity (Fig 3) showed a high dispersion of the data, although breed clustering was observed through PC1. Diet clustering can be observed through PC2, but only when features were analysed at lower taxonomic levels (i.e., genus and OTU levels). Centroid position also revealed that separation between breeds was higher than separation between diets. At OTU level, centroid distance is maximum for both breed and diet groups. Individual beta-diversities were calculated as the Euclidean distance to centroids for groups from mixed "breed-diet" effect. Average individual dispersion within groups was not significantly different between groups, except for family and OTU-glommed datasets. At OTU level, DU-C samples had an average beta-diversity index significantly lower compared to DU-O, IB-C and IB-O groups, although p-value in DU-C vs IB-C contrast (same diet in different breeds) was close to 0.05. (Fig 4). At family level, DU-C individual beta-diversities were also significantly lower than DU-O ones ($p = 0.047$).

PERMANOVA analysis confirmed that samples from different breeds are significantly different. Differences between diets (corrected by breed effect) are also significant at OTU, genus and family level, although these differences blur at phylum level (Table 3). Nonetheless, the differences between groups are moderate, as 10–12% of variance from our data can be explained by breed effect, and only 3–7% by diet effect.

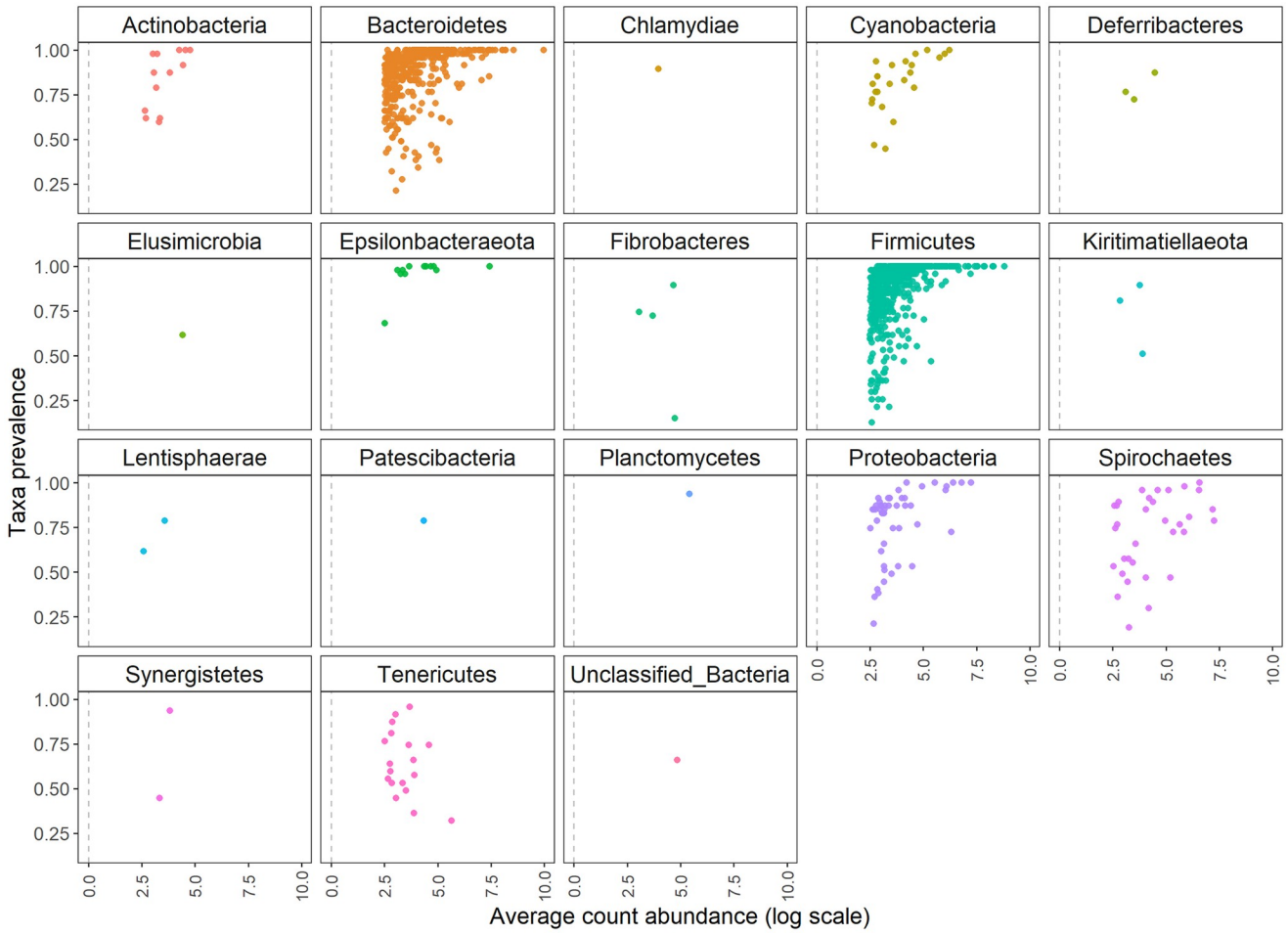

Fig 1. Phyla prevalence and abundance in dataset. Each dot represents an OTU, being the x-axis its average relative abundance and the y-axis the proportion of samples in which it is present. "Unclassified_Bacteria" is the denomination given to those OTUs with no phylum assignment.

https://doi.org/10.1371/journal.pone.0251804.g001

Differential abundance (DA) analysis

Two separate contrasts were made, for breed and diet factors, using an FDR cut-off of 0.05 and a fold-change minimum threshold of ±1.5 (log_2 *FC* ≈ ±0.59) for DESeq2 analysis, and a minimum median difference between CLR values of ±1 (*diff.btw* = ±1) for ALDEx2 analysis.

Fig 5 includes volcano plots for both contrasts with DESeq2 and ALDEx2 methods. Both breed and diet contrasts resulted in more restrictive results with ALDEx2 method. Breed contrast shows a higher number of over-abundant (**OA**) OTUs in IB pigs when done with DESeq2, while ALDEx2 breed contrast shows a higher number of DU OA OTUs. On the other hand, O diet presents a higher number of OA OTUs than under-abundant (**UA**) OTUs, compared to C diet, both with DESeq2 and ALDEx2. With DESeq2 pipeline, 468 DA OTUs from 119 different genera were found for DU vs IB contrast, mostly from *Ruminococcaceae*, *Prevotellaceae* and *Lachnospiraceae* families. On the other hand, 185 DA OTUs from 71 genera were found between C and O diet groups, predominated by the same families. With ALDEx2, 207 DA OTUs were found for breed contrast, whereas only 18 DA OTUs were found for diet test, mostly from *Prevotellaceae* family. Common DA OTUs between pipelines are counted in Table 4. For breed contrast, the most representative taxa in overlapping DA OTUs was

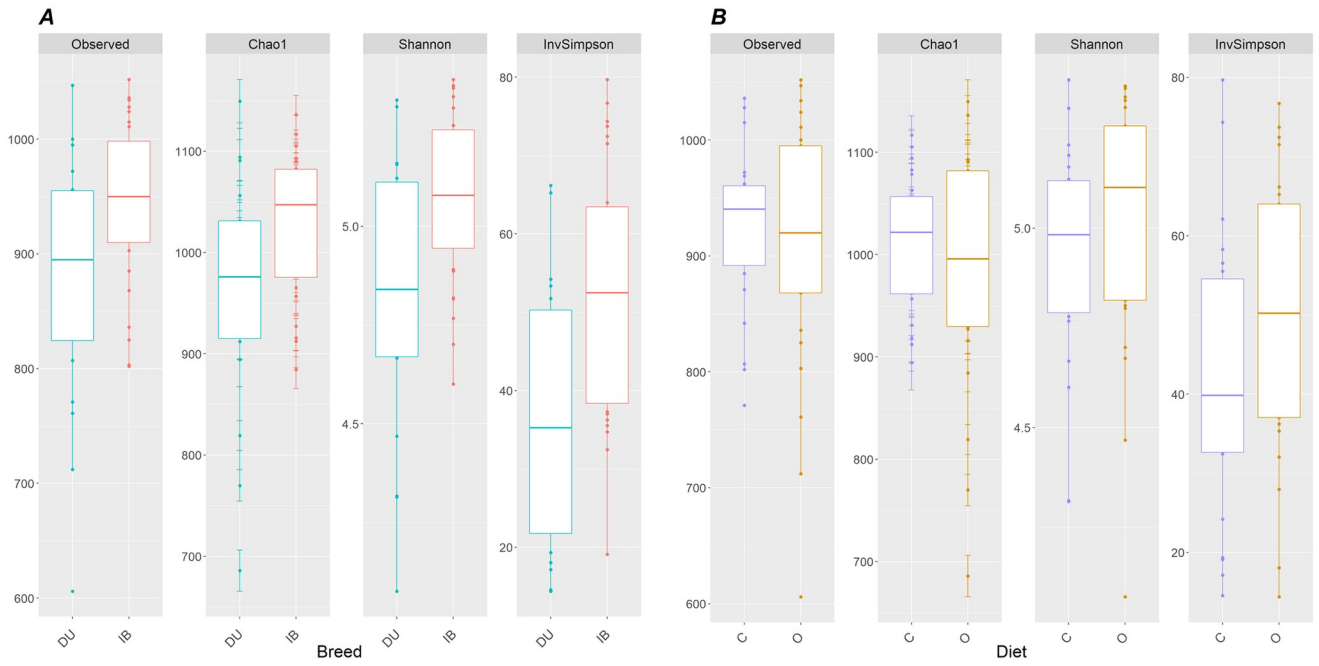

Fig 2. α-diversity boxplots. Alpha-diversity boxplots according to breed (left) and diet (right) groups in rarefied data.

https://doi.org/10.1371/journal.pone.0251804.g002

Prevotellaceae (mostly *Prevotella*) (27%), followed by *Lachnospiraceae* (16%), *Ruminococcaceae* (15%) and *Veillonellaceae* (mostly *Megasphaera* and *Anaerovibrio* OTUs) (11%). For diet contrast only 18 OTUs were common to both pipelines: 12 were high-oleic OA OTUs from *Prevotellaceae* (7 of them), *Lachnospiraceae*, *Muribaculaceae* and *Ruminococcaceae* families, while the remaining 6 high-oleic UA OTUs were 3 Rikenellaceae members and 3 OTUs classified as Clostridium s.s. 6, Erysipelotrichaceae UCG-004 and Desulfovibrio, respectively. When exploring DESeq2 results, most of the high-oleic OA OTUs were from the same families, but a high number of OA *Prevotella* and *Alloprevotella* OTUs was detected. Within high-oleic UA OTUs, a higher variety of families was present, with high representativity of *Prevotellaceae*, *Ruminococcaceae*, *Rikenellaceae* and *Erysipelotrichaceae*. Supplementary material with the complete DA-OTU list for all contrasts is also available (S3 File).

We evaluated coincident DA OTUs among contrasts and methods, as shown in the Venn diagram (S1 Fig). When comparing DESeq2 with ALDEx2, 151 DA OTUs were coincident in the breed contrast and 18 coincidences were found in the diet contrast. On the other hand, when comparing the results obtained for breed and diet effects within the same method, we detected some coincident OTUs (75 OTUs in DESeq2 and only 1 OTU in ALDEx2). Only one OTU is common to all four comparisons, the one named as *denovo378129*, from genus *Shuttleworthia*, which was OA in DU and O groups.

Significant interaction between breed and diet was detected for 26 OTUs ($P_{adj} < 0.05$), most of them being qualitative interactions (Fig 6). Most of these OTUs were members of *Corynebacterium*, *Lactobacillus* and *Ruminococcaceae* taxa.

Pairwise log-ratio analysis

sPLS-DA for OTU-grouped dataset confirmed a total of 141 relevant OTUs (VIP $\geq$ 1) as predictive of the "breed-diet" phenotype. This subcomposition has been used to calculate pairwise

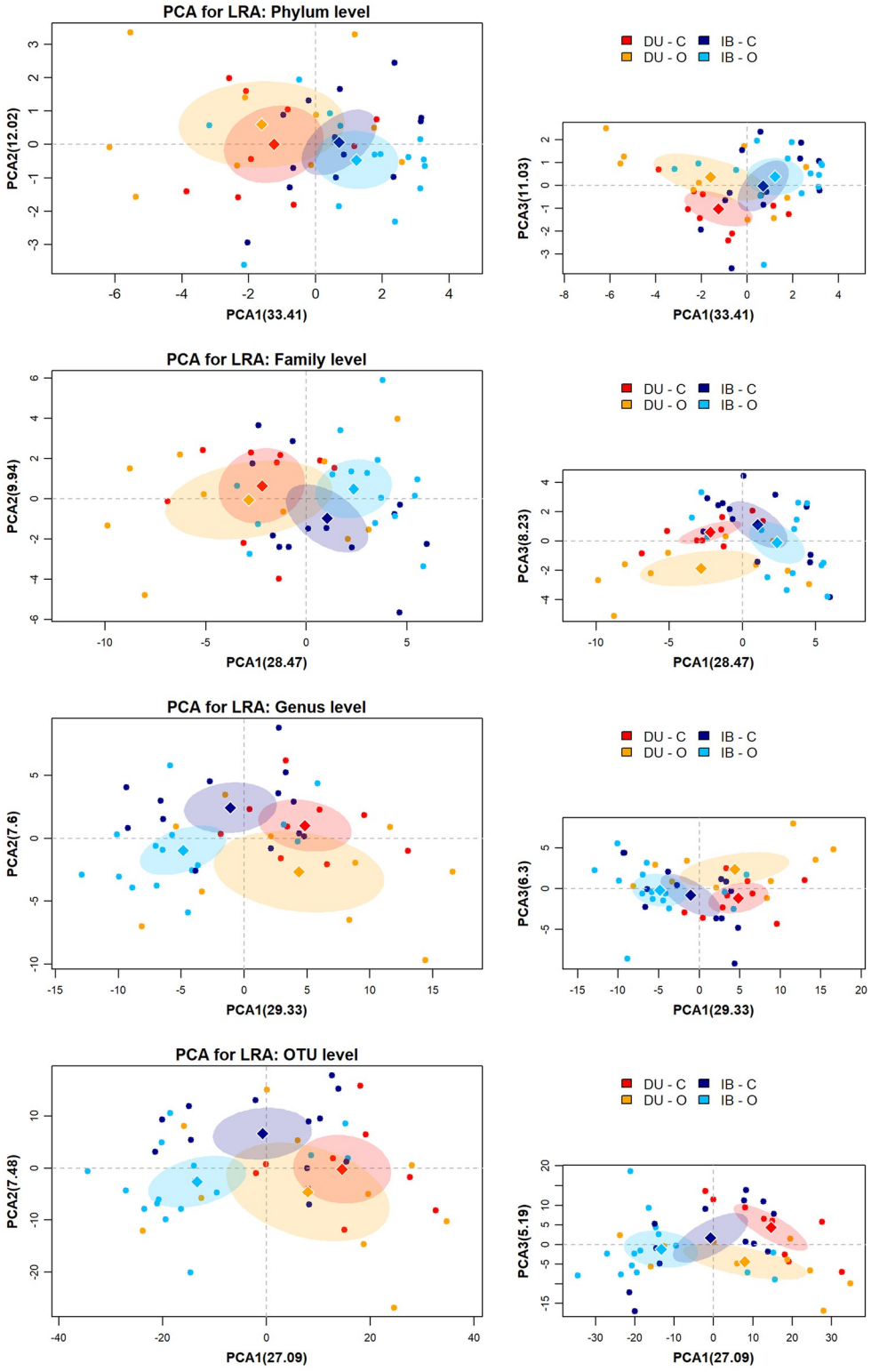

Fig 3. β-diversity plots. Beta-diversity plots from PCA of CLR-transformed data, at phylum, family, genus and OTU levels and variance percentage explained by each component. Ellipses show the Normal-theory confidence regions with alpha = 0.95. Centroids are represented by diamond-shaped points.

https://doi.org/10.1371/journal.pone.0251804.g003

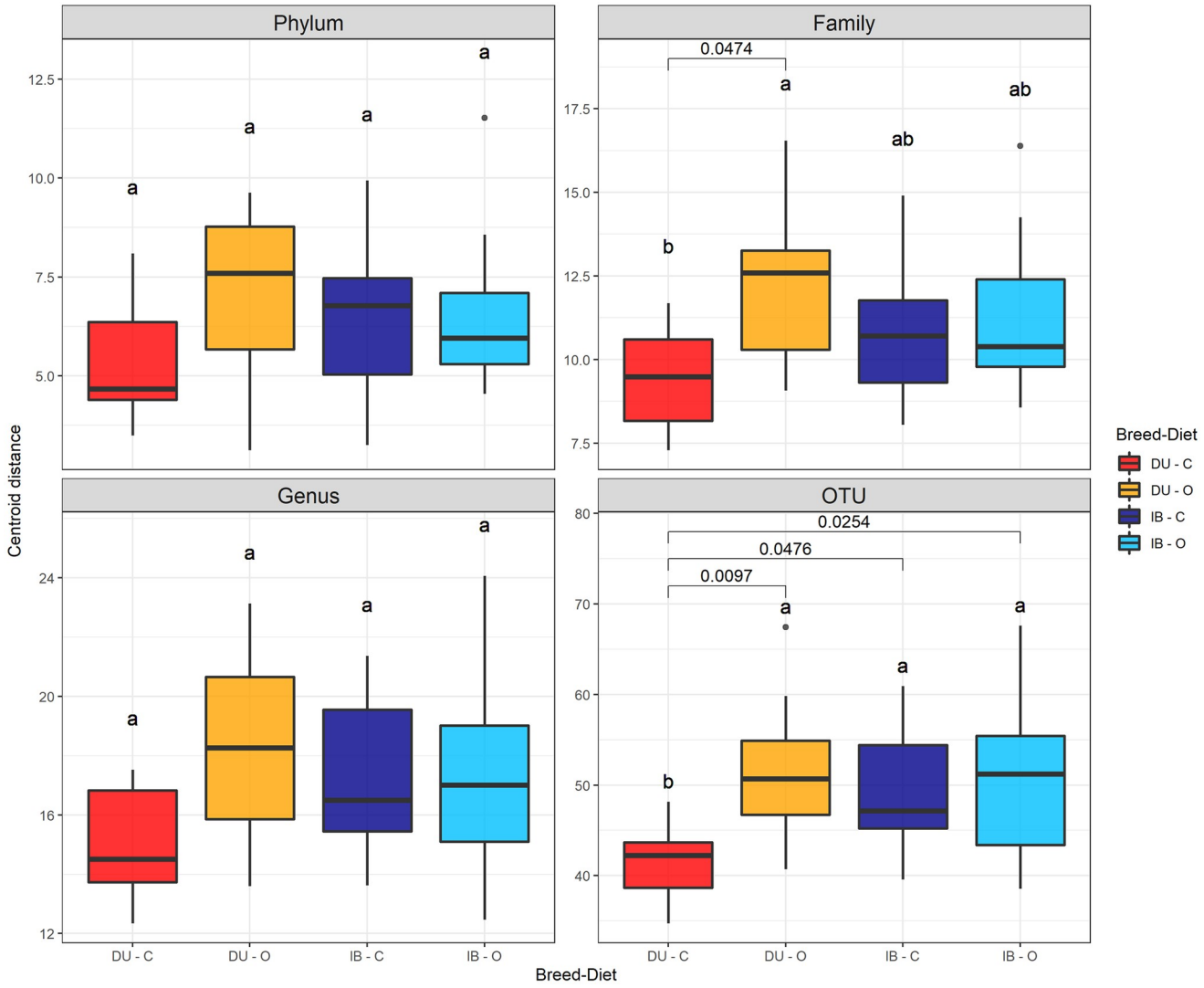

Fig 4. β-diversity indices. Beta-diversity indices by group at different feature glom levels. Tukey HSD test group letters and p-values for significant group comparisons ($p < 0.05$) are represented.

https://doi.org/10.1371/journal.pone.0251804.g004

log-ratios at OTU level. For genus, family and phylum grouped datasets, full table has been used for p-LR calculation. Recursive Partitioning and Regression Trees (**RPART**) were built for each p-LR matrix, as shown in Fig 7. We observed that the best classification occurs at family level, being the ratios *Corynebacteriaceae*/*Spirochaetaceae* (**COR/SPR**) and *Peptostreptococcaceae*/*Streptococcaceae* (**PEP/STR**) the most relevant to classify between breeds.

At phylum level diet differences were not clear inside each breed, as it was observed in beta-diversity analysis. *Actinobacteria*/*Epsilonbacteraeota* ratio was able to separate samples by breed, being reliable to classify as IB but not as DU, and was not capable of properly separate diet groups inside each breed.

At family level, three ratios were capable of separating animals in the four"breed-diet" groups with high precision. As mentioned, COR/SPR and PEP/STR ratios mainly clustered Iberian and Duroc separately, but also differentiated dietary groups inside the IB breed, while *Atopobiaceae*/*Enterobacteriaceae* (**ATO/ENT**) ratio separated diets within the DU animals.

Table 3. PERMANOVA statistics.

		F statistic	R^2	p-value
Phylum	*Breed*	5.05	0.10	0.0017*
	Diet	1.44	0.03	0.1776
	Interaction	1.32	0.03	0.2187
Family	*Breed*	5.96	0.11	0.0001*
	Diet	2.24	0.04	0.0283*
	Interaction	1.51	0.03	0.1247
Genus	*Breed*	6.15	0.12	0.0001*
	Diet	2.55	0.05	0.0118*
	Interaction	1.27	0.02	0.1859
OTU	*Breed*	5.39	0.10	0.0001*
	Diet	3.43	0.07	0.0021*
	Interaction	1.17	0.02	0.2313

F statistic, R^2 and p-values at different taxonomic levels.

* $p < 0.05$.

https://doi.org/10.1371/journal.pone.0251804.t003

These ratios are closely related with those responsible of group classification at genus level (*Anaerostipes/Corynebacterium_1* and *Family_Peptostreptococcaceae/Peptococcus*), except that these genera ratios can only discriminate between IB-C, IB-O and DU animals, but no ratio was able to differentiate diets within DU breed.

In the case of COR/SPR, both ratio components were responsible of variation. ALDEx2 DA analysis at family level showed that both *Corynebacteriaceae* and *Spirochaetaceae* were significantly DA between DU and IB groups (IB-UA: diff.btw = -3.23 and IB-OA: diff.btw = 1.13, respectively) although *Corynebacteriaceae* reduction in IB is stronger than *Spirochaetaceae* increment. As *Corynebacteriaceae* are *Actinobacteria*, breed separation at phylum level is reflected at lower taxonomic levels. Same occurs at the genus decision tree, whose first ratio *Anaerostipes/Corynebacterium_1* was able to differentiate breeds, with *Corynebacterium* having the highest weight, as DA analysis at genus level stated a *Corynebacterium_1* OA in DU animals genus (diff.btw = -2.78).

The PEP/STR ratio was able to identify IB animals, even separating C and O populations when accounting also COR/SPR ratio. DA analysis revealed that *Peptostreptococcaceae* family was more abundant in DU animals (diff.btw = -1.27), while *Streptococcaceae* family was not differentially abundant between breeds. At genus level, *Peptococcus* abundance played a similar role, as IB pigs had a higher RA (diff.btw = 1.14).

As indicated, ATO/ENT ratio was able to distinguish between C and O diets within DU animals, but only when applied after discriminating IB animals with former ratios (otherwise it would only discriminate between DU-O and the rest of"breed-diet" groups). *Enterobacteriaceae* were OA in DU animals (diff.btw = -1.92) but no abundance differences were detected between diet groups.

Finally, the two representative ratios at OTU level (*denovo219246/denovo273762* and *denovo277739/denovo351253*) were composed by OTUs assigned to *Rickenellaceae/ Ruminococcaceae* families and *Peptococcus/Lactobacillus* genera, respectively. Additionally, *denovo219246* and *denovo277739* appeared as DA between IB and DU animals (first one with DESeq2 pipeline and second with both DESeq2 and ALDEx2 pipelines), which gave them a substantial weight in breed differentiation.

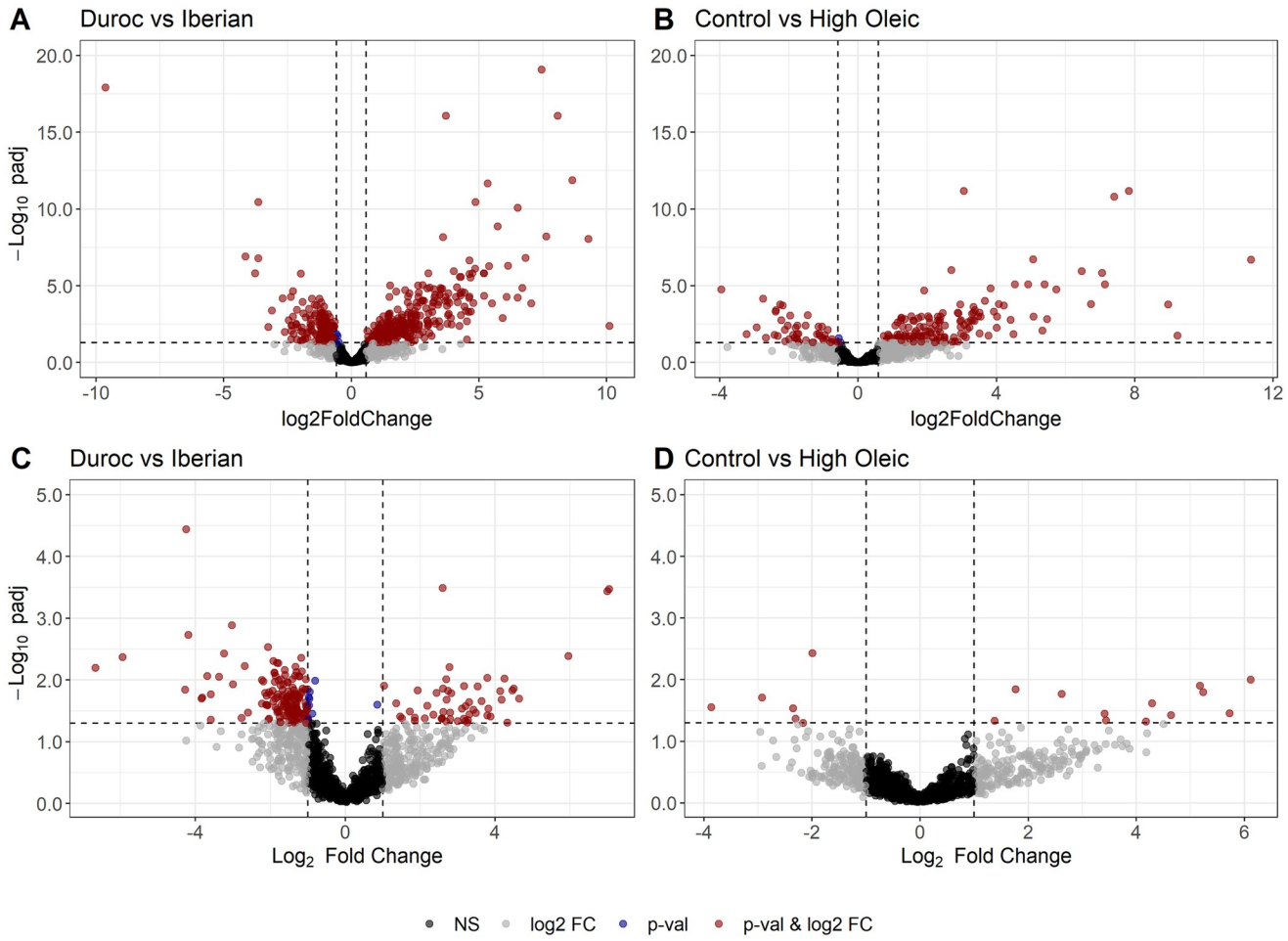

Fig 5. DA volcano plots. Volcano plots of differentially abundant OTUs using DESeq2 (A, B) and ALDEx2 (C, D), for breed (left) and diet (right) contrasts. Red dots represent OTUs with an adjusted p-value below the FDR cut-off (0.05) and a FC value above 1.5 or below -1.5 for DESeq2 contrasts, or a median difference between CLR values above 1 or below -1 for ALDEx2 contrasts. OTUs with Log2(FC) < 0 or diff.btw < 0 are more abundant in Duroc and Control groups, respectively.

https://doi.org/10.1371/journal.pone.0251804.g005

Association with phenotypic traits

PERMANOVA did not show any significant association between overall microbiota composition and the analysed phenotypic traits. When the effect of each individual OTU was evaluated

Table 4. DA OTUs by factor and pipeline.

		DA OTUs	DA genera	DA families	DA phyla
DESeq2	**Duroc vs Iberian**	468	119	44	13
	Control vs High Oleic	185	71	32	9
ALDEx2	**Duroc vs Iberian**	207	72	22	6
	Control vs High Oleic	18	13	8	3
Overlap	**Duroc vs Iberian**	151	63	20	6
	Control vs High Oleic	18	13	8	3

Number of DA OTUs in breed and diet contrasts using DESeq2 or ALDEx2 pipelines, and common DA OTUs between both pipelines. Number of genera, families and phyla grouping these OTUs are also shown.

https://doi.org/10.1371/journal.pone.0251804.t004

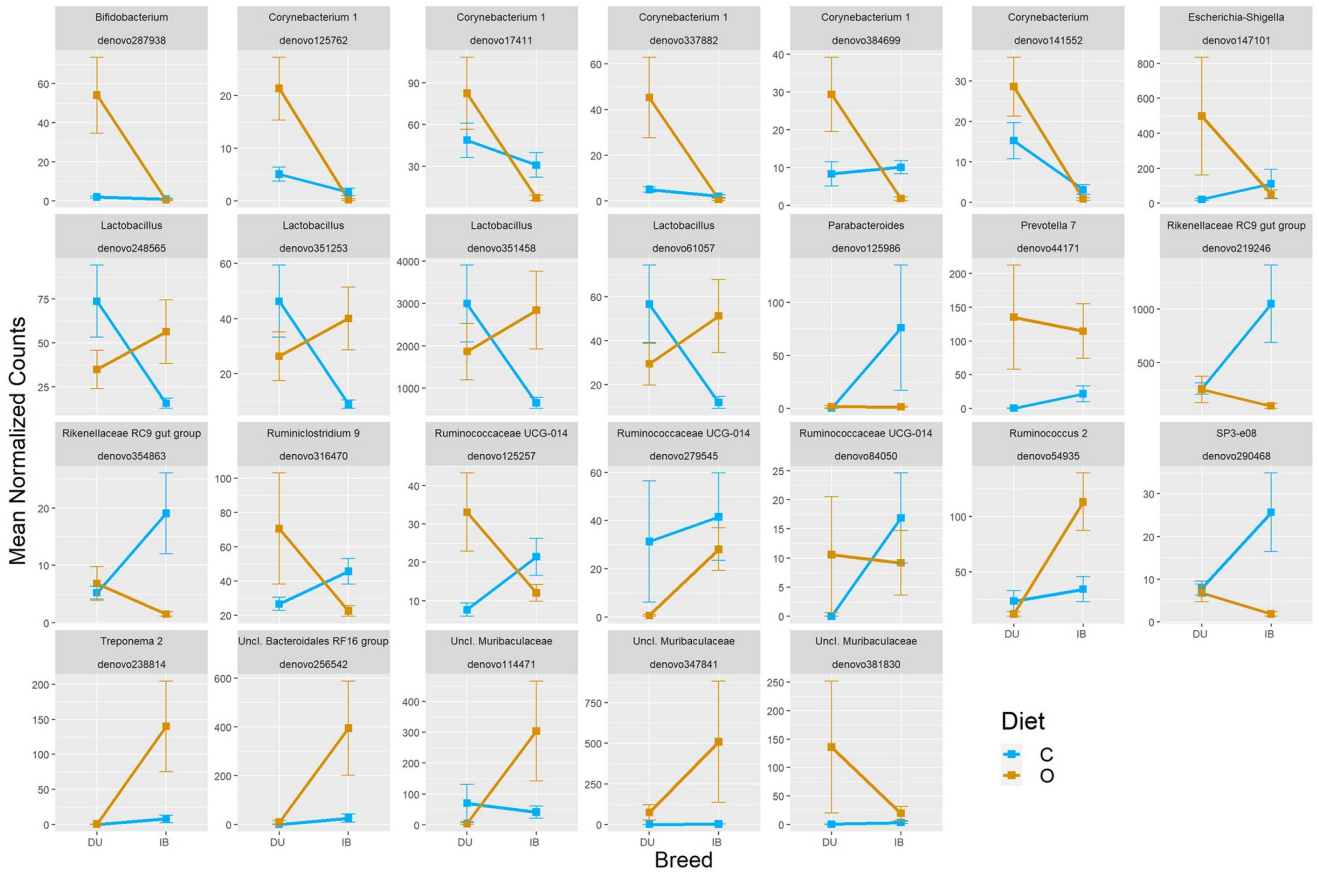

Fig 6. DA interaction plots (DESeq2). Interaction plots for each OTU with significant interaction between breed and diet factors for DA analysis with DESeq2. Normalized counts from DESeq2 algorithm and standard error of the mean (SEM) are represented in y-axis.

https://doi.org/10.1371/journal.pone.0251804.g006

through Limma regression different results were observed for each tested model. Without diet correction, only one OTU was significantly associated to lipidic traits in DU subset, *denovo283234* (*Treponema 2*), which was negatively associated to backfat MUFA and oleic acid content, while, a total of 87 individual OTUs were found as significantly associated to phenotypic traits in IB subset (Fig 8). Most of them belonged to *Alloprevotella* and other *Prevotella* groups and were associated to multiple traits. *Ruminococcaceae* OTUs were positively linked to oleic acid and negatively linked to palmitic acid proportions in backfat, while *Rikenellaceae* OTUs mainly appear negatively associated to oleic acid proportion, affecting in both cases overall MUFA and SFA proportions. Some relevant OTUs from *Shuttleworthia*, *Lactobacillus* and *Corynebacterium* genera also appeared associated to FA composition. When diet correction was applied almost every correlation disappeared, and only one OTU was negatively correlated to ham oleic acid proportion in IB subset, *denovo37314* (*Treponema 2*).

Discussion

In this study we explore the composition and disparities of the bacterial fraction of fecal microbiota populations between two different pig breeds (Iberian and Duroc) at 117 days old, fed diets with different energy source (carbohydrates *vs*. sunflower oil with high oleic acid content). Phenotypic characterization of these animals has been previously reported by Benítez *et al*. (2018) [38]. Briefly, breed effects affected feed intake, fatness and premium cuts' yield

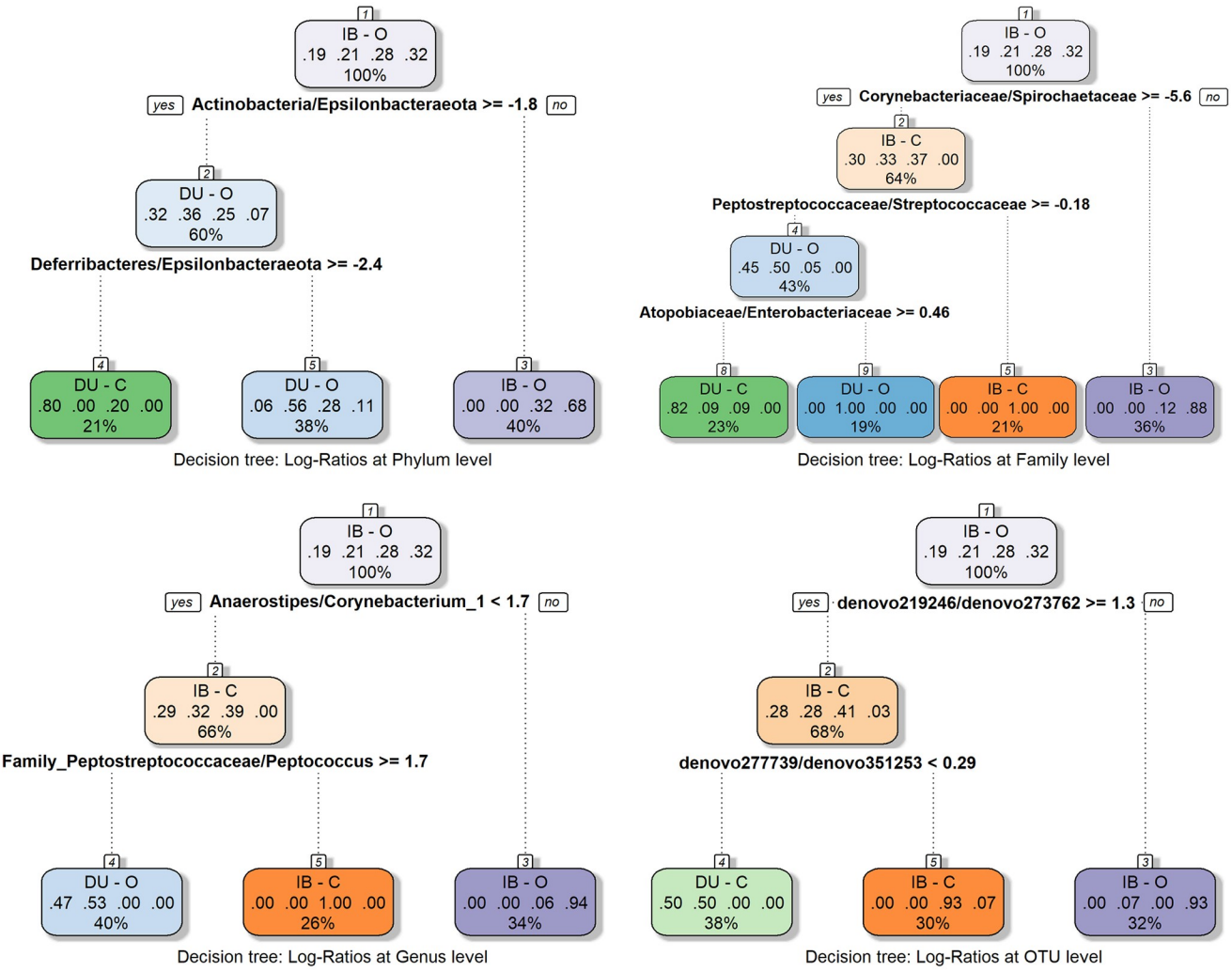

Fig 7. p-LR RPART trees. RPART trees from pairwise log-ratios at phylum, family, genus and OTU levels. Each decision level indicates the threshold value of one p-LR to classify a sample in one phenotypic group or another. Each node box displays the classification, the probability of each class at that node (i.e. the probability of the class conditioned on the node) and the percentage of observations used at that node. Class probabilities at each node are sorted as: DU-C, DU-O, IB-C and IB-O.

https://doi.org/10.1371/journal.pone.0251804.g007

with Iberian animals showing higher feed consumption, thicker backfat and lighter hams, with no differences in body weight. Regarding diet, differences were observed in FA composition, as the high-oleic group had higher MUFA and oleic acid, and lower SFA than the control group.

Overall taxonomic structure revealed *Bacteroidetes*, *Firmicutes* and *Proteobacteria* as the most abundant phyla, with an important predominance of *Prevotellaceae*, *Lachnospiraceae* and *Ruminococcaceae* families. *Prevotella* genus was the most abundant, with an average RA of 30% in Duroc animals and 22% in Iberian animals when considering all SILVA genus-clustering designations. *Prevotella* are anaerobic saccharolytic bacteria which produce acetate and succinate through fermentation [39] and their abundance is highly related to high-fiber long-term diets [40, 41]. Their presence in fecal microbiota is highly reported in human [42] and other animals such as ruminants, predominating in rumen microbiota [43–45]. Multiple studies have shown the dominance of *Prevotella* in pig gut microbiota [46] as well as their relevance

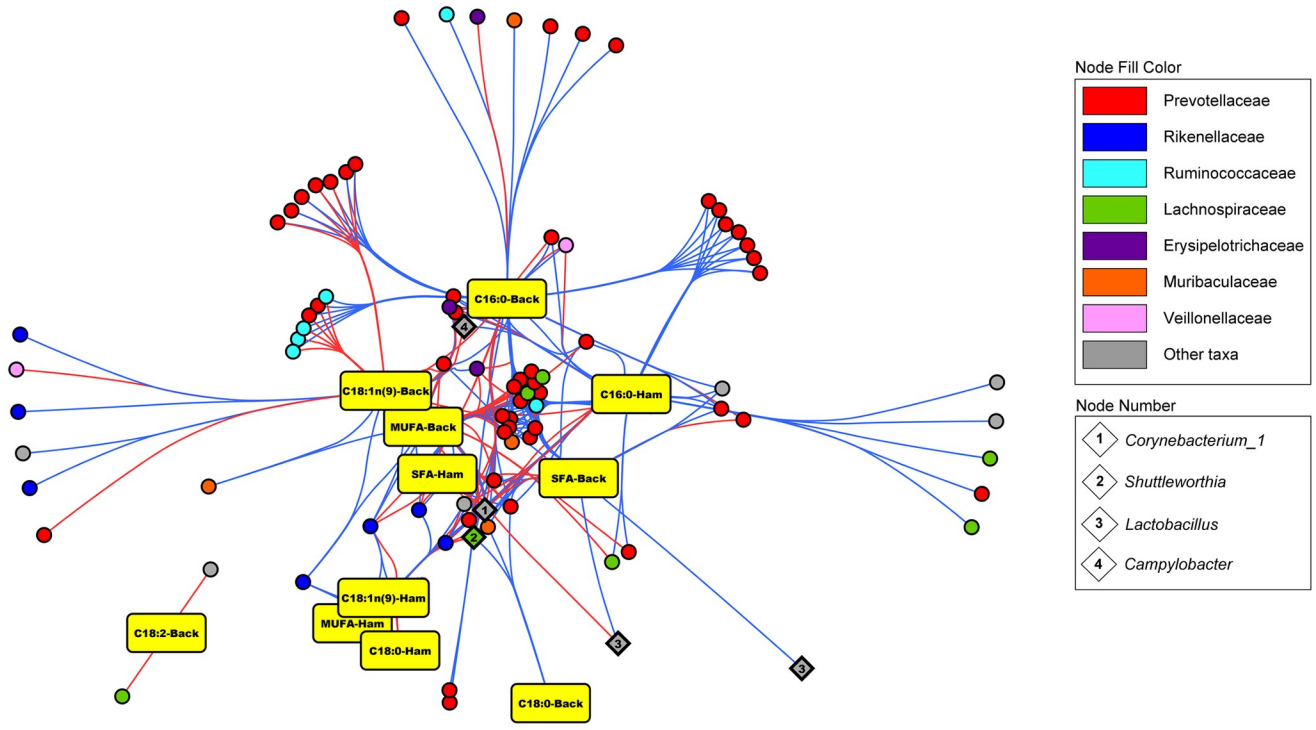

Fig 8. IB OTU-phenotype correlation network. Network representing individual OTU significant correlation with each phenotypic trait within IB subpopulation, from Limma regression analysis, when no diet correction was applied. Edge colour: red = positive correlation; blue = negative correlation. Central nodes have more connections (i.e., significantly associated to more phenotypic traits) than peripheral nodes. Numbered nodes represent OTUs classified as genera with high relevance due to their abundance patterns in our dataset.

https://doi.org/10.1371/journal.pone.0251804.g008

in enterotypes [47] or even immune response [48]. *Alloprevotella* is also one of the most representative genus in our data, with an average RA of 4.2% in DU and 11.6% in IB animals. This genus is also known for its saccharolytic activity [49], although it is less studied than *Prevotella* due to its lower abundance in gut communities. *Lachnospiraceae* and *Ruminococcaceae* families also comprehend well-known polysaccharide fermenters, which in addition participate in methanogenesis by producing H_2 and formate as substrates to archaeal communities [50]. Abundance balance of these three families in gut microbiota, through a number of abundance ratios, might be relevant for host traits related with feed efficiency, metabolism or health. For instance, *Prevotella*-to-*Bacteroides* ratio has been related with weight and fat loss in humans [51], and *Firmicutes*-to-*Bacteroidetes* ratio has been repeatedly associated to fiber metabolism or obesity in human populations [40, 52] and murine models [53], although there is controversy about these links [54]. Our data support that this core microbiota might also be shared by other monogastric species along all digestive tract. Similar findings are reported by Crespo-Piazuelo *et al.* [55] when analysing gut microbiota in different intestinal regions using the same IB-C pig population. However, they reported a range of 35 to 41% of *Prevotella* abundance within all colon regions. These discrepancies must be attributed to different intestinal region, different genotype and diet of the total animal population studied (colon from IB-C animals vs rectum from IB and DU fed C and O diets).

The pipeline used for analysing differences between experimental groups included a classic alpha-diversity evaluation and a compositional approach to beta-diversity. Differential abundance analysis was used to acquire a more specific insight about changes in microbiome

population structure. DESeq2 is a RNAseq-based, robust and well documented method broadly used in microbiome DA analyses. However, in recent years the concern about compositional nature of these data has increased, and some authors remark that RNAseq-based tools are not accounting for the compositional nature of microbiome data. For this reason, we also performed differential abundance analysis with ALDEx2 method, an ANOVA-based approach which accounts for this intrinsic compositionality and has a similar sensitivity than other methods, also reducing false positive rate near to zero [31]. The most conservative strategy would be to focus in the overlapping DA OTU set between DESeq2 and ALDEx2, but due to the restrictive behaviour of ALDEx2 we decided to consider also DESeq2 results for further exploration. Furthermore, pairwise log-ratio (p-LR) analysis was also used to deepen in composition differences, as ratios between OTUs can also be biologically relevant, beyond single association with OTU RAs. This approach allowed to detect those microorganism pairwise ratios whose abundances could separate animals by both breed and diet.

Our results revealed an important effect of the host genetic background on the fecal microbiota composition, both for bacterial richness and overall composition. Regarding richness, a higher number of different OTUs is observed in IB pigs, which can also explain the higher values of Shannon and Simpson indices calculated for this breed. These interesting findings might be related with the higher rusticity and resilience of Iberian pig breed compared with other commercial breeds [56], since higher ecological diversity is associated with a better response to environmental disturbances (response diversity) [57]. Regarding composition, fecal microbiome resulted more different between breeds than between diets, as PCA sample distribution, PERMANOVA and DA analyses revealed. DU animals have a higher abundance of *Actinobacteria*, specifically from *Corynebacterium* and *Bifidobacterium* genera, as well as multiple *Lachnospiraceae* (*Blautia*, *Roseburia*, *Lachnoclostridium*, *Fusicatenibacter*, and *Shuttleworthia*), multiple *Prevotellaceae* OTUs (mostly *Prevotella* and *Alloprevotella*) and other genera such as *Clostridium s.s. 1*, *Holdemanella*, *Megasphaera* or *Succinivibrio*, while Iberian OA OTUs were scarcer and belonged to more diverse taxa (see S3 File for complete OTU list). As far as we know, no former studies have compared Duroc and Iberian pigs' microbiota, so these findings are novel and might be important to define unique microbiome patterns associated to host genotypes. Still, gut microbiota contrasts between Duroc, Landrace and Large White pigs also revealed a higher presence of *Prevotella* in Duroc animals [15].

Focusing on Actinobacteria, we specifically found a higher abundance of *Bifidobacterium* and *Corynebacterium* in DU animals. *Bifidobacterium* are well known anaerobic lactic acid producer bacteria [58] whose abundance in fecal microbiota has been reported as negatively correlated with pig age, being more abundant at early life stages [59]. The transition from weaning to adult cereal-based diet is a key factor in the replacement of these early-colonizers by other microorganisms such as *Prevotella*, *Roseburia* or *Succinivibrio*, more typical of adult fecal microbiota [48]. Only one *Bifidobacterium* OTU was present in our data, presenting low abundance (mean RA = 0.01%) and majorly absent in IB animals. This might indicate a slower replacement of weaning to adult microbiota in DU pigs, which might as well be related to their lower alpha-diversity. *Corynebacterium* includes a wide variety of gram-positive aerobic and facultative species, some of them being fermentative organisms with special relevance in amino acid biosynthesis pathways, such as those of lysine or histidine [60, 61]. Lower protein deposition rates have been reported in Iberian pigs, compared to other commercial breeds [62]. In agreement, our Duroc pigs showed higher muscle deposition, as lower fat and higher ham yield was observed with similar body weight respect to Iberian pigs [38]. Thus, *Corynebacterium* might be contributing to a differential protein synthesis and degradation balance, through a differential bioavailability of essential amino acids in the gut, thus, their abundance could be correlated to animal protein deposition.

Differential abundance of other groups might as well be related with richness and resilience of Iberian pigs. In this sense, the lower abundance of *Shigella-Escherichia* in IB pigs might be related with the previously reported increase of microbial diversity, as a higher alpha-diversity has been associated to the prevention in the establishment of pathogenic microbes such as *Enterobacteriaceae*. However, this association has only been described in humans [40] and our data do not confirm this finding. Likewise, the mentioned increment in richness seems to be related with the lower abundance of multiple *Prevotella* OTUs in IB pigs, as it happens when comparing C and O diets. *Treponema* DA OTUs, which are virtually absent in Duroc animals, appear as moderately abundant in Iberian pigs (RA $\approx$ 1%), thus these fiber-fermenters might have settled the niche that *Prevotella* left. Finally, we detected an abundance opposition between *Prevotella* and *Treponema* taxa: we found a number of *Prevotella* OTUs with high RA being over-abundant in DU animals, while several *Treponema* OTUs were over-abundant in IB animals. As described by Crespo-Piazuelo *et al.* [55], this might suggest a competition for dietary fibre.

The diet effect on microbiome appears masked by the breed influence, but its impact on microbiome structure is also noticeable, since PERMANOVA revealed significant differences in distances between samples from different diets in relation to distances within each diet. It must be pointed out that O diet had a higher fiber content than C diet, as it was formulated to be isoenergetic in relation to C diet. This fact might be related to the greater presence of *Prevotella* and kindred taxa in O animals, since fiber content has already been linked to gut microbiota changes [63] and *Prevotella* OTUs have been defined as biomarkers for high-fiber diets [64]. Generally speaking, high-oleic OA OTUs were rare OTUs (i.e., RA $\leq$ 0.01% in C group), while control diet OA OTUs tended to be more common (i.e., RA $>$ 0.01% in C group), meaning that the abundance decrease of common OTUs due to oleic acid and fiber supplementation results in an abundance increment of marginal OTUs. This fact is supported by our Shannon index values showing an upward trend in O group compared to C diet, which, in this case, means a higher overall evenness in O group, as observed richness is not different between diets.

Regarding the combined effect of breed and diet, no significant interaction was detected between both factors for alpha and beta-diversity. Nevertheless, PCA showed a slightly different separation of C and O centroids within each breed, which could suggest a different response to diet between Duroc and Iberian microbial communities. Also, within-group dispersions revealed a possible variation in diet response based on breed, as DU-O dispersions tended to be higher than the rest, while DU-C dispersions were the lowest (these differences being significant at OTU level). Conversely, IB-C and IB-O dispersions were similar and intermediate. The higher dispersion level within DU-O cluster compared to DU-C is coincident with the increment in alpha-diversity variability detected for this group, especially when observing richness or Chao1 index values, and opposes to the sample homogeneity within DU-C. In contrast, IB-O animals did not show different alpha-diversity mean or dispersion values compared to IB-C. Thus, it can be hypothesised that Duroc animals might respond in a more heterogeneous way to dietary oleic acid increment.

We found a unique OTU, *denovo378129* (*Shuttleworthia*) as differentially abundant between all breed and diet groups. This bacterial group is relatively unknown, although it has been reported as a typical member of animal gut microbiome in several studies [59, 65]. According to our results, *Shuttleworthia* could be a key differentiator between Iberian and Duroc animals, having a relevant link either with fiber or with oleic acid dietary supply. Additionally, DESeq2 allowed us to test interaction effects between breed and diet factors. With this method we were able to find as much as 26 significant interactions, most of them being qualitative. According to DESeq2 DA analysis for diets (C vs O) using DU and IB subpopulations,

we were able to classify these interactions based on the different response to diets inside each breed: no response to O diet (NR), abundance increase due to O diet (AI) or abundance decrease due to O diet (AD). 12 OTUs were NR within DU animals (7 IB-AI and 5 IB-AD) while 7 OTUs were NR within IB animals (all of them DU-AI). Only 1 OTU (*denovo347841*) presented an abundance increase in response to O diet within both breeds, being its AI much larger for Iberian animals (quantitative interaction). The remaining 6 interaction-significant OTUs were not significantly DA between diets within each breed. The OTU *denovo347841* has been classified as *Unclassified_Muribaculaceae*, a relatively unknown *Bacteroidales* family which has been observed in multiple gut microbiomes through 16S or metagenomic analyses [66]. Although their main metabolic role remains unknown, their phylogenetic closeness to *Prevotellaceae* and *Bacteroidaceae* could reveal saccharolytic activity. In summary, most of the observed interactions are qualitative and reveal that one breed responds to diet but the other does not. No pure qualitative interactions (i.e., with opposite directions of diet effect according to breed) were observed. The presence of interactions for individual OTU abundances would be a clear indicator of differential effects of diet on the gut microbiota of Duroc and Iberian pigs. This fact can be related with our beta-diversity results, as our PCA showed a tendency to a higher centroid distance within IB-C and IB-O groups than within DU-C and DU-O, although the absence of significant interactions in PERMANOVA test throws some uncertainty about this conclusion. Also, former studies carried out with the same animals have proven the existence of a differential response to O diet for several productive traits and for adipose tissue transcriptome [67]. In summary, this differential response may exist for gut microbiome composition and structure, but might be not clearly noticed due to the low number of animals and to the relatively high resemblance between feed formulations and to the intrinsic difficulty in evaluating and interpreting interaction effects.

The p-LR approach must be mentioned apart, as it identifies ratios that could be useful as differentiators between breed and diet combinations, although distinction between Duroc and Iberian phenotypes is clearer. *Actinobacteria* abundance has been observed to be important when compared to *Epsilonbacteraeota* (Phyl. Nov.), a novel phylum proposed by Waite *et al.* [68] as a separate group from former *Epsilonproteobacteria* class, which was classified inside *Proteobacteria* phylum. Our dataset contained *Campylobacter* and *Helicobacter* members of *Epsilonbacteraeota*, two genera with diverse metabolisms more known by their pathogenic species affecting human and other animals. One out of our OTUs has been classified as *Campylobacter hyointestinalis*, a species name proposed to *Campylobacter* isolated from pigs with proliferative enteritis [67]. Interaction relationship of these pathogenic bacteria and *Actinobacteria* (*Corynebacterium*, *Bifidobacterium*, *Collinsella*, *Enterorhabdus* and non-classified *Atopobiaceae*) present in our database remains unknown. However, as *Epsilonbacteraeota* RAs remain fairly constant across samples, differentiation between IB and DU pigs might be due to lower *Actinobacteria* abundances in Iberian animals. Actinobacteria relevance in breed separation has been upheld when analysing ratios at lower taxonomic levels. Accordingly, *Corynebacteriaceae*/*Spirochaetaceae* and *Anaerostipes*/*Corynebacterium* were relevant in breed separation at family and genus levels, respectively. In any case, complete separation between breeds and diets, including the discrimination between Duroc diet groups, is only possible at family level making use of three pairwise ratios (COR/SPR, PEP/STR and ATO/ ENT), considering that *Enterobacteriaceae* are more abundant in DU-O group. This fact also reflects the masking effect that breed is causing to differentiate diet effects on microbial populations, previously commented.

This type of approach might be interesting to detect microbial biomarkers, given their importance as tools for genetic diagnosis or diet determination. In the case of Iberian pig industry this might be especially useful in order to address frauds due to crossbreeding and

dietary manipulation, which can importantly affect final products' quality [6, 7]. Nevertheless, additional studies including crossbred animals, different production systems and sampling and different developmental stages, including the standard slaughter age, are necessary in order to validate these potential biomarkers and establish causal relationships.

As an additional approach we related the individual OTU abundances with different fat composition traits, in order to deepen in the potential effects of the gut bacteria on host's metabolism. A former study with the same animals reported significant differences in FA composition between C and O diets [38]. Our first Limma approach (breed subsetting and no diet correction) was used to globally visualize the correlations between OTU abundances and lipidic composition, independently of their causal relationship. Again, differences were detected between DU and IB, as only one OTU abundance was significantly correlated to FA composition in DU animals. When the second Limma approach was applied (breed subsetting and diet correction) we saw that most of these correlations were caused by the diet influence. In IB subpopulation, *Prevotellaceae* OTUs mostly appear negatively correlated to SFA and/or positively correlated to MUFA (as the variation in one FA alters the proportion of the others). As they have a role in fiber digestion, these correlations might be caused by the higher fiber content in high-oleic diet, as mentioned before. Other correlations might be directly caused by increased supply of dietary oleic acid, such as the negative correlations between OTUs from *Lactobacillus* (*denovo2335*, *denovo274149*) and backfat SFA content, or between *Campylobacter* (*denovo141328*) and both ham and backfat SFA content. *Lactobacillus* are important fermenters with a crucial role in VFA biosynthesis [69], so its association with FA availability could be expected. Regarding *Campylobacter*, as it happens with *C. jejuni*, *C. hyointestinalis* might also lack common microbial metabolic pathways related to carbohydrate utilization, replacing them by a rapid utilization of amino acids and SCFA such as lactate and acetate [70]. Thus, its correlation with FA composition might be related to differential SCFA and amino acid bioavailability between dietary groups. Only one OTU was detected as correlated with FA after correcting by diet (second Limma approach), *denovo37314* (*Treponema 2*), negatively correlated with ham oleic acid content in IB. This suggests that direct (i.e., not caused by diet differences) association between gut microbiota and FA muscle accretion might also exist. Unfortunately, our experiment does not allow to detect these associations, as intra-group variability is very low for lipidic traits and sample size does not provide enough statistical power. Further experimental designs standardizing breed and diet might be adequate to focus on causal relationships between gut microbiota composition and host lipidic traits.

As our experiment has focused on the bacterial subset of fecal microbiota, interactions with other microbial clades, especially eukaryotes, must be studied for a better understanding of the real nature of microbial population changes. On the other hand, metabolite variations between breeds and diets must be analyzed in order to stablish the true relationship between microbiome and host phenotype. Mid-gut microbiota populations might also be interesting to observe dietary responses, as multiple digestion processes occur at superior gut sections. Finally, the complexity in analysing crossover designs must be taken into account, and additional experiments focused in a unique effect might help to clarify some of the differences found in this study, although interaction between effects must not be ignored.

Conclusions

Our study reveals that genetic background has an important impact on pig gut microbiota composition, while dietary changes have a smaller and more variable effect, and it also brings light to the complexity of microbial relationships. In this experiment, the O diet provided a high level of oleic acid content instead of carbohydrates as energy source, but also higher fiber

content was mandatory to keep both diets isoenergetic, thus the effects of diet on microbiota composition could be highly dependent on fiber. This work is the first comparison of breed and diet effects on gut microbiome involving Iberian pigs, providing novel and interesting results, but limitations due to the sample size must be taken into account. We report a higher OTU richness in Iberian pigs, which highlights the importance of genetic background (i.e., animal breed) in overall microbiome composition resemblance, and may suggest a potential relationship between gut microbiota diversity and composition and Iberian pigs' resilience. Dietary modifications had a small effect in microbiome composition, much less important than host genotype. Differential abundance revealed a complex picture of microbe relationships, with a considerable effect of the breed and multiple interactions between breed and diet. In spite of this complexity, we identified relevant DA taxa and taxa ratios as potentially associated to the metabolic differences between breeds and to a lower extent for the diet influence. Finally, an approach to the use of OTU pairwise ratios as a predictive tool has been performed, with some interesting results that should be subject of further investigations.

Supporting information

S1 Fig. DA Venn diagrams. Venn Diagram comparing DA OTUs in breed (red) and diet (blue) contrasts using both DESeq2 and ALDEx2.
(TIF)

S1 File. Diet composition. Nutrients, fatty acid composition and ingredient formulation for Control (C) and High-Oleic (O) diets.
(PDF)

S2 File. Relative abundance of genera. Tables containing overall relative abundance of identified genera and relative abundance by"breed-diet" groups. Mean and standard deviation of RA are shown for each OTU, as well as taxonomic classification. RA_Rank column represents the rareness of each OTU (High_RA: RA $\geq$ 1%; Med_RA: 1% $>$ RA $\geq$ 0.1%; Low_RA: RA $<$ 0.1%).
(XLSX)

S3 File. Differentially abundant OTUs. Tables containing differentially abundant OTUs for each factor (breed and diet) with both pipelines (DESeq2 and ALDEx2) and overlapping of both pipelines. Prevalence represents the proportion of samples in which each OTU is found. For DESeq2, normalized counts per group (Ncounts), fold change (log2FC) and adjusted p-value (padj) are shown. For ALDEx2, relative abundance per group (RA), median difference in CLR values between factor groups (diff.btw) and Welch's test adjusted p-value (we.eBH) are shown. In overlapping tables information of both pipelines is shown.
(XLSX)

Acknowledgments

The authors are grateful to the staff of the "Pig Test Center" ITACYL (Hontalbilla, Segovia, Spain) for technical assistance with husbandry and sampling. We also thank GABI-INRA for hosting the research stay for the development of bioinformatic pipelines.

Author Contributions

Conceptualization: Clemente López-Bote, Cristina Óvilo.

Data curation: Adrián López-García.

Formal analysis: Adrián López-García, Óscar González-Recio, Jordi Estellé.

Funding acquisition: Clemente López-Bote, Cristina Óvilo.

Investigation: Adrián López-García, Rita Benítez, Yolanda Núñez, Emilio Gómez-Izquierdo, Eduardo de Mercado, Juan M. García-Casco, Cristina Óvilo.

Methodology: Emilio Gómez-Izquierdo, Juan M. García-Casco, Cristina Óvilo.

Project administration: Clemente López-Bote, Cristina Óvilo.

Resources: Emilio Gómez-Izquierdo, Eduardo de Mercado.

Software: Adrián López-García, Óscar González-Recio, Jordi Estellé.

Supervision: Clemente López-Bote, Cristina Óvilo.

Visualization: Adrián López-García.

Writing – original draft: Adrián López-García, Cristina Óvilo.

Writing – review & editing: Rita Benítez, Yolanda Núñez, Emilio Gómez-Izquierdo, Eduardo de Mercado, Juan M. García-Casco, Óscar González-Recio, Clemente López-Bote, Jordi Estellé, Cristina Óvilo.

References

1. FAO. FAOSTAT Database [Internet]. Food and Agriculture Organization of the United Nations. 2020 [cited 2020 Aug 17]. Available from: http://www.fao.org/faostat/es/#data/TP/visualize
2. López-Bote CJ. Sustained utilization of the Iberian pig breed. Meat Science [Internet]. 1998 Jan [cited 2017 Jan 26]; 49:17–27. Available from: http://www.sciencedirect.com/science/article/pii/S0309174098900365 PMID: 22060709
3. Ventanas S, Ventanas J, Ruiz J. Sensory characteristics of Iberian dry-cured loins: Influence of cross-breeding and rearing system. Meat Science. 2007 Feb 1; 75(2):211–9. https://doi.org/10.1016/j.meatsci.2006.07.003 PMID: 22063652
4. Óvilo C, Benítez R, Fernández AI, Núñez Y, Ayuso M, Rodríguez C, et al. Longissimus dorsi transcriptome analysis of purebred and crossbred Iberian pigs differing in muscle characteristics. BMC Genomics [Internet]. 2014 May 31 [cited 2020 Aug 18]; 15(1):413. Available from: https://doi.org/10.1186/1471-2164-15-413 PMID: 24885501
5. Robina A, Viguera J, Perez-Palacios T, Mayoral AI, Vivo JM, Guillén MT, et al. Carcass and meat quality traits of Iberian pigs as affected by sex and crossbreeding with different Duroc genetic lines. Spanish Journal of Agricultural Research [Internet]. 2013 Nov 12 [cited 2020 Aug 18]; 11(4):1057–67. Available from: http://dx.doi.org/10.5424/sjar/2013114-4637
6. Muñoz M, García-Casco JMM, Alves E, Benítez R, Barragán C, Caraballo C, et al. Development of a 64 SNV panel for breed authentication in Iberian pigs and their derived meat products. Meat Science [Internet]. 2020 Sep 1 [cited 2020 Aug 18]; 167:108152. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0309174019311921 https://doi.org/10.1016/j.meatsci.2020.108152 PMID: 32361066
7. Daza A, Olivares A, Rey AI, Ruiz J, López-Bote CJ. Iberian pig production: The problems of success. Options Méditerranénnes [Internet]. 2008 [cited 2020 Oct 28]; A(78):163–71. Available from: http://om.ciheam.org/om/pdf/a78/00800259.pdf
8. Wood JD, Enser M, Fisher AV., Nute GR, Sheard PR, Richardson RI, et al. Fat deposition, fatty acid composition and meat quality: A review. Meat Science. 2008 Apr 1; 78(4):343–58. https://doi.org/10.1016/j.meatsci.2007.07.019 PMID: 22062452
9. Ayuso M, Fernández AI, Isabel B, Rey AI, Benítez R, Daza A, et al. Long term vitamin A restriction improves meat quality parameters and modifies gene expression in Iberian pigs. Journal of Animal Science. 2015 Jun 1; 93(6):2730–44. https://doi.org/10.2527/jas.2014-8573 PMID: 26115261
10. Ayuso M, Fernández AI, Núñez Y, Benítez R, Isabel B, Rey AI, et al. Developmental stage, muscle and genetic type modify muscle transcriptome in pigs: Effects on gene expression and regulatory factors involved in growth and metabolism. PLoS ONE [Internet]. 2016 Dec 1 [cited 2020 Oct 8]; 11(12):167858. Available from: /pmc/articles/PMC5148031/?report=abstract https://doi.org/10.1371/journal.pone.0167858 PMID: 27936208

11. Benítez R, Trakooljul N, Núñez Y, Isabel B, Murani E, de Mercado E, et al. Breed, diet, and interaction effects on adipose tissue transcriptome in iberian and duroc pigs fed different energy sources. Genes [Internet]. 2019 Aug 4 [cited 2020 Aug 11]; 10(8):589. Available from: https://www.mdpi.com/2073-4425/10/8/589

12. Hanning I, Diaz-Sanchez S. The functionality of the gastrointestinal microbiome in non-human animals. Microbiome [Internet]. 2015; 3:51. Available from: https://doi.org/10.1186/s40168-015-0113-6 PMID: 26552373

13. Khanal P, Maltecca C, Schwab C, Fix J, Tiezzi F. Microbiability of meat quality and carcass composition traits in swine. 2020;

14. Verschuren LMG, Calus MPL, Jansman AJM, Bergsma R, Knol EF, Gilbert H, et al. Fecal microbial composition associated with variation in feed efficiency in pigs depends on diet and sex. Journal of Animal Science [Internet]. 2018 Apr 14 [cited 2018 Dec 18]; 96(4):1405–18. Available from: https://academic.oup.com/jas/article/96/4/1405/4970691 https://doi.org/10.1093/jas/sky060 PMID: 29669075

15. Bergamaschi M, Tiezzi F, Howard J, Huang YJ, Gray KA, Schillebeeckx C, et al. Gut microbiome composition differences among breeds impact feed efficiency in swine. Microbiome [Internet]. 2020 Jul 22 [cited 2020 Aug 7]; 8(1):110. Available from: https://doi.org/10.1186/s40168-020-00888-9 PMID: 32698902

16. Rodríguez-Valdovinos C, García-Casco J, García-Ruiz F, Núñez-Moreno Y, Silió-López L. A further look of the genetic origin and singularity of the Torbiscal Iberian pig line. Spanish Journal of Agricultural Research. 2019; 17(1):1–13.

17. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research [Internet]. 2013 Jan 1 [cited 2019 Aug 26]; 41(1):e1–e1. Available from: https://academic.oup.com/nar/article/41/1/e1/1164457 https://doi.org/10.1093/nar/gks808 PMID: 22933715

18. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27(6):863–4. https://doi.org/10.1093/bioinformatics/btr026 PMID: 21278185

19. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods [Internet]. 2010; 7(5):335–6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3156573/ https://doi.org/10.1038/nmeth.f.303 PMID: 20383131

20. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010 Oct; 26 (19):2460–1. https://doi.org/10.1093/bioinformatics/btq461 PMID: 20709691

21. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research [Internet]. 2013 Jan [cited 2019 Aug 26]; 41(D1):D590–6. https://doi.org/10.1093/nar/gks1219 PMID: 23193283

22. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods [Internet]. 2013 Jan [cited 2019 Aug 26]; 10(1):57–9. https://doi.org/10.1038/nmeth.2276 PMID: 23202435

23. Fernandes AD, Reid JNS, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome [Internet]. 2014 May 5 [cited 2020 Feb 24]; 2(1):15. Available from: http://www.microbiomejournal.com/content/2/1/15

24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology [Internet]. 2014 Dec 5 [cited 2020 Feb 13]; 15(12):550. Available from: https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

25. Greenacre M. Compositional Data Analysis in Practice. Compositional Data Analysis in Practice. New York: Chapman and Hall/CRC; 2018.

26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research [Internet]. 2015 Jan 6 [cited 2021 Mar 16]; 43(7):e47. Available from: https://academic.oup.com/nar/article/43/7/e47/2414268 https://doi.org/10.1093/nar/gkv007 PMID: 25605792

27. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. Schneidman D, editor. PLoS Computational Biology [Internet]. 2017 Nov 3 [cited 2021 Mar 16]; 13(11):e1005752. Available from: https://doi.org/10.1371/journal.pcbi.1005752 PMID: 29099853

28. McMurdie PJ, Holmes SP. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE [Internet]. 2013; 8(4):e61217. Available from: https://doi.org/10.1371/journal.pone.0061217 PMID: 23630581

29. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: community ecology package. [Internet]. 2018. Available from: https://cran.r-project.org/web/packages/vegan/vegan.pdf

30. Palarea-Albaladejo J, Martín-Fernández JA. ZCompositions—R package for multivariate imputation of left-censored data under a compositional approach. Chemometrics and Intelligent Laboratory Systems. 2015 Apr 5; 143:85–96.

31. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Frontiers in Microbiology [Internet]. 2017 Nov 15 [cited 2019 May 21]; 8 (NOV):2224. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2017.02224/full

32. Greenacre M, Lewi P. Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements. Journal of Classification [Internet]. 2009 Apr 5 [cited 2019 Aug 26]; 26(1):29–54. Available from: http://link.springer.com/10.1007/s00357-009-9027-y

33. McArdle BH, Anderson MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology [Internet]. 2001 [cited 2020 Nov 11]; 82(1):290–7. Available from: https://sci-hub.se/https://esajournals.onlinelibrary.wiley.com/doi/10.1890/0012-9658%282001%29082%5B0290%3AFMMTCD%5D2.0.CO%3B2

34. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecology. 2001; 26(1):32–46.

35. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology [Internet]. 2010 Oct 27 [cited 2020 Feb 24]; 11(10):R106. Available from: https://doi.org/10.1186/gb-2010-11-10-r106 PMID: 20979621

36. Chen H, Boutros PC. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics [Internet]. 2011 Jan 26 [cited 2020 Jun 3]; 12(1):35. Available from: https://doi.org/10.1186/1471-2105-12-35 PMID: 21269502

37. Kamada T, Kawai S. An algorithm for drawing general undirected graphs. Information Processing Letters. 1989; 31(1):7–15.

38. Benítez R, Fernández A, Isabel B, Núñez Y, de Mercado E, Gómez-Izquierdo E, et al. Modulatory effects of breed, feeding status, and diet on adipogenic, lipogenic, and lipolytic gene expression in growing iberian and duroc pigs. International Journal of Molecular Sciences [Internet]. 2018 Jan 1 [cited 2020 Nov 23]; 19(1). Available from: /pmc/articles/PMC5795973/?report=abstract

39. Shah HN, Collins DM. Prevotella, a new genus to include Bacteroides melaninogenicus and related species formerly classified in the genus Bacteroides. International Journal of Systematic Bacteriology [Internet]. 1990 [cited 2020 Jun 17]; 40(2):205–8. Available from: https://www.microbiologyresearch.org/docserver/fulltext/ijsem/40/2/ijs-40-2-205.pdf?expires=1592395214&id=id&accname=guest&checksum=15A96C48B56E888D20977E5E29F1C5DF https://doi.org/10.1099/00207713-40-2-205 PMID: 2223612

40. de Filippo C, Cavalieri D, di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proceedings of the National Academy of Sciences of the United States of America [Internet]. 2010 Aug 17 [cited 2020 Jul 13]; 107(33):14691–6. https://doi.org/10.1073/pnas.1005963107 PMID: 20679230

41. Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker BR. Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. Scientific Reports [Internet]. 2017 Dec 1 [cited 2020 Jul 13]; 7(1). Available from: /pmc/articles/PMC5453967/?report=abstract https://doi.org/10.1038/s41598-017-02995-4 PMID: 28572676

42. Arumugam M, Raes J, Pelletier E, le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature [Internet]. 2011 May 12 [cited 2020 Jul 13]; 473(7346):174–80. Available from: https://doi.org/10.1038/nature09944 PMID: 21508958

43. Li J, Zhong H, Ramayo-Caldas Y, Terrapon N, Lombard V, Potocki-Veronese G, et al. A catalog of microbial genes from the bovine rumen reveals the determinants of herbivory. bioRxiv [Internet]. 2018;272690. Available from: https://www.biorxiv.org/content/10.1101/272690v2%0Ahttps://www.biorxiv.org/content/early/2018/02/27/272690

44. Lee HJ, Jung JY, Oh YK, Lee SS, Madsen EL, Jeon CO. 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. Applied and Environmental Microbiology. 2012 Sep; 78(17):5983–93. https://doi.org/10.1128/AEM.00104-12 PMID: 22706048

45. Lopes LD, de Souza Lima AO, Taketani RG, Darias P, da Silva LRF, Romagnoli EM, et al. Exploring the sheep rumen microbiome for carbohydrate-active enzymes. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology. 2015 Jul 8; 108(1):15–30.

46. Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, et al. A reference gene catalogue of the pig gut microbiome. Nature Microbiology [Internet]. 2016; 1(12):1–6. Available from: https://doi.org/10.1038/nmicrobiol.2016.161 PMID: 27643971

47. Ramayo-Caldas Y, Mach N, Lepage P, Levenez F, Denis C, Lemonnier G, et al. Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. ISME Journal [Internet]. 2016 Dec 1 [cited 2020 Nov 13]; 10(12):2973–7. Available from: www.nature.com/ismej https://doi.org/10.1038/ismej.2016.77 PMID: 27177190

48. Mach N, Berri M, Estellé J, Levenez F, Lemonnier G, Denis C, et al. Early-life establishment of the swine gut microbiome and impact on host phenotypes. Environmental Microbiology Reports [Internet]. 2015 Jun 1 [cited 2020 Nov 13]; 7(3):554–69. Available from: https://doi.org/10.1111/1758-2229.12285 PMID: 25727666

49. Downes J, Dewhirst FE, Tanner ACR, Wade WG. Description of Alloprevotella rava gen. nov., sp. nov., isolated from the human oral cavity, and reclassification of Prevotella tannerae Moore et al. 1994 as Alloprevotella tannerae gen. nov., comb. nov. International Journal of Systematic and Evolutionary Microbiology [Internet]. 2013 Apr 1 [cited 2020 Jul 13];63(PART4):1214–8. Available from: /pmc/articles/PMC3709537/?report=abstract

50. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection [Internet]. Vol. 36, Nature Biotechnology. Nature Research; 2018 [cited 2020 Jun 17]. p. 359–67. Available from: https://doi.org/10.1038/nbt.4110 PMID: 29553575

51. Hjorth MF, Blædel T, Bendtsen LQ, Lorenzen JK, Holm JB, Kiilerich P, et al. Prevotella-to-Bacteroides ratio predicts body weight and fat loss success on 24-week diets varying in macronutrient composition and dietary fiber: results from a post-hoc analysis. International Journal of Obesity [Internet]. 2019 Jan 1 [cited 2020 Nov 13]; 43(1):149–57. https://doi.org/10.1038/s41366-018-0093-2 PMID: 29777234

52. Xu P, Li M, Zhang J, Zhang T. Correlation of intestinal microbiota with overweight and obesity in Kazakh school children. BMC Microbiology [Internet]. 2012 [cited 2020 Nov 13]; 12. https://doi.org/10.1186/1471-2180-12-283 PMID: 23190705

53. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proceedings of the National Academy of Sciences of the United States of America [Internet]. 2005 Aug 2 [cited 2020 Nov 13]; 102(31):11070–5. Available from: www.arb-home.de https://doi.org/10.1073/pnas.0504978102 PMID: 16033867

54. Magne F, Gotteland M, Gauthier L, Zazueta A, Pesoa S, Navarrete P, et al. The firmicutes/bacteroidetes ratio: A relevant marker of gut dysbiosis in obese patients? [Internet]. Vol. 12, Nutrients. MDPI AG; 2020 [cited 2020 Nov 13]. Available from: /pmc/articles/PMC7285218/?report=abstract https://doi.org/10.3390/nu12051474 PMID: 32438689

55. Crespo-Piazuelo D, Estellé J, Revilla M, Criado-Mesas L, Ramayo-Caldas Y, Óvilo C, et al. Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. Scientific Reports. 2018 Dec 1; 8(1). https://doi.org/10.1038/s41598-018-30932-6 PMID: 30143657

56. Forero Vizcaíno FJ, Andrés AI. El cerdo ibérico una revisión transversal. Forero Vizcaíno FJ, editor. Sevilla: Consejería de Agricultura y Pesca (Andalucía); 2008. 382.

57. Elmqvist T, Folke C, Nystrom M, Peterson G, Bengtsson J, Walker B, et al. Response Diversity, Ecosystem Change, and Resilience. Frontiers in Ecology and the Environment [Internet]. 2003 Nov [cited 2020 May 29]; 1(9):488. Available from: http://doi.wiley.com/10.2307/3868116

58. Pokusaeva K, Fitzgerald GF, van Sinderen D. Carbohydrate metabolism in Bifidobacteria. Genes and Nutrition [Internet]. 2011 Aug [cited 2020 Oct 2]; 6(3):285–306. Available from: /pmc/articles/PMC3145055/?report=abstract https://doi.org/10.1007/s12263-010-0206-6 PMID: 21484167

59. Han GG, Lee JY, Jin GD, Park J, Choi YHYJ, Kang SK, et al. Tracing of the fecal microbiota of commercial pigs at five growth stages from birth to shipment. Scientific Reports [Internet]. 2018 Dec 16 [cited 2018 Dec 18]; 8(1):6012. Available from: https://doi.org/10.1038/s41598-018-24508-7 PMID: 29662088

60. Liu Y, White RH, Whitman WB. Methanococci Use the Diaminopimelate Aminotransferase (DapL) Pathway for Lysine Biosynthesis. Journal of Bacteriology [Internet]. 2010 Jul 1 [cited 2020 May 20]; 192(13):3304–10. Available from: https://jb.asm.org/content/192/13/3304 https://doi.org/10.1128/JB.00172-10 PMID: 20418392

61. Oliveira A, Oliveira LC, Aburjaile F, Benevides L, Tiwari S, Jamal SB, et al. Insight of genus Corynebacterium: Ascertaining the role of pathogenic and non-pathogenic species [Internet]. Vol. 8, Frontiers in Microbiology. Frontiers Media SA; 2017 [cited 2019 Aug 19]. p. 1937. https://doi.org/10.3389/fmicb.2017.01937 PMID: 29075239

62. Rivera-Ferre MG, Aguilera JF, Nieto R. Differences in whole-body protein turnover between Iberian and Landrace pigs fed adequate or lysine-deficient diets. In: Journal of Animal Science [Internet]. Oxford Academic; 2006 [cited 2020 Oct 5]. p. 3346–55. Available from: https://academic.oup.com/jas/article/84/12/3346/4776731 https://doi.org/10.2527/jas.2005-405 PMID: 17093227

63. Heinritz SN, Weiss E, Eklund M, Aumiller T, Louis S, Rings A, et al. Intestinal microbiota and microbial metabolites are changed in a pig model fed a high-fat/low-fiber or a low-fat/high-fiber diet. PLoS ONE. 2016; 11(4):154329. https://doi.org/10.1371/journal.pone.0154329 PMID: 27100182

64. le Sciellour M, Labussière E, Zemb O, Renaudeau D. Effect of dietary fiber content on nutrient digestibility and fecal microbiota composition in growing-finishing pigs. PLoS ONE [Internet]. 2018 [cited 2020 Nov 16]; 13(10). Available from: https://doi.org/10.1371/journal.pone.0206159.g001 https://doi.org/10.1371/journal.pone.0206159 PMID: 30356293

65. Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW, et al. The rumen microbial metagenome associated with high methane production in cattle. BMC Genomics [Internet]. 2015 Dec 23 [cited 2018 Jan 29]; 16(1):839. Available from: http://www.biomedcentral.com/1471-2164/16/839 https://doi.org/10.1186/s12864-015-2032-0 PMID: 26494241

66. Lagkouvardos I, Lesker TR, Hitch TCA, Gálvez EJC, Smit N, Neuhaus K, et al. Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. Microbiome [Internet]. 2019 Feb 19 [cited 2020 Aug 13]; 7(1):28. Available from: https://doi.org/10.1186/s40168-019-0637-2 PMID: 30782206

67. Gebhart CJ, Edmonds P, Ward GE, Kurtz HJ, Brenner DJ. "Campylobacter hyointestinalis" sp. nov.: a new species of Campylobacter found in the intestines of pigs and other animals. Journal of Clinical Microbiology. 1985; 21(5):715–20. https://doi.org/10.1128/JCM.21.5.715-720.1985 PMID: 3998100

68. Waite DW, Vanwonterghem I, Rinke C, Parks DH, Zhang Y, Takai K, et al. Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to epsilonbacteraeota (phyl. nov.). Frontiers in Microbiology. 2017 Apr 24; 8(APR).

69. Hammes WP, Vogel RF. The genus Lactobacillus. In: The Genera of Lactic Acid Bacteria [Internet]. Springer US; 1995 [cited 2020 Oct 8]. p. 19–54. https://link.springer.com/chapter/10.1007/978-1-4615-5817-0_3

70. Stahl M, Butcher J, Stintzi A. Nutrient acquisition and metabolism by Campylobacter jejuni. Frontiers in cellular and infection microbiology [Internet]. 2012 Feb 7 [cited 2020 Oct 28]; 2:5. Available from: http://journal.frontiersin.org/article/10.3389/fcimb.2012.00005/abstract https://doi.org/10.3389/fcimb.2012.00005 PMID: 22919597

Chapter 5: Fungal and ciliate protozoa are the main rumen microbes associated with methane emissions in dairy cattle

GigaScience
Q1 Computer Science Applications best quartile
SJR 2020 2.95 powered by scimagojr.com
GigaScience, 2022, **11**, 1–14
DOI: 10.1093/gigascience/giab088
RESEARCH

Fungal and ciliate protozoa are the main rumen microbes associated with methane emissions in dairy cattle

Adrián López-García[1], Alejandro Saborío-Montero[1,2], Mónica Gutiérrez-Rivas[1], Raquel Atxaerandio[3], Idoia Goiri[3], Aser García-Rodríguez[3], Jose A. Jiménez-Montero[4], Carmen González[1], Javier Tamames[5], Fernando Puente-Sánchez[5], Magdalena Serrano[1], Rafael Carrasco[6], Cristina Óvilo[1] and 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]NEIKER – Instituto Vasco de Investigación y Desarrollo Agrario. Basque Research and Technology Alliance (BRTA), Campus Agroalimentario de Arkaute s/n, 01192 Arkaute, Spain
[4]Confederación de Asociaciones de Frisona Española (CONAFE), Ctra. de Andalucía km 23600 Valdemoro, 28340 Madrid, Spain
[5]Departamento de Biología de Sistemas, Centro Nacional de Biotecnología, CSIC, Madrid, 28049 Madrid, Spain
[6]Departamento de Periodismo y Nuevos Medios, Universidad Complutense de Madrid, Ciudad Universitaria s/n, 28040 Madrid, 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
*Correspondence address. Óscar González-Recio. INIA-CSIC, Dpt Mejora Genética Animal, Crta. de la Coruña km 7.5, 28040 Madrid, Spain; E-mail: gonzalez.oscar@inia.es

Abstract

Background: Mitigating the effects of global warming has become the main challenge for humanity in recent decades. Livestock farming contributes to greenhouse gas emissions, with an important output of methane from enteric fermentation processes, mostly in ruminants. Because ruminal microbiota is directly involved in digestive fermentation processes and methane biosynthesis, understanding the ecological relationships between rumen microorganisms and their active metabolic pathways is essential for reducing emissions. This study analysed whole rumen metagenome using long reads and considering its compositional nature in order to disentangle the role of rumen microbes in methane emissions.

Results: The β-diversity analyses suggested a subtle association between methane production and overall microbiota composition ($0.01 < R^2 < 0.02$). Differential abundance analysis identified 36 genera and 279 KEGGs as significantly associated with methane production ($P_{adj} < 0.05$). Those genera associated with high methane production were Eukaryota from Alveolata and Fungi clades, while Bacteria were associated with low methane emissions. The genus-level association network showed 2 clusters grouping Eukaryota and Bacteria, respectively. Regarding microbial gene functions, 41 KEGGs were found to be differentially abundant between low- and high-emission animals and were mainly involved in metabolic pathways. No KEGGs included in the methane metabolism pathway (ko00680) were detected as associated with high methane emissions. The KEGG network showed 3 clusters grouping KEGGs associated with high emissions, low emissions, and not differentially abundant in either. A deeper analysis of the differentially abundant KEGGs revealed that genes related with anaerobic respiration through nitrate degradation were more abundant in low-emission animals.

Conclusions: Methane emissions are largely associated with the relative abundance of ciliates and fungi. The role of nitrate electron acceptors can be particularly important because this respiration mechanism directly competes with methanogenesis. Whole metagenome sequencing is necessary to jointly consider the relative abundance of Bacteria, Archaea, and Eukaryota in the statistical analyses. Nutritional and genetic strategies to reduce CH_4 emissions should focus on reducing the relative abundance of Alveolata and Fungi in the rumen. This experiment has generated the largest ONT ruminal metagenomic dataset currently available.

Keywords: dairy cattle, microbiome, rumen, methane, Nanopore, long reads

Introduction

Next-generation sequencing technologies have provided special relevance to microbial communities from different niches because they let their taxonomic and functional profile be identified. This has made it possible to unravel the relationships between host and microbiota, as well as the complex interactions between microbes, focusing on the special contribution of the role of digestive microbiome in complex traits both in humans [1] (e.g., Type 2 diabetes, cancer, mental diseases) and in domestic animals [2,3] (e.g., feed efficiency, methane emissions, animal health).

Microbial communities are of special relevance in livestock. In ruminants, one of the main microbial communities is found in the rumen, owing to its high diversity and large microbial mass [4] and its main role in feed fermentation to provide substrate to the animal, which is then transformed into product. Additionally, enteric methane is produced in the rumen by methanogenic microorganisms during feed fermentation [5] and is the main contribu-

Received: August 11, 2021. **Revised:** October 18, 2021. **Accepted:** November 30, 2021

tor of greenhouse gases from livestock, with 2.8–3.5 gigatonnes of CO_2-equivalent per year [6,7]. The ongoing climate emergency urgently calls for efficient strategies to mitigate the carbon footprint from all sectors, including agriculture and livestock farming. Previous studies have proven that complex traits in ruminants are usually influenced by global changes in ruminal microbial communities, more than by fluctuations in the abundance of specific microorganisms [8,9]. These global changes are usually due to the intricate interactions between different species in these communities (i.e., predation, competition of ecological niche, or codependency). Consequently, a better understanding of the interactions between microbial genes during methanogenesis is needed to propose strategies for reducing methane emissions. Promising strategies have been proposed to modulate the metagenome, nutrition, and genetics [10].

Classic statistical approaches do not allow the results of microbiome studies to be accurately assessed. The high sparsity of these data and their compositional nature generate multiple problems in statistical analysis, including subcompositional incoherence, increased false-positive rates in differential abundance analyses, and detection of spurious correlations [11].

As a consequence, new approaches considering both compositionality and multiple correlations are needed. It is also important to point out the advantages of whole-metagenome sequencing over metataxonomic studies because the latter cannot be used to determine functionality and because they pose some difficulties at simultaneously analysing different superkingdoms [12], which is necessary to account for the total variability of microbiomes and the interactions among their components. Different amplicons must be used to correctly classify Bacteria, Archaea, Protozoa, and Fungi, increasing the cost of the studies and involving additional bias due to PCR [13]. They pose the additional difficulty of a proper comparison between communities sequenced in different reactions with different primers. Nanopore sequencing offers a cost-efficient sequencing strategy for metagenomics studies, providing both taxonomical and functional information simultaneously and for microbes from all superkingdoms. This technology has been improved in recent years, allowing taxonomic and functional assignments to be performed with an accuracy comparable to Illumina [14].

The objective of this study was to characterize the taxonomical and functional composition of rumen microbiota using long sequence reads obtained with Nanopore technology, and their relationship with enteric methane emission.

Results

Taxonomy of microbial composition

After initial selection of core taxonomy, 6,394,671 reads with N50 = 4,022 bp were classified in 3,921 taxonomical features up to genus level. A filtering strategy was implemented to exclude low-abundance microbes while keeping the core microbiome relevant for methane emissions. This process removed 48,517 reads (<1%), which reduced the sparsity of the metagenome from 87% to 68%, although a large number of singleton and doubleton features remained (Supplementary Fig. S1). The final core subcomposition included a total of 6,318,344 reads, in 437 samples, classified in 1,240 taxonomical features: 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 relative abundance (RA) distribution by superkingdoms and phyla is summarized below.

Predominant microorganisms in this core rumen subcomposition were bacteria (mean 91.6% [SD 6.93] of total mean RA) from Bacteroidetes, Firmicutes, and Fibrobacteres (Fig. 1), representing a mean RA of 63%, 16%, and 5%, respectively. The Bacteroidetes fraction was majorly composed by *Prevotella* and was the main representative genus in the total community (19.4% mean 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% mean RA of the core metagenome. Eukaryotes represented a total mean RA of 8.2% (SD 6.95) of the core subcomposition. Predominant eukaryotic clades were those included in the SAR supergroup (Stramenopiles-Alveolata-Rhizaria) [15], accounting for 6% of total mean RA, followed by Fungi (1.3% of total mean 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 (mean 0.24% [SD 0.25] of total mean RA) consisted mostly of Methanomicrobia, Methanobacteria, and Thermoplasmata members. Yet, a large number of reads could not be assigned to a known genus. The relative abundance per animal of the most relevant taxonomic groups is depicted in Supplementary Fig. S2.

Functionality of microbial composition

A total of 30,326,550 reads with N50 = 5,720 bp were assigned to KEGGs. After prevalence filtering, a total of 84,219 reads (0.28%) were removed and the sparsity was reduced from 72% to 39% (Supplementary Fig. S1). The final KEGG table was composed by 30,145,459 reads from 437 samples, classified in 6,644 KEGGs. These KEGG pathways and BRITE hierarchies [16–18] were represented in a Treemap according to their mean RA (Fig. 2). We found that 26% of the rumen metagenome functions were in pathways that represent the metabolism of carbohydrates, amino acids, and other biological compounds, as well as of energy metabolism. In addition, 34% of functions were involved in cellular generic processes (cell growth [3%], transport and catabolism [4%], genetic and environmental information processing [23%], and other [4%]). KEGG BRITE classification showed a high presence of proteins involved in cellular processes (36%) and metabolism (26%).

β-diversity and PERMANOVA analysis

The β-diversity was represented in principal component analysis (PCA) between samples at 5 different taxonomic levels (phylum, class, order, family, and genus), as well as with KEGG, using centered log-ratio (CLR) transformed datasets. Then a permutational analysis of variance (PERMANOVA) was implemented [11], sequentially adding the effect of farm-batch (B), stage of lactation (SL), number of lactation (NL) and level of methane emissions (CH4) discretized in 4 groups (LOW, L-MID, H-MID, and HIGH). The visualization did not show a clear visual clustering of samples by methane emission levels (Fig. 3). However, a generalized additive model (GAM) smooth fitting allowed visualization of non-linear distribution patterns of the microbial samples according to CH_4 emissions inside the ordination at all taxonomic levels. The non-linear pattern was more evident at the phylum, class, and genus levels, although the proportion of methane variability explained was low ($\simeq$4.8% according to GAM model fitting). No relevant differences were visually appreciated using the KEGG information.

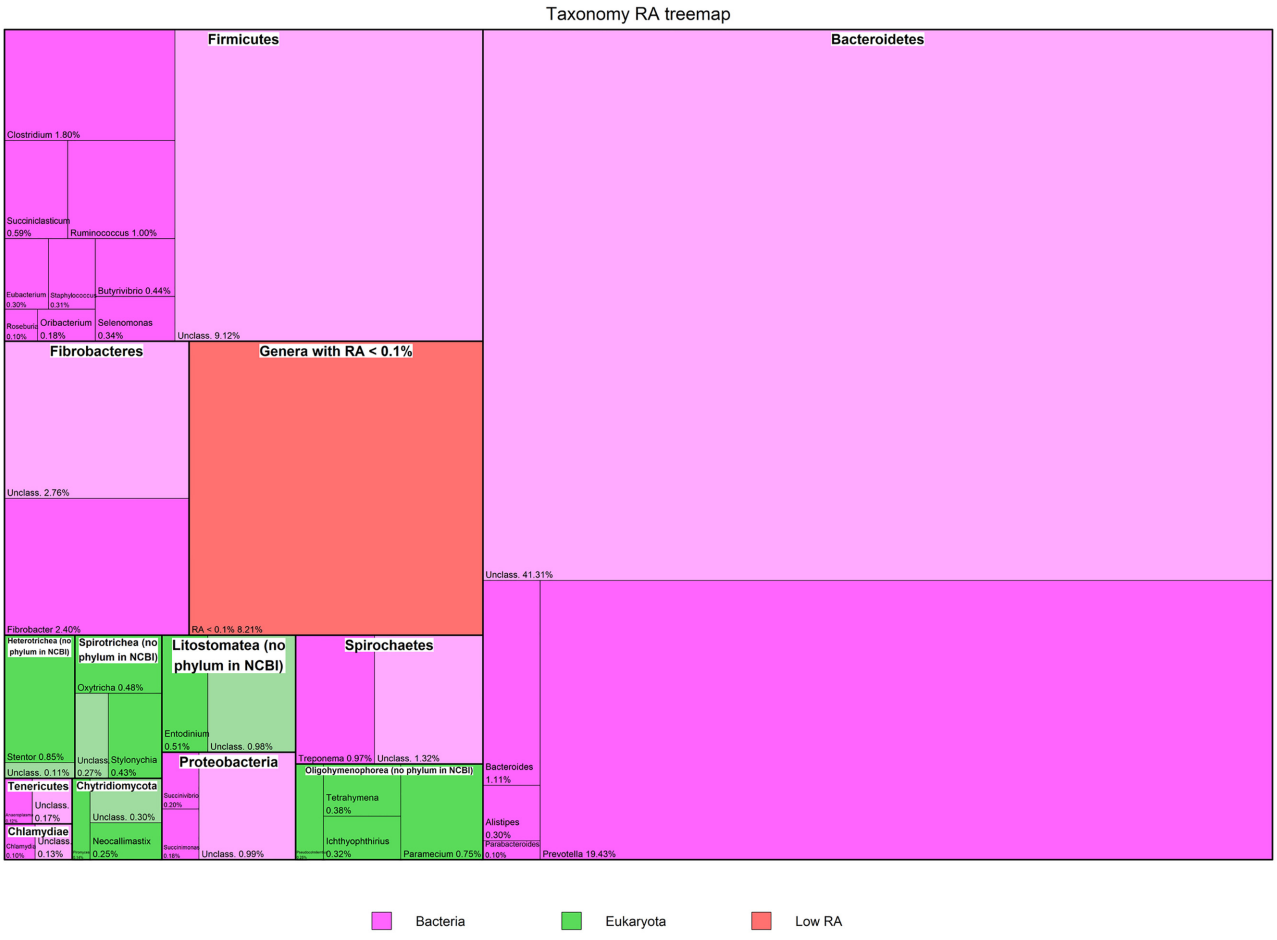

Figure 1: Mean relative abundance of genera. Mean relative abundance of core microbial taxa, including those classified only to family level (i.e., unclassified genera), which represent 60.2% of total abundance.

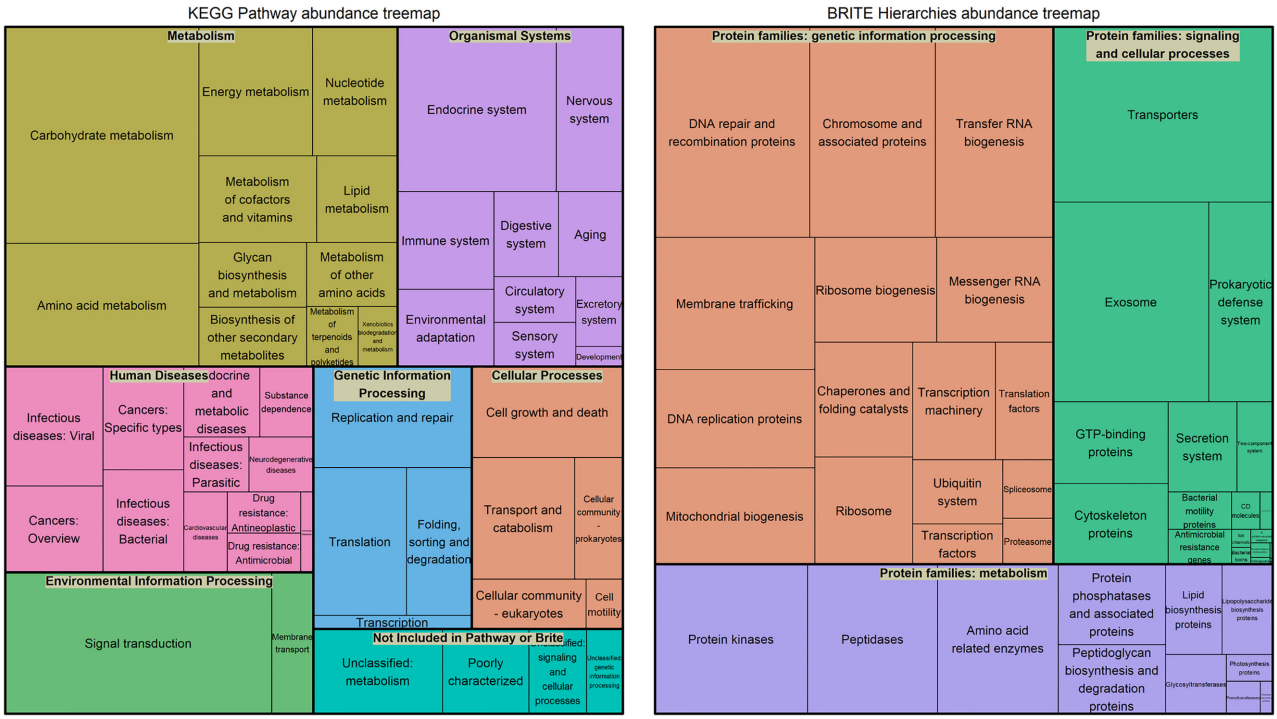

Figure 2: Metagenome functionality. TreeMap distribution of functionality abundances classified as KEGG pathways (left) and BRITE hierarchies (right) associated with core KEGG subcomposition.

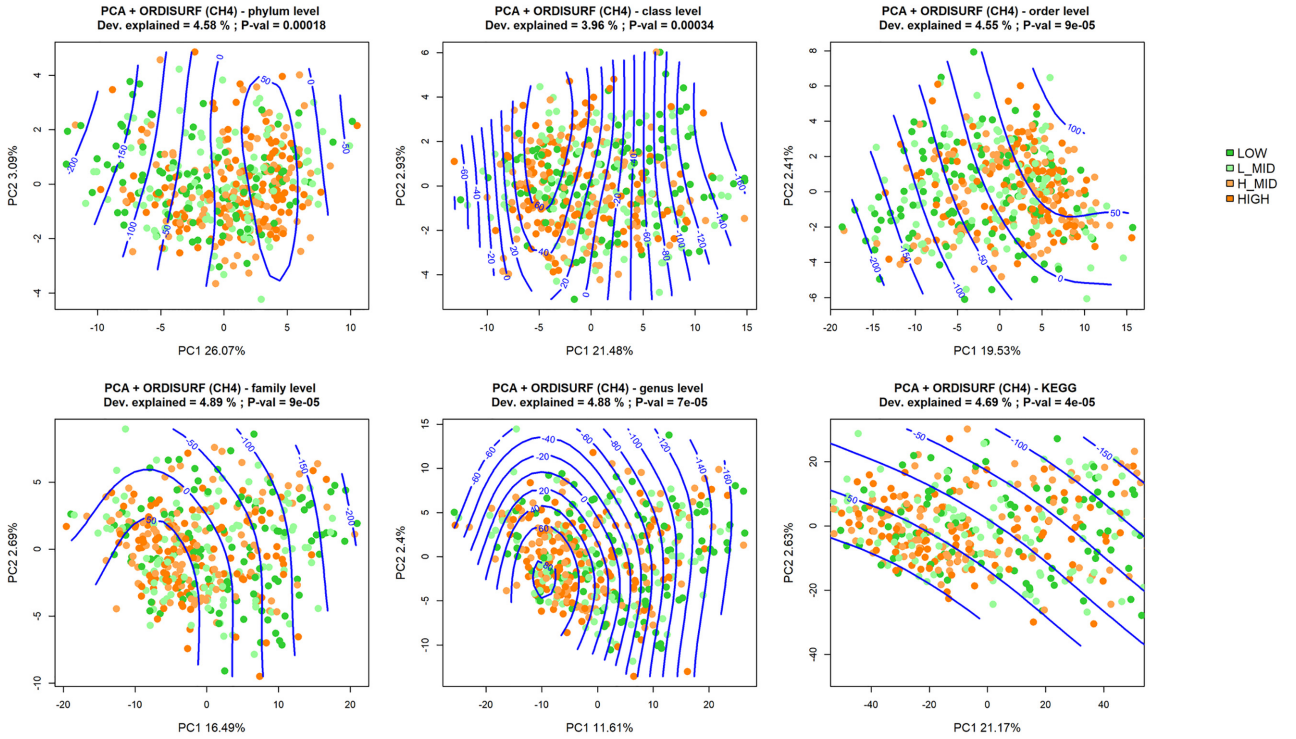

Figure 3: Fitted surface representation of principal component analysis. Dots represent the samples using Euclidean distances of CLR-transformed taxa abundances, coloured by CH4 levels. CH_4 emissions (ppm) corrected by number and stage of lactation are represented as smooth fitting following a generalized additive model (GAM) (green). Dev. Explained: variability explained by GAM; P-val: approximate significance of the smooth terms being 0 ($\alpha = 0.05$).

Table 1: F statistic and *P*-values for stage of lactation (SL), number of lactation (NL), and methane emission (CH4) variables (added sequentially) and *P*-values from PERMANOVA of the entire dataset (i.e., including all superkingdoms)

Category	Variable	F statistic	R^2	P-value*
Phylum	SL	6.1	0.014	<0.01
	NL	1.4	0.003	0.11
	CH4	2.8	0.019	<0.01
Class	SL	5.6	0.013	<0.01
	NL	1.5	0.003	0.07
	CH4	2.4	0.016	<0.01
Order	SL	5.4	0.012	<0.01
	NL	1.7	0.004	0.03
	CH4	2.3	0.016	<0.01
Family	SL	4.9	0.011	<0.01
	NL	1.6	0.004	0.03
	CH4	2.1	0.014	<0.01
Genus	SL	4.0	0.009	<0.01
	NL	1.4	0.003	0.03
	CH4	1.7	0.012	<0.01
KEGG	SL	5.3	0.012	<0.01
	NL	2.0	0.004	0.02
	CH4	2.4	0.016	<0.01

*Italics indicate a statistically significant finding ($P < 0.05$).

Nonetheless, some differences in the overall rumen microbiome composition between animals with different methane emissions were evidenced by the PERMANOVA analysis, both for taxonomy and functionality (Table 1). The results showed significant differences for the centroid distance between methane emission groups at every taxonomic level and also for KEGGs ($P < 0.01$), but they explained a low percentage of total variance ($0.01 < R^2 < 0.02$).

Rumen microbes associated with CH_4 emissions

The effect of taxonomic features on methane emission levels was evaluated through differential abundance (DA) analysis. Thirty-three genera were found to be differentially abundant (P_{adj} <0.05) between LOW and HIGH emitters (Fig. 4A), while 15 genera showed DA between LOW and H-MID emitters and 1 genus between LOW and L-MID emitters (Supplementary Data S1). Note that 13 of the 15 genera showing DA (P_{adj} <0.05) between LOW and H-MID groups were also significant in the LOW vs HIGH contrast but not in LOW vs L-MID contrast, indicating gradual abundance change from low to high emitters. Accounting for all contrasts and duplicated genera, 36 DA genera had statistically significant results. We classified these genera according to their respective overabundance (OA) in the LOW or HIGH emissions groups. Thus, 10 of them were more abundant in the LOW group (LOW-OA) and 1 in the L-MID group. The remaining 25 genera were OA in the HIGH groups (HIGH-OA): HIGH (12), HIGH and H-MID (11), or H-MID (2). HIGH-OA genera represented an overall RA of 4.15%, whereas LOW-OA genera accounted for 0.25% of total RA. The 2 genera overabundant in H-MID were *Dictyostelium* and *Unclassified Eimeriidae*, and the one associated to L-MID was classified as *Candidatus Izimaplasma* (Tenericutes). The $\log_2$FC values ranged between 0.7 and −0.7 in genera showing DA for methane emission levels, highlighting that the differences between groups were moderate.

Overall, DA results indicate that taxa associated with higher methane levels belong to the Eukaryota superkingdom, while those associated with lower emissions were bacteria. We found

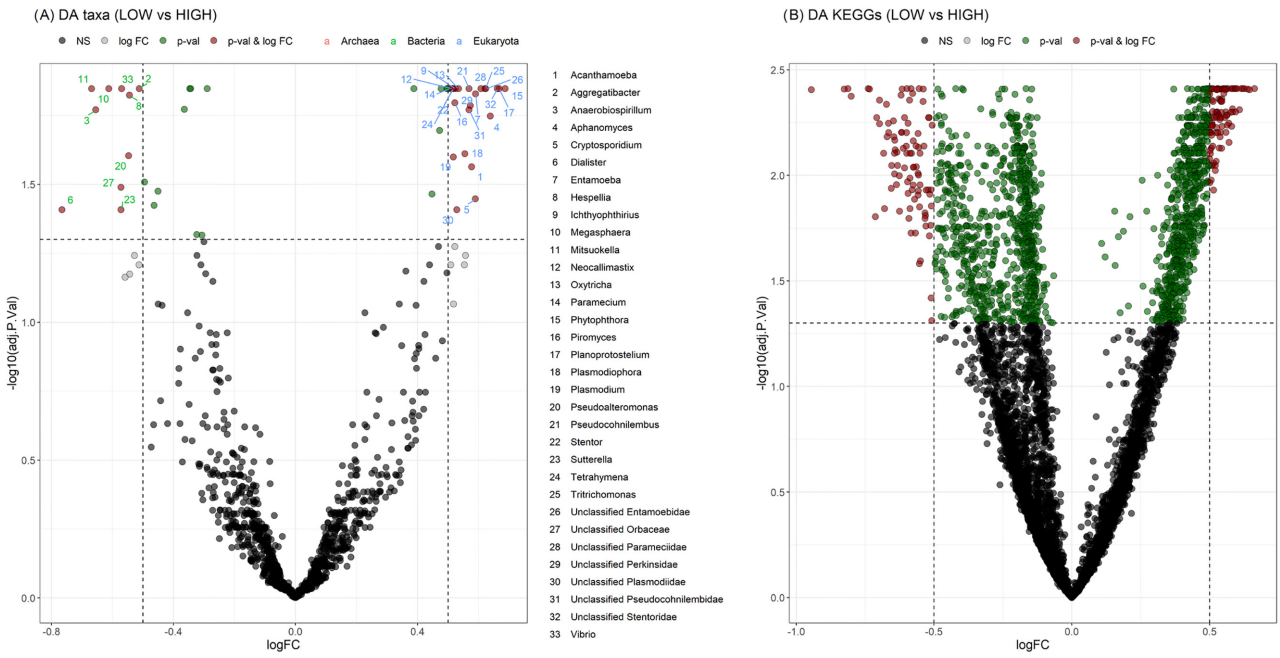

Figure 4: Volcano plots. Volcano plot representing the differential abundance (DA) of genera (A) and KEGGs (B) between LOW and HIGH groups from limma. Significance thresholds were established at $P_{adj} = 0.05$ and $\log_2 FC = \pm 0.5$. Red points: Significant features with DA above the fold change (FC) threshold. Green points: Significant features with DA below the FC threshold. Gray points: Non-significant features with DA above the FC threshold. Black points: Non-significant features with DA below the FC threshol.

multiple Ciliophora genera associated with the HIGH group (mostly Parameciidae, Stentoridae, and Pseudocohnilembidae members) but also Amoebozoa and some Fungi or pseudo-fungi. Other bacterial genera associated with lower methane production were *Hespellia*, from Clostridiales, and *Sutterella*, an asaccharolytic genus from Betaproteobacteria.

Microbial gene function associated with CH_4 emissions

DA analysis was also performed for KEGG features on methane emission levels. A total of 192 KEGGs were DA between the LOW and HIGH emissions groups (Fig. 4B). Differences were also found between the LOW and H-MID groups (Supplementary Data S1). As in the taxonomy dataset, some of the KEGGs presented significant DA in both LOW vs HIGH and LOW vs H-MID contrasts. Accounting for these duplicates and all the contrasts, 182 were overabundant in the high-emissions groups (HIGH-OA), whereas 97 KEGGs were overabundant in low-emissions groups (LOW-OA). The overall RA for HIGH-OA KEGGs was 2.31% and 0.64% for LOW-OA KEGGs. Of these, 13 HIGH-OA KEGGs and 28 LOW-OA KEGGs were assigned to metabolic pathways. No KEGGs from the ko00680 pathway were found as HIGH-OA. KEGGs related to inositol-phosphate metabolism (K00889, K01110, K18082, and K20279), starch and sucrose metabolism (K01203), or several lipid metabolism pathways were present in the HIGH-OA group. According to LOW-OA KEGGs, some of them were involved in volatile fatty acid metabolism (e.g., K00209 enoyl-[acyl-carrier protein] reductase [EC:1.3.1.9], K01902 succinyl-CoA synthetase alpha subunit [EC:6.2.1.5], and K01682 aconitate hydratase 2 [EC:4.2.1.3]) and the K09251 putrescine aminotransferase (EC:2.6.1.82) related to putrescine and cadaverine degradation to 4-amino-butanoate (GABA) or 2-oxoglutarate. Also, several KEGGs in the LOW-OA group were related to nitrogen metabolism (K00370 and K00371 nitrate reductase subunits [EC:1.7.5.1]), oxidative phosphorylation (K03885 NADH dehydrogenase [EC:1.6.99.3]), and to carbohydrate, lipid, or vitamin metabolism pathways. The ko00680 KEGG K13788 was also overabundant in the LOW emissions group.

Co-abundance of genera and KEGGs

Interaction networks were built using the previous results in order to visualize the association between taxa and genes using pairwise correlations between features. Pairwise proportionality correlation coefficients (ρ_p) were calculated on the CLR-transformed datasets for phylum, genus, and KEGG features to mitigate the effect of spurious correlations that can potentially surge in compositional data [19].

The most relevant pairwise proportionalities between genera and between KEGGs were visualized as interaction networks, classifying features as associated with high methane emissions (HIGH), low methane emissions (LOW), or not associated with methane emissions (N/A), according to the results from the differential abundance analyses. The interaction networks for genera and KEGGs are shown in Figs 5 and 6, respectively.

Eukaryotes clustered together in the network with large representation of the SAR supergroup and showed negative proportionality to bacteria. The genera that were associated with higher methane emissions belonged to the Eukaryota superkingdom (Ciliophora and Fungi), whereas Bacteria were associated with lower CH_4 production. The strongest inverse proportionalities between both subpopulations connected several eukaryotes with Unclassified Veillonellaceae and *Oribacterium* ($-0.64 < \rho_p < -0.53$); i.e., microbiomes with lower abundance of *Oribacterium* or Veillonellaceae tend to present larger abundances of protozoa and fungi and were therefore associated with larger emissions. Unclassified microbes from Neocallimastigaceae, Oxytrichidae, and Vibrionaceae families showed the highest centrality and a large degree of connectivity.

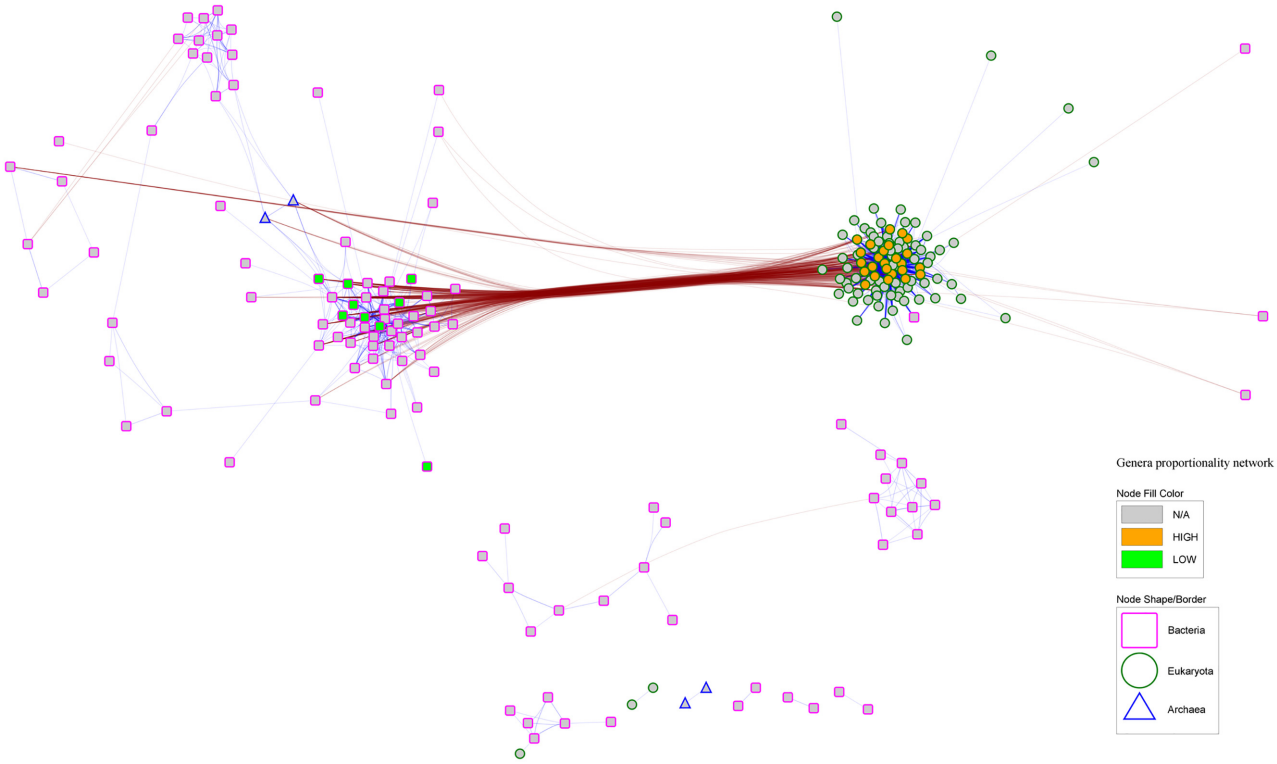

Figure 5: Taxonomy interaction network. Pairwise proportionalities between genera with $|\rho_p| \geq 0.4$. Node shapes and shape colour indicate superkingdom, and node fill colour, CH_4 association. Blue links indicate direct proportionality (>0), and brown links, inverse proportionality (<0).

The functional network showed 3 main clusters that grouped KEGGs associated with HIGH methane level (cluster H), KEGGs not related to methane emissions (cluster N), and a small one including KEGGs associated with lower emissions (cluster L). Connections between clusters were not symmetric: H cluster was connected to N cluster by inverse proportionalities between some of their components, but the L cluster seemed to be connected only to N cluster by direct proportionalities through non-clustered KEGGs. Also, most of the ko00680 KEGGs (i.e., directly involved in methanogenesis or participating in pathways leading to methanogenesis precursors) did not appear as DA between high-emission and low-emission cows.

Distribution of genes among clades

A traceback of genes' taxonomy was carried out, separately for ko00680 KEGGs and for DA KEGGs. A total of 30 of the 85 ko00680 KEGGs were predominant in Archaea groups, 1 predominated in Eukaryota (K05979), and the rest were predominant in Bacteria (Fig. 7). Although the RA distribution of these KEGGs was normally between 60% and 100% in the predominant superkingdom, 4 KEGGs were more evenly distributed between clades: K01007 and K00863 had RA < 60% in Bacteria and showed RA > 30% in Eukaryota; K05979 was the KEGG predominating in Eukaryota, but with RA near 60% (38% in Bacteria and 12% in Archaea); and K14080 had RA of 57% in Archaea and 43% in Bacteria. Regarding the DA KEGGs, those from the LOW-OA group showed larger abundance in Bacteria, mostly in genera from Proteobacteria, Bacteroidetes, and Firmicutes phyla. Different groups of bacteria also carried KEGGs from the HIGH-OA group, although these KEGGs were more abundant in eukaryotes. The HIGH-OA KEGGs were mainly mapped to unclassified eukaryotes, but those that could be classified belonged most often to Fungi and SAR supergroup (Fig. 8).

Discussion

In this study we assessed the composition of the ruminal microbiota using long reads from Nanopore sequencing technology. We observed predominance of Bacteroidetes, Firmicutes, and Fibrobacteres, as reported in previous studies [8,20]. Bacteroidetes and Firmicutes are common bacteria in all kind of ecosystems, including gut microbiota of multiple animals. The fraction of Bacteroidetes was mainly composed by *Prevotella*, a group of anaerobic gram-negative bacteria involved in saccharolytic processes [21]. Their large abundance in the digestive microbiota has been previously reported in ruminant [22–26] and monogastric species [27,28]. Firmicutes were less abundant, with a more diverse distribution of genera. Fibrobacteres, a small group of cellulose-degrading bacteria usually present in ruminant digestive systems [29], was mainly represented by the *Fibrobacter* genus. Eukaryotes also represented a relevant amount of the rumen core metagenome. This group has been reported to contribute up to 50% of total ruminal biomass [30]. The SAR supergroup and Fungi were the most relevant ones, which are found in a wide variety of ruminants and pseudoruminants [15,31]. Other eukaryotes included *Stentor* and *Paramecium*; the former are aquatic free-living heterotricheans that can be particle filtrators or predators of other protozoa and live symbiotically with some algae species [32,33], whereas the latter are well-known ciliates that predate bacteria and other microorganisms, including protozoa [34]. Archaeal fraction was mostly composed of strict methanogenic organisms from Methanomicrobia and Methanobacteria clades [35]

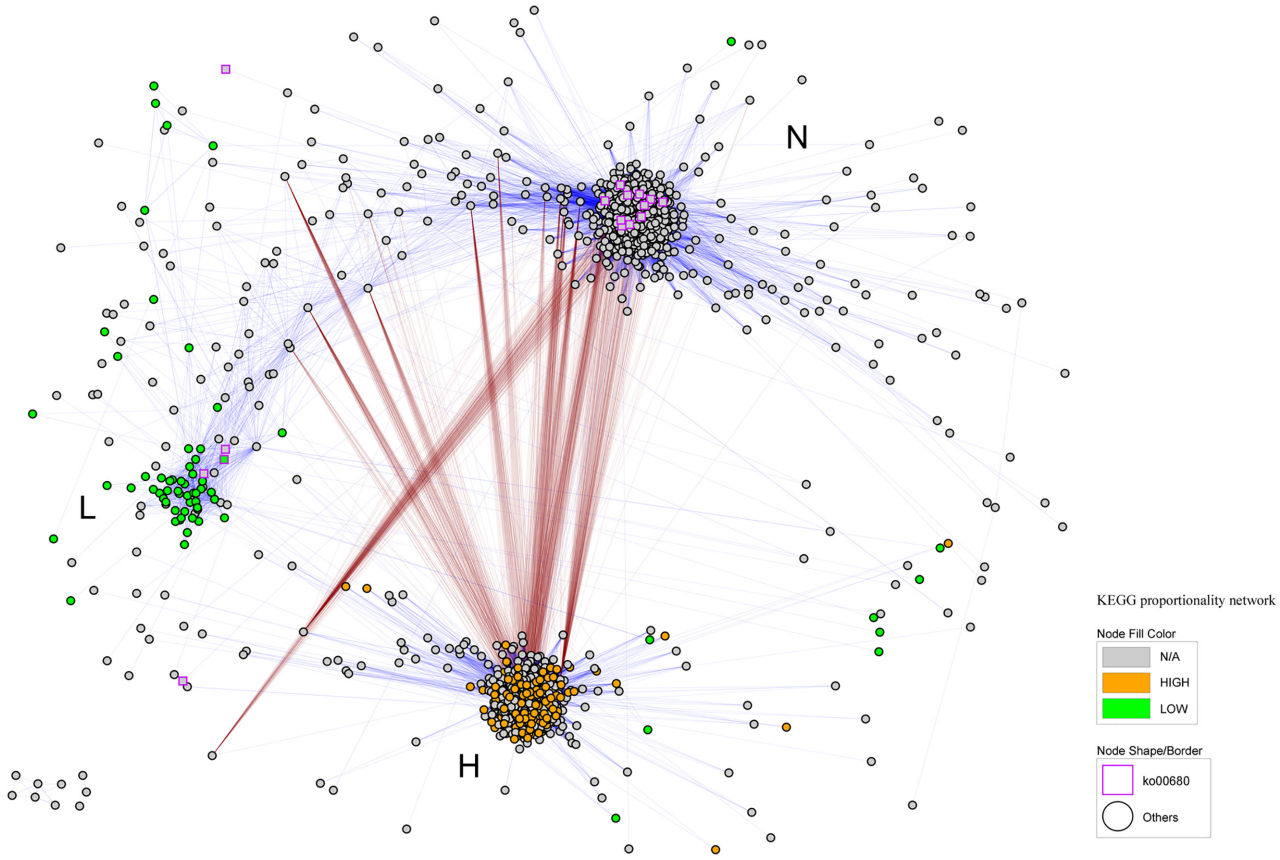

Figure 6: Functionality interaction network. Presented are pairwise proportionalities between KEGGs with $|\rho_p| \geq 0.7$. Node shapes and shape colour indicate participation in methane metabolism (ko00680 [direct or indirect participation] or other [no participation]), and node fill colour, CH_4 association. Blue links indicate direct proportionality (>0), and brown links, inverse proportionality (<0). Clusters are indicated as L (KEGGs associated with LOW methane), H (KEGGs associated with HIGH methane), and N (KEGGs not related to methane emissions).

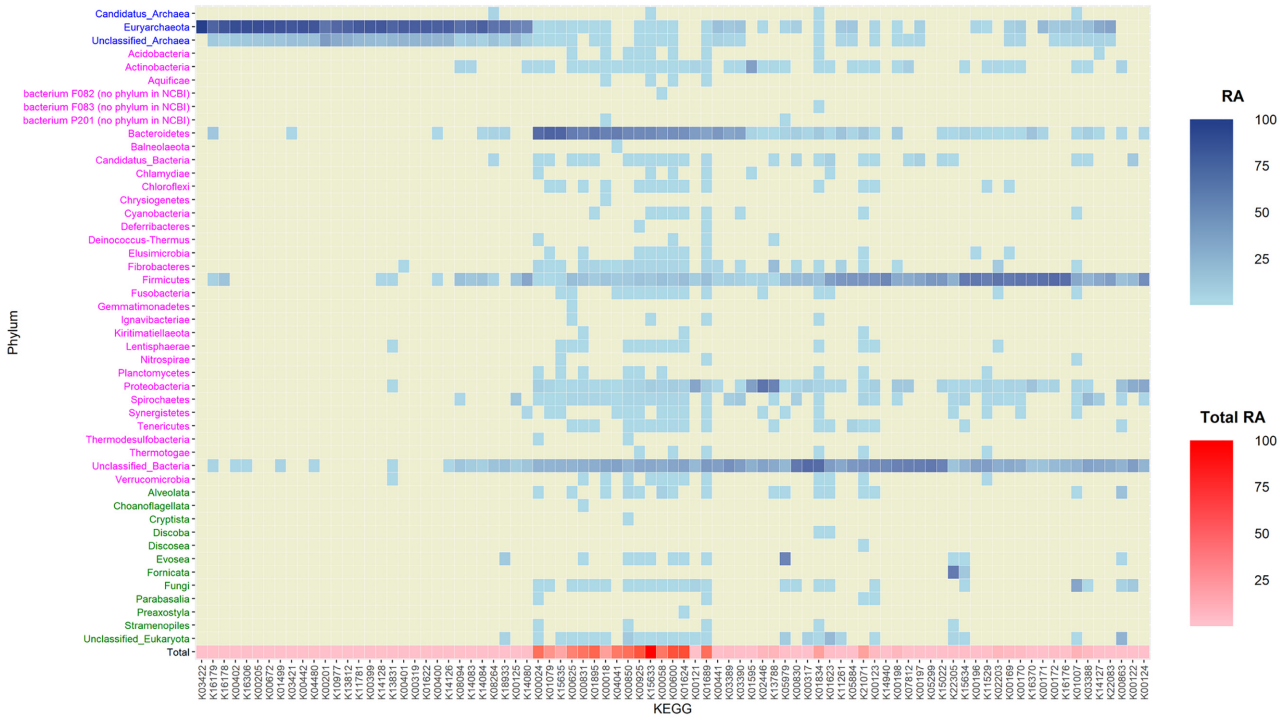

Figure 7: Taxonomy of ko00680 KEGGs. Relative abundance of KEGGs present in ko00680 pathway for each phylum in Archaea (blue), Bacteria (fuschia), and Eukaryota (green) superkingdoms. Relative abundance of each ko00680-KEGG with respect to the sum of reads mapped to all ko00680-KEGGs.

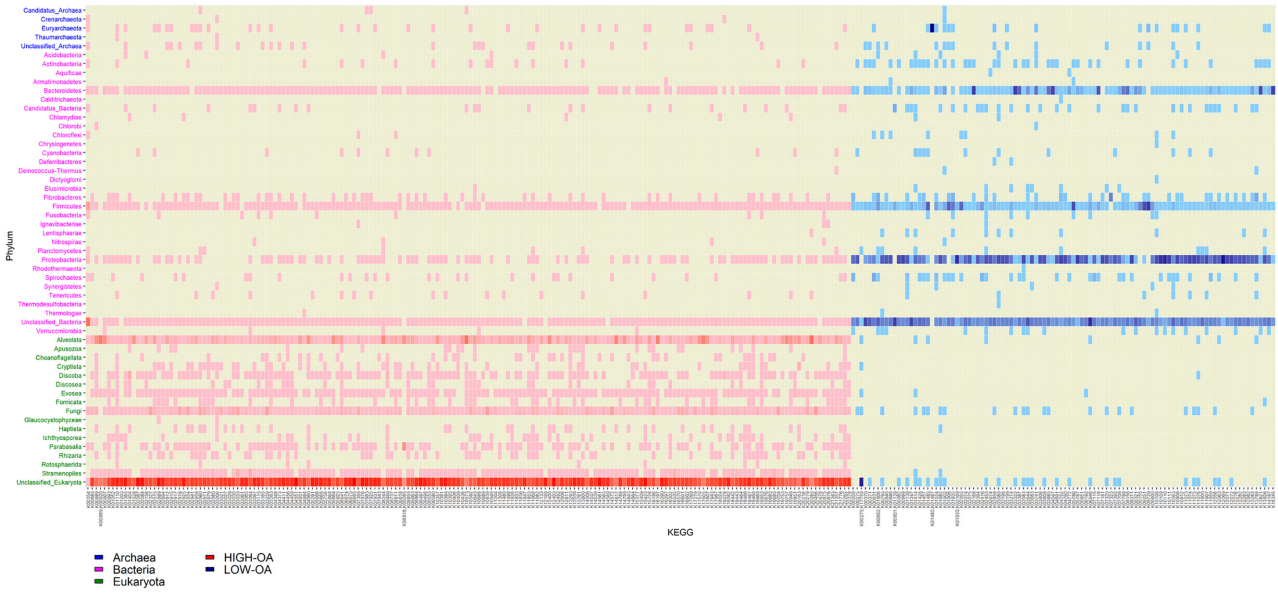

Figure 8: Taxonomic distribution of DA KEGGs. Red density scale represents KEGGs overabundant (OA) in HIGH emitters; Blue density scale represents KEGGs OA in LOW emitters. More intense colours mean a higher number of reads assigned to 1 phylum.

but also included Thermoplasmata, which are methylotrophic-methanogenic acidophilic organisms [36].

The DA analysis showed that ciliates, fungi, and pseudo-fungi were more abundant in cows with higher levels of methane emissions. Microbes associated with lower methane emissions were saccharolytic members of class Gammaproteobacteria (*Anaerobiospirillum* [37], *Vibrio* [38], or *Pseudoalteromonas* [39]), as well as Negativicutes genera from Veillonellaceae (*Dialister*, *Megasphaera*) and Selenomonadaceae (*Mitsuokella*). *Dialister* produce succinate decarboxylation, and *Megasphaera* ferment carbohydrate and lactate [40], while *Mitsuokella* are saccharolytic bacteria [41]. The low-emissions ruminotype had larger abundance of Proteobacteria and Firmicutes genera. Other authors also reported higher abundances of these bacterial phyla in low methane emissions animals [8]. Lactate and succinate producers have been reported to be more abundant in low-emitters as well [42], supporting the higher abundance of *Anaerobiospirillum* or *Megasphaera* in LOW animals.

Despite this association between methane and large taxonomic groups, it is of interest to find out which specific clades and microbial genes are participating directly or indirectly in methanogenesis. The genera co-abundance network showed a clear cluster of eukaryotes, with many of them being significantly more abundant in the high-emissions group. Other authors have already established a positive correlation between fungi abundance and methane emissions [8], as well as a close interdependence of protists and fungi. Although correlation between methane emissions and protozoa abundances is still under discussion [43,44], current meta-analyses point to a linear relationship between protozoa abundance and methane emissions ($r = 0.96$) [45].

Interestingly, no taxonomic group of methanogenic archaea showed association with methane emissions. The relationship between Archaea and methane production in rumen is not consistent in the literature. Some authors reported either individual relationships between methane emissions and some archaeal species [46,47] or correlations between overall archaeal gene abundance and methane emissions level [43,48]. However, other studies showed no relationship between methanogenic Archaea and methane [47,49]. All studies to date showed a low relative abundance of archaea in the rumen, compared to eukaryotes and bacteria [50]. However, the association between the abundance of rumen eukaryotes and methane emissions has been demonstrated through defaunation experiments, both *in vitro* [51,52] and *in vivo* [44,53], with lower emissions in defaunated animals [54]. This has been attributed to the tight link existing between methanogenic archaea abundance and some fungi and protozoa [50]. Specifically, ciliates and some Chytridiomycota (e.g., *Neocallimastix* sp.) are known to symbiotically engulf a variety of methanogenic archaea. They provide the archaea with substrate for methane production from H_2 produced in their hydrogenosomes, as well as protection against oxygen toxicity [30,55,56]. Thus, free-living methanogens might represent a low fraction of microbial population [45], and CH_4 biosynthesis might be more influenced by endosymbiotic methanogens [55]. Hence, a larger methanogenesis activity is expected to be correlated with a larger abundance of eukaryotes, especially ciliates, which are more abundant and better represented. Another partial explanation for the low abundance of free archaea, and thereby for the lack of association between Archaea and methane emissions in previous studies [10], is that lysis of archaea cell walls often requires specific protocols during DNA extraction, and they might be under-represented in metagenomics studies [57].

In terms of Gene Ontology, the KEGGs were associated with several metabolic functions and cellular processes (nutrient metabolism and biosynthesis, cellular transport, cell growth, or genetic information processing). Pathways related to pathogenic activity were also found, in agreement with the RA of several genera that include known pathogenic species (e.g., *Vibrio*, *Haemophilus*, *Trypanosoma*, or *Staphylococcus*), although not every species from these genera is pathogenic, but opportunistic or commensal organisms. Besides, pathogenic activity presence in our dataset might be biased owing to a larger representation of human-related diseases in the databases. The KEGGs were classified according to their presence or absence in ko00680 pathway (methane metabolism), as a way to evaluate their direct involvement in methanogenesis or an indirect involvement in

pathways leading to biosynthesis of precursor compounds. Although we found several ko00680 KEGGs, which are presumably involved in the biosynthesis of methanogenesis precursors, most of them were not associated with methane emissions (i.e., not differentially abundant between methane groups). Most of these KEGGs were mainly present in bacteria or eukaryotes and might be functioning in metabolic pathways not related to methanogenesis. For instance, some of the KEGGs inside the methane metabolism pathway can also be involved in glycine, serine, and threonine metabolism (e.g., K00058, K00831, K01079, and K00600), pyruvate and propanoate metabolism (e.g., K00625 and K13788), glycolysis (e.g., K01689, K15633, K01624, and K02446), or anaerobic carbon fixation (e.g., K00198) [16–18]. Another group of ko00680 KEGGs is exclusive from Archaea, but the under-representation of this clade in our dataset might obscure statistical significance.

Other detected KEGGs could be indirectly related to methanogenesis through biosynthesis of precursor compounds. For instance, K00209 and K13788 are involved in butyrate and propanoate biosynthesis, being essentially carried by primary fermentative bacteria [58]. Then the volatile fatty acids can be used by secondary fermenters to produce methanogenesis precursors such as H_2, CO_2, acetate, and formate [59,60]. In fact, K13788 is a phosphate acetyltransferase (EC:2.3.1.8) that can be involved in the biosynthesis of acetate from acetyl-CoA [61]. Also, K09251 is involved in biosynthesis of GABA and 2-oxoglutarate. GABA has been related to a volatile fatty acid concentration increment [62], while 2-oxoacid compounds can be used by Archaea to synthesize coenzyme M and coenzyme B, which are essential in methane production [63]. However, all these KEGGs were observed as overabundant in the LOW methane group, suggesting a strong presence of fermentative bacteria in these animals, not directly correlated with methane production.

Other KEGGs that were overabundant in LOW emitters might offer an explanation of the lower presence of active methanogenesis processes through competence mechanisms (e.g., LOW-OA KEGGs K01682, K01902, and K13788 are involved in citrate cycle and pyruvate metabolism, related to respiration). The K00370 and K00371 are nitrate oxidoreductase subunits playing a role in anaerobic respiration using nitrate as electron acceptor. This enzyme uses nitrate as electron acceptor, a process that has been reported as a competitive inhibitor of methanogenesis [64,65]. Nitrate supplementation has proven to be an useful strategy to mitigate methane emissions [66]. Nitrite produced by the nitrate-reductases has a known antimicrobial effect and toxicity to animal cells [67–69], which might also reduce the proportion of free archaea in LOW animals, although toxicity to archaea must be further studied [70]. However, the role of ciliates and fungi must be clarified because their abundance is also lower in LOW emitters. We hypothesize that the predatory nature of these eukaryotes might be a control mechanism for bacterial populations, and their lower relative abundance in LOW animals might allow overgrowth of related bacteria. Nevertheless, there is the possibility that a higher proportion of facultative anaerobes using nitrate as acceptor might affect ciliate populations by toxicity, thus reducing the presence of endosymbiotic methanogenic archaea.

The SqueezeMeta software [71] uses a last common ancestor (LCA) algorithm, which assigns to 1 read the lowest-level taxon common to all hits, using a stringent cut-off identity value for each taxonomic rank. On its part, functional assignments are done with the fun3 algorithm, which by default assigns the hit with the highest mean bitscore compared to the n first hits passing the e-value, identity, and coverage filters. This LCA approach ensures that reads have a large probability of being correctly classified, at the expense of a large number of reads remaining unclassified, which explains the larger number of reads assigned to a known KEGG than to taxa. Despite this strict requirement, this composition is consistent with other populations reported before [2,3,20]. Most studies to date report large abundance of Bacteroidetes and Firmicutes, with *Prevotella spp*. as the most prevalent genus. Some minor discrepancies with other studies were observed in the RA of the core subcomposition. For example, Wallace *et al.* [20] showed a higher presence of Proteobacteria and Euryarchaeota, although using amplicons instead of whole-metagenome sequencing.

Our statistical approach evidenced the difficulty of inferring a phenotypic association between microbiome composition and methane production, with an important role of environmental factors that mask the statistical signal. However, a meaningful relationship between the microbiome composition and methane emissions could be uncovered yet, emphasizing the role of the different phyla, with the Eukaryota superkingdom being of particular relevance. Previous studies also revealed a link between ruminal microbiota and methane production. Difford *et al.* [3] showed different clusters of high and low methane emitters according to their bacterial and archaeal subcomposition. Danielsson *et al.* [46] also found clustering for low and high methane emitters within prokaryotic rumen subcompositions. Wallace *et al.* [20] found that a core set of rumen microbiome was capable of explaining up to 30% of methane emissions variability, mostly formed by prokaryotes. The aforementioned studies used different methodologies, like amplicon analysis and operational taxonomic unit clustering, contrasting with our full-metagenome genus-clustering protocol, which increases the information entropy. Stewart *et al.* [72] used Nanopore sequencing and found significant differences between low and high methane emitter sheep, with clear clustering between groups, but using a lower number of microbial groups and animals in the same farm with similar management practices.

Conclusions

The full metagenome compositional analysis used in this study provided novel insights in the association between the microbiota and CH_4 emissions through differential abundance analysis, pairwise correlation, and interaction networks. Our approach evidenced a phenotypic association between microbiome composition and methane production, regardless of the challenges posed by the microbiome complexity and the compositional nature of the data. This association is mainly driven by the relative abundance of ciliates and fungi, which carry host-specific genetic functions providing substrate to the methanogenic archaea. On the other side, we detected some bacterial groups that performed a more efficient feed digestion, leaving less hydrogen available to archaea and hence associated with lower methane emissions.

This study generated the largest ruminal metagenomic dataset sequenced using ONT and grants free access to a publicly available dataset. The complexity of the rumen microbiome and the compositional nature of their sequencing data require proper statistical methods to allow disentangling the role of microbes and their genes in host complex traits such as methane emissions. Future nutritional and genetic strategies to reduce CH_4 emissions should focus on reducing the relative abundance of Alveolata and Fungi in the rumen, without impairing other important metabolic processes for an efficient feed digestion in ruminants.

Methods

Animal housing and feeding

Our cohort included 439 Holstein lactating cows sampled at 14 different herds from northern Spain (Cantabria, Euskadi, Navarra, and Girona regions). The animals received total mixed ration diet differently formulated for each individual herd, although most of them were based on maize and grass silage plus concentrate. Cows were fed *ad libitum*, with concentrate supplementation in the automatic milking station (AMS) during milking.

Methane measurement

Methane concentration was individually recorded through breath sampling during each cow visit to the AMS (3–7 times daily) in a period of 2–3 weeks. Eructation peaks were recorded using a non-dispersive infrared methane detector (Guardian NG infrared gas monitor, Edinburgh Sensors, Livingston, UK) as described by Rey *et al.* [73]. Each cow's peaks were then averaged to get a unique methane record per cow, as described by López-Paredes *et al.* [74]. Animals were distributed in groups according to number of lactation (NL) and stage of lactation (SL) criteria. Furthermore, quartile-based qualitative categories were created for CH_4 recordings (ppm), resulting in a methane factor (CH4) with 4 levels (LOW, L-MID, H-MID, and HIGH methane emissions).

Ruminal content sampling

Ruminal fluid was sampled using an oral tube (18 mm diameter and 160 mm long) connected to a 1,000-mL Erlenmeyer flask and continued to a mechanical pump (Vacubrand ME 2SI, Wertheim, Germany), with all the material contacting the cow being carefully cleaned between cows. Each animal was moved to an individual stall for this process. The solid fraction of the ruminal content was discarded by filtering through 4 layers of sterile cheesecloth, while the outcoming liquid fraction was instantly frozen using liquid nitrogen and then stored at −80°C until DNA extraction.

DNA extraction and sequencing

Genomic DNA was extracted from 250 μL of each thawed and homogenized ruminal content sample, using the "DNeasy Power Soil" commercial kit (Qiagen, Valencia, CA, USA). Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and Nanodrop ND-1000 UV/Vis spectrophotometer (Nanodrop Technologies Inc., Wilmington, DE, USA) were used to measure DNA concentration and purity. The 260/280 and 260/230 ratios were ∼1.8 and ∼2.0, respectively. Oxford Nanopore Technologies (ONT) SQK-LSK109 Ligation Sequencing kit was used for multiplexed sequencing in MinION automatic sequencer. The 1D Native barcoding ONT kit (EXP-NBD104 or EXP-NBD114) was used for multiplexing the samples, pooling barcoded DNA from 12 samples for each run. Pooling was done using a 1.5-mL DNA LoBind tube to perform adapter ligation and sequenced using a R9.4.1 flow cell.

Read processing, mapping, and filtering

Guppy toolkit (ONT) was used for basecalling. A quality control was then applied removing sequences with QS <7 and length <150 bp. Sequence analysis was performed using SqueezeMeta (SQM) pipeline for long reads [71], which performs Diamond Blastx against GenBank nr taxonomic database and against COG and KEGG functional databases, then identifying and annotating open reading frames using the LCA method for taxonomy and the fun3 algorithm for functional annotation (based on e-value and identity scores). This tool is specifically designed to process long reads from ONT.

A total of 49,718,901 reads were processed in Blastx by SQM longreads pipeline. Blastx mapped 25,750,755 reads (51.79%) to taxonomy (NCBI-nr database) or function (KEGG database). All sequences mapped as non-microbial (i.e., virus, animals, and plants) were discarded. Microbial sequences were then filtered by prevalence to reduce data sparsity and sequencing errors (Supplementary Data S2). A first estimation of sample sparsity and reads distribution was assessed using R. Two animals were then withdrawn from the filtered dataset, one owing to low read coverage and another owing to lack of host information, leaving 437 animals in the final dataset.

Genera were divided into superkingdom groups (Archaea, Bacteria, or Eukaryota), and KEGGs were sorted by their involvement in methane metabolism (MP): KEGGs included in the KEGG orthology pathway ko00680 (Methane metabolism) were labeled as "ko00680," while the rest were identified as "Other."

Compositional data

Considering the compositional nature of metagenomic data, a CLR method [75] was applied using the unweighted option of the CLR function from the easyCODA R package [76] as follows:

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

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

Being $\mathbf{x} = [x_1, x_2, \ldots, x_D]$ a vector of counted features (taxa or KEGGs) in 1 sample and $G(x)$ the geometric mean of $\mathbf{x}$. Count zero values in the initial data frame were imputed through the Geometric Bayesian Multiplicative procedure, using the zCompositions R package [77] cmultRepl function, so that logarithms could be computed.

β-diversity and PERMANOVA analysis

The CLR-transformed data (at phylum, class, order, family, genus, and KEGG levels) were used to explore β-diversity in the samples through PCA using the prcomp function in R. Fitted smooth surface of methane emissions corrected by SL and NL was included for principal components 1 and 2 using the ordisurf function from the vegan R package [78]. A generalized additive model smooth fitting (GAM) was used to elucidate non-linear distribution of samples in PCA according to methane emissions. Differences between centroid distances using methane as grouping variable (CH4) were determined through PERMANOVA [79,80] following this model and using the matrix of Aitchison distances between samples (i.e., the Euclidean distance on CLR-transformed data) as input variable:

$$D_{jkl ni} = \mu + B_j + \mathrm{SL}_k + \mathrm{NL}_l + \mathrm{CH4}_n + e_{jkl ni},$$

with B_j being the farm-batch effect ($j = 24$ levels), SL_k being the stage of lactation at the day of sampling ($k = 3$ levels), NL_l the number of lactation ($l = 2$ levels), and $\mathrm{CH4}_n$ the methane emission level ($n = 4$ levels: LOW, L-MID, H-MID, HIGH), and $e_{jkl ni}$ was the corresponding residual term.

Association between microbiota and methane production

Differential abundance of genera and KEGGs between samples regarding the different methane emissions levels was addressed through linear regression using Limma [81]. Count normalization and log-transformation were addressed using CLR-transformed data as inputs. *P*-values were adjusted by the Benjamini-Hochberg

method, to control the false discovery rate (FDR). Differential abundance threshold was set to $|\log_2 FC| \geq 0.5$ and the adjusted significance threshold was set to $\alpha = 0.05$.

Pairwise proportionality analysis

Pairwise correlations between phyla, genera, and KEGGs were calculated as described in the propr R package [82]. Proportionality coefficient ρ_p [83] under CLR data transformation was chosen. Thresholds were selected according to 2 conditions: (i) representing the maximum number of proportionalities avoiding computational issues; (ii) FDR <1%. The thresholds used were $|\rho_p| \geq 0.4$ for genera proportionalities and $|\rho_p| \geq 0.7$ for KEGG proportionalities.

Microbial networks

Microbial networks for taxonomy (at the genus level) and functionality were built from the proportionality matrices described above. Input edges were defined from the cytoscape function in the propr package in R, which converts a propr object into a data frame of node connections compatible with Cytoscape software (v. 3.8.0). Results from the DA analyses were used to associate each feature (node) to high or low methane emissions levels. Significantly overabundant genera and KEGGs in the low methane emitters group (i.e., more abundant in LOW than in HIGH or H-MID groups) were designated as LOW-associated, while those on the contrary overabundant in high methane emitters were appointed as HIGH-associated. Non-DA features were classified as N/A (not associated). In addition, SK and MP factors were included as node attributes for genera and KEGGs, respectively. For graph visualization, Kamada-Kawai algorithm (edge-weighted spring embedded layout) was set [84], using ρ_p coefficient as force parameter.

Data Availability

The raw data underlying this article are available at the ENA website and can be accessed with bioproject accession No. PRJNA789746 [85] . The filtered sequences in fastq format and their metadata are available from the *GigaScience* database [86]. Other data can be requested from the METALGEN project [87].

SqueezeMeta software is available at [88]. Guppy basecaller software was used to convert fast5 raw signals to fastq files [89]. The R environment and packages used are available from [90]. Correspondence and material requests should be addressed to O.G.R. Other data further supporting this work (including methane measurements) are openly available in the *GigaScience* repository, GigaDB [86].

Additional Files

Supplementary Figure S1. Feature counts distribution. Features with zero counts, singletons, doubletons and 3 or more counts per sample. A) Count distribution in raw taxonomy table (87% sparsity); B) Count distribution in filtered taxonomy table (68% sparsity); C) Count distribution in raw KEGG table (72% sparsity); D) Count distribution in filtered KEGG table (39% sparsity). Filtering processes removed less than 1% of total reads in both datasets.

Supplementary Figure S2. Phyla relative abundance per sample. Samples are sorted from lowest to highest RA of Bacteroidetes.

Supplementary Data S1. Differential Abundance of genera and KEGGs. Tables containing significant differentially abundant (DA) genera and KEGGs for methane emissions levels. Contrast.AvsB column represents the CH4 levels for which the feature is significantly DA. Grp.A and Grp.B columns represent the average number of reads per contrast group. For KEGGs, Association column shows if the specific KEGG participates in methanogenesis pathway (k00680) (Methane) or not (Others).

Supplementary Data S2. Additional methods. Detailed information about the taxonomy and prevalence filters applied to mapped reads.

Abbreviations

AMS: automatic milking station; BLAST: Basic Local Alignment Search Tool; bp: base pairs; CLR: centered log-ratio; DA: differential abundance; FC: fold change; FDR: false discovery rate; GAM: generalized additive model; KEGG: Kyoto Encyclopedia of Genes and Genomes; LCA: last common ancestor; NADH: nicotinamide adenine dinucleotide; NCBI: National Center for Biotechnology Information; OA: overabundance; ONT: Oxford Nanopore Technologies; PCA: principal component analysis; PERMANOVA: permutational analysis of variance; RA: relative abundance.

Ethics 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 28 March 2017.

Competing Interests

The authors state that they have no competing interests.

Funding

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). A.L.G. was funded by FPI-INIA grant with reference FPI-SGIT2016-06.

Authors' Contributions

A.L.G. and A.S.M. filtered and prepared the data, implemented the statistical analyses, and prepared the data visualization and the first draft of the manuscript. M.G.R. and C.G. performed the DNA extraction and sequencing. O.G.R. supervised the DNA sequencing and contributed to the statistical analyses. R.C. contributed to develop interaction networks. O.G.R., A.G.R., R.A., and I.G. conceived the study and designed the experiments. O.G.R., A.G.R., J.J.M., C.O. and M.S. contributed to the acquisition of funding and resources. J.T. and F.P.S. developed the computational pipelines for the metagenome and assisted on its analyses. A.L.G., A.S.M., and O.G.R. wrote the manuscript. All authors helped writing and configuring the last version of the manuscript.

Acknowledgments

We thank the regional Holstein Associations and farmers collaborating in the project. Computational support from the High-Performance Computing Centre in Galicia (CESGA, Spain) is acknowledged.

References

1. Wang, B, Yao, M, Lv, L, *et al.* The human microbiota in health and disease. *Engineering* 2017;**3**(1):71–82. doi:10.1016/J.ENG.2017.01.008
2. Delgado, B, Bach, A, Guasch, I, *et al.* Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. *Sci Rep* 2019;**9**(1): doi:10.1038/s41598-018-36673-w.
3. Difford, GF, Plichta, DR, Løvendahl, P, *et al.* Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet* 2018;**14**(10):e1007580. doi:10.1371/journal.pgen.1007580
4. Weimer, PJ. Redundancy, resilience, and host specificity of the ruminal microbiota: Implications for engineering improved ruminal fermentations. *Front Microbiol* 2015;**6**: doi:10.3389/fmicb.2015.00296.
5. Knapp, JR, Laur, GL, Vadas, PA, *et al.* Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *J Dairy Sci* 2014;**97**(6):3231–61. doi:10.3168/JDS.2013-7234
6. Gerber, PJ, Steinfeld, H, Mottet, O, *et al.* Tackling climate change through livestock. A global assessment of emissions and mitigation opportunities. In: Rome, Italy: Food and Agriculture Organization of the United Nations (FAO); http://www.fao.org/3/a-i3437e.pdf. 2013. Accessed 5 January 2022.
7. Food and Agriculture Organization of the United Nations (FAO). *Global Livestock Environmental Assessment Model (GLEAM)*. http://www.fao.org/gleam/results/en/. 2019. Accessed 4 August 2021.
8. Martínez-Álvaro, M, Auffret, MD, Stewart, RD, *et al.* Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. *Front Microbiol* 2020;**11**:doi:10.3389/fmicb.2020.00659.
9. Malmuthuge, N, Guan, LL. Understanding host-microbial interactions in rumen: Searching the best opportunity for microbiota manipulation. *J Anim Sci Biotechnol* 2017;**8**(1): doi:10.1186/s40104-016-0135-3.
10. Negussie, E, de Haas, Y, Dehareng, F, *et al.* 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* 2017;**100**(4):2433–53. doi:10.3168/jds.2016-12030
11. Gloor, GB, Macklaim, JM, Pawlowsky-Glahn, V, *et al.* Microbiome datasets are compositional: And this is not optional. *Front Microbiol* 2017;**8**: doi:10.3389/fmicb.2017.02224.
12. Liu, YX, Qin, Y, Chen, T, *et al.* A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* 2021;**12**(5):315–30. doi:10.1007/s13238-020-00724-8
13. Bonk, F, Popp, D, Harms, H, *et al.* PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. *J Microbiol Methods* 2018;**153**:139–47. doi:10.1016/J.MIMET.2018.09.015
14. Brandt, C, Bongcam-Rudloff, E, Müller, B. Abundance tracking by long-read nanopore sequencing of complex microbial communities in samples from 20 different biogas/wastewater plants. *Appl Sci* 2020;**10**(21):7518. doi:10.3390/app10217518
15. Burki, F, Shalchian-Tabrizi, K, Minge, M, *et al.* Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2007;**2**(8):e790. doi:10.1371/journal.pone.0000790
16. Kanehisa, M, Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;**28**(1):27–30. doi:10.1093/nar/28.1.27
17. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019;**28**(11):1947–51. doi:10.1002/pro.3715
18. Kanehisa, M, Sato, Y, Furumichi, M, *et al.* New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 2019;**47**(D1):D590–5. doi:10.1093/nar/gky962
19. Lovell, D, Pawlowsky-Glahn, V, Egozcue, JJ, *et al.* Proportionality: A valid alternative to correlation for relative data. *PLoS Comput Biol* 2015;**11**(3):e1004075. doi:10.1371/journal.pcbi.1004075
20. Wallace, RJ, Sasson, G, Garnsworthy, PC, *et al.* A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci Adv* 2019;**5**(7):eaav8391. doi:10.1126/sciadv.aav8391
21. Shah, HN, Collins, DM. *Prevotella*, a new genus to include *Bacteroides melaninogenicus* and related species formerly classified in the genus *Bacteroides*. *Int J Syst Bacteriol* 1990;**40**(2):205–8. doi:10.1099/00207713-40-2-205
22. Pitta, DW, Pinchak, WE, Dowd, SE, *et al.* Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets. *Microb Ecol* 2010;**59**(3): 511–22. doi:10.1007/s00248-009-9609-6
23. Lee, HJ, Jung, JY, Oh, YK, *et al.* 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* 2012;**78**(17):5983–93. doi:10.1128/AEM.00104-12
24. Lopes, LD, de Souza Lima, AO, Taketani, RG, *et al.* Exploring the sheep rumen microbiome for carbohydrate-active enzymes. *Antonie Van Leeuwenhoek* 2015;**108**(1):15–30. doi:10.1007/s10482-015-0459-6
25. Li, J, Zhong, H, Ramayo-Caldas, Y, *et al.* A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading environment. *Gigascience* 2020;**9**(6): doi:10.1093/gigascience/giaa057.
26. Seshadri, R, Leahy, SC, Attwood, GT, *et al.* Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol* 2018;**36**(4):359–67. doi:10.1038/nbt.4110
27. Crespo-Piazuelo, D, Estellé, J, Revilla, M, *et al.* Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. *Sci Rep* 2018;**8**(1): doi:10.1038/s41598-018-30932-6.
28. Han, GG, Lee, JY, Jin, GD, *et al.* Tracing of the fecal microbiota of commercial pigs at five growth stages from birth to shipment. *Sci Rep* 2018;**8**(1):6012. doi:10.1038/s41598-018-24508-7
29. Ransom-Jones, E, Jones, DL, McCarthy, AJ, *et al.* The Fibrobacteres: An important phylum of cellulose-degrading bacteria. *Microb Ecol* 2012;**63**(2):267–81. doi:10.1007/s00248-011-9998-1
30. Newbold, CJ, De la Fuente, G, Belanche, A, *et al.* The role of ciliate protozoa in the rumen. *Front Microbiol* 2015;**6**:1313. doi:10.3389/fmicb.2015.01313
31. Williams, AG, Coleman, GS. The rumen microbial ecosystem. In: PN Hobson, GS Coleman, eds. *The Rumen Protozoa*. Dordrecht: Springer; 1997: doi:10.1007/978-94-009-1453-7_3.
32. Johnson, HP. A contribution to the morphology and biology of the Stentors. *J Morphol* 1893;**8**(3):467–562. doi:10.1002/jmor.1050080303
33. Tartar, V. *The Biology of Stentor*. Oxford, London, New York, Paris: Pergamon; 1961. doi:10.1016/c2013-0-01654-4
34. Wichterman, R. *The Biology of Paramecium*. New York: Springer; 1986. doi:10.1007/978-1-4757-0372-6
35. Cavalier-Smith, T. The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial

megaclassification. *Int J Syst Evol Microbiol* 2002;**52**(1):7–76. doi:10.1099/00207713-52-1-7

36. Poulsen, M, Schwab, C, Borg Jensen, B, *et al.* Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* 2013;**4**(1): doi:10.1038/ncomms2432.
37. Santos E de, O, Thompson, F. *et al.*. The family Succinivibrionaceae. In: E Rosenberg, EF DeLong, S Lory, eds. *The Prokaryotes: Gammaproteobacteria*. Berlin, Heidelberg: Springer; 2014: doi:10.1007/978-3-642-38922-1_368.
38. Gomez-Gil, B, Thompson, CC, Matsumura, Y, *et al.* The family vibrionaceae. In: E Rosenberg, EF DeLong, S Lory, eds. *The Prokaryotes: Gammaproteobacteria*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-38922-1_225
39. Ivanova, EP, Ng, HJ, Webb, HK, *et al.* The family Pseudoalteromonadaceae. In: E Rosenberg, EF DeLong, S Lory, eds. *The Prokaryotes: Gammaproteobacteria*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-38922-1_229
40. . In: Marchandin, H, Jumas-Bilak, E. Rosenberg, E, DeLong, EF, Lory, S *et al.* The family Veillonellaceae, eds. *The Prokaryotes: Firmicutes and Tenericutes*. Berlin, Heidelberg: Springer; 2014: doi:10.1007/978-3-642-30120-9_361.
41. Shah, HN, Collins, MD. Reclassification of *Bacteroides multiacidus *(Mitsuoka, Terada, Watanabe and Uchida) in a new genus *Mitsuokella*, as *Mitsuokella multiacidus comb. nov*. *Zentralblatt fur Bakteriol Angew und Okol Microbiol Abt1 OrigC Hyg* 1982: doi:10.1016/S0721-9571(82)80006-9.
42. Kittelmann, S, Pinares-Patiño, CS, Seedorf, H, *et al.* Two different bacterial community types are linked with the Low-Methane emission trait in sheep. *PLoS One* 2014;**9**(7): e103171. doi:10.1371/JOURNAL.PONE.0103171
43. Wallace, RJ, Rooke, JA, Duthie, CA, *et al.* Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci Rep* 2014;**4**:5892. doi:10.1038/srep05892
44. Morgavi, DP, Martin, C, Jouany, JP, *et al.* Rumen protozoa and methanogenesis: Not a simple cause-effect relationship. *Br J Nutr* 2012;**107**(3):388–97. doi:10.1017/S0007114511002935
45. Tapio, I, Snelling, TJ, Strozzi, F, *et al.* The ruminal microbiome associated with methane emissions from ruminant livestock. *J Anim Sci Biotechnol* 2017;**8**(1): doi:10.1186/s40104-017-0141-0.
46. Danielsson, R, Dicksved, J, Sun, L, *et al.* Methane production in dairy cows correlates with rumen methanogenic and bacterial community structure. *Front Microbiol* 2017;**8**: doi:10.3389/fmicb.2017.00226.
47. Shi, W, Moon, CD, Leahy, SC, *et al.* Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* 2014;**24**(9):1517–25. doi:10.1101/gr.168245.113
48. Wallace, RJ, Rooke, JA, McKain, N, *et al.* The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 2015;**16**(1): doi:10.1186/s12864-015-2032-0.
49. Aguinaga Casañas, MA, Rangkasenee, N, Krattenmacher, N, *et al.* Methyl-coenzyme M reductase A as an indicator to estimate methane production from dairy cows. *J Dairy Sci* 2015;**98**(6):4074–83. doi:10.3168/jds.2015-9310
50. Janssen, PH, Kirs, M. Structure of the archaeal community of the rumen. *Appl Environ Microbiol* 2008;**74**(12):3619–25. doi:10.1128/AEM.02812-07
51. Newbold, CJ, Lassalas, B, Jouany, JP. The importance of methanogens associated with ciliate protozoa in ruminal methane production in vitro. *Lett Appl Microbiol* 1995;**21**(4):230–4. doi:10.1111/j.1472-765X.1995.tb01048.x
52. Qin, WZ, Li, CY, Kim, JK, *et al.* Effects of defaunation on fermentation characteristics and methane production by rumen microbes in vitro when incubated with starchy feed sources. *Asian-Australas J Anim Sci* 2012;**25**(10):1381–8. doi:10.5713/ajas.2012.12240
53. Schönhusen, U, Zitnan, R, Kuhla, S, *et al.* Effects of protozoa on methane production in rumen and hindgut of calves around time of weaning. *Arch Anim Nutr* 2003;**57**(4):279–95. doi:10.1080/00039420310001594423
54. Guyader, J, Eugène, M, Nozière, P, *et al.* Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. *Animal* 2014;**8**(11):1816–25. doi:10.1017/S1751731114001852
55. Belanche, A, de la Fuente, G, Newbold, CJ. Study of methanogen communities associated with different rumen protozoal populations. *FEMS Microbiol Ecol* 2014;**90**(3):663–77. doi:10.1111/1574-6941.12423
56. Hackstein, JHP, de Graaf, RM. Anaerobic ciliates and their methanogenic endosymbionts. In: JHP Hackstein, ed. *(Endo)symbiotic Methanogenic Archaea*. 2nd ed. Cham: Springer; 2018: doi:10.1007/978-3-319-98836-8_2.
57. Roopnarain, A, Mukhuba, M, Adeleke, R, *et al.* Biases during DNA extraction affect bacterial and archaeal community profile of anaerobic digestion samples. *3 Biotech* 2017;**7**(6):375. doi:10.1007/s13205-017-1009-x
58. Esquivel-Elizondo, S, Ilhan, ZE, Garcia-Peña, EI, *et al.* Insights into butyrate production in a controlled fermentation system via gene predictions. *mSystems* 2017;**2**(4): doi:10.1128/msystems.00051-17.
59. Crable, BR, Plugge, CM, McInerney, MJ, *et al.* Formate formation and formate conversion in biological fuels production. *Enzyme Res* 2011;**2011**:532536. doi:10.4061/2011/532536
60. Hao, L, Michaelsen, TY, Singleton, CM, *et al.* Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME J* 2020;**14**(4):906–18. doi:10.1038/s41396-019-0571-0
61. Brown, TDK, Jones-Mortimer, MC, Kornberg, HL. The enzymic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*. *J Gen Microbiol* 1977;**102**(2):327–36. doi:10.1099/00221287-102-2-327
62. Xie, M, Chen, HH, Nie, SP, *et al.* Gamma-aminobutyric acid increases the production of short-chain fatty acids and decreases pH values in mouse colon. *Molecules* 2017;**22**(4):653. doi:10.3390/molecules22040653
63. Graham, DE. 2-oxoacid metabolism in methanogenic CoM and CoB biosynthesis. *Methods Enzymol* 2011;**494**:301–26. doi:10.1016/B978-0-12-385112-3.00015-9
64. Klüber, HD, Conrad, R. Effects of nitrate, nitrite, NO and N2O on methanogenesis and other redox processes in anoxic rice field soil. *FEMS Microbiol Ecol* 2006;**25**(3):301–18. doi:10.1111/j.1574-6941.1998.tb00482.x
65. Yang, C, Rooke, JA, Cabeza, I, *et al.* Nitrate and inhibition of ruminal methanogenesis: Microbial ecology, obstacles, and opportunities for lowering methane emissions from ruminant livestock. *Front Microbiol* 2016;**7**: doi:10.3389/fmicb.2016.00132.
66. Lee, C, Beauchemin, KA. Une revue de l'ajout de nitrate dans l'alimentation des ruminants : Toxicité aux nitrates, émissions de méthane et performance de production. *Can J Anim Sci* 2014;**94**(4):557–70. doi:10.4141/CJAS-2014-069
67. Cui, X. The bacteriocidal effects of transition metal complexes containing the NO+ group on the food-spoilage bacterium *Clostridium sporogenes*. *FEMS Microbiol Lett* 1992;**98**(1-3):67–70. doi:10.1016/0378-1097(92)90133-9

68. Zumft, WG. The biological role of nitric oxide in bacteria. *Arch Microbiol* 1993;**160**(4):253–64. doi:10.1007/BF00292074
69. Reed, B, Dinn, J, McCann, S, *et al.* Alterations in mammalian cells induced by inactivation of vitamin B12 with nitrous oxide. In: B Zagalak, B Friedrich, eds. *Vitamin B12. Proceedings of the 3rd European Symposium on Vitamin B12 and Intrinsic Factor* Zürich: De Gruyter; 1979: doi:10.1515/9783111510828-120.
70. Klüber, HD, Conrad, R. Inhibitory effects of nitrate, nitrite, NO and N2O on methanogenesis by *Methanosarcina barkeri* and *Methanobacterium bryantii*. *FEMS Microbiol Ecol* 1998;**25**(4):331–9. doi:10.1111/j.1574-6941.1998.tb00484.x
71. Tamames, J, Puente-Sánchez, F. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front Microbiol* 2019;**9**: doi:10.3389/fmicb.2018.03349.
72. Stewart, RD, Auffret, MD, Warr, A, *et al.* Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol* 2019;**37**(8):953–61. doi:10.1038/s41587-019-0202-3
73. Rey, J, Atxaerandio, R, Ruiz, R, *et al.* Comparison between noninvasive methane measurement techniques in cattle. *Animals* 2019;**9**(8):563. doi:10.3390/ani9080563
74. López-Paredes, J, Goiri, I, Atxaerandio, R, *et al.* 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* 2020;**103**(8):7199–7209. doi:10.3168/jds.2019-17597
75. Aitchison, J. The statistical analysis of compositional data. *J R Stat Soc Ser B* 1982: doi:10.1111/j.2517-6161.1982.tb01195.x.
76. Greenacre, M. *Compositional Data Analysis in Practice*. Chapman & Hall/CRC; 2018. doi:10.1201/9780429455537
77. Palarea-Albaladejo, J, Martín-Fernández, JA. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. *Chemometr Intell Lab Syst* 2015;**143**:85–96.doi:10.1016/j.chemolab.2015.02.019
78. Oksanen, J, Blanchet, FG, Friendly, M, *et al. vegan: Community Ecology Package. R package version* 2.5-6. https://CRAN.R-project.org/package=vegan. 2019. Accessed 5 January 2022
79. Anderson, MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;**26**(1):32–46. doi:10.1111/j.1442-9993.2001.01070.pp.x
80. McArdle, BH, Anderson, MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* 2001;**82**(1):290–7. doi:10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2
81. Ritchie, ME, Phipson, B, Wu, D, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**(7):e47. doi:10.1093/nar/gkv007
82. Quinn, TP, Richardson, MF, Lovell, D, *et al.* Propr: An R-package for identifying proportionally abundant features using compositional data analysis. *Sci Rep* 2017;**7**(1): doi:10.1038/s41598-017-16520-0.
83. Erb, I, Notredame, C. How should we measure proportionality on relative gene expression data? *Theor Biosci* 2016;**135**(1-2): 21–36. doi:10.1007/s12064-015-0220-8
84. Kamada, T, Kawai, S. An algorithm for drawing general undirected graphs. *Inf Process Lett* 1989;**31**(1):7–15. doi:10.1016/0020-0190(89)90102-6
85. *ENA project for metagenome data*. https://www.ebi.ac.uk/ena/browser/view/PRJNA789746. Accessed 5 January 2022.
86. López-García, A, Saborío-Montero, A, Gutiérrez-Rivas, M, *et al.* Supporting data for "Fungal and ciliate protozoa are the main rumen microbes associated with methane emissions in dairy cattle." *GigaScience Database* 2021. http://dx.doi.org/10.5524/100950.
87. *METALGEN project*. https://www.metalgen.es. Accessed 5 January 2022.
88. *SqueezeMeta software*. https://github.com/jtamames/SqueezeMeta. Accessed 5 January 2022.
89. *Guppy Basecaller software*. https://community.nanoporetech.com/downloads. Accessed 5 January 2022.
90. *CRAN package repository*. https://cran.r-project.org/. Accessed 5 January 2022.

Chapter 6: General discussion

6.1. Recapitulation

In this thesis, the work undertaken was focused on testing different methodological strategies to evaluate the relationships of microbiota with interesting phenotypic traits of livestock species. An amplicon sequencing approach has been used for characterizing the faecal microbiota of pigs from different breeds (Duroc and Iberian) fed diets differing in oleic acid content. The goal of this experiment was to explore the effects of genetic background and monounsaturated fatty acid supplementation on gut microbiota, in the context of the Iberian pig rearing systems, which involves crossbreeding with Duroc and diet modulation to imitate *montanera*. Preliminarily, an additional experiment to select the optimal bioinformatic procedure for post-processing of amplicon sequences was done, testing the most used amplicon reference databases (GreenGenes and SILVA) and pipelines (QIIME and Mothur). On the other hand, full metagenome sequencing using ONT long reads has been used to characterize the rumen microbiota of a large population of Spanish Holstein dairy cows. This third experiment is an exploratory approach aiming at the identification of the rumen microbes whose abundance and/or metabolic activity are highly associated with methane production.

Furthermore, the compositional nature of microbiota data has been taken into account for downstream statistical analysis. This approach is rarely applied in microbiome analysis, which makes this research innovative and offers an alternative to classical statistical approaches that might lead to misinterpretation of the results.

6.2. Comparing pipelines for amplicon sequencing analysis

The first study (Chapter 3) compares popular AS post-processing pipelines and reference databases to test their clustering and classification performance. The current comparison was useful to determine the major factor conditioning AS post-processing. Even though it was observed that the database is considerably important for getting a correct microbial identification, with SILVA notably overcoming GreenGenes, the pipeline chosen was less decisive. Although Mothur seemed able to detect rarer OTUs in higher depth than QIIME, differences were less pronounced than those attributed to the reference database. Moreover, the lack of an available mock community for RM composition impeded to assess which composition is closer to the reality.

Current knowledge points to significant differences between pipelines, although no consensus has been reached about which one performs better. Background information is scarce, and generally reports the existence of differences between pipelines, though these differences do not seem critical. Plummer *et al.* analysed human GM data, comparing MG-RAST, Mothur and QIIME old version and using SILVA

database. They report similar and comparable results between pipelines, detecting similar abundances for frequent phyla, although more differences were found at genus level. QIIME and Mothur resulted more powerful than MG-RAST **(Plummer *et al.*, 2015)**. Allali *et al.* compared different sequencing platforms combined with either several QIIME approaches (both *de novo* and open-reference OTU-picking), UPARSE or DADA2, in *Salmonella*-infected chicken. They concluded that both sequencing platform and bioinformatic pipeline are important sources of variation, though platform is a more important factor **(Allali *et al.*, 2017)**. However, the utilization of GreenGenes database might be a major issue, since, as we observed, differences are more substantial when this database is used, in contrast to SILVA.

On the other hand, the recent paradigm shift from OTU to ASV clustering has resulted in a reconversion of these pipelines, with the implementation of QIIME2 and the updating of Mothur, among others. Although, as formerly said, OTU approach is still valid, current research trends to ASV analysis. Following this trend, recent comparative research has been done, mostly for human microbiota. Prodan *et al.* used HMP data for comparing six different pipelines, three using OTUs (QIIME, Mothur and UPARSE) and three using ASVs (DADA2, QIIME2-Deblur and UNOISE3), including also a mock community. This work reported QIIME to produce spurious OTUs, however, QIIME was run using GreenGenes while the other OTU-picking tools used SILVA, thus the bias from different databases was not taken into account **(Prodan *et al.*, 2020)**. Another study was carried out by Marizzoni *et al.*, comparing two ASV pipelines (QIIME2 and Bioconductor) and two OTU pipelines (Mothur and UPARSE), where they reported significant differences in read assignment and taxa richness and relative abundance, with a higher diversity in ASV pipelines **(Marizzoni *et al.*, 2020)**. This conclusion is interesting, since it might be caused either by a wrong OTU assignment of distinct genera or by an overestimation due to ASV false positives. Based on these findings, new comparisons for both ruminal and faecal microbiota are necessary to inquire if new ASV-generating pipelines are equally valid, for which building the corresponding mock communities would be mandatory.

6.3. Association of pig gut microbiota composition with host genotype and diet

The second study (Chapter 4) focuses on the differences in faecal microbiota composition between Duroc and Iberian pigs, fed diets differing in energy source (carbohydrates vs oleic acid). This work was done under an AS approach, using V3-V4 regions of 16S rRNA gene.

As mentioned, the first study covered in this thesis was used as a guide to select the best procedure for sequence processing in future AS experiments. In agreement with these previous results, SILVA

reference database was the most logical choice, whereas the pipeline selection was less clear. While Mothur possesses high plasticity and a good quality control protocol, QIIME has a simpler implementation and performs better regarding time and memory consumption. On the other hand, according to previous background and the reported results, we found no reasonable justification for rejecting QIIME as a valid pipeline. However, closed-reference OTU-picking has been repeatedly stated as the less optimal solution due to its lower assignment rate – as it entirely depends on the reference database completeness – and its higher identity bias **(Bhat *et al.*, 2019)**. For these reasons, the final choice was QIIME pipeline with *de novo* OTU-picking.

The first relevant conclusion withdrawn from this experiment is that genotypic background has an important influence on GM composition. Further analyses have revealed that Iberian and Duroc pigs seem largely differentiated in overall proportion of primary fermenters, with a high weight from *Prevotella* OTUs, which represent an overall RA of 30% in Duroc against a 21% in Iberian pigs. From a compositional perspective, other genera log-ratios differing between Duroc and Iberian pigs have been found, such as the ratios *Bifidobacterium*/*Ruminiclostridium* (higher in Duroc pigs) or *Anaerostipes/Bifidobacterium* (lower in Duroc pigs). Hence, breed differences might be attributed to the combination of two factors: a different genetic background, and a different transition rate from pre-weaning to adult microbial composition. Both the overall OTU alpha-diversity and the relative abundance of core genera, such as *Prevotella*, might be related to the animal metabolic characteristics and/or resilience **(Elmqvist *et al.*, 2003)**, either as a cause or as a consequence. Accordingly, Iberian pigs possess higher OTU richness, less *Prevotella* and more abundance of alternative fibre fermenters such as *Treponema*, *Alloprevotella* and *Ruminiclostridium* **(Downes *et al.*, 2013; Norris *et al.*, 2015; Wu & Cheng, 2015)**. However, it is remarkable that the abundance of total fibre fermenters is lower in Iberian pigs, which could lead to a lower digestibility. On the other hand, *Escherichia-Shigella* and *Bifidobacterium*, usually appearing on early stages of colonization **(Mach *et al.*, 2015; Fouhse *et al.*, 2016)**, are practically absent in Iberian pigs, which might reflect a faster microbial transition.

The different diets supplied had a smaller impact on microbiota composition than the breed, which reveals that these specific diet modifications are not strong enough to trigger relevant shifts in microbiota composition. Since diet is understood as a crucial factor for gut microbiomes, the smaller effect on microbial composition compared to genetic background might result unexpected. In addition, it must be taken into account that dietary oleic acid increase required a fibre adjustment to maintain a similar energy content than control diet. For this reason, while high-oleic/high-fibre (HO/HF) diet did not have a significant impact on total species richness and evenness, it did for the abundance of fibre fermenters. Indeed, though the total abundance of fibre fermenters was not stimulated by the dietary

fibre increase, the abundance of some specific clades was different between animals fed control or HO/HF diets. For instance, overall *Prevotella* abundance tended to be lower in animals fed HO/HF diet, although several *Prevotella* OTUs resulted significantly over-abundant in HO/HF group. Despite sounding contradictory, this fact can be related to the expansion of other competitors performing the same role, such as *Treponema* or *Ruminococcus*, which are present in lower abundance than *Prevotella* but seem to proliferate in HO/HF animals. This involves in turn a potential positive effect on alpha-diversity, as Shannon and Simpson indices showed higher average values in HO/HF diet, though differences were not significant. Interestingly, the additional supply of oleic acid and fibre also seem to have an effect on the abundances of early colonizers. As stated before, *Escherichia-Shigella* and *Bifidobacterium* are only present in Duroc pigs, but their distribution is not even, since they mainly appear in HO/HF pigs. This breed-diet interaction effect for *Bifidobacterium* abundance could be related to the occupation level of the different ecological niches by other specialized microbes. Duroc animals possess a higher abundance of fibre fermenters than Iberians, while the HO/HF diet would make more oligosaccharides available, thereby allowing bifidobacteria to survive in late life stages, as they can utilize simple sugars as well **(Pokusaeva *et al.*, 2011)**. Meanwhile, Iberian pigs possess higher abundance of oligosaccharide fermenters such as *Anaerostipes* **(Bui *et al.*, 2019)**, which fill this niche and replace the bifidobacteria. As for *Escherichia-Shigella*, they are opportunistic facultative anaerobes which establish in pre-weaning stages, triggering pathogenic activity if early dysbiosis occurs **(Mach *et al.*, 2015)**. Its increased abundance in Duroc animals fed HO/HF remains unclear, and further research is needed to clarify it.

Other breed-diet interactions were found as well, mostly for OTUs classified as *Corynebacterium*, *Lactobacillus*, *Treponema* and several *Ruminococcaceae* and *Muribaculaceae* genera. Regarding lactobacilli, former works have reported somewhat contradictory results. Some studies reported a higher abundance of lactobacilli in lean pigs fed diets rich in soluble sugars: Petry *et al*. observed that progeny of Camborough gilts fed xylanase supplementation, which release arabinoxylane-oligosaccharides, had more abundance of *Lactobacillus* **(Petry *et al.*, 2021)**; Pan *et al*. observed the same in Duroc x Landrace x Yorkshire pigs supplemented with xylo-oligosaccharides **(Pan *et al.*, 2019)**; also, Tian *et al*. observed that Duroc x Landrace x Yorkshire pigs fed fibre-free diets had more *Lactobacillus* than those fed diets containing different types of fibre **(Tian *et al.*, 2017)**. However, other studies reported that lean or fatty pigs (Ossabaw or German Landrace x Piétrain) fed low-fat and high-fibre diets have shown higher abundance of *Lactobacillus* **(Pedersen *et al.*, 2013; Heinritz *et al.*, 2016)**. In the present experiment, several *Lactobacillus* OTUs are more abundant in HO/HF group, but only in the fatty Iberian pigs, while its abundance is higher – with a more marked difference – in

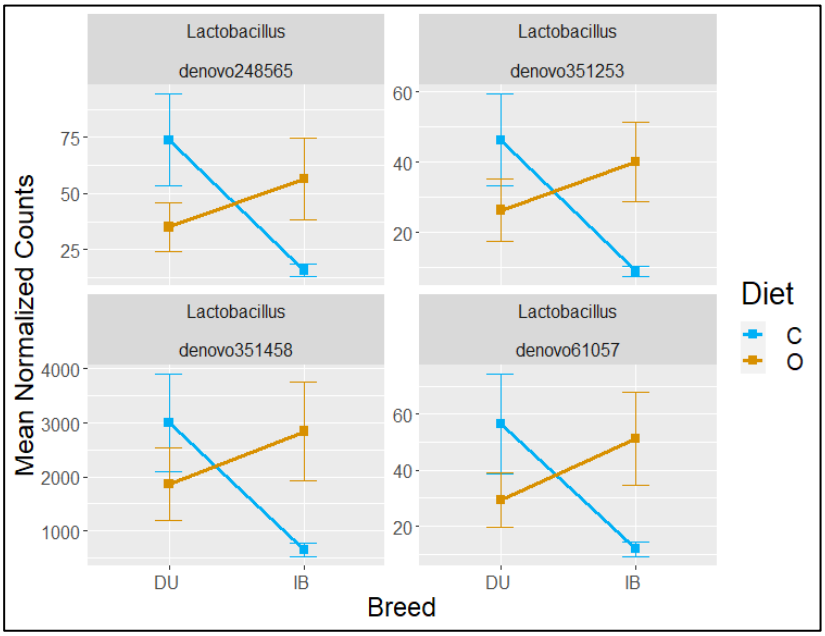

Figure 15: *Interaction plots for Lactobacillus OTUs with significant interaction between breed and diet for DESeq2 DA analysis. Y-axis: normalized counts with standard error of the mean (SEM).*

control-fed Duroc pigs, which are leaner (**Figure 15**). These results are difficult to interpret, but agree with most of the previous findings, and may be explained by complex genotype x diet interactions, which occur both in these previous studies and in the current experiment. For a bigger challenge, different *Lactobacillus* strains are more adapted to diets rich in specific sugars, and might be associated either to obesity or to weight gain protection **(Drissi *et al.*, 2014)**. Since our experiment cannot discern to strain level, the specific metabolism of the *Lactobacillus* OTUs found cannot be fully accounted. As for *Corynebacterium*, HO/HF diet increased their relative abundance in Duroc pigs, but decreased or kept it at minimum in Iberian pigs, opposite to the interaction effect for lactobacilli (**Figure 16**). Several studies have stated a negative correlation between the abundance of *Lactobacillus* and *Corynebacterium* **(Zhang *et al.*, 2019; Duarte & Kim, 2021)**. These qualitative interactions require additional research, as both genera are important for volatile fatty acid or amino acid bioavailability, which could explain part of the differences in protein and fat deposition **(Jiao *et al.*, 2020)**, as well as in general metabolism, between breeds.

In summary, we observed multiple abundance shifts between microbial pairs that could provide additional pairwise ratios useful to discriminate between breeds or

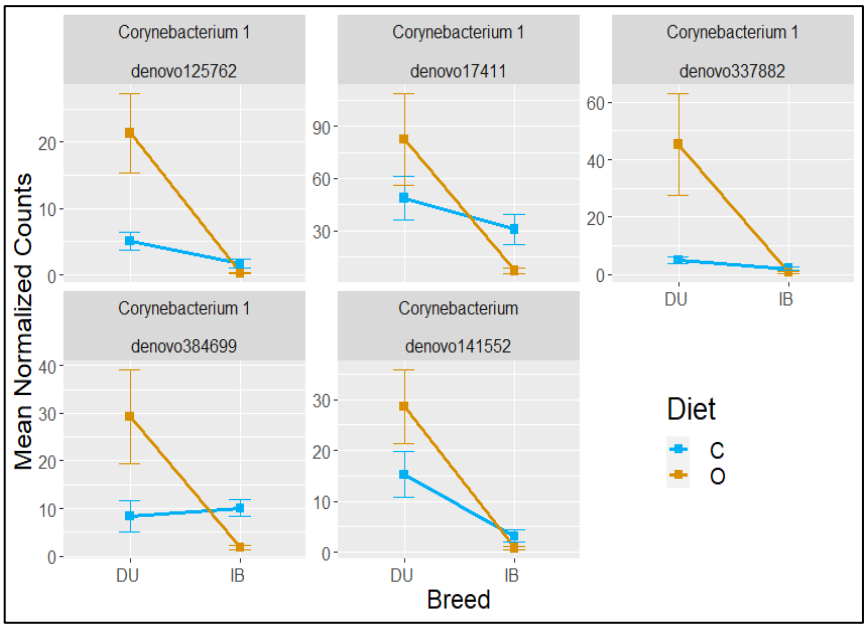

Figure 16: *Interaction plots for Corynebacterium OTUs with significant interaction between breed and diet for DESeq2 DA analysis.*

between diets with different fibre content. These new pairwise ratios should be considered together with those reported in Chapter 4 for further experiments, as these OTUs and/or genera could potentially explain most of the microbiota-associated phenotypic variability in Iberian and Duroc pigs.

Regarding the association between pig productive traits and GM composition, a previous study carried out by Benítez *et al.* collects the differences in performance and fat composition in the same Iberian and Duroc population **(Benítez *et al.*, 2018)**. They reported a higher average feed intake in Iberian pigs compared to Duroc, which might result in higher differences in the relative intake of sugar/fibre between diet groups within Iberian pigs, compared to Duroc group, thus explaining differential responses to diet and interactions in differential abundance of fibre fermenters. Also, previously reported variations in FA composition due to dietary changes are in consonance with abundance variations for some bacteria, though only within the Iberian group. For instance, several OTUs classified as fibre fermenters (*Prevotellaceae*, *Ruminococcaceae*, *Lachnospiraceae*) are positively correlated to oleic acid content and negatively correlated to SFA content in ham fat or backfat, which is expected, since high-oleic diet also contains high fibre. On the other hand, *Lactobacillus* appear as negatively correlated to backfat SFA content, the latter reported as higher in control animals than in HO/HF animals. Some strains of *Lactobacillus* have been reported as potential probiotics for reducing obesity **(Crovesy *et al.*, 2017)**, so their potential effect reducing SFA deposition might be expected. Although the exact mechanism is poorly understood, a higher concentration of VFA has been related to a decrease in feed intake via regulation of GH, leptin, PYY and other related hormones **(Wang *et al.*, 2017; Jiao *et al.*, 2020)**. The present study suggests a possible association between the abundance of *Lactobacillus* and FA composition; however, the complexity of the experimental design and the limited statistical power of the current dataset make it difficult to separate the effect of breed, dietary fibre and the influence of other microbes from the lactobacilli effect.

Moreover, another recent publication deepens in the differences in muscular transcriptome **(Benítez *et al.*, 2021)**, and reports the differential expression of genes involved in muscular development and growth (*FOS*, *MSTN*, *MYBL2* or *MYOD1*) between animals of different age and genotype. These results, obtained from the same animals employed in this thesis, support the hypothesis of a higher precocity of the Iberian pigs. Interestingly, these findings are consistent with the faster transition rate of gut microbiota composition, from pre-weaning to adult, observed in the Iberian population. Since gut microbiota provides nutrients to the host, it is reasonable to expect a direct association between both phenomena, and therefore additional studies should be carried out.

This study sheds light on the composition variations in gut microbiota between different genetic backgrounds based on differential abundance. Other authors have formerly found associations between host genetic variants and the relative abundance of some GM genera **(Crespo-Piazuelo *et al.*, 2019)** in Iberian crossbred pigs. It might be interesting to follow this approach with current data, trying to relate described SNPs to differentially abundant genera, in order to validate that microbial variation is linked to structural genetic variation.

Even though this experiment reports interesting outcomes related to Iberian pig's gut microbiota, inferences about the nutritional relationships between bacteria, as well as about host nutrient intake, are made based exclusively on microbial abundances. Metabolomics analyses would be useful to reveal the true extent of microbial fermentative activity on the host metabolism. Moreover, using faecal microbiome as a proxy of fore and mid-gut microbiomes has proven useful – as well as cheaper and faster – to explore overall differences in gut microbiota composition, but specific changes related to dietary metabolites should be directly accounted in both proximal and distal gut regions. Finally, this thesis underscores the importance of encompassing the full microbial community in order to consider all the possible ecological interactions. Therefore, WMS approaches might be useful to detect protozoa or fungi, as well as unknown biodiversity.

6.4. Association of cow rumen microbiota composition with methane production

In the third experiment (Chapter 5), whole metagenome sequencing was used for characterizing rumen microbiota composition in Holstein dairy cows with different methane production. Metagenomic data were generated from ONT long reads with a mapping-based processing approach, and both taxonomy and functionality were considered.

The most important novelty in this study was the utilization of ONT-based metagenomics, instead of using AS **(Bayat *et al.*, 2018; Difford *et al.*, 2018; Bach *et al.*, 2019; Ramayo-Caldas *et al.*, 2020)** or Illumina short-read WMS **(Martínez-Álvaro *et al.*, 2020)**, which are more extended for rumen microbiome characterization. Long-read technologies are relatively novel and in constant improvement, therefore rumen microbiota studies implementing them are still scarce. Some important works have focused on assembling complete microbial genomes, using short or long reads, in order to characterize the taxonomic and functional structure of rumen microbiome, even leading to the generation of curated and rumen-specific microbial genomes and gene catalogues **(Seshadri *et al.*, 2018; Stewart *et al.*, 2019; Watson, 2021; Xie *et al.*, 2021)**. However, these studies do not take into account the association between rumen microbiome and methane production, and therefore the

experiment covered in this thesis entails one of the first attempts to apply long reads technology for finding microbial biomarkers linked to methane emissions.

The combined analysis of all the microbiota life domains has allowed to establish associations among different microorganisms without amplification bias. In this sense, a negative correlation between protozoa and some bacterial groups has been detected, with more presence of the former in high emissions animals. Though this is not a novel result, since the importance of ciliate protozoa has been repeatedly reported in both comparative and defaunation studies **(Newbold *et al.*, 2015; Tapio *et al.*, 2017)**, the combination of WMS techniques and compositional treatment of the data has reinforced these findings, with the addition of a possible interaction between ciliates and some bacteria not caused by false negative correlations. On the other hand, no big differences were detected according to the overall abundance of fibre fermenters, however, low methane emitters had higher abundance of soluble sugar fermenters, mostly succinate or propionate producers (*Anaerobiospirillum*, *Megasphaera*, *Hespellia* or *Mitsuokella*) **(Malnick, 2015; Marchandin *et al.*, 2015; Whitehead *et al.*, 2015; Willems & Collins, 2015)**. Propionate production competes with methane production, since metabolic pathways driving to propionate production require a net intake of H_2 **(Janssen, 2010)**, thus, these bacteria might be acting as competitors of methanogenic archaea. *Mitsuokella* have also been described as hydrogenotrophic fumarate reducers, directly competing with methanogens for H_2 **(Mamuad *et al.*, 2014)**. Drops in the rumen pH have been as well related to a reduction of the activity of methanogens, in favour of a higher propionate formation **(Janssen, 2010)**. *Anaerobiospirillum* fermentative activity is known to shift depending on the pH conditions, with an increase in succinate production when pH drops **(Samuelov *et al.*, 1991)**. On the other hand, the higher abundance of *Megasphaera* could be related to a drop in pH, since they can degrade lactate, which is more abundant under low-pH conditions **(Chen *et al.*, 2019)**. However, the involvement of rumen pH conditions in the composition of RM cannot be verified with current data, since pH measures were not taken in this study.

Another interesting outcome is the negative association between *Oribacterium* and several ciliate and fungi genera, reaching rho index values up to -0.65. *Oribacterium* have been described as strictly anaerobic low-fermentative bacteria, which utilize simple sugars to produce lactate and acetate **(Sizova *et al.*, 2014)**. Its relationship with ciliates (*Tetrahymena*, *Stentor*, *Entodinium* or *Paramecium*) has not been described elsewhere, though it can be speculated that it might be related to the predatory nature of ciliates. However, the negative association *Oribacterium*/*Neocallimastix* is less explainable, as these fungi are cellulose degraders, thus facilitators of bacterial fermentative activity **(Cox *et al.*, 2021)**.

This approach also allowed to generate a functional profile of rumen microbiome, which can help to identify the metabolic paths more expressed in high methane producers. It is clear that gene abundance data must be carefully interpreted, considering that gene abundance is not equal to gene expression, and that the vast majority of genes are redundant in multiple organisms, carrying out the same or very similar functions, as observed. Nonetheless, some interesting findings have been reported, taking into account the taxonomic distribution of each mapped gene. A higher abundance of genes involved in the biosynthesis of methanogenesis precursors is reported in low methane emitters, thus suggesting that an increase in secondary fermentative bacteria in the rumen is not directly correlated to methane production. The most relevant finding has to do with the over-abundance of nitrate-oxidoreductase genes in low emissions animals, since this enzyme participates in nitrate-mediated anaerobic respiration, a competitive inhibitor of methanogenesis **(Klüber & Conrad, 2006; Yang *et al.*, 2016a)** and the preferred respiration mechanism by bacteria in absence of oxygen **(Gennis & Stewart, 1996)**. These KEGGs were present mostly in *Proteobacteria*, as shown in **Figure 17**.

Although a higher nitrate reduction metabolism would explain a lower presence of free *Archaea* in low methane emitters, further studies targeting this pathway would be useful to determine the potential involvement of ciliates and fungi in this phenomenon.

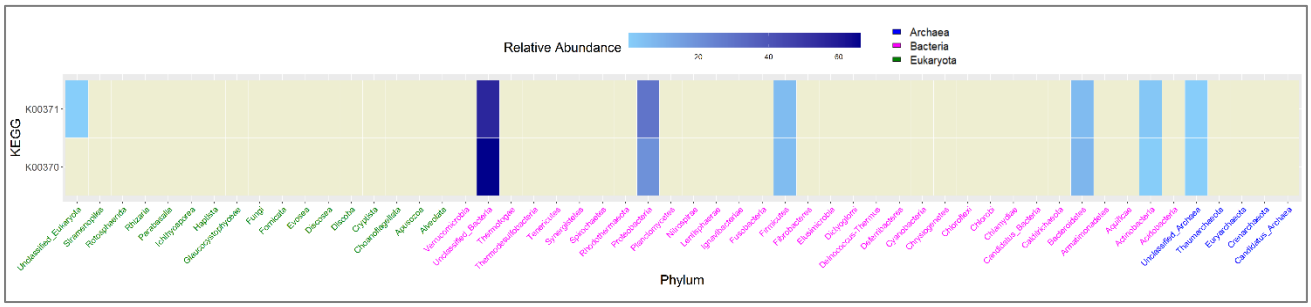

Figure 17: *Heatmap showing the taxonomic distribution of nitrate-oxidoreductase KEGGs.*

This study carries some limitations as well. As reported, standard metagenomics DNA extraction protocols are not properly prepared for lysing archaeal cell walls, and thereby the abundance of free archaea might be underestimated in most of the metagenome studies up to date **(Roopnarain *et al.*, 2017)**. For fully accounting their true diversity, ONT protocols must be re-adapted to this circumstance. On the other hand, the chosen mapping-based sequence analysis could pose additional bias. Although the assembly of ONT reads is reliable precisely because of the larger read length, pipelines must be carefully adjusted to reach an adequate read mapping percentage. Direct read mapping skips the assembly issues while keeping a reliable classification, since SqueezeMeta algorithms work with very strict thresholds **(Tamames & Puente-Sánchez, 2019)**, but this approach

could be underestimating the true diversity of the microbiome, as it is fully dependent on the reference database **(Quince *et al.*, 2017)**. Future studies must take into account these issues, since knowing the real implication of free methanogens in methane production is crucial for addressing the contribution of cattle to global warming.

6.5. Compositional approach to microbiota analysis

The compositional treatment of these data has been useful to extract conclusions while minimizing the risk of taking spurious correlations or differential abundance false positives as true differences between groups. Preliminary comparatives between non-compositional and CLR approaches applied to this thesis' data, which have not been included, revealed that subcompositional incoherence is a crucial issue which can severely affect the distribution of the samples in multivariate techniques, and also that differential abundance can be influenced by false positive rate when accounting for minor differences or rare features, though the strongest differences can be detected by both methods. For this reason, the general conclusions reported in this thesis are usually in consonance with previous bibliography, whether they used compositional or classical approaches.

On the other hand, current methodology for compositional analysis still carries some limitations. First, the treatment of zeros needs to be polished, as current methods cannot correctly account for microbial absence, though this is a problem coming from sequencing. Second, compositional transformations perform worse than quantitative transformations – i.e., those accounting sample microbial loads – **(Lloréns-Rico *et al.*, 2021)**. However, cell counting is difficult to implement and could introduce additional bias. Third, while CLR somehow mitigates subcompositional incoherence, it cannot be completely defined as subcompositionally coherent, since removing one feature changes the geometric means **(Greenacre *et al.*, 2021a)**. This could be fixed by using ALR, but this approach is conditioned to find an optimal reference feature **(Greenacre *et al.*, 2021b)**. And fourth, the available tools for differential abundance analysis possess different limitations. ANOVA-like tools (ANCOM and ALDEx2) are designed to account for compositionality – and even zero origin in the case of ANCOM – but they are computationally expensive when implementing complex models (i.e., multifactorial or with covariates), and either have low power to detect differences (ALDEx2) **(Nearing *et al.*, 2022)** or do not show p-values and confidence intervals for pairwise comparisons (ANCOM) **(Lin & Peddada, 2020)**. By contrast, RNAseq-like tools (MetagenomeSeq, DESeq2 and Limma) are quite plastic and allow fast processing of complex models, but only raw sequences can be used as input, while their normalization methods do not account for compositionality **(Gloor *et al.*, 2017)**. The exception to this is Limma, which offers the option of skipping normalization and using a transformed matrix as input,

but it was designed for microarray analysis and hence it assumes a normal distribution of the data **(Ritchie *et al.*, 2015)**. It is therefore mandatory to create a standard procedure to analyse microbiome data, which considers their true nature but also fixes the weaknesses of the used compositional method, added to the optimization of tools for correctly detecting differences in microbial abundances.

6.6. Final thoughts

The vast diversity and complexity of gut microbiomes might be the main reason for several findings reported in this thesis. Highly diverse ecosystems usually possess high functional redundancy, either due to the existence of multiple species competing for the same ecological niche **(Weimer, 2015)**, or due to a high conservation degree of some metabolic functions in different microbes, even if their products act as precursors in different metabolic pathways. For this reason, although microbial species might vary among individuals, the abundance of genes related to their representative metabolic pathways might remain stable. This phenomenon has been observed in several microbial and functional balances within this thesis' experiments. The balance between *Prevotella* and *Treponema* in faecal microbiota of pigs indicates a competitive relationship between both genera for fibre degradation. Indeed, this relationship has been previously described **(Ramayo-Caldas *et al.*, 2016; Crespo-Piazuelo *et al.*, 2018)**. As for current results, this contraposition has been linked to genetic background as well, since Duroc pigs lack *Treponema* OTUs. Similar competitions might occur between *Prevotella* and *Ruminococcus* or *Alloprevotella*, as suggested by their differences in relative abundance. Further studies must be carried out in order to unveil the relationships among fibre degraders in Duroc and Iberian pigs. On the other hand, *Lactobacillus* and *Corynebacterium*, which have been reported as negatively correlated in Duroc faecal microbiota, possess lactic fermentative activity under anaerobic conditions **(Hammes & Vogel, 1995; Inui *et al.*, 2004)**, and so lactate metabolism genes might not be differentially represented. Another example can be seen for serine and glycine biosynthesis genes (*serA*, *serB*, *serC* and *glyA*) in rumen microbiome. These genes are present in archaea, having a role in hydrogenotrophic methanogenesis through the conversion of serine to glycine and 5,10-THMPT, the latter being an intermediate in the formation of coenzyme-M, the main carbon transporter in methanogenesis **(Wolfe, 1993)**. Since genes belonging to the same orthologous groups are also present in bacteria and eukaryotes, methane production might not appear as associated with the genes involved in the corresponding metabolic pathway, even if taxonomy composition does. It similarly happens for genes involved in VFA metabolism or glycolysis, among others.

This thesis addresses a comprehensive exploration of gut microbiome, including different sequencing strategies, sequence processing and classification pipelines, and statistical analysis approaches, applied

to different goals of interest in two relevant livestock species. This broad study has tried to show the importance of gut microbiome on host metabolism and development, successfully identifying groups of microbes as targets to tackle current problems in the industry of animal husbandry. Future research lines should focus on standardizing metagenome sequencing and analysis protocols, in order to encompass as much biodiversity as possible and to identify the core microbial subpopulations with the greatest influence on host phenotypic traits. Also, implementing multi-omics approaches would help to reach a better understanding of all the mechanisms controlling a specific phenotypic trait. In this sense, it would be interesting to deepen in the genetic control of the host on the microbiome, either implementing the genomic information in the statistic models or searching relationships between host's genomic structural variants and microbial abundances. Metatranscriptomics and meta-metabolomics approaches would as well result helpful to reach a more complete overview of microbiome functionality, both through the expression of genes or through the presence of metabolites derived from their activity.

Chapter 7: Conclusions

1. The comparison of different pipelines for amplicon sequence processing indicated that the reference database is the main factor affecting the identification of OTUs, with SILVA outperforming GreenGenes. Although Mothur was able to detect a greater number of rare OTUs than QIIME, the absence of a mock community makes these differences insufficient for choosing one tool over the other.
2. The gut microbiota profile is different between Duroc and Iberian pig breeds. The higher rusticity of Iberian pigs might be associated with the higher OTU richness in their GM, which results in a higher response diversity.
3. Diet enrichment with high-oleic sunflower oil and addition of fibre has a limited effect on pig gut microbiome, although it has an effect on populations of primary fermenters, reflected in an increase of low-abundant fibre degraders replacing some *Prevotella* OTUs. The response of GM to diet is conditional on breed, showing relevant differences in Duroc and Iberians.
4. Several bacteria have been identified as potential markers discriminating either between Duroc and Iberian breeds or between diets with different fat and fibre supply. These microbes are mainly complex carbohydrate fermenters (*Prevotella*, *Alloprevotella*, *Treponema*, *Ruminococcus* or *Ruminiclostridium*) or soluble sugar fermenters (*Lactobacillus*, *Corynebacterium* or *Anaerostipes*).
5. The metabolic characteristics of the Iberian pigs, such as their higher precocity and feed intake, and the fatty acid composition of their intramuscular fat, agree with the observed gut microbiota composition differences between breeds, as well as with the differential responses to diet.
6. Ciliates and fungi are strongly associated with the production of methane in rumen, as substrate providers for methanogenic archaea. Conversely, *Proteobacteria* and *Firmicutes* genera producers of succinate and propionate are more abundant in low emissions animals. Two sharply differentiated ruminotypes have been described, which are associated with methane emissions: high eukaryote abundance and high bacterial abundance.
7. No association between the abundance of archaea and methane emissions has been observed, possibly due to their low abundance or their under-representation in the current study.

8. The abundance of genes involved in methanogenesis KEGG pathway is not associated with methane emissions, due to the gene redundancy phenomenon. However, KEGGs corresponding to nitrate-oxidoreductase subunits are over-abundant in low emissions animals, reflecting an increase of nitrate-mediated anaerobic respiration and the inhibition of methanogenesis.
9. The ecological structure of microbiomes involves numerous and complex interactions between microorganisms, making it difficult to establish causal relationships between host phenotypic traits and the presence or abundance of specific microbial species.
10. Ecological redundancy is a key phenomenon in microbiomes which contributes to maintain functional balance among alternative microbial compositions.

References

Abbas, W., Howard, J. T., Paz, H. A., Hales, K. E., Wells, J. E., Kuehn, L. A., et al. (2020). Influence of host genetics in shaping the rumen bacterial community in beef cattle. *Sci. Rep.* 10, 1–14. doi:10.1038/s41598-020-72011-9.

Albanese, D., Fontana, P., De Filippo, C., Cavalieri, D., and Donati, C. (2015). MICCA: A complete and accurate software for taxonomic profiling of metagenomic data. *Sci. Rep.* 5, 1–7. doi:10.1038/srep09743.

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi:10.1093/nar/gkz935.

Ali, M. M., Li, F., Zhang, Z., Zhang, K., Kang, D. K., Ankrum, J. A., et al. (2014). Rolling circle amplification: A versatile tool for chemical biology, materials science and medicine. *Chem. Soc. Rev.* 43, 3324–3341. doi:10.1039/c3cs60439j.

Allali, I., Arnold, J. W., Roach, J., Cadenas, M. B., Butz, N., Hassan, H. M., et al. (2017). A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiol.* 17, 1–16. doi:10.1186/s12866-017-1101-8.

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., et al. (2019). A new genomic blueprint of the human gut microbiota. *Nature* 568, 499–504. doi:10.1038/s41586-019-0965-1.

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., et al. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* 39, 105–114. doi:10.1038/s41587-020-0603-3.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Alves, E., Óvilo, C., Rodríguez, M. C., and Silió, L. (2003). Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations. *Anim. Genet.* 34, 319–324. doi:10.1046/j.1365-2052.2003.01010.x.

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., et al. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* 2. doi:10.1128/msystems.00191-16.

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *N. Biotechnol.* 25, 195–203. doi:10.1016/j.nbt.2008.12.009.

Ardui, S., Ameur, A., Vermeesch, J. R., and Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Res.* 46, 2159–2168. doi:10.1093/nar/gky066.

Ayuso, M., Óvilo, C., Rodríguez-Bertos, A., Rey, A. I., Daza, A., Fenández, A., et al. (2015). Dietary vitamin A restriction affects adipocyte differentiation and fatty acid composition of intramuscular fat in Iberian pigs. *Meat Sci.* 108, 9–16. doi:10.1016/j.meatsci.2015.04.017.

Bach, A., López-García, A., González-Recio, O., Elcoso, G., Fàbregas, F., Chaucheyras-Durand, F., et al. (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.* doi:10.3168/jds.2018-16105.

Badal, V. D., Vaccariello, E. D., Murray, E. R., Yu, K. E., Knight, R., Jeste, D. V., et al. (2020). The Gut Microbiome, Aging, and Longevity: A Systematic Review. *Nutrients* 12, 1–25. doi:10.3390/NU12123759.

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

Bara, M. R., McGowan, M. R., O'Boyle, D., and Cameron, R. D. (1993). A study of the microbial flora of the anterior vagina of normal sows during different stages of the reproductive cycle. *Aust. Vet. J.* 70, 256–259. doi:10.1111/j.1751-0813.1993.tb08043.x.

Barnett, J. A. (1998). A history of research on yeasts 1: Work by chemists and biologists 1789-1850. *Yeast* 14, 1439–1451. doi:10.1002/(SICI)1097-0061(199812)14:16<1439::AID-YEA339>3.0.CO;2-Z.

Barnett, J. A. (2000). A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880. *Yeast* 16, 755–771. doi:10.1002/1097-0061(20000615)16:8<755::AID-YEA587>3.0.CO;2-4.

Barton, E. S., White, D. W., Cathelyn, J. S., Brett-McClellan, K. A., Engle, M., Diamond, M. S., et al. (2007). Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature* 447, 326–329. doi:10.1038/nature05762.

Basarab, J. A., Beauchemin, K. A., Baron, V. S., Ominski, K. H., Guan, L. L., Miller, S. P., et al. (2013). Reducing GHG emissions through genetic improvement for feed efficiency: effects on economically important traits and enteric methane production. *Animal* 7 Suppl 2, 303–315. doi:10.1017/S1751731113000888.

Batzoglou, S., Jaffe, D. B., Stanley, K., Butler, J., Gnerre, S., Mauceli, E., et al. (2002). ARACHNE: A whole-genome shotgun assembler. *Genome Res.* 12, 177–189. doi:10.1101/gr.208902.

Bayat, A. R., Tapio, I., Vilkki, J., Shingfield, K. J., and Leskinen, H. (2018). Plant oil supplements reduce methane emissions and improve milk fatty acid composition in dairy cows fed grass silage-based diets without affecting milk yield. *J. Dairy Sci.* 101, 1136–1151. doi:10.3168/jds.2017-13545.

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 10. doi:10.7554/eLife.65088.

Behnsen, J., Perez-Lopez, A., Nuccio, S. P., and Raffatellu, M. (2015). Exploiting host immunity: The Salmonella paradigm. *Trends Immunol.* 36, 112–120. doi:10.1016/j.it.2014.12.003.

Benítez, R., Fernández, A., Isabel, B., Núñez, Y., De Mercado, E., Gómez-Izquierdo, E., et al. (2018). Modulatory effects of breed, feeding status, and diet on adipogenic, lipogenic, and lipolytic gene expression in growing iberian and duroc pigs. *Int. J. Mol. Sci.* 19, 22. doi:10.3390/ijms19010022.

Benítez, R., Núñez, Y., Ayuso, M., Isabel, B., Fernández-Barroso, M. A., De Mercado, E., et al. (2021). Changes in Biceps femoris transcriptome along growth in Iberian pigs fed different energy sources and comparative analysis with Duroc breed. *Animals* 11, 3505. doi:10.3390/ANI11123505.

Benítez, R., Trakooljul, N., Núñez, Y., Isabel, B., Murani, E., De Mercado, E., et al. (2019). Breed, diet, and interaction effects on adipose tissue transcriptome in iberian and duroc pigs fed different energy sources. *Genes (Basel).* 10, 589. doi:10.3390/genes10080589.

Bergamaschi, M., Tiezzi, F., Howard, J., Huang, Y. J., Gray, K. A., Schillebeeckx, C., et al. (2020). Gut microbiome composition differences among breeds impact feed efficiency in swine. *Microbiome* 8, 1–15. doi:10.1186/s40168-020-00888-9.

Bhat, A. H., Prabhu, P., and Balakrishnan, K. (2019). A critical analysis of state-of-the-art metagenomics OTU clustering algorithms. *J. Biosci.* 44. doi:10.1007/s12038-019-9964-5.

Bhatt, P., Verma, A., Verma, S., Anwar, M. S., Prasher, P., Mudila, H., et al. (2020). Understanding phytomicrobiome: A potential reservoir for better crop management. *Sustain.* 12, 5446. doi:10.3390/su12135446.

Blankenship, S. (2017). Complete Genomics Sequencing. Available at: https://silo.tips/download/complete-genomics-sequencing [Accessed August 10, 2021].

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019).

Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi:10.1038/s41587-019-0209-9.

Bonk, F., Popp, D., Harms, H., and Centler, F. (2018). PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. *J. Microbiol. Methods* 153, 139–147. doi:10.1016/J.MIMET.2018.09.015.

Breider, I. S., Wall, E., and Garnsworthy, P. C. (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.

Brito, L. F., Schenkel, F. S., Oliveira, H. R., Cánovas, A., and Miglior, F. (2018). Meta-analysis of heritability estimates for methane emission indicator traits in cattle and sheep. *World Congr. Genet. Appl. to Livest. Prod.* 1, 700. Available at: https://www.researchgate.net/publication/323229072 [Accessed December 1, 2021].

Buermans, H. P. J., and den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1842, 1932–1941. doi:10.1016/j.bbadis.2014.06.015.

Bui, T. P. N., Schols, H. A., Jonathan, M., Stams, A. J. M., de Vos, W. M., and Plugge, C. M. (2019). Mutual Metabolic Interactions in Co-cultures of the Intestinal Anaerostipes rhamnosivorans With an Acetogen, Methanogen, or Pectin-Degrader Affecting Butyrate Production. *Front. Microbiol.* 10. doi:10.3389/fmicb.2019.02449.

Cai, R., Cheng, C., Chen, J., Xu, X., Ding, C., and Gu, B. (2020). Interactions of commensal and pathogenic microorganisms with the mucus layer in the colon. *Gut Microbes* 11, 680–690. doi:10.1080/19490976.2020.1735606.

Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639–2643. doi:10.1038/ismej.2017.119.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi:10.1038/nmeth.3869.

Camon, E. B., Barrell, D. G., Dimmer, E. C., Lee, V., Magrane, M., Maslen, J., et al. (2005). An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. *BMC Bioinformatics* 6, S17. doi:10.1186/1471-2105-6-S1-S17.

Cani, P. D., and de Vos, W. M. (2017). Next-generation beneficial microbes: The case of Akkermansia muciniphila. *Front. Microbiol.* 8, 1765. doi:10.3389/fmicb.2017.01765.

Caporaso, J. G., Bittinger, K., Bushman, F. D., Desantis, T. Z., Andersen, G. L., and Knight, R. (2010a). PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi:10.1093/bioinformatics/btp636.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335. doi:10.1038/NMETH.F.303.

Carabotti, M., Scirocco, A., Maselli, M. A., and Severi, C. (2015). The gut-brain axis: Interactions between enteric microbiota, central and enteric nervous systems. *Ann. Gastroenterol.* 28, 203–209. PMID:25830558.

Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M., and Owen, L. J. (2015). Dysbiosis of the gut microbiota in disease. *Microb. Ecol. Heal. Dis.* 26. doi:10.3402/mehd.v26.26191.

Chassard, C., and Lacroix, C. (2013). Carbohydrates and the human gut microbiota. *Curr. Opin. Clin. Nutr. Metab. Care* 16, 453–460. doi:10.1097/MCO.0B013E3283619E63.

Chen, L., Shen, Y., Wang, C., Ding, L., Zhao, F., Wang, M., et al. (2019). Megasphaera elsdenii Lactate degradation pattern shifts in rumen acidosis models. *Front. Microbiol.* 10, 162.

doi:10.3389/fmicb.2019.00162.

Clark, R. I., and Walker, D. W. (2018). Role of gut microbiota in aging-related health decline: insights from invertebrate models. *Cell. Mol. Life Sci.* 75, 93–101. doi:10.1007/s00018-017-2671-1.

Clauss, M., and Hofmann, R. R. (2014). "The digestive system of ruminants, and peculiarities of (wild) cattle," in *Ecology, Evolution and Behaviour of Wild Cattle: Implications for Conservation* (Cambridge University Press), 57–62. doi:10.1017/CBO9781139568098.008.

Cohen, J. (2017). "The Evolution of Koch's Postulates," in *Infectious Diseases*, eds. J. Cohen, W. G. Powderly, and S. M. Opal (Elsevier), 1-3.e1. doi:10.1016/b978-0-7020-6285-8.00001-0.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633. doi:10.1093/nar/gkt1244.

Copeland, J. K., Yuan, L., Layeghifard, M., Wang, P. W., and Guttman, D. S. (2015). Seasonal community succession of the phyllosphere microbiome. *Mol. Plant-Microbe Interact.* 28, 274–285. doi:10.1094/MPMI-10-14-0331-FI.

Cox, M. S., Deblois, C. L., and Suen, G. (2021). Assessing the Response of Ruminal Bacterial and Fungal Microbiota to Whole-Rumen Contents Exchange in Dairy Cows. *Front. Microbiol.* 12, 1386. doi:10.3389/fmicb.2021.665776.

Crable, B. R., Plugge, C. M., McInerney, M. J., and Stams, A. J. M. (2011). Formate formation and formate conversion in biological fuels production. *Enzyme Res.* 2011, 1–8. doi:10.4061/2011/532536.

Crespo-Piazuelo, D., Estellé, J., Revilla, M., Criado-Mesas, L., Ramayo-Caldas, Y., Óvilo, C., et al. (2018). Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. *Sci. Rep.* 8. doi:10.1038/s41598-018-30932-6.

Crespo-Piazuelo, D., Migura-Garcia, L., Estellé, J., Criado-Mesas, L., Revilla, M., Castelló, A., et al. (2019). Association between the pig genome and its gut microbiota composition. *Sci. Rep.* 9, 1–11. doi:10.1038/s41598-019-45066-6.

Crovesy, L., Ostrowski, M., Ferreira, D. M. T. P., Rosado, E. L., and Soares-Mota, M. (2017). Effect of Lactobacillus on body weight and body fat in overweight subjects: A systematic review of randomized controlled clinical trials. *Int. J. Obes.* 41, 1607–1614. doi:10.1038/ijo.2017.161.

Dahiya, D. K., Renuka, Puniya, M., Shandilya, U. K., Dhewa, T., Kumar, N., et al. (2017). Gut microbiota modulation and its relationship with obesity using prebiotic fibers and probiotics: A review. *Front. Microbiol.* 8, 563. doi:10.3389/fmicb.2017.00563.

Daza, A., Olivares, A., Rey, A. I., Ruiz, J., and López-Bote, C. J. (2008). Iberian pig production: The problems of success. *Options Méditerranénnes* A, 163–171. ISBN:2-85352-374-8.

Daza, A., Rey, A. I., Isabel, B., and Lopez-Bote, C. J. (2005). Effect of dietary vitamin E and partial replacement of poly- with monounsaturated fat on fatty acid patterns of backfat and intramuscular fat in heavy pigs (Iberian x Duroc). *J. Anim. Physiol. Anim. Nutr. (Berl).* 89, 20–28. doi:10.1111/j.1439-0396.2004.00499.x.

de Haas, Y., Veerkamp, R. F., de Jong, G., and Aldridge, M. N. (2021). Selective breeding as a mitigation tool for methane emissions from dairy cattle. *Animal*, 100294. doi:10.1016/j.animal.2021.100294.

de Haas, Y., Windig, J. J., Calus, M. P. L., Dijkstra, J., de Haan, M., Bannink, A., et al. (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.

Dearing, M. D., and Kohl, K. D. (2017). Beyond fermentation: Other important services provided to endothermic herbivores by their gut microbiota. *Integr. Comp. Biol.* 57, 723–731. doi:10.1093/icb/icx020.

Derrien, M., Belzer, C., and de Vos, W. M. (2017). Akkermansia muciniphila and its role in regulating host functions. *Microb. Pathog.* 106, 171–181. doi:10.1016/J.MICPATH.2016.02.005.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi:10.1128/AEM.03006-05.

Difford, G. F., Plichta, D. R., Løvendahl, P., Lassen, J., Noel, S. J., Højberg, O., et al. (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.

Dillon, R. J., and Charnley, K. (2002). Mutualism between the desert locust Schistocerca gregaria and its gut microbiota. *Res. Microbiol.* 153, 503–509. doi:10.1016/S0923-2508(02)01361-X.

Dillon, R. J., and Dillon, V. M. (2004). The Gut Bacteria of Insects: Nonpathogenic Interactions. *Annu. Rev. Entomol.* 49, 71–92. doi:10.1146/ANNUREV.ENTO.49.061802.123416.

Downes, J., Dewhirst, F. E., Tanner, A. C. R., and Wade, W. G. (2013). Description of Alloprevotella rava gen. nov., sp. nov., isolated from the human oral cavity, and reclassification of Prevotella tannerae Moore et al. 1994 as Alloprevotella tannerae gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 63, 1214–1218. doi:10.1099/ijs.0.041376-0.

Drissi, F., Merhej, V., Angelakis, E., Kaoutari, A. El, Carrière, F., Henrissat, B., et al. (2014). Comparative genomics analysis of Lactobacillus species associated with weight gain or weight protection. *Nutr. Diabetes* 4, e109. doi:10.1038/nutd.2014.6.

Duarte, M. E., and Kim, S. W. (2021). Intestinal microbiota and its interaction to intestinal health in nursery pigs. *Anim. Nutr.* doi:10.1016/j.aninu.2021.05.001.

Ebler, J., Haukness, M., Pesout, T., Marschall, T., and Paten, B. (2019). Haplotype-aware diplotyping from noisy long reads. *Genome Biol. 2019 201* 20, 1–16. doi:10.1186/S13059-019-1709-0.

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 1–19. doi:10.1186/1471-2105-5-113.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi:10.1093/bioinformatics/btq461.

Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*, 081257. doi:10.1101/081257.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi:10.1093/bioinformatics/btr381.

Elmqvist, T., Folke, C., Nystrom, M., Peterson, G., Bengtsson, J., Walker, B., et al. (2003). Response Diversity, Ecosystem Change, and Resilience. *Front. Ecol. Environ.* 1, 488. doi:10.2307/3868116.

Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., and Sogin, M. L. (2015). Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* 9, 968–979. doi:10.1038/ismej.2014.195.

Ervynck, A., Dobney, K., Hongo, H., and Meadow, R. (2001). Born Free ? New Evidence for the Status of Sus scrofa at Neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey). *Paléorient* 27, 47–73. doi:10.3406/paleo.2001.4731.

FAO (2017). Global Livestock Environmental Assessment Model (GLEAM). Available at: www.fao.org/gleam/en/ [Accessed October 14, 2021].

FAO (2020). FAOSTAT Database. *Food Agric. Organ. United Nations*. Available at: http://www.fao.org/faostat/es/#data/TP/visualize [Accessed August 17, 2020].

FAO (2021). Domestic Animal Diversity Information System (DAD-IS). *Food Agric. Organ. United Nations*. Available at: http://www.fao.org/dad-is/dataexport/en/ [Accessed August 17, 2020].

Farré-Maduell, E., and Casals-Pascual, C. (2019). The origins of gut microbiome research in Europe: From Escherich to Nissle. *Hum. Microbiome J.* 14, 100065. doi:10.1016/j.humic.2019.100065.

Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Res.* 40. doi:10.1093/nar/gkr1178.

Fonty, G., Joblin, K., Chavarot, M., Roux, R., Naylor, G., and Michallon, F. (2007). Establishment and development of ruminal hydrogenotrophs in methanogen-free lambs. *Appl. Environ. Microbiol.* 73, 6391–6403. doi:10.1128/AEM.00181-07.

Forero Vizcaíno, F. J., and Andrés, A. I. (2008). *El cerdo ibérico una revisión transversal.* , ed. F. J. Forero Vizcaíno Sevilla: Consejería de Agricultura y Pesca (Andalucía) ISBN:848474244X.

Fouhse, J. M., Zijlstra, R. T., and Willing, B. P. (2016). The role of gut microbiota in the health and disease of pigs. *Anim. Front.* 6, 30–36. doi:10.2527/af.2016-0031.

Fouts, D. E., Szpakowski, S., Purushe, J., Torralba, M., Waterman, R. C., MacNeil, M. D., et al. (2012). Next Generation Sequencing to Define Prokaryotic and Fungal Diversity in the Bovine Rumen. *PLoS One* 7, e48289. doi:10.1371/journal.pone.0048289.

Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., et al. (1980). The phylogeny of prokaryotes. *Science (80-.).* 209, 457–463. doi:10.1126/science.6771870.

Frantz, L., Meijaard, E., Gongora, J., Haile, J., Groenen, M. A. M., and Larson, G. (2016). The evolution of suidae. *Annu. Rev. Anim. Biosci.* 4, 61–85. doi:10.1146/annurev-animal-021815-111155.

Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., Fitzhugh, W., et al. (2002). The genome of M. acetivorans reveals extensive metabolic and physiological diversity. *Genome Res.* 12, 532–542. doi:10.1101/gr.223902.

Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I., and Koonin, E. V (2018). Microbial genome analysis: The COG approach. *Brief. Bioinform.* 20, 1063–1070. doi:10.1093/bib/bbx117.

García-Casco, J. M., Muñoz, M., Lemonnier, G., Babilliot, J. M., Bouchez, O., Fernández, A. I., et al. (2019). The gut microbiota composition at slaughter as a potential certification tool for the Iberian pig traditional farming system. in *37th International Society for Animal Genetics Conference* (ISAG), 60. Available at: https://www.isag.us/proceedings.asp.

Garnett, T., Godde, C., Muller, A., Röös, E., Smith, P., De Boer, I., et al. (2017). Grazed and confused? Ruminating on cattle, grazing systems, methane, nitrous oxide, the soil carbon sequestration question - and what it all means for greenhouse gas emissions. Available at: https://www.oxfordmartin.ox.ac.uk/downloads/reports/fcrn_gnc_report.pdf [Accessed December 10, 2021].

Gennis, R. B., and Stewart, V. (1996). "Respiration," in *Escherichia coli and Salmonella: Cellular and Molecular Biology (second edition)*, ed. F. C. Neidhardt (Washington, DC.: ASM Press), 217–261. ISBN:1555811647.

Gest, H. (2004). The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, fellows of the Royal Society. *Notes Rec. R. Soc.* 58, 187–201. doi:10.1098/rsnr.2004.0055.

Ghurye, J. S., Cepeda-Espinoza, V., and Pop, M. (2016). Metagenomic assembly: Overview, challenges and applications. *Yale J. Biol. Med.* 89, 353–362. PMID:27698619.

Giuffra, E., Kijas, J. M. H., Amarger, V., Carlborg, Ö., Jeon, J. T., and Andersson, L. (2000). The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics* 154, 1785–1791. doi:10.1093/genetics/154.4.1785.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* 8, 2224. doi:10.3389/fmicb.2017.02224.

Gondret, F., Lefaucheur, L., Juin, H., Louveau, I., and Lebret, B. (2006). Low birth weight is associated with enlarged muscle fiber area and impaired meat tenderness of the longissimus muscle in pigs. *J. Anim. Sci.* 84, 93–103. doi:10.2527/2006.84193x.

Gonzalez-Añover, P., Encinas, T., Torres-Rovira, L., Pallares, P., Muñoz-Frutos, J., Gomez-Izquierdo, E., et al. (2011). Ovulation rate, embryo mortality and intrauterine growth retardation in obese swine with gene

polymorphisms for leptin and melanocortin receptors. *Theriogenology* 75, 34–41. doi:10.1016/j.theriogenology.2010.07.009.

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

González-Recio, O., Zubiria, I., García-Rodríguez, A., Hurtado, A., and Atxaerandio, R. (2018). Short communication: Signs of host genetic regulation in the microbiome composition in 2 dairy breeds: Holstein and Brown Swiss. *J. Dairy Sci.* 101, 2285–2292. doi:10.3168/jds.2017-13179.

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

Greenacre, M., Grunsky, E., and Bacon-Shone, J. (2021a). A comparison of isometric and amalgamation logratio balances in compositional data analysis. *Comput. Geosci.* 148, 104621. doi:10.1016/j.cageo.2020.104621.

Greenacre, M., Martínez-Álvaro, M., and Blasco, A. (2021b). Compositional Data Analysis of Microbiome and Any-Omics Datasets: A Validation of the Additive Logratio Transformation. *Front. Microbiol.* 12, 2625. doi:10.3389/fmicb.2021.727398.

Gresse, R., Durand, F. C., Dunière, L., Blanquet-Diot, S., and Forano, E. (2019). Microbiota composition and functional profiling throughout the gastrointestinal tract of commercial weaning piglets. *Microorganisms* 7. doi:10.3390/microorganisms7090343.

Grond, K., Sandercock, B. K., Jumpponen, A., and Zeglin, L. H. (2018). The avian gut microbiota: community, physiology and function in wild birds. *J. Avian Biol.* 49, e01788. doi:10.1111/JAV.01788.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi:10.1101/gr.112730.110.

Hackstein, J. H. P., and de Graaf, R. M. (2018). "Anaerobic Ciliates and Their Methanogenic Endosymbionts," in *(Endo)symbiotic Methanogenic Archaea*, ed. J. H. P. Hackstein (Berlin: Springer Berlin), 13–23. doi:10.1007/978-3-319-98836-8_2.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., and Troedsson, C. (2014). Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS One* 9, 87624. doi:10.1371/journal.pone.0087624.

Haider, B., Ahn, T. H., Bushnell, B., Chai, J., Copeland, A., and Pan, C. (2014). Omega: An Overlap-graph de novo Assembler for Metagenomics. *Bioinformatics* 30, 2717–2722. doi:10.1093/bioinformatics/btu395.

Haisan, J., Sun, Y., Guan, L. L., Beauchemin, K. A., Iwaasa, A., Duval, S., et al. (2014). The effects of feeding 3-nitrooxypropanol on methane emissions and productivity of Holstein cows in mid lactation. *J. Dairy Sci.* 97, 3110–3119. doi:10.3168/jds.2013-7834.

Hammes, W. P., and Vogel, R. F. (1995). "The genus Lactobacillus," in *The Genera of Lactic Acid Bacteria* (Springer US), 19–54. doi:10.1007/978-1-4615-5817-0_3.

Hanning, I., and Diaz-Sanchez, S. (2015). The functionality of the gastrointestinal microbiome in non-human animals. *Microbiome* 3, 51. doi:10.1186/s40168-015-0113-6.

Hao, X., Jiang, R., and Chen, T. (2011). Clustering 16S rRNA for OTU prediction: A method of unsupervised Bayesian clustering. *Bioinformatics* 27, 611–618. doi:10.1093/bioinformatics/btq725.

Hartmann, A., Rothballer, M., and Schmid, M. (2008). Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research. *Plant Soil* 312, 7–14. doi:10.1007/s11104-007-9514-z.

He, Y., Caporaso, J. G., Jiang, X.-T., Sheng, H.-F., Huse, S. M., Rideout, J. R., et al. (2015). Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. *Microbiome* 3, 1–10. doi:10.1186/s40168-015-0081-x.

Heinritz, S. N., Weiss, E., Eklund, M., Aumiller, T., Louis, S., Rings, A., et al. (2016). Intestinal microbiota and microbial metabolites are changed in a pig model fed a high-fat/low-fiber or a low-fat/high-fiber diet. *PLoS One* 11, e0154329. doi:10.1371/journal.pone.0154329.

Heras-Molina, A., Estellé, J., López-García, A., Pensantez-Pacheco, J. L., Astiz, S., Garcia-Contreras, C., et al. (2021). The impact of host genetics, independently of environmental factors, on porcine gut microbiota composition. in *38th International Society for Animal Genetics Conference* (ISAG), 31. Available at: https://www.isag.us/proceedings.asp.

Hernandez, D., François, P., Farinelli, L., Østerås, M., and Schrenzel, J. (2008). De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Res.* 18, 802–809. doi:10.1101/gr.072033.107.

Hernández Fernández, M., and Vrba, E. S. (2005). A complete estimate of the phylogenetic relationships in Ruminantia: A dated species-level supertree of the extant ruminants. *Biol. Rev. Camb. Philos. Soc.* 80, 269–302. doi:10.1017/S1464793104006670.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004). Bellerophon: A program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20, 2317–2319. doi:10.1093/bioinformatics/bth226.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi:10.1093/nar/gky1085.

Hungate, R. E. (1966). *The Rumen and its Microbes*. New York: Academic Press ISBN:1483263622.

Hungate, R. E., and Macy, J. (1973). The Roll-Tube Method for Cultivation of Strict Anaerobes. *Bull. Ecol. Res. Comm.* 17, 123–126. Available at: https://www.jstor.org/stable/20111550 [Accessed December 10, 2021].

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

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* 12, e1004957. doi:10.1371/journal.pcbi.1004957.

Inui, M., Murakami, S., Okino, S., Kawaguchi, H., Vertès, A. A., and Yukawa, H. (2004). Metabolic analysis of Corynebacterium glutamicum during lactate and succinate productions under oxygen deprivation conditions. *J. Mol. Microbiol. Biotechnol.* 7, 182–196. doi:10.1159/000079827.

IPCC (2000). Good Practice and Uncertainty Management in National Greenhouse Gas Inventories. , eds. J. Penman, D. Kruger, I. Galbally, T. Hiraishi, B. Nyenzi, S. Emmanul, et al. Japan ISBN:4-88788-000-6.

IPCC (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. , eds. V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. ISBN:9781119130536.

ITIS (2021). Integrated Taxonomic Information System (ITIS) on-line database. doi:https://doi.org/10.5066/F7KH0KBK.

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi:10.1038/nbt.4060.

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

Jašarević, E., Howerton, C. L., Howard, C. D., and Bale, T. L. (2015). Alterations in the vaginal microbiome by maternal stress are associated with metabolic reprogramming of the offspring gut and brain. *Endocrinology* 156, 3265–3276. doi:10.1210/en.2015-1177.

Jiao, A., Yu, B., He, J., Yu, J., Zheng, P., Luo, Y., et al. (2020). Short chain fatty acids could prevent fat

deposition in pigs: via regulating related hormones and genes. *Food Funct.* 11, 1845–1855. doi:10.1039/c9fo02585e.

Kamada, N., Chen, G. Y., Inohara, N., and Núñez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. *Nat. Immunol.* 14, 685–690. doi:10.1038/ni.2608.

Kamath, G. M., Shomorony, I., Xia, F., Courtade, T. A., and Tse, D. N. (2017). HINGE: Long-read assembly achieves optimal repeat resolution. *Genome Res.* 27, 747–756. doi:10.1101/gr.216465.116.

Kanagal-Shamanna, R. (2016). "Emulsion PCR: Techniques and applications," in *Methods in Molecular Biology* (New York: Humana Press, New York), 33–42. doi:10.1007/978-1-4939-3360-0_4.

Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* 96, 317–323. doi:10.1016/s1389-1723(03)90130-7.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47, D590–D595. doi:10.1093/nar/gky962.

Kaul, A., Mandal, S., Davidov, O., and Peddada, S. D. (2017). Analysis of microbiome data in the presence of excess zeros. *Front. Microbiol.* 8, 2114. doi:10.3389/fmicb.2017.02114.

Ke, S., Fang, S., He, M., Huang, X., Yang, H., Yang, B., et al. (2019). Age-based dynamic changes of phylogenetic composition and interaction networks of health pig gut microbiome feeding in a uniformed condition. *BMC Vet. Res.* 15, 1–13. doi:10.1186/s12917-019-1918-5.

Keegan, K. P., Glass, E. M., and Meyer, F. (2016). MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. *Methods Mol. Biol.* 1399, 207–233. doi:10.1007/978-1-4939-3369-3_13.

Kelly, J., Daly, K., Moran, A. W., Ryan, S., Bravo, D., and Shirazi-Beechey, S. P. (2017). Composition and diversity of mucosa-associated microbiota along the entire length of the pig gastrointestinal tract; dietary influences. *Environ. Microbiol.* 19, 1425–1438. doi:10.1111/1462-2920.13619.

Khanal, P., Maltecca, C., Schwab, C., Fix, J., and Tiezzi, F. (2021). Microbiability of meat quality and carcass composition traits in swine. *J. Anim. Breed. Genet.* 138, 223–236. doi:10.1111/jbg.12504.

Klüber, H. D., and Conrad, R. (2006). Effects of nitrate, nitrite, NO and N2O on methanogenesis and other redox processes in anoxic rice field soil. *FEMS Microbiol. Ecol.* 25, 301–318. doi:10.1111/j.1574-6941.1998.tb00482.x.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. doi:10.1111/mec.12481.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi:10.1038/s41587-019-0072-8.

Kono, N., and Arakawa, K. (2019). Nanopore sequencing: Review of potential applications in functional genomics. *Dev. Growth Differ.* 61, 316–326. doi:10.1111/DGD.12608.

Konstantinidis, T., Tsigalou, C., Karvelas, A., Stavropoulou, E., Voidarou, C., and Bezirtzoglou, E. (2020). Effects of antibiotics upon the gut microbiome: A review of the literature. *Biomedicines* 8, 1–15. doi:10.3390/biomedicines8110502.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi:10.1101/gr.215087.116.

Laehnemann, D., Borkhardt, A., and McHardy, A. C. (2016). Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief. Bioinform.* 17, 154–179. doi:10.1093/bib/bbv029.

Lane-Petter, W. (1962). The Provision and Use of Pathogen-Free Laboratory Animals. *J. R. Soc. Med.* 55, 253–263. doi:10.1177/003591576205500402.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U. S. A.* 82, 6955–6959. doi:10.1073/pnas.82.20.6955.

Lang, K., Schuldes, J., Klingl, A., Poehlein, A., Daniel, R., and Brune, A. (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.

Larson, G., and Fuller, D. Q. (2014). The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* 45, 115–136. doi:10.1146/annurev-ecolsys-110512-135813.

Lassen, J., and Løvendahl, P. (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.

Leong, I., Skinner, J., and Love, D. (2014). Application of Massively Parallel Sequencing in the Clinical Diagnostic Testing of Inherited Cardiac Conditions. *Med. Sci.* 2, 98–126. doi:10.3390/MEDSCI2020098.

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi:10.1093/bioinformatics/btv033.

Li, W., and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi:10.1093/bioinformatics/btl158.

Lin, H., and Peddada, S. Das (2020). Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* 11, 1–11. doi:10.1038/s41467-020-17041-7.

Lin, Y., Yuan, J., Kolmogorov, M., Shen, M. W., Chaisson, M., and Pevzner, P. A. (2016). Assembly of long error-prone reads using de Bruijn graphs. *Proc. Natl. Acad. Sci. U. S. A.* 113, E8396–E8405. doi:10.1073/pnas.1604560113.

Liu, H., Brettell, L. E., and Singh, B. (2020). Linking the Phyllosphere Microbiome to Plant Health. *Trends Plant Sci.* 25, 841–844. doi:10.1016/j.tplants.2020.06.003.

Lloréns-Rico, V., Vieira-Silva, S., Gonçalves, P. J., Falony, G., and Raes, J. (2021). Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases. *Nat. Commun.* 12, 1–12. doi:10.1038/s41467-021-23821-6.

López-Paredes, J., Goiri, I., Atxaerandio, R., García-Rodríguez, A., Ugarte, E., Jiménez-Montero, J. A., et al. (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.

Lorenzen, E., Kudirkiene, E., Gutman, N., Grossi, A. B., Agerholm, J. S., Erneholm, K., et al. (2015). The vaginal microbiome is stable in prepubertal and sexually mature Ellegaard Göttingen Minipigs throughout an estrous cycle. *Vet. Res.* 46, 1–13. doi:10.1186/s13567-015-0274-0.

Lu, J., Idris, U., Harmon, B., Hofacre, C., Maurer, J. J., and Lee, M. D. (2003). Diversity and Succession of the Intestinal Bacterial Community of the Maturing Broiler Chicken. *Appl. Environ. Microbiol.* 69, 6816–6824. doi:10.1128/AEM.69.11.6816-6824.2003.

Lucy, M. C. (2001). Reproductive loss in high-producing dairy cattle: Where will it end? *J. Dairy Sci.* 84, 1277–1293. doi:10.3168/jds.s0022-0302(01)70158-0.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1, 18. doi:10.1186/2047-217X-1-18.

Mach, N., Berri, M., Estellé, J., Levenez, F., Lemonnier, G., Denis, C., et al. (2015). Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ. Microbiol. Rep.* 7, 554–569. doi:10.1111/1758-2229.12285.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search

and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi:10.1093/nar/gkz268.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ* 2014, e593. doi:10.7717/peerj.593.

Malnick, H. (2015). "Anaerobiospirillum," in *Bergey's Manual of Systematics of Archaea and Bacteria* (John Wiley & Sons, Ltd), 1–4. doi:10.1002/9781118960608.gbm01084.

Maltecca, C., Bergamaschi, M., and Tiezzi, F. (2020). The interaction between microbiome and pig efficiency: A review. *J. Anim. Breed. Genet.* 137, 4–13. doi:10.1111/JBG.12443.

Mamuad, L., Kim, S. H., Jeong, C. D., Choi, Y. J., Jeon, C. O., and Lee, S. S. (2014). Effect of fumarate reducing bacteria on in vitro rumen fermentation, methane mitigation and microbial diversity. *J. Microbiol.* 52, 120–128. doi:10.1007/s12275-014-3518-1.

MAPAMA (2019). Registro informativo de organismos independientes de control del ibérico (RIBER). Available at: https://www.mapa.gob.es/es/alimentacion/temas/control-calidad/mesa-iberico/riber-publico/ [Accessed November 9, 2021].

Marchandin, H., Juvonen, R., and Haikara, A. (2015). "Megasphaera," in *Bergey's Manual of Systematics of Archaea and Bacteria* (John Wiley & Sons, Ltd), 1–16. doi:10.1002/9781118960608.gbm00697.

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* 6, 287–303. doi:10.1146/annurev-anchem-062012-092628.

Margulis, L. (1990). Words as battle cries--symbiogenesis and the new field of endocytobiology. *Bioscience* 40, 673–677. doi:10.2307/1311435.

Marizzoni, M., Gurry, T., Provasi, S., Greub, G., Lopizzo, N., Ribaldi, F., et al. (2020). Comparison of Bioinformatics Pipelines and Operating Systems for the Analyses of 16S rRNA Gene Amplicon Sequences in Human Fecal Samples. *Front. Microbiol.* 11, 1262. doi:10.3389/fmicb.2020.01262.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10. doi:10.14806/ej.17.1.200.

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

Massacci, F. R., Berri, M., Lemonnier, G., Guettier, E., Blanc, F., Jardet, D., et al. (2020a). Late weaning is associated with increased microbial diversity and Faecalibacterium prausnitzii abundance in the fecal microbiota of piglets. *Anim. Microbiome* 2. doi:10.1186/s42523-020-0020-4.

Massacci, F. R., Clark, A., Ruet, A., Lansade, L., Costa, M., and Mach, N. (2020b). Inter-breed diversity and temporal dynamics of the faecal microbiota in healthy horses. *J. Anim. Breed. Genet.* 137, 103–120. doi:10.1111/jbg.12441.

Maxam, A. M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 560–564. doi:10.1073/pnas.74.2.560.

McAllister, T. A., and Newbold, C. J. (2008). Redirecting rumen fermentation to reduce methanogenesis. *Aust. J. Exp. Agric.* 48, 7–13. doi:10.1071/EA07218.

Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7, 1–9. doi:10.1038/ncomms11257.

Miller, J. R., Delcher, A. L., Koren, S., Venter, E., Walenz, B. P., Brownley, A., et al. (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24, 2818–2824. doi:10.1093/bioinformatics/btn548.

Miller, J. R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315–327. doi:10.1016/j.ygeno.2010.03.001.

Miquel, S., Martín, R., Rossi, O., Bermúdez-Humarán, L. G., Chatel, J. M., Sokol, H., et al. (2013). Faecalibacterium prausnitzii and human intestinal health. *Curr. Opin. Microbiol.* 16, 255–261. doi:10.1016/j.mib.2013.06.003.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi:10.1093/nar/gkaa913.

Mizrahi, I., Wallace, R. J., and Moraïs, S. (2021). The rumen microbiome: balancing food security and environmental impacts. *Nat. Rev. Microbiol.* 19, 553–566. doi:10.1038/s41579-021-00543-6.

Montoya, J. (1982). Efecto del arbolado de las dehesas sobre los factores ecológicos que actúan al nivel del sotobosque. *An. INIA/Serv. For.* 5, 61–85.

Mulligan, F. J., and Doherty, M. L. (2008). Production diseases of the transition cow. *Vet. J.* 176, 3–9. doi:10.1016/j.tvjl.2007.12.018.

Muñoz, M., Bozzi, R., García, F., Núñez, Y., Geraci, C., Crovetti, A., et al. (2018). Diversity across major and candidate genes in European local pig breeds. *PLoS One* 13, e0207475. doi:10.1371/journal.pone.0207475.

Murase, K., Watanabe, T., Arai, S., Kim, H., Tohya, M., Ishida-Kuroki, K., et al. (2019). Characterization of pig saliva as the major natural habitat of Streptococcus suis by analyzing oral, fecal, vaginal, and environmental microbiota. *PLoS One* 14, e0215983. doi:10.1371/journal.pone.0215983.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., et al. (2000). A whole-genome assembly of Drosophila. *Science (80-.).* 287, 2196–2204. doi:10.1126/science.287.5461.2196.

Nakano, M., Komatsu, J., Matsuura, S. I., Takashima, K., Katsura, S., and Mizuno, A. (2003). Single-molecule PCR using water-in-oil emulsion. *J. Biotechnol.* 102, 117–124. doi:10.1016/S0168-1656(03)00023-3.

Nearing, J. T., Douglas, G. M., Hayes, M. G., MacDonald, J., Desai, D. K., Allward, N., et al. (2022). Microbiome differential abundance methods produce different results across 38 datasets. *Nat. Commun.* 13, 2021.05.10.443486. doi:10.1038/s41467-022-28034-z.

Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453. doi:10.1016/0022-2836(70)90057-4.

Neefs, J. marc, Van De Peer, Y., De Rijk, P., Chapelle, S., and de Wachter, R. (1993). Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Res.* 21, 3025–3049. doi:10.1093/nar/21.13.3025.

Neel, J. V. (1962). Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am. J. Hum. Genet.* 14, 353–362. PMID:13937884.

Newbold, C. J., De la Fuente, G., Belanche, A., Ramos-Morales, E., and McEwan, N. R. (2015). The role of ciliate protozoa in the rumen. *Front. Microbiol.* 6. doi:10.3389/fmicb.2015.01313.

Nieto, R., García-Casco, J., Lara, L., Palma-Granados, P., Izquierdo, M., Hernandez, F., et al. (2019). “Ibérico (Iberian) Pig,” in *European Local Pig Breeds - Diversity and Performance. A study of project TREASURE* (IntechOpen). doi:10.5772/intechopen.83765.

Nijman, I. J., Otsen, M., Verkaar, E. L. C., De Ruijter, C., Hanekamp, E., Ochieng, J. W., et al. (2003). Hybridization of banteng (Bos javanicus) and zebu (Bos indicus) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity (Edinb).* 90, 10–16. doi:10.1038/sj.hdy.6800174.

Nilsson, R. H., Abarenkov, K., Veldre, V., Nylinder, S., De Wit, P., Brosche, S., et al. (2010). An open source chimera checker for the fungal ITS region. *Mol. Ecol. Resour.* 10, 1076–1081. doi:10.1111/j.1755-0998.2010.02850.x.

Norris, S. J., Paster, B. J., and Smibert, R. M. (2015). “Treponema,” in *Bergey’s Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (John Wiley & Sons, Inc., Bergey’s Manual Trust), 1–42. doi:10.1002/9781118960608.gbm01249.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). MetaSPAdes: A new versatile

metagenomic assembler. *Genome Res.* 27, 824–834. doi:10.1101/gr.213959.116.

O'Neill, C. J., Swain, D. L., and Kadarmideen, H. N. (2010). Evolutionary process of Bos taurus cattle in favourable versus unfavourable environments and its implications for genetic selection. *Evol. Appl.* 3, 422–433. doi:10.1111/j.1752-4571.2010.00151.x.

Olmos Soto, J. (2021). Feed intake improvement, gut microbiota modulation and pathogens control by using Bacillus species in shrimp aquaculture. *World J. Microbiol. Biotechnol.* 37, 1–7. doi:10.1007/s11274-020-02987-z.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206. doi:10.1093/nar/gkt1226.

Óvilo, C., Benítez, R., Fernández, A., Núñez, Y., Ayuso, M., Fernández, A. I., et al. (2014). Longissimus dorsi transcriptome analysis of purebred and crossbred Iberian pigs differing in muscle characteristics. *BMC Genomics* 15. doi:10.1186/1471-2164-15-413.

Pan, J., Yin, J., Zhang, K., Xie, P., Ding, H., Huang, X., et al. (2019). Dietary xylo-oligosaccharide supplementation alters gut microbial composition and activity in pigs according to age and dose. *AMB Express* 9, 1–10. doi:10.1186/s13568-019-0858-6.

Pardo, Z., Mateos, I., Campos, R., Francisco, A., Lachica, M., Ranilla, M. J., et al. (2020). Heat stress increases in vitro hindgut fermentation of distinct substrates in iberian pigs. *Animals* 10, 1–13. doi:10.3390/ani10112173.

Pariente, N. (2019). Nature milestones: Human microbiota research. *Nature*, 1–44. Available at: https://www.nature.com/immersive/d42859-019-00041-z/index.html [Accessed December 10, 2021].

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996. doi:10.1038/nbt.4229.

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., et al. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 176, 649-662.e20. doi:10.1016/j.cell.2019.01.001.

Pasteur, L. (1870). *Études sur la maladie des vers à soie : moyen pratique assuré de la combattre et d'en prévenir le retour*. , ed. Gauthier-Villars Paris doi:10.5962/bhl.title.119544.

Pedersen, R., Ingerslev, H. C., Sturek, M., Alloosh, M., Cirera, S., Christoffersen, B., et al. (2013). Characterisation of Gut Microbiota in Ossabaw and Göttingen Minipigs as Models of Obesity and Metabolic Syndrome. *PLoS One* 8. doi:10.1371/journal.pone.0056612.

Peña-Ocaña, B. A., Ovando-Ovando, C. I., Puente-Sánchez, F., Tamames, J., Servín-Garcidueñas, L. E., González-Toril, E., et al. (2021). Metagenomic and metabolic analyses of poly-extreme microbiome from an active crater volcano lake. *Environ. Res.*, 111862. doi:10.1016/j.envres.2021.111862.

Petry, A. L., Patience, J. F., Koester, L. R., Huntley, N. F., Bedford, M. R., and Schmitz-Esser, S. (2021). Xylanase modulates the microbiota of ileal mucosa and digesta of pigs fed corn-based arabinoxylans likely through both a stimbiotic and prebiotic mechanism. *PLoS One* 16, e0246144. doi:10.1371/journal.pone.0246144.

Pitt, D., Sevane, N., Nicolazzi, E. L., MacHugh, D. E., Park, S. D. E., Colli, L., et al. (2019). Domestication of cattle: Two or three events? *Evol. Appl.* 12, 123–136. doi:10.1111/eva.12674.

Plummer, E., Twin, J., Bulach, D. M., Garland, S. M., and Tabrizi, S. N. (2015). A Comparison of Three Bioinformatics Pipelines for the Analysis of Preterm Gut Microbiota using 16S rRNA Gene Sequencing Data. *J. Proteomics Bioinform.* 8, 283–291. doi:10.4172/jpb.1000381.

Pokusaeva, K., Fitzgerald, G. F., and Van Sinderen, D. (2011). Carbohydrate metabolism in Bifidobacteria. *Genes Nutr.* 6, 285–306. doi:10.1007/s12263-010-0206-6.

Poore, J., and Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science (80-.).* 360, 987–992. doi:10.1126/science.aaq0216.

Porter, J. R. (1973). Agostino Bassi bicentennial (1773-1973). *Bacteriol. Rev.* 37, 284–288. doi:10.1128/mmbr.37.3.284-288.1973.

Prakash, T., and Taylor, T. D. (2012). Functional assignment of metagenomic data: Challenges and applications. *Brief. Bioinform.* 13, 711–727. doi:10.1093/bib/bbs033.

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., and Levin, E. (2020). Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One* 15, e0227434. doi:10.1371/journal.pone.0227434.

Pszczola, M., Rzewuska, K., Mucha, S., and Strabel, T. (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.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590-6. doi:10.1093/nar/gks1219.

Quince, C., Lanzen, A., Davenport, R. J., and Turnbaugh, P. J. (2011). Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics* 12, 38. doi:10.1186/1471-2105-12-38.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844. doi:10.1038/NBT.3935.

Quiniou, N., Dagorn, J., and Gaudré, D. (2002). Variation of piglets' birth weight and consequences on subsequent performance. *Livest. Prod. Sci.* 78, 63–70. doi:10.1016/S0301-6226(02)00181-1.

Ramayo-Caldas, Y., Mach, N., Lepage, P., Levenez, F., Denis, C., Lemonnier, G., et al. (2016). Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J.* 10, 2973–2977. doi:10.1038/ismej.2016.77.

Ramayo-Caldas, Y., Zingaretti, L., Popova, M., Estellé, J., Bernard, A., Pons, N., et al. (2020). Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. *J. Anim. Breed. Genet.* 137, 49–59. doi:10.1111/jbg.12427.

Rauw, W. M., Kanis, E., Noordhuizen-Stassen, E. N., and Grommers, F. J. (1998). Undesirable side effects of selection for high production efficiency in farm animals: A review. *Livest. Prod. Sci.* 56, 15–33. doi:10.1016/S0301-6226(98)00147-X.

Rawls, J. F., Samuel, B. S., and Gordon, J. I. (2004). Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4596–4601. doi:10.1073/pnas.0400706101.

Rehfeldt, C., and Kuhn, G. (2006). Consequences of birth weight for postnatal growth performance and carcass quality in pigs as related to myogenesis. *J. Anim. Sci.* 84 Suppl, E113–E123. doi:10.2527/2006.8413_supplE113x.

Rendón-Huerta, J. A., Pinos-Rodríguez, J. M., and Kebreab, E. (2018). Animal nutrition strategies to reduce greenhouse gas emissions in dairy cattle. *Acta Univ.* 28, 34–41. doi:10.15174/au.2018.1766.

Ribani, A., Utzeri, V. J., Geraci, C., Tinarelli, S., Djan, M., Veličković, N., et al. (2019). Signatures of de-domestication in autochthonous pig breeds and of domestication in wild boar populations from MC1R and NR6A1 allele distribution. *Anim. Genet.* 50, 166–171. doi:10.1111/age.12771.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007.

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (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, e1005846. doi:10.1371/journal.pgen.1005846.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. doi:10.7717/peerj.2584.

Romero, J., Ringø, E., and Merrifield, D. L. (2014). The gut microbiota of fish. *Aquac. Nutr. Gut Heal. Probiotics Prebiotics*, 75–100. doi:10.1002/9781118897263.ch4.

Roopnarain, A., Mukhuba, M., Adeleke, R., and Moeletsi, M. (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.

Roque, B. M., Venegas, M., Kinley, R. D., De Nys, R., Duarte, T. L., Yang, X., et al. (2021). Red seaweed (Asparagopsis taxiformis) supplementation reduces enteric methane by over 80 percent in beef steers. *PLoS One* 16, e0247820. doi:10.1371/journal.pone.0247820.

Rosengaus, R. B., Zecher, C. N., Schultheis, K. F., Brucker, R. M., and Bordenstein, S. R. (2011). Disruption of the termite gut microbiota and its prolonged consequences for fitness. *Appl. Environ. Microbiol.* 77, 4303–4312. doi:10.1128/AEM.01886-10.

Roslund, M. I., Puhakka, R., Grönroos, M., Nurminen, N., Oikarinen, S., Gazali, A. M., et al. (2020). Biodiversity intervention enhances immune regulation and health-associated commensal microbiota among daycare children. *Sci. Adv.* 6. doi:10.1126/sciadv.aba2578.

Rothschild, M. F., and Ruvinsky, A. (2011). *The genetics of the pig.* , eds. M. F. Rothschild and A. Ruvinsky CABI ISBN:9781845937560.

Saborío-Montero, A., Gutiérrez-Rivas, M., López-García, A., García-Rodríguez, A., Atxaerandio, R., Goiri, I., et al. (2021). 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.

Samuelov, N. S., Lamed, R., Lowe, S., and Zeikus, I. G. (1991). Influence of CO2-HCO3- levels and pH on growth, succinate production, and enzyme activities of Anaerobiospirillum succiniciproducens. *Appl. Environ. Microbiol.* 57, 3013–3019. doi:10.1128/aem.57.10.3013-3019.1991.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467. doi:10.1073/pnas.74.12.5463.

Scanlan, P. D., Hall, A. R., Blackshields, G., Friman, V. P., Davis, M. R., Goldberg, J. B., et al. (2015). Coevolution with bacteriophages drives genome-wide host evolution and constrains the acquisition of abiotic-beneficial mutations. *Mol. Biol. Evol.* 32, 1425–1435. doi:10.1093/molbev/msv032.

Schadt, E. E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum. Mol. Genet.* 19, R227–R240. doi:10.1093/hmg/ddq416.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi:10.1128/AEM.01541-09.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi:10.1093/bioinformatics/btr026.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U. S. A.* 109, 6241–6246. doi:10.1073/pnas.1117018109.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi:10.1038/nmeth.2066.

Sela-Adler, M., Ronen, Z., Herut, B., Antler, G., Vigderovich, H., Eckert, W., et al. (2017). Co-existence of

methanogenesis and sulfate reduction with common substrates in sulfate-rich estuarine sediments. *Front. Microbiol.* 8. doi:10.3389/fmicb.2017.00766.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* 14. doi:10.1371/journal.pbio.1002533.

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* 36, 359–367. doi:10.1038/nbt.4110.

Sizova, M. V., Muller, P. A., Stancyk, D., Panikov, N. S., Mandalakis, M., Hazen, A., et al. (2014). Oribacterium parvum sp. nov. and Oribacterium asaccharolyticum sp. nov., obligately anaerobic bacteria from the human oral cavity, and emended description of the genus Oribacterium. *Int. J. Syst. Evol. Microbiol.* 64, 2642–2649. doi:10.1099/ijs.0.060988-0.

Slifierz, M. J., Friendship, R. M., and Weese, J. S. (2015). Longitudinal study of the early-life fecal and nasal microbiotas of the domestic pig. *BMC Microbiol.* 15, 1–12. doi:10.1186/s12866-015-0512-7.

Stewart, C. S., Fonty, G., and Gouet, P. (1988). The establishment of rumen microbial communities. *Anim. Feed Sci. Technol.* 21, 69–97. doi:10.1016/0377-8401(88)90093-4.

Stewart, R. D., Auffret, M. D., Warr, A., Walker, A. W., Roehe, R., and Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* 37, 953–961. doi:10.1038/s41587-019-0202-3.

Stoler, N., and Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics Bioinforma.* 3. doi:10.1093/nargab/lqab019.

Sun, Y., Cai, Y., Liu, L., Yu, F., Farrell, M. L., Mckendree, W., et al. (2009). ESPRIT: Estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res.* 37, e76. doi:10.1093/nar/gkp285.

Sutton, G. G., White, O., Adams, M. D., and Kerlavage, A. R. (1995). TIGR Assembler: A New Tool for Assembling Large Shotgun Sequencing Projects. *Genome Sci. Technol.* 1, 9–19. doi:10.1089/gst.1995.1.9.

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

Tapio, I., Snelling, T. J., Strozzi, F., and Wallace, R. J. (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.

Tatusova, T., Ciufo, S., Federhen, S., Fedorov, B., McVeigh, R., O'Neill, K., et al. (2015). Update on RefSeq microbial genomes resources. *Nucleic Acids Res.* 43, D599–D605. doi:10.1093/nar/gku1062.

Tejeda, J. F., Hernández-Matamoros, A., and González, E. (2020). Free-range and low-protein concentrated diets in Iberian pigs: Effect on plasma insulin and leptin concentration, lipogenic enzyme activity, and fatty acid composition of adipose tissue. *Animals* 10, 1–14. doi:10.3390/ani10101917.

The Integrative HMP Research Network Consortium (2014). The integrative human microbiome project: Dynamic analysis of microbiome-host omics profiles during periods of human health and disease corresponding author. *Cell Host Microbe* 16, 276–289. doi:10.1016/j.chom.2014.08.014.

Thompson, J. F., and Steinmann, K. E. (2010). Single molecule sequencing with a HeliScope genetic analysis system. *Curr. Protoc. Mol. Biol.* 92, 7.10.1-7.10.14. doi:10.1002/0471142727.mb0710s92.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi:10.1038/nature24621.

Tian, G., Wu, X., Chen, D., Yu, B., and He, J. (2017). Adaptation of gut microbiome to different dietary nonstarch polysaccharide fractions in a porcine model. *Mol. Nutr. Food Res.* 61, 1700012. doi:10.1002/mnfr.201700012.

Toro, M. A., Rodrigañez, J., Silio, L., and Rodriguez, C. (2000). Genealogical analysis of a closed herd of black hairless Iberian pigs. *Conserv. Biol.* 14, 1843–1851. doi:10.1111/j.1523-1739.2000.99322.x.

Ungerfeld, E. M., Rust, S. R., and Burnett, R. (2003). Use of some novel alternative electron sinks to inhibit ruminal methanogenesis. *Reprod. Nutr. Dev.* 43, 189–202. doi:10.1051/rnd:2003016.

Upadhyay, M. R., Chen, W., Lenstra, J. A., Goderie, C. R. J., Machugh, D. E., Park, S. D. E., et al. (2017). Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. *Heredity (Edinb).* 118, 169–176. doi:10.1038/hdy.2016.79.

Valiquette, L., and Laupland, K. B. (2013). Something old, something new, something borrowed... *Can. J. Infect. Dis. Med. Microbiol.* 24, 63–64. doi:10.1155/2013/514130.

van Gastelen, S., Dijkstra, J., and Bannink, A. (2019). Are dietary strategies to mitigate enteric methane emission equally effective across dairy cattle, beef cattle, and sheep? *J. Dairy Sci.* 102, 6109–6130. doi:10.3168/jds.2018-15785.

Van Wesemael, D., Vandaele, L., Ampe, B., Cattrysse, H., Duval, S., Kindermann, M., et al. (2019). Reducing enteric methane emissions from dairy cattle: Two ways to supplement 3-nitrooxypropanol. *J. Dairy Sci.* 102, 1780–1787. doi:10.3168/jds.2018-14534.

van Zijderveld, S. M., Gerrits, W. J. J., Apajalahti, J. A., Newbold, J. R., Dijkstra, J., Leng, R. A., et al. (2010). Nitrate and sulfate: Effective alternative hydrogen sinks for mitigation of ruminal methane production in sheep. *J. Dairy Sci.* 93, 5856–5866. doi:10.3168/jds.2010-3281.

Vaughan, H. W. (1950). *Breeds of live stock in America*. Columbus, Ohio: Long's College Book Company.

Vázquez-Gómez, M., Estellé, J., Garcia-Contreras, C., Benitez, R., Astiz, S., Pesantez-Pacheco, J. L., et al. (2019). Gut microbiome, birth weight, and productive traits in pigs in Iberian × Duroc crossbred pigs. in *37th International Society for Animal Genetics Conference* (Lleida, Spain: ISAG), 60. Available at: https://www.isag.us/proceedings.asp.

Ventanas, S., Ventanas, J., and Ruiz, J. (2007). Sensory characteristics of Iberian dry-cured loins: Influence of crossbreeding and rearing system. *Meat Sci.* 75, 211–219. doi:10.1016/j.meatsci.2006.07.003.

Voelkerding, K. V, Dames, S. A., and Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clin. Chem.* 55, 641–658. doi:10.1373/clinchem.2008.112789.

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., van Elsas, J. D., et al. (2009). TerraGenome: A consortium for the sequencing of a soil metagenome. *Nat. Rev. Microbiol.* 7, 252. doi:10.1038/NRMICRO2119.

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C. A., Hyslop, J. J., Ross, D. W., et al. (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., Sasson, G., Garnsworthy, P. C., Tapio, I., Gregson, E., Bani, P., et al. (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.

Wang, H., Ni, X., Qing, X., Zeng, D., Luo, M., Liu, L., et al. (2017). Live probiotic Lactobacillus johnsonii BS15 promotes growth performance and lowers fat deposition by improving lipid metabolism, intestinal development, and gut microflora in broilers. *Front. Microbiol.* 8, 1073. doi:10.3389/fmicb.2017.01073.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi:10.1128/AEM.00062-07.

Wang, X., Tsai, T., Deng, F., Wei, X., Chai, J., Knapp, J., et al. (2019). Longitudinal investigation of the swine gut microbiome from birth to market reveals stage and growth performance associated bacteria. *Microbiome* 7, 1–18. doi:10.1186/s40168-019-0721-7.

Warren, R. L., Sutton, G. G., Jones, S. J. M., and Holt, R. A. (2007). Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* 23, 500–501. doi:10.1093/bioinformatics/btl629.

Watson, M. (2021). New insights from 33,813 publicly available metagenome-assembled-genomes (MAGs)

assembled from the rumen microbiome. *bioRxiv*, 2021.04.02.438222. doi:10.1101/2021.04.02.438222.

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.

Weiner, P. (1973). Linear pattern matching algorithms. in *14th Annual IEEE Symposium on Switching and Automata Theory* (Institute of Electrical and Electronics Engineers (IEEE)), 1–11. doi:10.1109/SWAT.1973.13.

Westcott, S. L., and Schloss, P. D. (2015). De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ* 2015. doi:10.7717/peerj.1487.

Westcott, S. L., and Schloss, P. D. (2017). OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere* 2. doi:10.1128/mspheredirect.00073-17.

Whipps, J. M., Lewis, K., and Cooke, R. C. (1988). "Mycoparasitism and plant disease control," in *Fungi in Biological Control Systems*, ed. M. Burge (Manchester & New York: Manchester University Press), 161–187. ISBN:0-7190-1979-6.

White, B. A., Cann, I. K. O., Kocherginskaya, S. A., Aminov, R. I., Thill, L. A., Mackie, R. I., et al. (1999). Molecular Analysis of Archaea, Bacteria and Eucarya Communities in the Rumen. *Asian-Australasian J. Anim. Sci.* 12, 129–138. doi:10.5713/ajas.1999.129.

Whitehead, T. R., Lawson, P. A., and Cotta, M. A. (2015). "Hespellia," in *Bergey's Manual of Systematics of Archaea and Bacteria* (John Wiley & Sons, Ltd), 1–5. doi:10.1002/9781118960608.gbm00644.

Willems, A., and Collins, M. D. (2015). "Mitsuokella," in *Bergey's Manual of Systematics of Archaea and Bacteria* (John Wiley & Sons, Ltd), 1–9. doi:10.1002/9781118960608.gbm00698.

Williams, A. G., and Coleman, G. S. (1992). *The Rumen Protozoa*. , eds. A. G. Williams and G. S. Coleman New York, NY: Springer New York doi:10.1007/978-1-4612-2776-2.

Winogradsky, S. (1890). Recherches sur les organismes de al nitrification. *Ann Inst Pasteur* 4, 213-231,257-275,760-771.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090. doi:10.1073/pnas.74.11.5088.

Wolfe, R. S. (1993). *Methanogenesis: Ecology, Physiology, Biochemistry & Genetics*. , ed. J. G. Ferry Boston, MA: Springer US doi:10.1007/978-1-4615-2391-8.

Wu, K., and Cheng, L. (2015). "Ruminiclostridium," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (John Wiley & Sons, Inc., Bergey's Manual Trust), 1–11. doi:10.1002/9781118960608.gbm01924.

Xiao, L., Estellé, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., et al. (2016). A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* 1, 1–6. doi:10.1038/nmicrobiol.2016.161.

Xie, F., Jin, W., Si, H., Yuan, Y., Tao, Y., Liu, J., et al. (2021). An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* 9, 1–20. doi:10.1186/s40168-021-01078-x.

Xu, C., Zhu, H., and Qiu, P. (2019). Aging progression of human gut microbiota. *BMC Microbiol.* 19, 1–10. doi:10.1186/s12866-019-1616-2.

Xue, M. Y., Sun, H. Z., Wu, X. H., Liu, J. X., and Guan, L. L. (2020). Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. *Microbiome* 8, 1–19. doi:10.1186/s40168-020-00819-8.

Yang, C., Rooke, J. A., Cabeza, I., and Wallace, R. J. (2016a). Nitrate and inhibition of ruminal methanogenesis: Microbial ecology, obstacles, and opportunities for lowering methane emissions from ruminant livestock. *Front. Microbiol.* 7, 132. doi:10.3389/fmicb.2016.00132.

Yang, G., Zhang, S., Li, Z., Huang, J., Liu, Y., Liu, Y., et al. (2020). Comparison Between the Gut Microbiota in Different Gastrointestinal Segments of Large-Tailed Han and Small-Tailed Han Sheep Breeds with High-Throughput Sequencing. *Indian J. Microbiol.* 60, 436–450. doi:10.1007/s12088-020-00885-4.

Yang, H., Huang, X., Fang, S., Xin, W., Huang, L., and Chen, C. (2016b). Uncovering the composition of microbial community structure and metagenomics among three gut locations in pigs with distinct fatness. *Sci. Rep.* 6, 1–11. doi:10.1038/srep27427.

Yang, H., Xiang, Y., Robinson, K., Wang, J., Zhang, G., Zhao, J., et al. (2018). Gut Microbiota Is a Major Contributor to Adiposity in Pigs. *Front. Microbiol.* 9, 3045. doi:10.3389/fmicb.2018.03045.

Zhang, K., Fujita, Y., Chang, L., Qu, Y., Pu, Y., Wang, S., et al. (2019). Abnormal composition of gut microbiota is associated with resilience versus susceptibility to inescapable electric stress. *Transl. Psychiatry* 9, 1–9. doi:10.1038/s41398-019-0571-x.

Zhang, R. C. (2000). "Interspecies Hybridization between Yak, Bos taurus and Bos indicus and Reproduction of the Hybrids," in *Recent Advances in Yak Reproduction*, eds. X. X. Zhao and R. C. Zhang (International Veterinary Information Service). Available at: https://www.ivis.org/library/recent-advances-yak-reproduction/interspecies-hybridization-between-yak-bos-taurus-and-bos [Accessed October 11, 2021].

Zhao, W., Wang, Y., Liu, S., Huang, J., Zhai, Z., He, C., et al. (2015). The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments. *PLoS One* 10, e0117441. doi:10.1371/journal.pone.0117441.