



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

UAB
Universitat Autònoma
de Barcelona



MÁSTER INTERUNIVERSITARIO EN MEJORA GENÉTICA
ANIMAL Y BIOTECNOLOGÍA DE LA REPRODUCCIÓN

EFECTOS MATERNOS SOBRE EL MICROBIOMA

Trabajo Final de Máster

Mónica Mora Fenoll

Universitat Politècnica de València

Julio 2020

Directoras:

Noelia Ibañez Escriche

Pilar Hernández Pérez

ABSTRACT

The study of the gastrointestinal microbiome is an emerging research field in the animal genetics domain. In recent years, the reduction in the costs of massive sequencing coupled with the availability of biological databases and bioinformatics tools has made this type of study possible. The gastrointestinal microorganisms are involved in the regulation of important barrier, metabolic, absorption, immunological and vitamin synthesis functions that influence the phenotype of an individual. However, the microbial composition can vary between individuals giving rise to very different microbial profiles. This variation may be due to external factors such as diet or use of antibiotics, but also to internal factors such as the individual's genetics or the vaginal microbial composition that the individual inherits from the mother. Usually, these last two factors are confused and there is very little information on the maternal effect on the individual's microbiome without including the genetic effect of the individual. Therefore, the objective of this work was to study the importance of the maternal effect in the microbiome, separating it from the genetic effect. For this, individuals from two lines divergently selected for intramuscular fat (GIM) were used along ten generations. In the last generation, random embryo transfers were performed between the two diverging lines. The females of both lines received embryos from both the high line (A) and from the low line (B). These individuals (75 rabbits) were genotyped at birth and classified into four groups: AA, BA, AB and BB, with the first letter being the line to which the mother belonged (A or B) and the second letter being the line to which the individual belonged. Nine-week of age samples were taken from the cecum content of rabbits and bacterial DNA was extracted. Bacterial genome sequencing was performed by paired-end (150x2) in Illumina NextSeq 550. A total of 2988 KEGG variables for functional annotation and 1138 taxonomic units for the gender hierarchical level were identified in the 75 individuals with the SqueezeMeta bioinformatics tool.

In both the functional and taxonomic databases, an initial filtering of outlier samples detected with a principal component analysis was performed, as well as low prevalence variables (percentage of data that are not zeros) using the PIME package. Subsequently, the zeros were imputed and the weighted center log ratio (wCLR) normalization was applied to treat the compositional data. Finally, the multivariate analysis of the projection of discriminant latent structures (PLS-DA) was performed to identify the KEGG variables and the taxa that were differentiated between groups in each of the comparisons. Statistical analysis was performed for three comparisons: AA vs BA (maternal effect when the individual is A), AB vs BB (maternal effect when the individual is B), and AA vs BB (maternal-genetic effect). The PLS-DA results showed a clear separation between groups with an $R^2 > 0.8$ and $Q^2 > 0.7$ in most cases. At the functional level, functions related to fatty acid and amino acid metabolism stood out without differences between comparisons. However, the KEGG variables responsible for these functions were different between the comparisons. At phylum level, *Firmicutes* (70%) were responsible for intergroup discrimination in all three comparisons. Most of the genera within this phylum belonged to the *Clostridia* class (55%), with the *Ruminococcae* family (25%) being the most abundant in all three comparisons. Our study provided evidence that there was a maternal effect at both a functional and taxonomic level, allowing a better understanding of the interaction between host genetics and the microbiome. However, to extend these findings to livestock interaction will require further research and improved methodologies to define how use this information to modify the microbiome.

Keywords: microbiome, maternal effect, intramuscular fat, PLS-DA

RESUMEN

El estudio del microbioma gastrointestinal es un campo de investigación emergente en el campo de la genética animal. En los últimos años, la reducción de los costes de la secuenciación masiva unido a la disponibilidad de bases de datos biológicas y herramientas bioinformáticas ha hecho posible este tipo de estudios. Los microorganismos gastrointestinales participan en la regulación de importantes funciones de barrera, metabólicas, de absorción, inmunológicas y de síntesis de vitaminas que influyen en el fenotipo de un individuo. Sin embargo, la composición microbiana puede variar entre individuos dando lugar a perfiles microbianos muy diferentes. Esta variación puede ser debida a factores externos como la alimentación o uso de antibióticos, pero también a factores internos como la genética del individuo o la composición microbiana vaginal que hereda el individuo de la madre. Habitualmente, estos dos últimos factores están confundidos y hay muy poca información del efecto materno en el microbioma de individuo sin que esté incluido el efecto genético del individuo. Por lo tanto, el objetivo de este trabajo fue estudiar la importancia del efecto materno en el microbioma separándolo del efecto genético del individuo. Para ello, se utilizaron individuos de dos líneas seleccionadas divergentemente para grasa intramuscular (GIM) durante diez generaciones. En la última generación de selección, se realizaron transferencias embrionarias aleatorias entre las dos líneas divergentes. Las hembras de ambas líneas recibieron tanto embriones de la línea alta (A) como de la línea baja (B). Estos individuos (75 gazapos) se genotiparon al nacer y se clasificaron en cuatro grupos: AA, BA, AB y BB, siendo la primera letra la línea a la que pertenecía la madre (A o B) y la segunda letra la línea a la que pertenece cada individuo (A o B). A las nueve semanas de edad se tomaron muestras del contenido del ciego de los conejos y se extrajo el ADN bacteriano. La secuenciación del genoma bacteriano se realizó por paired-end (150x2) en Illumina NextSeq 550. Un total de 2988 variables KEGG para la anotación funcional y 1138 unidades taxonómicas para el nivel jerárquico género fueron identificadas en los 75 individuos con la herramienta bioinformática SqueezeMeta.

En ambas bases de datos, la funcional y la taxonómica, se realizó un filtrado inicial de muestras "outliers" identificadas a través de un análisis de componentes principales y de variables de baja prevalencia (porcentaje de datos que no son ceros) utilizando el paquete PIME. Posteriormente se imputaron los ceros y se aplicó la normalización weighted center log ratio (wCLR) para tratar los datos composicionales. Finalmente, se realizó el análisis multivariante de la proyección de estructuras latentes discriminantes (PLS-DA) con el objetivo de identificar las variables KEGG y las taxas que se diferenciaban entre grupos en cada una de las comparaciones. El análisis estadístico se realizó para tres comparaciones: AA vs BA (efecto materno cuando el individuo es A), AB vs BB (efecto materno cuando el individuo es B) y AA vs BB (efecto materno-genético). Los resultados del PLS-DA mostraron una clara separación entre grupos con una $R^2 > 0,8$ y $Q^2 > 0,7$ en la mayoría de los casos. Las variables KEGG responsables de la separación entre grupos fueron distintas entre las comparaciones. No obstante, las funciones KEGG más importantes relacionadas con estos KEGG no difirieron entre las comparaciones, destacando funciones relacionadas con el metabolismo de ácidos grasos y de aminoácidos. De manera similar, las taxas a nivel de género responsables de las separaciones de grupo fueron diferentes entre las comparaciones. A nivel de filo, los *Firmicutes* (70%) fueron los responsables de la discriminación entre grupos en las tres comparaciones. La mayoría de los géneros dentro de este filo pertenecían a la clase *Clostridia* (55%), siendo la familia *Ruminococcae* (25%) la más abundante en las tres comparaciones. Nuestro estudio proporcionó evidencia de la existencia de un efecto materno tanto a nivel funcional como taxonómico, permitiendo entender mejor la

interacción entre la genética huésped y el microbioma. Sin embargo, la aplicación de estos conocimientos en producción animal requiere de una mayor investigación que permita definir cómo usar esta información para modificar el microbioma de los individuos.

Palabras clave: microbioma, efecto materno, grasa intramuscular, PLS-DA

RESUM

L'estudi del microbioma gastrointestinal és un camp d'investigació emergent en el camp de la genètica animal. En els últims anys, la reducció dels costos de la seqüenciació massiva unit a la disponibilitat de bases de dades biològiques i eines bioinformàtiques ha fet possible aquest tipus d'estudis. Els microorganismes gastrointestinals participen en la regulació d'importants funcions de barrera, metabòliques, d'absorció, immunològiques i de síntesis de vitamines que influeixen en el fenotip d'un individu. No obstant això, la composició microbiana pot variar entre individus donant lloc a perfils microbians molt diferents. Aquesta variació pot ser deguda a factors externs com l'alimentació o ús d'antibiòtics, però també a factors interns com la genètica de l'individu o la composició microbiana vaginal que hereta l'individu de la mare. Habitualment, aquests dos últims factors estan confosos i hi ha molt poca informació de l'efecte matern en el microbioma d'individu sense que estiga inclòs l'efecte genètic de l'individu. Per tant, l'objectiu d'aquest treball va ser estudiar la importància de l'efecte matern en el microbioma separant-lo de l'efecte genètic de l'individu. Per a això, es van utilitzar individus de dues línies seleccionades divergentment per a greix intramuscular (GIM) durant deu generacions. En l'última generació de selecció, es van realitzar transferències embrionàries aleatòries entre les dues línies divergents. Les femelles de totes dues línies van rebre tant embrions de la línia alta (A) com de la línia baixa (B). Aquests individus (75 llogrons) es van genotipar en néixer i es van classificar en quatre grups: AA, BA, AB i BB, sent la primera lletra la línia a la qual pertanyia la mare (A o B) i la segona lletra la línia a la qual pertany cada individu (A o B). A les nou setmanes d'edat es van prendre mostres del contingut del cec dels conills i es va extraure l'ADN bacterià. La seqüenciació del genoma bacterià es va realitzar per paired-end (150x2) en Illumina NextSeq 550. Un total de 2988 variables KEGG per a l'anotació funcional i 1138 unitats taxonòmiques per al nivell jeràrquic gènere van ser identificades en els 75 individus amb l'eina bioinformàtica SqueezeMeta.

En totes dues bases de dades, la funcional i la taxonòmica, es va realitzar un filtrat inicial de mostres "outliers" identificades a través d'una anàlisi de components principals i de variables de baixa prevalença (percentatge de dades que no són zeros) utilitzant el paquet PIME. Posteriorment es van imputar els zeros i es va aplicar la normalització weighted center log ràtio (wCLR) per a tractar les dades composicionals. Finalment, es va realitzar l'anàlisi multivariant de la projecció d'estructures latents discriminants (PLS-DA) amb l'objectiu d'identificar les variables KEGG i les taxes que es diferenciaven entre grups en cadascuna de les comparacions. L'anàlisi estadística es va realitzar per a tres comparacions: AA vs BA (efecte matern quan l'individu és A), AB vs BB (efecte matern quan l'individu és B) i AA vs BB (efecte matern-genètic). Els resultats del PLS-DA van mostrar una clara separació entre grups amb una $R^2 > 0,8$ i $Q^2 > 0,7$ en la majoria dels casos. Les variables KEGG responsables de la separació entre grups van ser diferents entre les comparacions. No obstant això, les funcions KEGG més importants relacionades amb aquests KEGG no van diferir entre les comparacions, destacant funcions relacionades amb el metabolisme d'àcids grassos i d'aminoàcids. De manera similar, les taxes a nivell de gènere responsables de les separacions de grup van ser diferents entre les comparacions. A nivell de filo, els Firmicutes (70%) van ser els responsables de la discriminació entre grups en les tres comparacions. La majoria dels gèneres dins d'aquest tall pertanyien a la classe Clostridia (55%), sent la família Ruminococcae (25%) la més abundant en les tres comparacions. El nostre va proporcionar evidència de l'existència d'un efecte matern tant a nivell funcional com taxonòmic, permetent entendre millor la interacció entre la genètica del hoste i el microbioma. No obstant això, l'aplicació d'aquests coneixements en producció animal requereix d'una major investigació que permeti definir com usar aquesta informació per a modificar el microbioma dels individus.

Paraules claus: microbioma, efecte matern, greix intramuscular, PLS-DA

ÍNDICE

1	Introducción	1
1.1	El Microbioma Gastrointestinal	1
1.2	Metagenómica	2
1.2.1	Control de calidad	4
1.2.2	Ensamblaje de las lecturas y predicción génica	4
1.2.3	Anotación funcional y anotación taxonómica	5
1.3	Análisis de Datos Metagenómicos	5
1.3.1	Datos Composicionales	5
1.3.2	Análisis de Datos Multivariante	6
1.4	La Grasa Intramuscular (GIM)	8
2	Objetivos	10
3	Materiales y Métodos	11
3.1	Material Animal	11
3.2	Secuenciación del Metagenoma	12
3.3	Bioinformática y Análisis Estadístico	12
3.3.1	Análisis Bioinformático	13
3.3.2	Análisis preliminares	14
3.3.3	Tratamiento de Datos	15
4	Resultados	17
4.1	Anotación Funcional y Taxonómica	17
4.2	Análisis Exploratorios	17
4.3	Filtrado de Variables e Imputación de Ceros	20
4.4	Análisis Estadístico	21
4.4.1	Anotación funcional	22
4.4.2	Anotación taxonómica	26
5	Discusión	29

6	Conclusión	33
7	Bibliografía	34

ÍNDICE DE TABLAS

Tabla 1. Distribución de muestras por línea genética de madre receptora y línea genética del individuo. Entre paréntesis, el número de hembras receptoras.....	11
Tabla 2. Distribución del número de variables para la anotación funcional y para cada nivel de la anotación taxonómica.	17
Tabla 3. Número de variables (Variables), porcentaje de ceros (% Ceros) y error de clasificación (OOB) asociado al intervalo de prevalencia obtenido con el paquete PIME (Prevalence Interval for Microbiome Evaluation) de R para los filtrados de prevalencia del 10%, 40% y 70%.	21
Tabla 4. Resultados de la proyección de estructuras latentes discriminantes (PLS-DA) para el filtrado del 10%, 40% y 70% para cada una de las comparaciones. Número de variables después del filtrado (variables), número de variables que tienen un VIP (Variable Importance in Projection) mayor que 1 después del primer PLS-DA (Variables VIP >1), la bondad de ajuste (R^2) y la bondad de predicción (Q^2) asociado al PLS-DA con las variables VIP>1.....	22
Tabla 5. Resultados de la proyección de estructuras latentes discriminantes (PLS-DA) para la anotación taxonómica de género. Número de variables después del filtrado (variables), número de variables que tienen un VIP (<i>Variable Importance in Projection</i>) mayor que 1 después del primer PLS-DA (Variables VIP >1), la bondad de ajuste (R^2) y la bondad de predicción (Q^2) asociado al PLS-DA con las variables VIP>1 últimas.....	26

ÍNDICE DE FIGURAS

- Figura 1. Esquema del análisis bioinformático para el estudio de datos metagenómicos.** Procedimiento desde la extracción de ADN de las muestras a estudiar hasta la obtención de la anotación taxonómica y funcional. 3
- Figura 2. Gráfico de la primera (izquierda) y segunda (derecha) componente principal y proyección de las observaciones sobre las componentes.**..... 7
- Figura 3. Esquema del plan de trabajo seguido.** Pada cada una de las etapas principales del trabajo: análisis bioinformático, análisis preliminares y tratamiento de datos se muestran los pasos a seguir. 13
- Figura 4. Ejemplo del resultado del análisis bioinformático.** En las columnas se encuentran las muestras y en las filas las variables (el identificador KEGG o el taxón). 14
- Figura 5. Diagrama de barras del porcentaje de anotaciones funcionales con ceros (rojo) y sin ceros (azul) a lo largo de todas las muestras.** El eje X corresponde a las muestras y el eje Y al porcentaje de datos con y sin ceros. a) Base de datos de anotación funcional. b) Base de datos de anotación taxonómica para género. 18
- Figura 6. Diagramas de cajas de la distribución de ceros por grupos.** El eje X corresponde a los grupos AA, BA, AB y BB y eje Y al porcentaje de ceros. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja. a) Base de datos de anotación funcional. b) Base de datos de anotación taxonómica para género. 18
- Figura 7. Análisis de componentes principales de la comparación materno-genética (AA vs BB) para la base de datos de anotación funcional.** Cada punto corresponde a una muestra. El eje de la X representa la componente principal 1 (PC1) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente. El eje de la Y representa la componente principal 2 (PC2) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente. 19
- Figura 8. Análisis de componentes principales para la base de datos de anotación funcional después del filtrado del 95% para cada una de las comparaciones a), b) y c).** Cada punto corresponde a una muestra. El eje de la X representa la componente principal 1 (PC1) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente. El eje de la Y representa la componente principal 2 (PC2) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja. a) Efecto materno cuando el individuo es de la línea A (AA vs BA). b) Efecto materno cuando el individuo es de la línea B (AB vs BB). c) Efecto materno-genético (AA vs BB). 20

Figura 9. Distribución de la variable KEGG K07450 (35% de ceros). a) Antes de la imputación de ceros. b) Después del a imputación de ceros. 21

Figura 10. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal con el filtrado del 70% para la base de datos de anotación funcional. Cada punto corresponde a una muestra. CP: Componente principal. AA: individuos y madre de línea alta, BA: individuos de la línea alta y madre de baja, AB: individuos de la línea baja y madre de línea alta y BB: individuos de la línea baja y madre de la línea baja. a) Efecto materno cuando el individuo es de la línea A (AA vs BA). b) Efecto materno cuando el individuo es de la línea B (AB vs BB). c) Efecto materno-genético (AA vs BB). 23

Figura 11. Diagrama de Venn entre el efecto materno cuando el individuo fue de la línea A (AA vs BA), efecto materno cuando el individuo fue de la línea B (AB vs BB) y el efecto materno-genético (AA vs BB). AA: individuos de la línea alta y madre de la línea alta, BA: individuos de la línea alta y madre de la línea baja, AB: individuos de la línea baja y madre de la línea alta y BB: individuos de la línea baja y madre de la línea baja. 24

Figura 12. Diagrama de barras de las cinco subcategorías con un mayor VIP acumulado y su distribución en las tres comparaciones. AA vs BA: Efecto materno cuando el individuo es de la línea A. AB vs BB: Efecto materno cuando el individuo es de la línea B. AA vs BB: Efecto materno-genético. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja. 25

Figura 13. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal para la comparación del efecto materno-genético (AA vs BB) para la base de datos de anotación taxonómica para género. Cada punto corresponde a una muestra. CP: Componente principal. AA individuos y madre de línea alta y BB individuos y madre de la línea baja. 27

Figura 14. Diagrama de barras de las siete familias con un mayor VIP acumulado y su distribución en las tres comparaciones. AA vs BA: Efecto materno cuando el individuo es de la línea A. AB vs BB: Efecto materno cuando el individuo es de la línea B. AA vs BB: Efecto materno-genético. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja. 28

1 INTRODUCCIÓN

1.1 EL MICROBIOMA GASTROINTESTINAL

El microbioma se define como el conjunto de microorganismos (bacterias, arqueas, eucariotas superiores e inferiores y virus), sus genomas y las condiciones ambientales que los rodean (Marchesi and Ravel, 2015). Estos microorganismos habitan en diferentes partes de nuestro cuerpo como la piel, la boca, la vagina o el intestino (Kumar and Chordia, 2017). Sin embargo, el microbioma gastrointestinal es el más investigado en biomedicina. El genoma de los microorganismos gastrointestinales está considerado el segundo genoma en animales y humanos (Zhu et al., 2010). Además, su importancia en el desarrollo de enfermedades como la obesidad (Castaner et al., 2018), el cáncer o enfermedades autoinmunes como la esclerosis múltiple (Kirby and Ochoa-rep, 2018) es cada vez más evidente.

En producción animal el estudio del microbioma gastrointestinal también ha suscitado un gran interés (Turner, 2018). Los microorganismos gastrointestinales, generalmente bacterias, forman interacciones complejas entre ellos y su huésped (Nicholson et al., 2012). De esta forma estos microorganismos realizan funciones beneficiosas que el huésped no puede realizar por sí solo, como digerir polisacáridos complejos (Yan et al., 2016), producir vitaminas y prevenir de la colonización de patógenos (Caballero and Pamer, 2015). Además, participan en la regulación de caracteres productivos de interés como el tamaño de camada (Wang et al., 2018) o el índice de conversión (Tan et al., 2018). Asimismo, las poblaciones bacterianas intestinales modulan la respuesta inmune del huésped modificando la composición del microbioma para mantener la homeostasis intestinal (Kamada et al., 2013). Una alteración en la composición del microbioma puede desencadenar en el desarrollo de patologías (Littman and Pamer, 2012).

La composición microbiana puede variar entre individuos dando lugar a perfiles microbianos muy diferentes (Wen and Duffy, 2017). Esta variación se puede deber a factores externos como la alimentación o el uso de medicamentos (generalmente antibióticos) pero también a otros factores internos como la edad y la genética del individuo o de la madre (Wen and Duffy, 2017). Conocer estos factores y cómo afectan a la composición del microbioma de cada individuo podría permitir, en un futuro, modificar el microbioma para producir beneficios tanto en la salud como en los caracteres productivos de interés. Entre los factores externos, la alimentación es un factor de gran influencia en el microbioma gastrointestinal (Senghor et al., 2018), ya que los nutrientes son los principales sustratos para las poblaciones microbianas (Xu and Knight, 2015). Un equilibrio en la proporción de carbohidratos, fibras y proteínas en la dieta es clave para mantener la composición de la microbiota (Shahid et al., 2017). Los antibióticos también pueden modular la composición del microbioma (Danzeisen et al., 2015), los cuales se utilizan en alimentación animal para tratar enfermedades clínicas y para prevenir y controlar enfermedades comunes (Vanderhaeghen and Dewulf, 2017). No obstante, pueden alterar el perfil microbiano del intestino conduciendo a la expansión de poblaciones patogénicas (Bäumler and Sperandio, 2016), además de contribuir a la difusión de genes de resistencia a antibióticos (Zhang and Chen, 2019). En la actualidad hay un gran interés por la utilización de probióticos y prebióticos como alternativa a los antibióticos en producción animal, permitiendo así mejorar la salud animal y sus índices productivos sin necesidad de utilizar antibióticos (Cheng et al., 2014). Los probióticos

son alimentos o suplementos que contienen microorganismos vivos que proporcionan beneficios al organismo mientras que los prebióticos son sustancias (generalmente carbohidratos) que actúan como nutrientes para las bacterias intestinales (Floch, 2014).

En cuanto a los factores internos que afectan a la composición del microbioma, la composición microbiana varía a lo largo de la vida de un individuo, siendo las etapas claves la colonización inicial al nacimiento, la lactación y el destete (Tanaka and Nakayama, 2017). Como se ha mencionado anteriormente, además de la edad, la composición del microbioma de un individuo puede estar determinada por el genoma del individuo donde se hospeda (Blekhman et al., 2015; Bonder et al., 2016). Variantes genómicas del huésped contribuyen a la variabilidad en la diversidad de la composición y funcionalidad del microbioma (Awany et al., 2019). Por otra parte, en mamíferos, el microbioma se transmite a la descendencia a través del canal vaginal en el nacimiento y durante la lactancia (Treichel et al., 2019). Experimentos en humanos han visto que la forma de parto, vía vaginal o cesárea, tiene un gran impacto en la composición del microbioma (Azad et al., 2013; Jakobsson et al., 2014). Estos factores refuerzan la idea de que, en el período neonatal temprano, el microbioma de un individuo está influenciado por el microbioma de la madre e incluso puede afectar a su evolución posterior (Dunn et al., 2017). Conocer este efecto materno supone entender la composición del microbioma en las etapas iniciales de un individuo, además de mejorar la salud en las etapas posteriores de los animales. No obstante, el estudio del efecto materno es complicado ya que frecuentemente se encuentra confundido con el efecto genético del individuo.

En los últimos años los estudios centrados en el conocimiento y en el análisis del microbioma han aumentado drásticamente debido a la mejora de las tecnologías de secuenciación genómica. Anteriormente, la microbiología tradicional se basaba en técnicas de cultivo (Ravin et al., 2015). Estas técnicas eran limitadas ya que más del 99% de los microorganismos presentes en muchos ambientes naturales no son fácilmente cultivables y, por lo tanto, no eran accesibles para la investigación (Amann et al., 1995). El desarrollo de la metagenómica, (el análisis genómico de una población de microorganismos), ha supuesto una herramienta fundamental para el conocimiento de estos microorganismos (Handelsman, 2004).

1.2 METAGENÓMICA

A lo largo de los últimos años las tecnologías de secuenciación han ido evolucionando y desplazando, de forma gradual, la clásica secuenciación Sanger por las tecnologías de secuenciación masiva (*Next-generation sequencing, NGS*) (Thomas et al., 2012). Sin embargo, la secuenciación Sanger continúa siendo un estándar en secuenciación por su baja tasa de error y por su capacidad de secuenciar fragmentos largos (>700 pares de bases). El inconveniente de Sanger es su volumen de secuenciación, mientras que Sanger secuenciar solo un fragmento de ADN a la vez, NGS secuenciar de forma paralela millones de fragmentos al mismo tiempo (Reuter et al., 2015). En la actualidad, la reducción de los costes de la secuenciación masiva unido a la disponibilidad de bases de datos biológicas han convertido la secuenciación en el procedimiento estándar para analizar la estructura y la funcionalidad de los microorganismos (Thomas et al., 2012).

A la hora de secuenciar los genes de los microorganismos existen dos aproximaciones principales: secuenciación del gen ribosomal 16S o secuenciación del

genoma completo (*Whole Genome Sequencing, WGS*) (Jovel et al., 2016). El gen 16S se utiliza principalmente para estudios filogenéticos ya que se encuentra altamente conservado entre las diferentes especies de organismos procariontas. La secuenciación del genoma completo supone estudiar los genes de todos los microorganismos presentes en una muestra, el estudio de esta colección de genes se conoce como metagenómica (Marchesi and Ravel, 2015). El metagenoma es más costoso y difícil de estudiar que el gen 16S pero presenta la ventaja de mostrar todos los genes presentes en una muestra ambiental y las rutas metabólicas en las que participan. Además del objetivo final y del coste, estas dos metodologías se diferencian también en la tecnología NGS empleada. Entre las tecnologías NGS, destacan *454/Roche System* para la secuenciación del gen 16S (Kröber et al., 2009) e *Illumina/Solexa* para la secuenciación del genoma completo (Yang et al., 2016). *454/Roche System* genera lecturas de 600 pares de bases mientras que *Illumina/Solexa* de 300 pares de bases, sin embargo *Illumina/Solexa* produce más de 10 veces el número de lecturas *per run* que *454/Roche* (Thomas et al., 2012). Los datos generados de estas dos aproximaciones requieren metodologías computacionales sofisticadas y grandes recursos de *hardware* para su análisis (Gevers et al., 2012). Esto supone un reto en la investigación en el campo de la microbiología. Afortunadamente, diferentes herramientas bioinformáticas de uso libre se han desarrollado para el análisis de los datos de secuenciación NGS.

Una vez secuenciados los genomas de los microorganismos de una muestra las lecturas resultantes se analizan de forma bioinformática. Para ello se han desarrollado herramientas como Smash (Arumugam et al., 2010), Anvi'o (Eren et al., 2015), FMAP (Kim et al., 2016), MetaWrap (Uritskiy et al., 2018) y SqueezeMeta (Tamames and Puente-sánchez, 2019), entre otras. Todas estas herramientas siguen unos pasos comunes descritos en los apartados siguientes: control de calidad, ensamblaje o no de lecturas, predicción génica, anotación funcional y anotación taxonómica de los genes predichos (Figura 1).

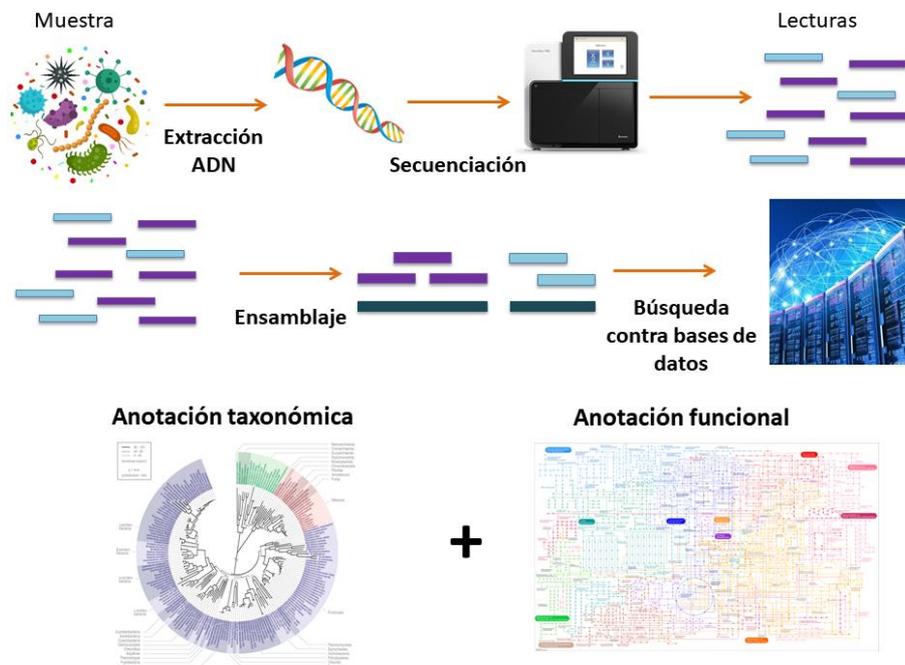


Figura 1. Esquema del análisis bioinformático para el estudio de datos metagenómicos. Procedimiento desde la extracción de ADN de las muestras a estudiar hasta la obtención de la anotación taxonómica y funcional. Adaptado de Breitwieser et al. (2017).

1.2.1 Control de calidad

Las lecturas procedentes de las tecnologías NGS requieren un primer paso de control de calidad antes de ser utilizadas en análisis posteriores. Este control de calidad consiste en la identificación y en la eliminación de artefactos de la secuenciación. Estos artefactos pueden ser lecturas de baja calidad debido a errores del propio instrumento o contaminantes causados por una mala preparación de las muestras, pudiendo introducir genomas distintos de los microbios en las muestras metagenómicas (Zhou et al., 2014). Entre los softwares utilizados para el control de calidad destacan *FastQC* (Andrews, 2010), *Cutadapt* (Martin, 2011) y *Trimmomatic* (Bolger et al., 2014).

Otro filtrado necesario cuando se trabaja con lecturas metagenómicas es el filtrado del hospedador. Si el microbioma proviene de un hospedador cuyo genoma se encuentra secuenciado es importante identificar las posibles lecturas correspondientes al hospedador y eliminarlas de análisis posteriores (Breitwieser et al., 2017). Para ello las lecturas se mapean contra el genoma del hospedador, las lecturas que no mapeen se utilizan en análisis posteriores. Un software utilizado para este tipo de filtrado es *Bowtie2* (Langmead and Salzberg, 2012)

1.2.2 Ensamblaje de las lecturas y predicción génica

El ensamblaje consiste en el alineamiento de muchos fragmentos de ADN para obtener fragmentos más largos (*contigs*) y llegar a reconstruir la secuencia original. Después de la eliminación de lecturas de baja calidad o contaminantes se puede realizar o no el ensamblaje de las lecturas. Esta decisión depende de la complejidad de la comunidad microbiana, la profundidad de secuenciación, la longitud de las lecturas y el objetivo del experimento (Singh et al., 2017).

En el ensamblaje de lecturas de un único organismo todos los fragmentos de ADN proceden del mismo genoma. Sin embargo, este no es el caso del ensamblaje metagenómico, donde el ensamblaje de lecturas de diferentes organismos es todo un reto (Ma et al., 2014). El coste computacional del ensamblaje puede llegar a ser elevado y costoso en algunas situaciones. En la actualidad, se han desarrollado ensambladores metagenómicos para la reconstrucción de genomas como *metaSPAdes* (Nurk et al., 2017) y *Megahit* (Li et al., 2015). Estos ensambladores, generan *contigs* largos que mejoran los pasos posteriores de predicción génica. Sin embargo, una profundidad de secuenciación baja o de lecturas cortas dificultan este proceso haciendo el ensamblaje casi inviable debido a la incapacidad de ensamblar algunas lecturas (Singh et al., 2017). En cambio, una adecuada profundidad de secuenciación y lecturas largas permiten un buen ensamblado.

En el no ensamblado de las lecturas, las lecturas no se mapean contra un genoma de referencia y se utilizan las lecturas crudas en los pasos posteriores de predicción génica. Este procedimiento tiene la ventaja de que el coste computacional del análisis bioinformático es menor, por ello muchos investigadores prefieren utilizar herramientas bioinformáticas que no realicen este paso como Anvi'o (Eren et al., 2015) o FMAP (Kim et al., 2016). No obstante, si las lecturas son demasiado cortas la búsqueda de genes se va a ver comprometido dificultando las etapas posteriores del análisis (Wommack et al., 2008; Carr and Borenstein, 2014).

Después del ensamblado o no ensamblado el siguiente paso es la identificación de genes, ya sea en los *contigs* o en las lecturas. Para la predicción génica se utilizan softwares como *Prodigal* (Hyatt et al., 2010), *MetaGene Annotator* (MGA) (Yok and Rosen, 2011) y *FragGeneScan* (Rho et al., 2014). Estos softwares se basan en la búsqueda de fragmentos de ADN que se traducirán y darán lugar a una proteína, estos fragmentos se conocen como *Open Reading frame* (ORF) y se caracterizan por empezar por una secuencia de tres nucleótidos (codón de inicio) y terminar con un codón de parada.

1.2.3 Anotación funcional y anotación taxonómica

Una vez realizada la predicción génica se realiza la anotación funcional y la anotación taxonómica. La anotación funcional y taxonómica consiste en asignar a las secuencias génicas predichas en el paso anterior, funciones y taxones. Para ello, se hace una búsqueda por homología en diferentes bases de datos. Existen diferentes bases de datos disponibles para la anotación funcional, como KEGG (Kanehisa et al., 2004), eggNOG (Muller et al., 2010), PFAM (Finn et al., 2014) y TIGRFAM (Selengut et al., 2007). La anotación taxonómica se realiza con bases de datos como *Kaiju* (Menzel et al., 2016) o Genbank (Sayers et al., 2018). La taxonomía clasifica a los seres vivos en ocho niveles jerárquicos denominados taxones: especie, género, familia, orden, clase, filo, reino y dominio (ordenados del nivel más bajo al más alto). Esta búsqueda por homología contra las diferentes bases de datos se realiza con diferentes algoritmos para que el proceso no sea computacionalmente costoso, ya que los conjuntos de datos metagenómicos suelen ser muy grandes (Wang et al., 2009; Ye et al., 2011).

Por último, se calcula el número de lecturas que hay para cada función y para cada taxón por muestra, a este número de lecturas se le conoce como abundancia. Para calcular la abundancia se mapean las lecturas crudas contra los *contigs* resultantes del ensamblaje. Una misma función o taxón puede estar asignada en diferentes *contigs*, por lo que se extrae el número total de lecturas crudas que mapean contra cada función o taxón. En este proceso se utilizan software como *Bowtie2* (Langmead and Salzberg, 2012) seguido de *Bedtools* (Quinlan and Hall, 2010).

1.3 ANÁLISIS DE DATOS METAGENÓMICOS

1.3.1 Datos Composicionales

Actualmente, existe una creciente conciencia de que las bases de datos generadas por las tecnologías de secuenciación masiva son datos composicionales. La abundancia de cada componente (una función o un taxón) para una muestra solo es coherentemente interpretable en relación con otros componentes de la misma muestra (Quinn et al., 2018). Esta propiedad surge por la tecnología de secuenciación, donde el número total de lecturas para una muestra es un número arbitrario impuesto por el instrumento (Gloor et al., 2017). Los datos de composición miden cada muestra como una composición, un vector de valores positivos distintos de cero con información relativa (Aitchison, 1982). Las componentes de una composición se denominan partes, y cada una de estas partes se representa con la letra j siendo J el número total de partes.

Generalmente los datos composicionales son tratados con transformaciones logarítmicas para facilitar el trabajo con ratios de componentes. El uso de logaritmos

permite aumentar la escala y la corrección de asimetrías, pero impide la presencia de ceros en las bases de datos lo que obliga a imputarlos. Existen diferentes transformaciones logarítmicas a considerar por sus propiedades teóricas y prácticas: *additive logratio transformation* (ALR), *isometric logratio transformation* (IRL) y *center logratio transformation* (CLR) (Greenacre, 2019).

ALR es la transformación logarítmica más simple, se realizan ratios entre las diferentes partes eligiendo como denominador siempre a la misma parte. Por ejemplo, denominando a las partes X_1, X_2, \dots, X_J y asumiendo que la última parte se elige como denominador, ALR se define como:

$$ARL: \log\left(\frac{X_j}{X_J}\right) = \log(X_j) - \log(X_J), \quad j = 1, \dots, J - 1 \quad (1)$$

Una desventaja de este tipo de transformación es que la elección de la parte que se utiliza como denominador (X_J) es subjetiva y podría ser cualquiera de las partes.

IRL es la propuesta de transformación logarítmica más atractiva teóricamente, pero la más complicada de interpretar. La definición de IRL se basa en separar las partes en dos subconjuntos (J_1 y J_2) expresando las medias geométricas de los dos subconjuntos de partes entre sí. Luego la definición de IRL en estos dos subconjuntos es:

$$IRL: \sqrt{\frac{J_1 J_2}{J_1 + J_2}} \left(\frac{1}{J_1} \sum_{j \in J_1} \log(X_j) - \frac{1}{J_2} \sum_{j \in J_2} \log(X_j) \right) \quad (2)$$

CLR no requiere elegir una parte j como referencia lo que supone una ventaja. Además, cada parte puede recibir el mismo peso o no (*unweighted* o *weighted center log ratio*). La definición de CLR es la siguiente:

$$CLR: \log \frac{x_j}{[\prod_j x_j]^{1/J}} = \log(x_j) - \frac{1}{J} \sum_j \log(x_j), \quad j = 1, \dots, J \quad (3)$$

CLR es una transformación de logaritmos centrados con respecto a la media geométrica a lo largo de las partes. Este es el *unweighted* CLR, donde cada parte recibe el mismo peso $1/J$. El *weighted* CLR es preferido, donde cada parte recibe un peso distinto. Este peso se calcula como la media en abundancias relativas de cada parte a lo largo de todas las muestras.

1.3.2 Análisis de Datos Multivariante

Los análisis metagenómicos generan miles de datos para cada muestra analizada, donde habitualmente hay más variables que observaciones. Los análisis de datos multivariantes permiten analizar este tipo de bases de datos a partir de la reducción de dimensionalidad (Varmuza and Filzmoser, 2009). El objetivo es transformar la matriz $X_{n \times J}$, donde cada fila (n) corresponde a una muestra y cada columna (J) contiene las variables de estudio (funciones o taxones), en una nueva matriz de dimensiones $n \times a$, donde $a \ll J$, conservando la mayor cantidad de información posible de los datos originales. Las a nuevas dimensiones se denominan variables latentes (o componentes principales) y se calculan de forma que no estén correlacionadas entre sí (son ortogonales). Estas variables latentes forman un nuevo hiperplano en el espacio donde las n observaciones son proyectadas, a estas proyecciones se les conoce como *scores*.

Existen diferentes metodologías de análisis multivariante, cada una de ellas con objetivos diferentes. En el presente trabajo nos centramos en el análisis de componentes principales (PCA) y en la proyección de estructuras latentes discriminantes (PLS-DA).

El PCA se emplea principalmente como análisis exploratorio ya que permite ver patrones, tendencias y detectar *outliers* en los datos. En este caso, las componentes principales son rectas en el espacio que maximizan la varianza de las observaciones proyectadas (*scores*, t). De esta manera, la primera variable latente (p_1) captura la mayor fuente de variación, la segunda variable latente (p_2) es ortogonal a la primera y captura la segunda mayor fuente de variación, y así sucesivamente. Los elementos de cada vector p se denominan “cargas” o “pesos” de cada variable en esa variable latente. De esta forma, la matriz X se descompone en pesos (p) y *scores* (t), una descomposición en vectores y valores propios de la matriz varianza-covarianzas de X .

El número de componentes óptimos se estiman con los métodos de validación cruzada (*cross-validation*, *CV*). Estos métodos se basan en dividir los datos en un número de grupos, seguidamente se desarrollan modelos con la eliminación de un grupo de los datos y el cálculo de un nuevo modelo a partir de los datos restantes y la predicción de los valores de los datos eliminados. Este proceso se realiza un número de veces elevado y desarrollado el modelo, se calcula la bondad de predicción (Q^2) y la bondad de ajuste (R^2). Estos parámetros son unos buenos indicadores para decidir hasta que componente incluir en el modelo. Un aumento de la R^2 acompañado de una disminución o estabilización de la Q^2 indica que hay un sobreajuste, lo que reduce su capacidad predictiva (Wold et al., 2001).

El PCA captura la mayor fuente de variación de los datos, pero sin saber a qué se debe dicha variación. El PLS-DA tiene como objetivo discriminar el conjunto de datos de X en dos o más clases. Para ello, además de la matriz X , se utiliza un vector de variables categóricas Y que clasifica las observaciones por grupos. A diferencia del PCA, en el PLS-DA las componentes principales son rectas en el espacio que maximizan la covarianza entre las observaciones proyectadas en el espacio X (*scores*, t) y el espacio Y (*scores*, u). Como en el PCA, el primer par de vectores t_1 y u_1 (componente principal 1) están correlacionados linealmente y capturan la mayor fuente de variación y correlación entre X e Y . El segundo par de vectores t_2 y u_2 (componente principal 2), también están correlacionados linealmente, aunque su grado de relación lineal suele ser inferior al de la primera componente (Figura 2)

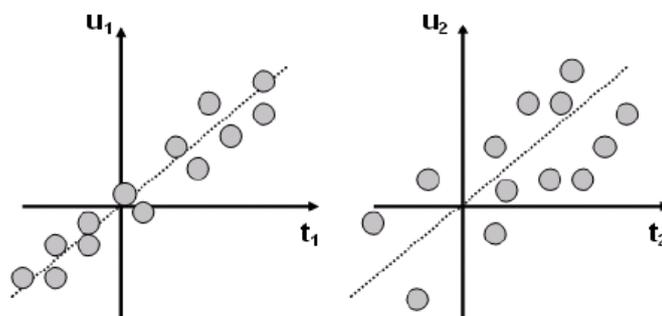


Figura 2. Gráfico de la primera (izquierda) y segunda (derecha) componente principal y proyección de las observaciones sobre las componentes.

La estima del número de componentes óptimos se realiza de la misma forma que en el PCA, utilizando los métodos de validación cruzada. Además, en el PLS-DA se puede utilizar parámetros que miden la contribución a la discriminación de cada una de las variables. De esta forma, se pueden seleccionar las variables más importantes de la discriminación. El parámetro VIP (*Variable Importance in Projection*) (Galindo-Prieto et al., 2014) es el parámetro más utilizado y mide la importancia relativa de cada una de las variables en la definición del espacio de la proyección en conjunto.

Los gráficos que permiten observar las diferencias entre las observaciones son los gráficos de los *scores t/t*. Estos gráficos muestran las observaciones proyectadas en el hiperplano definido por las nuevas componentes. Las coordenadas de los individuos en cada una de las componentes que definen el nuevo hiperplano vienen recogidas por el vector de *scores* de cada una de las componentes t_a . En estos gráficos pueden representarse regiones de confianza elípticas al 95% basada en la T^2 de Hotelling, generalización multivariante de la *t* de Student. La elipse T^2 de Hotelling provee una región por fuera de la cual se considera que las observaciones forman parte del 5% de los casos extremos.

1.4 LA GRASA INTRAMUSCULAR (GIM)

La deposición de grasa es un proceso biológico complejo, regulado por una gran cantidad de enzimas, hormonas y metabolitos que regulan el metabolismo y la diferenciación de las células grasas (Moisá et al., 2014). En humanos, la deposición excesiva de grasa conduce a la obesidad, uno de los principales problemas de salud de los últimos años (Fleming et al., 2014). Además, la obesidad está relacionada con otros problemas de salud como la diabetes y enfermedades cardiovasculares (Csige et al., 2018). En producción animal conocer los mecanismos de deposición de grasa también suscita un gran interés. Por un lado, la acumulación de grasa de depósito, está relacionado con una menor eficiencia alimentaria y menor rendimiento de la canal (Kim and Kim, 2017). Por otro lado, la grasa intramuscular (GIM) tiene un gran efecto en las propiedades organolépticas de la carne, afectando a la jugosidad, la ternura y el sabor de la misma (Hocquette et al., 2010). Además, debido a su alta heredabilidad y moderada variabilidad genética, el contenido en GIM es fácilmente modificable (Sapp et al., 2002; Schwab et al., 2009).

Los conejos son un buen modelo animal para estudios genéticos en otras especies ganaderas debido a su corto intervalo de generación y a su bajo coste. Además, al incluirse en la categoría de animales pequeños queda bajo el ámbito del comité ético local, un proceso menos costoso y menos estricto que en animales grandes (Mapara et al., 2012). En la Universitat Politècnica de València (UPV) se ha desarrollado un experimento de selección divergente para GIM en conejos durante 10 generaciones (Zomeño et al., 2013; Martínez-Álvaro et al., 2016). Estos experimentos de selección divergente tienen la ventaja de generar dos poblaciones que solo difieren en GIM y en los caracteres correlacionados. De esta forma, los cambios ambientales afectan a las dos líneas por igual por término medio. Este experimento presentó una alta respuesta a la selección, donde la diferencia entre la línea de alto contenido en GIM y la de bajo contenido en GIM era de 2.4 desviaciones típicas en la octava generación y donde ambas líneas presentaban una tendencia simétrica (Martínez Álvaro et al., 2016). Estudios recientes en estas líneas han mostrado que aparte del efecto genético del individuo el efecto de la camada explica una parte de la variabilidad de la GIM. Este efecto se debe en gran parte al efecto materno durante la gestación y la lactación. En

los últimos años, el estudio del microbioma y su participación en los procesos de deposición de grasa ha ganado importancia ya que los mecanismos claves de esta relación siguen sin estar claros (Rastelli et al., 2018). Diferentes especies bacterianas y rutas metabólicas han sido descritas en función del contenido en grasa en vacuno (Kim et al.) y porcino (Fang et al., 2017), siendo las especies relacionadas con el metabolismo de carbohidratos, de energía y de aminoácidos las más asociadas con un alto contenido en GIM. Por otra parte, hay evidencias de la existencia de un efecto materno que modula el perfil microbiano de los individuos al nacer y en lactación (Treichel et al., 2019). Conocer la contribución del microbioma gastrointestinal en el contenido en GIM y su efecto materno podría ser fundamental, en un futuro, para mejorar el contenido en GIM a través de la modulación del microbioma gastrointestinal.

2 OBJETIVOS

El **objetivo** de este trabajo es estudiar la influencia del efecto materno en el microbioma de dos líneas de conejos seleccionadas divergentemente para el carácter grasa intramuscular. Para ello, se llevaron a cabo los siguientes objetivos específicos:

- Implementar análisis multivariantes para el análisis de datos metagenómicos.
- Caracterizar los genes y taxones responsables del efecto materno en el microbioma de las dos líneas de conejos seleccionadas divergentemente por GIM.

3 MATERIALES Y MÉTODOS

3.1 MATERIAL ANIMAL

Todos los procedimientos experimentales con conejos fueron aprobados por el Comité de Ética de Investigación de la Universitat Politècnica de València, de acuerdo con las directivas del consejo 2010/63/EU (Directiva de la Comisión Europea, 2010).

El material animal utilizado para este estudio proviene de un experimento de selección divergente por grasa intramuscular (GIM) de la Universitat Politècnica de València (Zomeño et al., 2013; Martínez-Álvaro et al., 2016), realizado durante 10 generaciones en conejos. Se crearon dos líneas, una de alto contenido de GIM y otra de bajo contenido de GIM. El experimento de selección empezó con una población base de 13 machos y 83 hembras. En la décima generación, las líneas seleccionadas para alta (A) y baja (B) GIM tenían 10 machos y 60 hembras. El contenido de GIM se midió a las nueve semanas de edad en el músculo *Longissimus thoracis et lumborum* (LTL) en dos hijos (un macho y una hembra) de cada coneja. El contenido en GIM se midió usando Espectroscopia de Reflectancia de Infrarrojo cercano (NIRS) en el primer parto y excepcionalmente en el segundo y tercer parto. Todas las madres se clasificaron de acuerdo con el promedio de los dos valores fenotípicos de GIM obtenidos de su descendencia y el 20% de las mejores madres proporcionaron todas las hembras de la siguiente generación. Cada padre se apareó con seis conejas y se clasificaron los apareamientos. Para reducir la endogamia, solo se seleccionó un macho del mejor apareamiento como padre de la siguiente generación. El procedimiento de selección completo se describe en Martínez-Álvaro et al. (2016).

Para estimar el efecto materno en el microbioma se llevó a cabo un experimento en el que se realizaron transferencias embrionarias aleatorias entre las dos líneas divergentes de la décima generación. Las hembras de ambas líneas recibieron embriones tanto de la línea alta (A) como de la línea baja (B). Un total de 14 hembras de la línea alta y 9 hembras de la línea baja se utilizaron como donantes de embriones y 8 hembras de la línea alta y 6 de la línea baja como receptoras de embriones de las dos líneas. Una media de 6 embriones por cuerno uterino fueron transferidos siguiendo el protocolo descrito por Vicente and García-Ximénez (1996). Al nacer, se identificó la línea a la que pertenecía cada gazapo mediante su genotipado a través del análisis HRM (*High resolution melting*). La distribución de los animales utilizados y de madres receptoras para este trabajo queda descrita en la Tabla 1.

Tabla 1. Distribución de muestras por línea genética de madre receptora y línea genética del individuo. Entre paréntesis, el número de hembras receptoras.

	AA	BA	AB	BB
Madre	A (8)	B (6)	A (8)	B (5)
Individuos	A	A	B	B
Nº Datos	20	19	21	15

A línea alta, B línea baja. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja.

Los conejos de la misma camada fueron criados en jaulas de 9 animales desde el destete hasta el sacrificio, a las 4 y a las 9 semanas de edad respectivamente. Su alimentación fue *ad libitum* con dieta comercial compuesta de 16,0% de proteína bruta, 2,4% de aceites y grasa bruta y 16,5% de fibra bruta. Previo al sacrificio, se realizó un ayuno de 4 horas. El sacrificio se realizó mediante exanguinación previo aturdimiento eléctrico. Inmediatamente después del sacrificio se tomaron muestras del contenido cecal. Estas muestras se congelaron con nitrógeno líquido y se almacenaron a -80°C hasta el momento de su análisis. Después del sacrificio los canales se refrigeraron durante 24 horas a 4°C. Posteriormente se procedió a la disección de los músculos LTL.

La grasa intramuscular se midió en el músculo LTL. Para ello, el músculo se picó y se liofilizó previo al análisis por espectroscopia de infrarrojo cercano (NIRS, modelo 5000, FOSS NIRSystems Inc., Hillerod, Dinamarca). Un 10% de las muestras se analizaron químicamente mediante extracción con éter dietílico (Soxtec 1043 extraction unit, Tecator, Höganäs, Suecia) previa hidrólisis ácida (Soxcap 2047, Tecator, Höganäs, Suecia), como control para confirmar las mediciones de NIRS. El contenido de GIM se obtuvo gracias a las ecuaciones de calibración previamente desarrolladas por Zomeño et al., (2012).

3.2 SECUENCIACIÓN DEL METAGENOMA

La extracción del ADN bacteriano de las muestras del ciego se realizó mediante una homogenización mecánica con bolas de sílice para liberar el material genético, y con el kit DNeasy powersoil N. 12888-100 para aislar el ADN. Las muestras fueron conservadas a -80°C hasta el momento de su análisis.

La secuenciación masiva para determinar la composición del microbiota fue llevada a cabo por la empresa Sistemas Genómicos (Valencia, España). La preparación de las librerías se realizó siguiendo los protocolos de Agilent's y recomendaciones en "SureSelectQXT Whole Genome Library Prep for Illumina Multiplexed Sequencing-Featuring Transposase-Based Library Prep Technology". La cantidad y la calidad de las librerías fue analizada en 4200 Tape Station, High Sensitivity assay. El pool de librerías fue secuenciado por paired-end (150x2) en Illumina NextSeq 550, High-Output kit.

3.3 BIOINFORMÁTICA Y ANÁLISIS ESTADÍSTICO

En la Figura 3 se muestra un esquema del proceso de trabajo realizado desde la obtención de los ficheros fastq de la secuenciación completa del metagenoma hasta los análisis estadísticos.

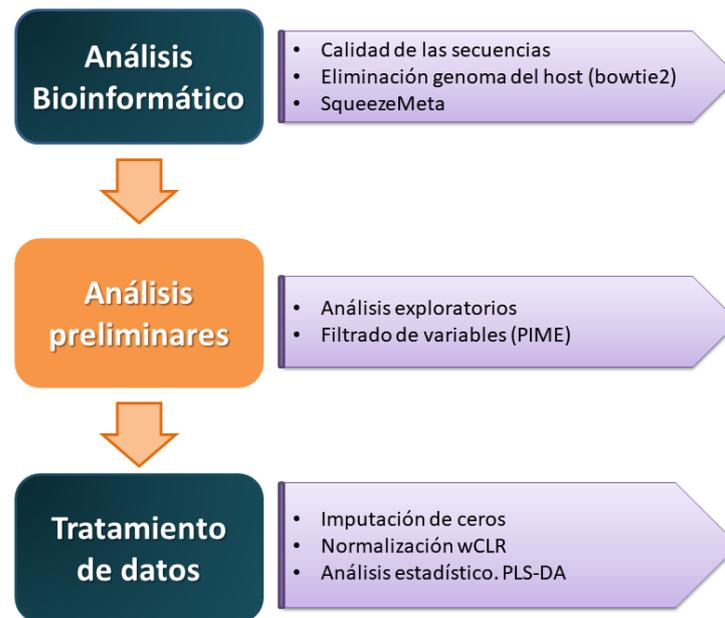


Figura 3. Esquema del plan de trabajo seguido. Pada cada una de las etapas principales del trabajo: análisis bioinformático, análisis preliminares y tratamiento de datos se muestran los pasos a seguir.

3.3.1 Análisis Bioinformático

El análisis bioinformático se realizó en el cluster Rigel de la Universidad Politécnica de Valencia. Los ficheros *fastq* procedentes de la secuenciación se mapearon contra el genoma de referencia del conejo (OryCun v2.0) utilizando el *software bowtie2* (v2.3.4.3) (Langmead and Salzberg, 2012) con las opciones por defecto. Las lecturas que no mapearon contra el genoma del conejo se asumieron como secuencias no contaminadas por el hospedador y se analizaron con el método secuencial de SqueezeMeta (Tamames and Puente-sánchez, 2019), donde cada uno de los metagenomas se trata de forma individual. Esta *pipeline* cubre todos los pasos para el análisis metagenómico: filtrado de calidad a través del programa *Trimmomatic* (Bolger et al., 2014), donde se eliminaron las lecturas con un *score* de calidad menor a 8 y con una longitud mínima menor a 30 (opciones por defecto). El ensamblaje de las lecturas se realizó con Megahit, un ensamblador con un buen balance entre el rendimiento óptimo y el uso de memoria (Li et al., 2015), la predicción génica sobre los *contigs* con Prodigal (Hyatt et al., 2010), y la búsqueda de homología contra las bases de datos taxonómica y funcional con el *software* Diamond (Buchfink et al., 2015). La base de datos funcional que se empleó fue la última versión disponible de KEGG (Kanehisa and Goto, 2000) y la base de datos para la anotación taxonómica Genbank (Sayers et al., 2018)

El análisis bioinformático dio como resultado las tablas correspondientes a la anotación funcional y a la anotación taxonómica. Estas tablas tienen como filas las muestras y como columnas las variables (el identificador KEGG o el taxón), donde para cada variable y para cada muestra se representa el número de lecturas que hay en una muestra para una determinada variable (Figura 4).

	S1	S2	S3	S4	S5	...
KEGG_1	Counts	Counts	Counts	Counts	Counts	...
KEGG_2	Counts	Counts	Counts	Counts	Counts	...
KEGG_3	Counts	Counts	Counts	Counts	Counts	...
KEGG_4	Counts	Counts	Counts	Counts	Counts	...
KEGG_5	Counts	Counts	Counts	Counts	Counts	...
KEGG_6	Counts	Counts	Counts	Counts	Counts	...
KEGG_7	Counts	Counts	Counts	Counts	Counts	...
...

Figura 4. Ejemplo del resultado del análisis bioinformático. En las columnas se encuentran las muestras y en las filas las variables (el identificador KEGG o el taxón).

3.3.2 Análisis preliminares

El objetivo de los análisis exploratorios fue observar la cantidad y la distribución del número de lecturas por muestra y por grupo, así como los ceros, de manera que se puedan identificar patrones anormales en los datos y/o *outliers*. Para ello se realizaron los gráficos con el lenguaje de programación *R*, usando funciones propias y el paquete *ggplot2*. El comportamiento de las muestras se analizó con diagramas de barras, de esta forma, se pudo estudiar la distribución de los datos en cada una de las muestras. Además, se realizaron diagramas de cajas para observar la distribución de ceros en cada uno de los grupos (AA, AB, BA, BB). También como parte de los análisis exploratorios, se realizó la detección de la presencia de posibles *outliers* y patrones en los datos con el análisis de componentes principales (PCA). Este análisis se realizó para tres filtrados en función de la prevalencia (porcentaje de datos sin ceros).

Los datos utilizados para estudiar comunidades microbianas se caracterizan por tener muchos ceros y pocos datos, lo que impide la identificación de diferencias reales entre grupos. Un punto clave en los análisis con datos metagenómicos es el filtrado de ceros, que afecta de manera considerable a los resultados finales (Kaul et al., 2017). En este trabajo el filtrado de ceros se realizó con el paquete PIME (*Prevalence Interval for Microbiome Evaluation*) de *R* (Roesch et al., 2019), una herramienta que filtra variables en función del porcentaje de datos (prevalencia) que tenga. El paquete PIME realiza un filtrado de variables para cada intervalo de prevalencia, empezando por el 5% y con incrementos del 5% de prevalencia en cada paso. Para cada intervalo de prevalencia se realizan análisis de *random forest* con el objetivo de determinar el nivel de prevalencia que proporciona el mejor modelo para predecir diferencias entre grupos. Cada intervalo de prevalencia tiene asociado un error de clasificación (*OOB error rate*), de forma que, el mejor filtrado corresponde al intervalo de prevalencia asociado a la tasa de error más baja. Sin embargo, debe haber un balance entre el mínimo error y el máximo de variables con elevada prevalencia.

El filtrado de variables se realizó para tres intervalos de prevalencia cuya finalidad era elegir el filtrado óptimo:

- Filtrado suave (10% de prevalencia)
- Filtrado medio (40% de prevalencia)
- Filtrado alto (70% de prevalencia)

3.3.3 Tratamiento de Datos

Las bases de datos de la anotación funcional y taxonómica fueron analizadas utilizando técnicas de análisis de datos composicionales. Las técnicas de datos composicionales involucran la transformación de logarítmica, lo que hace inviable utilizar los ceros como dato. Para evitar esta limitación, los ceros son reemplazados intentando no alterar la estructura de los datos. En este estudio, la imputación de aquellos datos que contenían ceros, después del filtrado de variables se realizó a través de *Bayesian-Multiplicative Replacement* (Martín-Fernández et al., 2015) con la función *cmultRepl* del paquete de *R zCompositions* (Palarea-Albaladejo and Martín-Fernández, 2015). Esta técnica reemplaza los ceros por valores cercanos y permite la aplicación de transformaciones logarítmicas. Posteriormente, se aplicó la transformación con *weighted center log ratio* (wCLR) (Greenacre, 2019) para la normalización del número de lecturas (counts) de los KEGG y las taxas como se describe a continuación:

$$wCLR(x_{ij}) = \log \frac{x_{ij}}{\prod_1^n x_{ij}} = \log(x_{ij}) - \sum_{j=1}^J w_j \log(x_{ij}) \quad (4)$$

Donde x_{ij} es el número de lecturas para una muestra i una variable j y el peso $w_j = \frac{\sum_{i=1}^n x_{ij}}{n}$ se define como la media en abundancias relativas para cada variable a lo largo de las n muestras (75)

En anotación matriz-vector la base de datos completa wCLR se puede escribir como:

$$wCLR: \begin{bmatrix} 1 - w_1 & -w_1 & -w_1 & \dots & -w_1 \\ -w_2 & 1 - w_2 & -w_2 & \dots & -w_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -w_j & -w_j & -w_j & \dots & 1 - w_j \end{bmatrix} \begin{bmatrix} \log(X_1) \\ \log(X_2) \\ \vdots \\ \log(X_j) \end{bmatrix} \quad (5)$$

La normalización wCLR consiste en logaritmos centrados donde cada variable recibe un peso distinto (w_j). Esta normalización se realizó con el paquete de *R easyCODA* con la función *CLR*, utilizando como *input* la matriz de datos de ceros imputados.

La identificación de KEGG y taxas que se diferencian debido al efecto materno o materno y genético, se realizó aplicando el análisis discriminante de mínimos cuadrados parciales (PLS-DA) (Lee et al., 2018) con el *software* SIMCA. El análisis se aplicó para las siguientes comparaciones:

- Estudio del efecto materno cuando el individuo fue de la línea A: AA vs BA (gazapos genéticamente de la línea alta que se diferencian en la madre)
- Estudio del efecto materno cuando el individuo fue de la línea B: AB vs BB (gazapos genéticamente de la línea baja que se diferencian en la madre)
- Efecto materno-genético: AA vs BB.

Para realizar el PLS-DA se utilizó la matriz de datos imputada y normalizada como variable dependiente (X) y un vector que clasifica las observaciones por grupos (por ejemplo línea genética del individuo) como variable independiente (Y). Para cada comparación se realizaron dos PLS-DA. El primer PLS-DA se realizó con todas las variables y solo aquellas variables con VIP (*Variable Importance in Projection*) (Galindo-Prieto et al., 2014) mayores que uno fueron consideradas importantes en la proyección de estructuras latentes (Chong and Jun, 2005) y por lo tanto seleccionadas para utilizar en un segundo PLS-DA.

Los gráficos utilizados para observar la proyección de los individuos fueron los gráficos de dispersión de *scores*. En estos gráficos se pudo observar cómo se distribuían los individuos en cada una de las comparaciones en función de su grupo (AA, BA, AB o BB). El número de componentes óptimos se estimaron con los métodos de validación cruzada (*cross-validation*, CV). De esta forma, también se calculó la bondad de ajuste (R^2) y la bondad de predicción (Q^2).

4 RESULTADOS

Una media de 4.500.400 lecturas fue generada durante la secuenciación para los 4 grupos, AA, AB, BA y BB. El grupo AA fue el que presentó mayor número de lecturas medio con 4.682.241, mientras que los grupos AB, BA y BB presentaron un número de lecturas ligeramente inferior (4.544.546, 4.520.623, y 4.542.409 respectivamente). Posteriormente, los resultados del alineamiento de estas lecturas contra el genoma de referencia del conejo (*OryCun v2.0*) mostraron un bajo nivel de contaminación (0,17%) por el genoma del huésped en todos los grupos.

4.1 ANOTACIÓN FUNCIONAL Y TAXONÓMICA

En la Tabla 2 se muestra el número de variables para la anotación funcional y para cada uno de los niveles de la anotación taxonómica obtenidos a través del programa SqueezeMeta con las bases de datos *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (Kanehisa et al., 2002) y Genbank (Sayers et al., 2018). La anotación funcional (2988) mostró un mayor número de variables que la anotación taxonómica. Dentro de los niveles jerárquicos de la taxonomía, fue el género el nivel más bajo y con más variables identificadas (1138). Por lo tanto, los análisis posteriores se realizaron para los KEGG y el nivel taxonómico de género.

Tabla 2. Distribución del número de variables para la anotación funcional y para cada nivel de la anotación taxonómica.

Anotación	Nivel	Variables
Funcional	KEGG	2988
	Espece	854
Taxonómica	Género	1138
	Familia	402
	Orden	244
	Clase	140
	Filo	159
	Reino	4

4.2 ANÁLISIS EXPLORATORIOS

Tras obtener las bases de datos funcionales y taxonómicas se procedió a realizar análisis exploratorios de la cantidad y distribución de ceros e identificación de posibles *outliers*. La distribución del porcentaje de presencia y ausencia de ceros para cada una de las muestras se presenta en la Figura 5. En el caso de la base de datos de anotación

funcional (Figura 5a) se observó que la mayoría de las muestras tienen más del 40% de las anotaciones funcionales KEGG con *counts*, siendo el menor porcentaje del 35%. Sin embargo, en la base de datos de anotación taxonómica para género (Figura 5b) se observó que la mayoría de las muestras tienen tan solo el 20% de las anotaciones taxonómicas de género con *counts*. Esto indica que el porcentaje de datos con cero en esta última base de datos es claramente mayor que en la base de datos de anotación funcional.

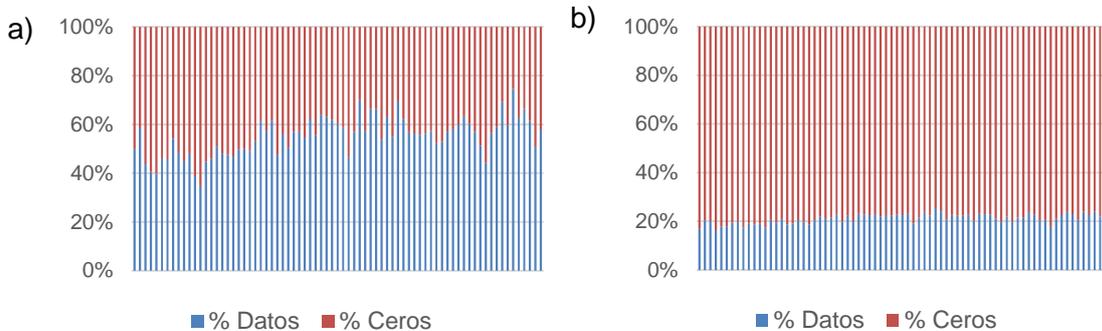


Figura 5. Diagrama de barras del porcentaje de anotaciones funcionales con ceros (rojo) y sin ceros (azul) a lo largo de todas las muestras. El eje X corresponde a las muestras y el eje Y al porcentaje de datos con y sin ceros. a) Base de datos de anotación funcional. b) Base de datos de anotación taxonómica para género.

La distribución de ceros dentro de cada grupo, AA, BA, AB y BB se muestra en la Figura 6. Tanto la base de datos de anotación funcional como la base de datos de anotación taxonómica para género mostraron una distribución similar a lo largo de los diferentes grupos. No obstante, la distribución del porcentaje de ceros por grupos es más variable en la base de datos de anotación funcional que en la base de datos de anotación taxonómica, siendo los grupos más variables AA y AB (individuos con madres de la línea alta). En total, la matriz de datos de anotación funcional presentó el 45% de datos con ceros, siendo del 44,5% y 44,14% de ceros para los grupos de individuos con madre de la línea alta (AA y AB) y del 45,31% y 46,11% para los individuos con madre de la línea baja (BA y BB). En el caso de la base de datos de anotación taxonómica para género, el porcentaje de datos medio con ceros en toda la matriz fue superior al 78%, siendo del 78,88% para AA, 78,2% para AB, 78,7% para BA y 79,66% para BB.

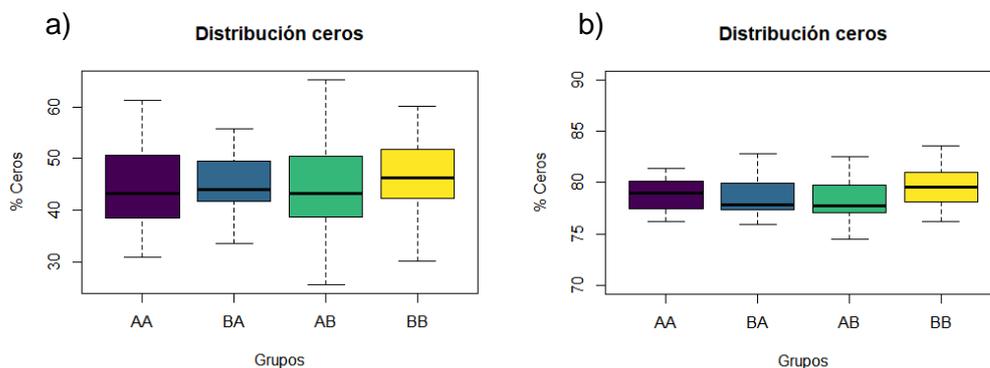


Figura 6. Diagramas de cajas de la distribución de ceros por grupos. El eje X corresponde a los grupos AA, BA, AB y BB y eje Y al porcentaje de ceros. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja. a) Base de datos de anotación funcional. b) Base de datos de anotación taxonómica para género.

La detección de posibles *outliers* se realizó con PCA (Análisis de componentes principales). Para ello se trabajó con la base de datos de anotación funcional. Esta base de datos presenta menos datos faltantes que la base de anotación taxonómica de género, por lo que la detección de muestras *outliers* es más evidente. En un primer PCA, con todas las variables, no se observaron potenciales *outliers* en ninguna de las comparaciones. A modo de ejemplo en la Figura 7 se muestra el PCA del primer y segundo componente principal para la comparación materno-genético (AA vs BB), donde no se separan las muestras en función de si son de la línea alta o de la línea baja y donde ninguna muestra se encuentra claramente alejada del resto.

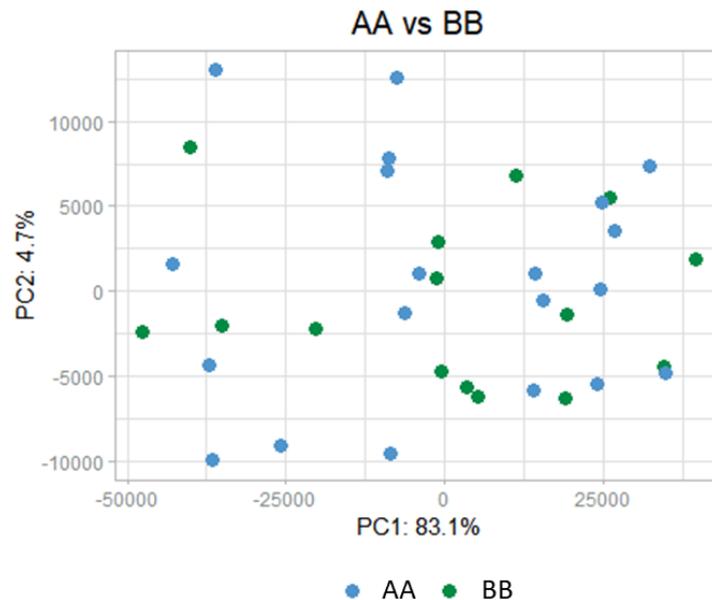


Figura 7. Análisis de componentes principales de la comparación materno-genética (AA vs BB) para la base de datos de anotación funcional. Cada punto corresponde a una muestra. El eje de la X representa la componente principal 1 (PC1) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente. El eje de la Y representa la componente principal 2 (PC2) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente.

Posteriormente, para observar mejor la presencia de posibles *outliers* se hicieron tres filtrados en función de la prevalencia (porcentaje de datos sin ceros). En un primer filtrado se eliminaron aquellas variables cuya prevalencia era menor del 10%, en el segundo filtrado aquellas variables cuya prevalencia era menor al 50% y en el último aquellas variables cuya prevalencia era menor al 95%. En el último filtrado se observaron *outliers* de manera evidente. En la Figura 8 se muestra los resultados del PCA después del filtrado del 95% para cada una de las comparaciones. Un total de 6 muestras se identificaron como posibles *outliers*. Estas muestras presentaron más del 60% de datos con ceros antes del filtrado del 95% y por lo tanto fueron eliminadas para los análisis posteriores.

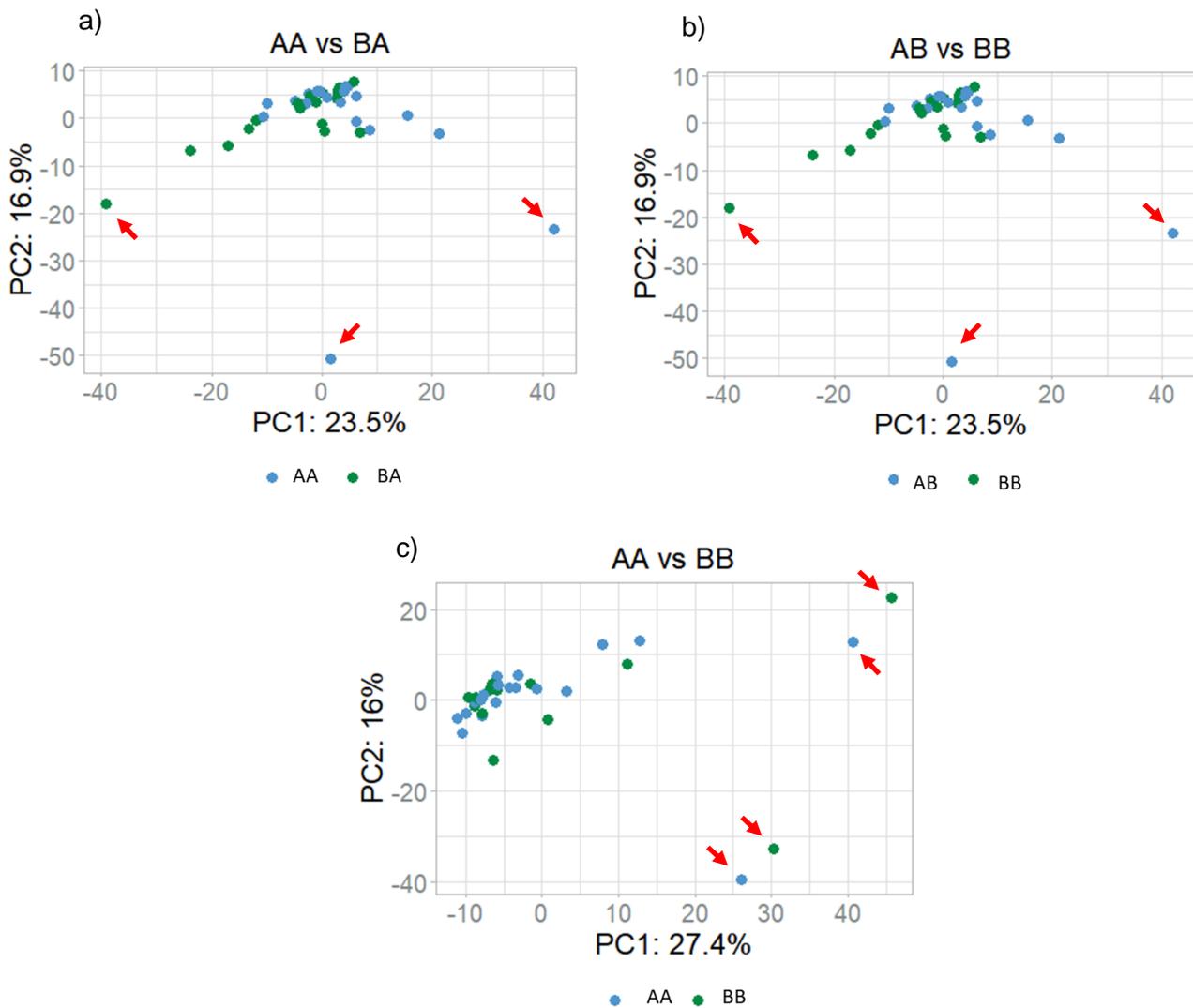


Figura 8. Análisis de componentes principales para la base de datos de anotación funcional después del filtrado del 95% para cada una de las comparaciones a), b) y c). Cada punto corresponde a una muestra. Las flechas rojas indican aquellos puntos considerados *outliers*. El eje de la X representa la componente principal 1 (PC1) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente. El eje de la Y representa la componente principal 2 (PC2) y el porcentaje de variabilidad explicada entre las muestras debida a esta componente. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja. a) Efecto materno cuando el individuo es de la línea A (AA vs BA). b) Efecto materno cuando el individuo es de la línea B (AB vs BB). c) Efecto materno-genético (AA vs BB).

Después de la eliminación de los *outliers*, la nueva distribución de las muestras por la línea genética del individuo y de su madre receptora fue la siguiente: 18 AA, 18 BA, 20 AB y 13 BB.

4.3 FILTRADO DE VARIABLES E IMPUTACIÓN DE CEROS

Una vez eliminados los *outliers* de la base de datos de anotación funcional y de la base de datos de taxonomía para género se realizó un filtrado suave (10%), uno medio (40%) y uno alto (70%) de prevalencia de las variables (porcentaje de datos sin cero), en al menos uno de los grupos.

En la Tabla 3 se muestra los resultados obtenidos para cada uno de estos filtrados aplicados en la base de datos de anotación funcional.

Tabla 3. Número de variables (Variables), porcentaje de ceros (% Ceros) y error de clasificación (OOB) asociado al intervalo de prevalencia obtenido con el paquete PIME (Prevalence Interval for Microbiome Evaluation) de R para los filtrados de prevalencia del 10%, 40% y 70%.

	Filtrado 10%	Filtrado 40%	Filtrado 70%
Variables	2364	1824	1566
% Ceros	30%	13%	6%
OOB	75%	20%	0%

En el caso de la base de datos taxonómica de género solo se realizó un tipo de filtrado y se eligió el intervalo prevalencia asociado al 0% de OOB. En este caso, el intervalo de prevalencia asociado al 0% de OOB fue del 50% (al menos un grupo de los individuos debía tener más del 50% de datos que no fuesen cero para cada una de las variables). De esta forma, reducimos el número de variables de 1138 a 234 y el porcentaje de ceros de 78% a 19%. La imputación de ceros se realizó en las bases de datos ya filtradas sin alterar la estructura de los datos. En la Figura 9 se muestra un ejemplo de la distribución de una variable de la base de datos de anotación funcional antes y después de la imputación de ceros. El resto de las variables mostraron el mismo comportamiento.

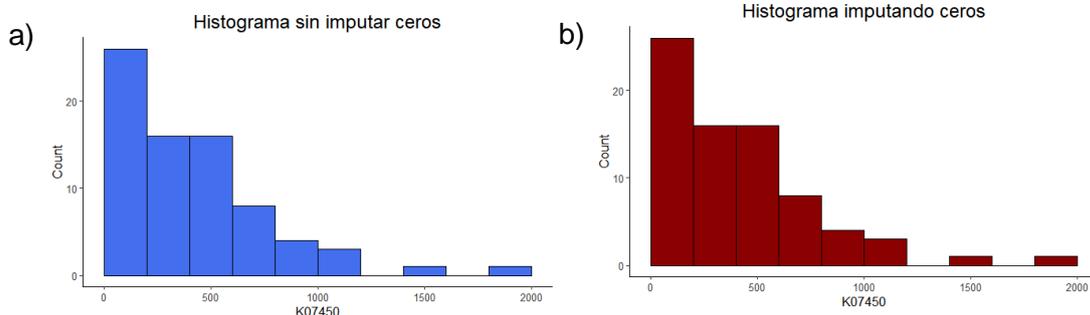


Figura 9. Distribución de la variable KEGG K07450 (35% de ceros). a) Antes de la imputación de ceros. b) Después de la imputación de ceros.

4.4 ANÁLISIS ESTADÍSTICO

Los resultados de los análisis estadísticos multivariantes de estructuras latentes discriminantes (PLS-DA) tras la imputación de ceros y la transformación *weighted center log ratio* (wCLR) para las variables KEGG y taxonómicas a nivel de género en los tres grupos de comparación (AA vs BA, AB vs BB y AA vs BB) se presentan a continuación. Los resultados presentados se obtuvieron realizando un primer PLS-DA en el cual se seleccionaron las variables con $VIP > 1$ con las que se realizó un segundo PLS-DA.

4.4.1 Anotación funcional

En la Tabla 4 se muestra los resultados del PLS-DA para cada uno de los tres niveles de filtrado (10%, 40% y 70% de prevalencia). Los resultados obtenidos del PLS-DA fueron similares en términos de bondad de ajuste (R^2) y predicción (Q^2) para los tres filtrados (10%, 40% y 70%) y para las tres comparaciones (AA vs BA, AB vs BB y AA vs BB), en todos los casos la R^2 fue mayor de 0,9 y la Q^2 mayor de 0,7. No obstante, análisis posteriores mostraron que las variables responsables de la discriminación entre grupos en los filtrados del 10% y del 40% eran variables de baja prevalencia dentro de grupo.

Tabla 4. Resultados de la proyección de estructuras latentes discriminantes (PLS-DA) para el filtrado del 10%, 40% y 70% para cada una de las comparaciones. Número de variables después del filtrado (variables), número de variables que tienen un VIP (Variable Importance in Projection) mayor que 1 después del primer PLS-DA (Variables VIP >1), la bondad de ajuste (R^2) y la bondad de predicción (Q^2) asociado al PLS-DA con las variables VIP>1.

	Filtrado	AA vs BA	AB vs BB	AA vs BB
Variables	10%	2364	2364	2364
	40%	1824	1824	1824
	70%	1566	1566	1566
Variables VIP>1	10%	1017	870	906
	40%	757	697	740
	70%	623	625	635
R²	10%	0,994	0,996	0,992
	40%	0,994	0,987	0,995
	70%	0,993	0,995	0,992
Q²	10%	0,871	0,886	0,883
	40%	0,912	0,774	0,908
	70%	0,916	0,714	0,859

AA: individuos de la línea alta y madre de la línea alta, BA: individuos de la línea alta y madre de la línea baja, AB: individuos de la línea baja y madre de la línea alta y BB: individuos de la línea baja y madre de la línea baja. AA vs BA: Efecto materno cuando el individuo es de la línea A. AB vs BB: Efecto materno cuando el individuo es de la línea. AA vs BB: Efecto materno-genético.

Los gráficos de dispersión de los *scores* para las tres comparaciones con el filtrado del 70% muestran una clara separación entre grupos. En la Figura 10 se muestra los gráficos de dispersión para la comparación del efecto materno cuando el individuo fue de la línea A (AA vs BA), efecto materno cuando el individuo fue de la línea B (AB vs BB) y el efecto materno-genético (AA vs BB).

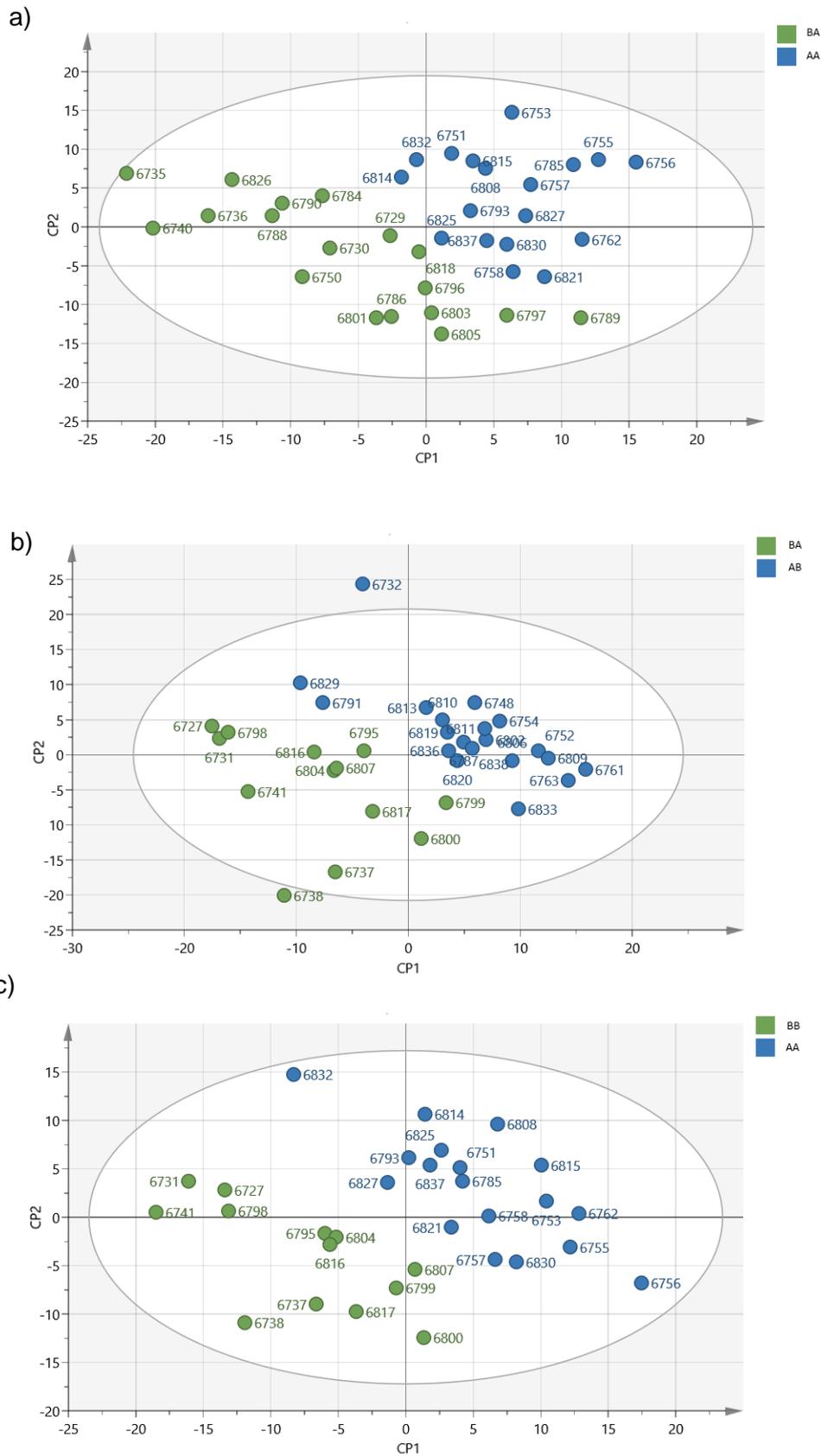


Figura 10. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal con el filtrado del 70% para la base de datos de anotación funcional. Cada punto corresponde a una muestra. CP: Componente principal. AA: individuos y madre de línea alta, BA: individuos de la línea alta y madre de baja, AB: individuos de la línea baja y madre de línea alta y BB: individuos de la línea baja y madre de la línea

baja. a) Efecto materno cuando el individuo es de la línea A (AA vs BA). b) Efecto materno cuando el individuo es de la línea B (AB vs BB). c) Efecto materno-genético (AA vs BB).

Un número similar de variables KEGG se identificaron como responsables de la discriminación entre grupos en cada una de las comparaciones. Sin embargo, estas variables fueron diferentes para cada grupo de comparación. De las variables KEGG con un $VIP > 1$, solo el 15% de las variables KEGG fueron coincidentes entre las comparaciones del efecto materno (AA vs BA y AB vs BB) y únicamente el 5% de las variables fueron las mismas en las 3 comparaciones. El resto de las variables KEGG fueron diferentes entre las comparaciones.

A pesar de las diferencias entre las comparaciones a nivel de variables KEGG, se observaron similitudes a nivel de la subcategoría a la que pertenecían las variables KEGG con $VIP > 1$. En total, se detectaron 30, 30 y 31 subcategorías para el efecto materno cuando el individuo es de la línea A (AA vs BA), el efecto materno cuando el individuo es de la línea B (AB vs BB) y el efecto materno-genético (AA vs BB) respectivamente. De ellas 24 fueron coincidentes entre las 3 comparaciones (Figura 11), entre las que destacaron subcategorías como: metabolismo de lípidos y carbohidratos, metabolismo de aminoácidos y metabolismo energético.

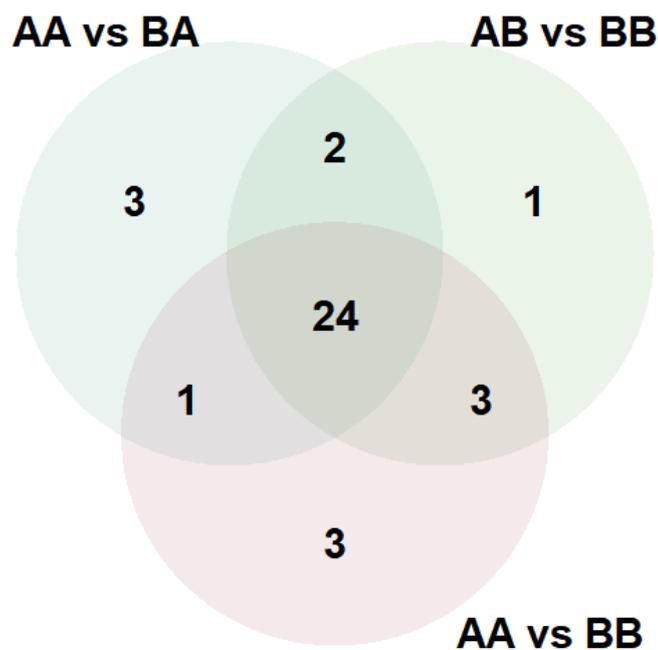


Figura 11. Diagrama de Venn entre el efecto materno cuando el individuo fue de la línea A (AA vs BA), efecto materno cuando el individuo fue de la línea B (AB vs BB) y el efecto materno-genético (AA vs BB). AA: individuos de la línea alta y madre de la línea alta, BA: individuos de la línea alta y madre de la línea baja, AB: individuos de la línea baja y madre de la línea alta y BB: individuos de la línea baja y madre de la línea baja.

En la Figura 12 se muestran las subcategorías con un VIP acumulado mayor a 20 de las 24 subcategorías coincidentes entre las tres comparaciones, y su distribución en cada una de ellas. Las cinco subcategorías que tuvieron un VIP acumulado mayor a 20 corresponden a procesos celulares y señalización, metabolismo, procesamiento de la información genética, metabolismo de carbohidratos, lipídico y de energía y metabolismo de aminoácidos. Además, estas subcategorías fueron las más importantes de forma individual en cada una de las comparaciones. Esto indica que a nivel de subcategoría no se observaron diferencias entre el efecto materno cuando los individuos

fueron de la línea A (AA vs BA), el efecto materno cuando los individuos fueron de la línea B (AB vs BB) y el efecto materno-genético (AA vs BB).

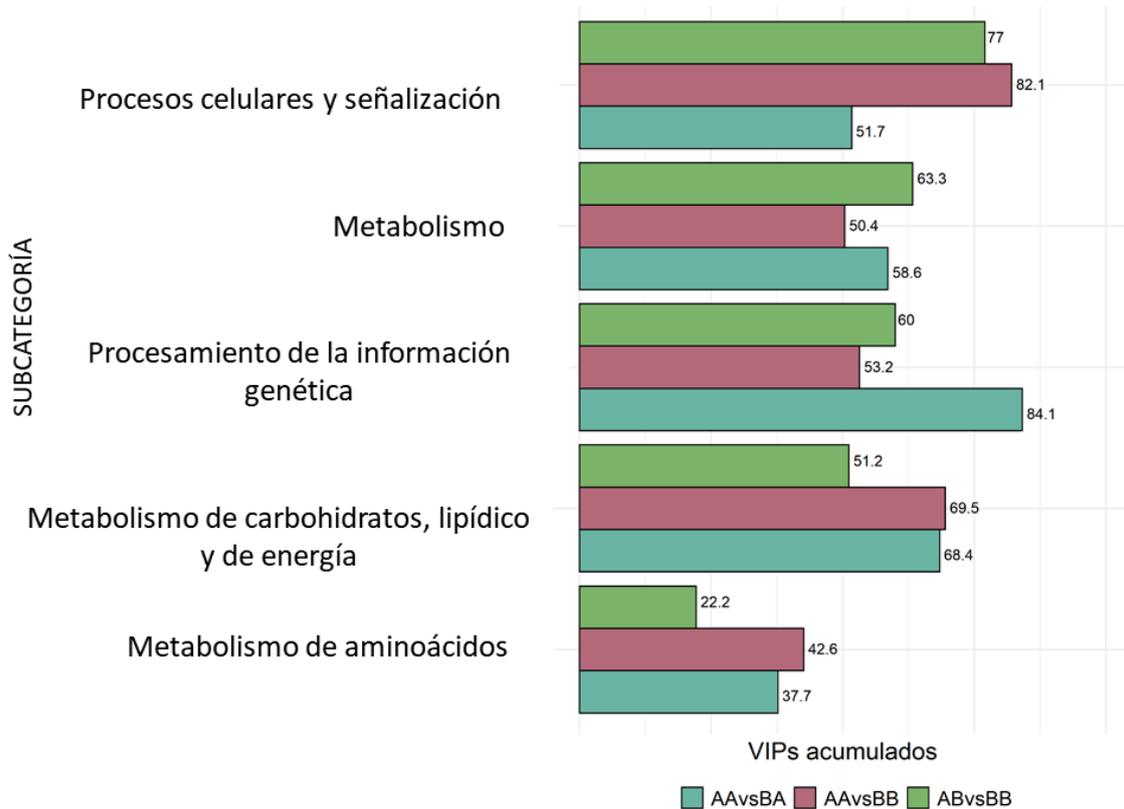


Figura 12. Diagrama de barras de las cinco subcategorías con un mayor VIP acumulado y su distribución en las tres comparaciones. AA vs BA: Efecto materno cuando el individuo es de la línea A. AB vs BB: Efecto materno cuando el individuo es de la línea B. AA vs BB: Efecto materno-genético. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja.

Las variables KEGG que pertenecen a la subcategoría “Metabolismo” son variables que no están clasificadas en una ruta metabólica concreta. Las variables KEGG agrupadas en las categorías de procesos celulares y señalización y procesamiento de la información genética participan en procesos de transmisión de señales como K07670 para el efecto materno cuando el individuo es de la línea B o de la reparación del ADN como K15342 para el efecto materno cuando el individuo es de la línea A o K04067 para el efecto materno-genético.

Entre las 20 variables KEGG más importantes que realizan funciones de metabolismo de carbohidratos, lipídico y de energía destacan para el efecto materno cuando el individuo es de la línea A: K12308 y K04072. La primera de ellas es una beta-galactosidasa que participa en el metabolismo de la galactosa y la segunda de ellas es un alcohol deshidrogenasa que participa en procesos importantes como la glucólisis y la degradación de los ácidos grasos. En el efecto materno cuando el individuo es de la línea B las dos variables KEGG más importantes que pertenecen a esta subcategoría son: K01201 que se encarga de la degradación de glicanos y K00284 que participan en el metabolismo del glioxilato y en el metabolismo de nitrógeno como fuente de energía.

Por último, para el efecto materno-genético destacan K00848 que se encarga del metabolismo de la fructosa y manosa y K07029 del metabolismo de glicerolípidos.

Dentro de las funciones de metabolismo de aminoácidos destacan en todas las comparaciones las variables KEGG relacionadas con el metabolismo de diferentes aminoácidos como histidina (K01755), lisina (K01929), asparagina (K01953) o K11358 que participa en la biosíntesis de arginina y en el metabolismo de cisteína, metionina, arginina y fenilalanina entre otros. Además de participar en el metabolismo de aminoácidos otras variables como K01666, participa también en la degradación y metabolismo de xenobióticos.

4.4.2 Anotación taxonómica

Los resultados del PLS-DA para la base de datos de anotación taxonómica de género se muestran en la Tabla 5. Los valores de R^2 y Q^2 fueron ligeramente más bajos para todas las comparaciones comparado con los resultados del PLS-DA para la anotación funcional, pero se observó una clara separación entre los grupos. A modo de ejemplo se puede observar en la Figura 13 el gráfico de *scores* de la comparación materno-genética (AA vs BB) para la base de datos de anotación taxonómica.

Tabla 5. Resultados de la proyección de estructuras latentes discriminantes (PLS-DA) para la anotación taxonómica de género. Número de variables después del filtrado (variables), número de variables que tienen un VIP (*Variable Importance in Projection*) mayor que 1 después del primer PLS-DA (Variables VIP >1), la bondad de ajuste (R^2) y la bondad de predicción (Q^2) asociado al PLS-DA con las variables VIP>1 últimas.

	AA vs BA	AB vs BB	AA vs BB
Variables	234	234	234
Variables VIP>1	77	86	82
R^2	0,756	0,911	0,916
Q^2	0,468	0,788	0,766

AA: individuos de la línea alta y madre de la línea alta, BA: individuos de la línea alta y madre de la línea baja, AB: individuos de la línea baja y madre de la línea alta y BB: individuos de la línea baja y madre de la línea baja. AA vs BA: Efecto materno cuando el individuo es de la línea A. AB vs BB: Efecto materno cuando el individuo es de la línea. AA vs BB: Efecto materno-genético.

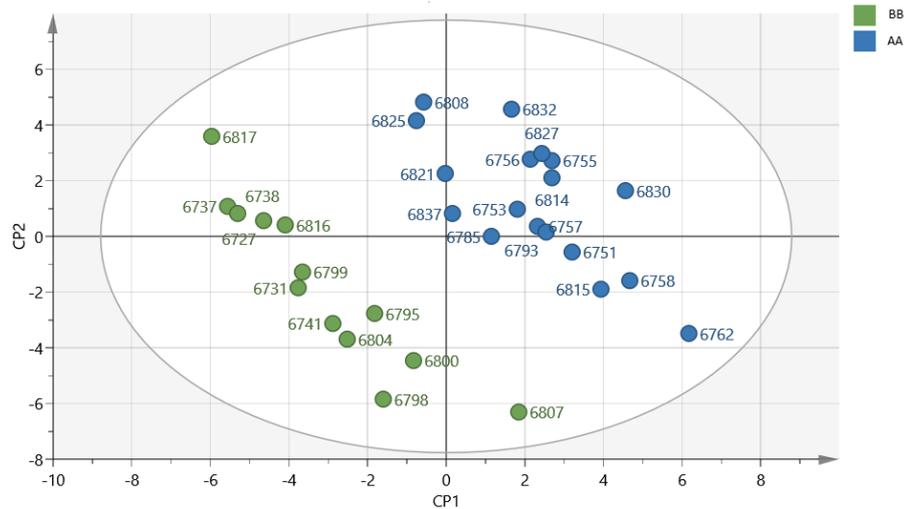


Figura 13. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal para la comparación del efecto materno-genético (AA vs BB) para la base de datos de anotación taxonómica para género. Cada punto corresponde a una muestra. CP: Componente principal. AA individuos y madre de línea alta y BB individuos y madre de la línea baja.

Así como en el análisis con KEGG, un número similar de géneros se identificaron responsables de la discriminación entre grupos en cada una de las comparaciones. No obstante, estos géneros fueron distintos entre las comparaciones. De las variables género con $VIP > 1$, solo el 8% fueron las mismas en ambas comparaciones del efecto materno y únicamente el 5% de las variables género coincidieron en las tres comparaciones. Sin embargo, al igual que en la base de datos de anotación funcional, se observaron similitudes en los niveles taxonómicos superiores a los que pertenecían estos géneros: filo, clase y familia. El filo más predominante para las tres comparaciones (AA vs BA, AB vs BB y AA vs BB) fue Firmicutes (presente en un porcentaje del 62% para efecto materno cuando el individuo es de la línea A, del 76% para el efecto materno cuando el individuo es de la línea B y del 70% para el efecto materno-genético). El segundo filo más representado, también para las tres comparaciones, fue el filo Proteobacteria con un 10% para el efecto cuando el individuo es de la línea A y un 14% para el efecto materno cuando el individuo es de la línea B y el efecto materno-genético. Por último, el tercer filo más representado fue Bacteroidetes con un 10% para el efecto materno cuando el individuo es de la línea A y un 5% para el efecto materno cuando el individuo es de la línea B y el efecto materno-genético. Dentro del filo Firmicutes, la clase taxonómica más representada para las tres comparaciones fue Clostridia. Sin embargo, hubo diferencias en abundancias relativas entre las comparaciones. La comparación del efecto materno cuando el individuo es de la línea B y del efecto materno-genético presentó alrededor del 60% de los géneros asociados a la clase Clostridia, mientras que la comparación del efecto materno cuando el individuo es de la línea A presentó una abundancia del 37%. En esta última comparación aparecían otras clases que no estaban representadas en las otras comparaciones como Alphaproteobacteria y Tissierellia. No obstante, estas últimas taxas se encontraron representadas en baja proporciones (5% y 8% respectivamente) y por lo tanto su importancia en cuanto al VIP era también baja, por lo que el resultado era el mismo en todas las comparaciones. Por último, a nivel de familia el comportamiento fue similar que a nivel de clase. En la Figura 14 se muestran las primeras familias más importantes con respecto al VIP y como se distribuyeron en cada una de las comparaciones. El taxón más representado e importante fue Ruminococcae para las tres comparaciones. Su

importancia (VIP acumulado) fue de 12, 12,1 y 6,1 para la comparación del efecto materno cuando el individuo es de la línea B, para la comparación materno-genético y para la comparación del efecto materno cuando el individuo es de la línea A respectivamente. Estas diferencias se deben a que en el efecto materno cuando el individuo fue de la línea A aparecían otras familias, en baja proporción, que no aparecían en las otras dos comparaciones, como Peptoniphilaceae y Muribaculaceae. Además de Ruminococcae, la familia Lachnospiraceae fue la segunda más importante para las tres comparaciones. El resto de familias fueron de baja importancia y se encontraron en baja proporción entre las comparaciones.

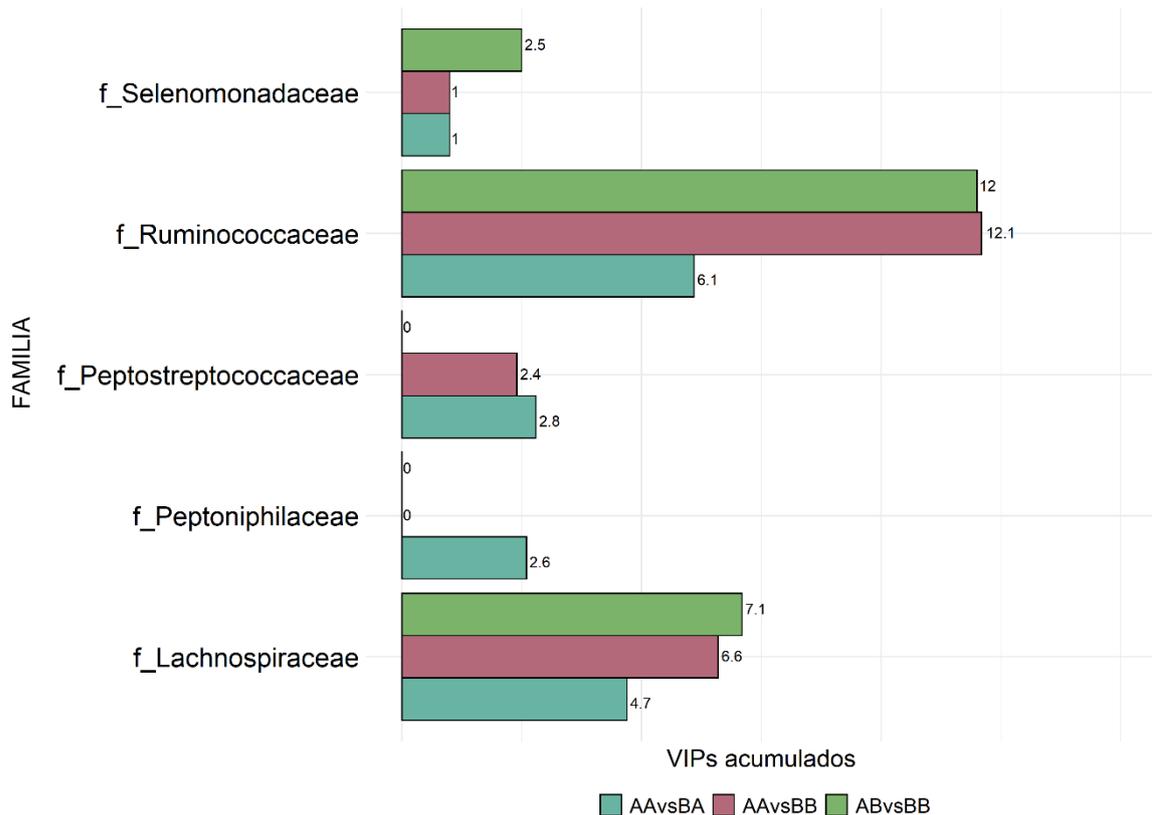


Figura 14. Diagrama de barras de las siete familias con un mayor VIP acumulado y su distribución en las tres comparaciones. AA vs BA: Efecto materno cuando el individuo es de la línea A. AB vs BB: Efecto materno cuando el individuo es de la línea B. AA vs BB: Efecto materno-genético. AA: individuos de la línea alta y madre de la línea alta. BA: individuos de la línea alta y madre de la línea baja. AB: individuos de la línea baja y madre de la línea alta. BB: individuos de la línea baja y madre de la línea baja

5 DISCUSIÓN

Numerosos trabajos se han desarrollado a lo largo de los años para conocer la composición de la microbiota gastrointestinal de diferentes animales de producción como las vacas (Mao et al., 2015), cerdos (Crespo-piazuelo et al., 2018) y conejos (Velasco-galilea et al., 2018) entre otros. La mayoría de estos estudios están enfocados en estudiar la composición microbiana a lo largo del tracto digestivo para conocer el papel que desempeña. Sin embargo, algunos de ellos, también evidencian la importancia de factores externos como la nutrición en la composición del microbioma (Senghor et al., 2018) o de factores internos como la influencia de la genética del hospedador (Chen et al., 2018; Li et al., 2019). En este trabajo, dimos un paso más e investigamos el efecto materno en el microbioma gastrointestinal del conejo, no estudiado hasta ahora. Específicamente, lo hicimos en líneas de conejo seleccionadas divergentemente por GIM, en las que se secuenció el metagenoma de individuos de la línea alta (A) y de la línea baja (B) provenientes de madres receptoras de la misma o diferente línea que ellos. De esta forma, a través de las comparaciones entre grupos, se pudo estudiar el efecto materno cuando el individuo era de la línea A (AA vs BA), el efecto materno cuando el individuo era de la línea B (AB vs BB), el efecto materno-genético (AA vs BB) y observar si el efecto era el mismo en las tres comparaciones.

Un paso clave después del análisis bioinformático de los resultados de secuenciación es el filtrado de variables. De esta forma, se eliminan aquellas variables que tienen un porcentaje de ceros alto y que impiden la identificación de diferencias reales entre grupos. Además, la imputación de ceros es necesaria para aplicar las transformaciones logarítmicas que tratan los datos composicionales, pero no se puede realizar si el porcentaje de ceros en la matriz es elevado. La elección del filtrado óptimo es complicada debido a que, dependiendo del filtrado, los resultados pueden variar (Kaul et al., 2017). Por ello, en este estudio probamos con diferentes intervalos de prevalencia (10%, 40% y 70%). Los resultados mostraron que, aunque el filtrado del 10% y del 40% de prevalencia dieron resultados ligeramente superiores en términos de capacidad predictiva (Q^2) para las comparaciones AB vs BB y AA vs BB, con estos filtrados se seleccionaron variables cuyos datos superaron el 50% de ceros dentro de cada comparación. Estas variables, dada la gran cantidad de ceros dentro de grupo (por ejemplo, en el grupo AA 18 individuos con ceros y 0 con counts y en el grupo BB 10 individuos con ceros y 3 con counts), no las consideramos como claras responsables de la discriminación entre grupos. Por tanto, se consideró como mejor intervalo de prevalencia el del 70%.

Las funciones KEGG y los taxones responsables de la discriminación entre grupos en cada una de las comparaciones se identificaron a través de técnicas de análisis multivariantes. Estas técnicas permiten la reducción de dimensionalidad de las grandes bases de datos generadas en el análisis bioinformático y facilitan el análisis y la interpretación de la gran cantidad de variables generadas. En este trabajo se utilizó la técnica multivariante de proyección de estructuras latentes discriminantes (PLS-DA) para buscar la dirección de máxima discriminación entre dos grupos. Los resultados de estos análisis mostraron una clara separación entre grupos de efecto materno y materno-genético con valores de $R^2 > 0,8$ y $Q^2 > 0,7$ en la mayoría de los casos. Estos resultados, sugerirían la existencia de un efecto materno en el microbioma gastrointestinal en conejo y estarían de acuerdo con estudios previos en humanos donde se observaron diferencias en la microbiota de recién nacidos en función de la

forma de parto (Dominguez-bello et al., 2010; McDonald and McCoy, 2019) o la lactancia (Elsen et al., 2019).

En la mayoría de estudios de microbioma se secuencian el ADN Ribosomal 16S (Osman et al., 2018) y por tanto solo se obtiene la anotación taxonómica. En este trabajo se realizó la secuenciación completa, lo que permitió realizar los análisis de discriminación tanto para la base de datos de anotación funcional de KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (Kanehisa and Goto, 2000) como para la base de datos de anotación taxonómica (Genbank) (Sayers et al., 2018). La utilización de las dos anotaciones (KEGG y Genbank) nos permitió conocer y estudiar los microorganismos y las funciones responsables del efecto materno.

A nivel funcional, las variables KEGG responsables de la discriminación entre grupos fueron distintas cuando el individuo era de la línea A, de cuando era de la línea B y de cuando la madre y el individuo eran de la misma línea. No obstante, se detectaron subcategorías comunes en las tres comparaciones siendo las más importantes: procesos celulares y señalización, metabolismo, procesamiento de la información genética, metabolismo de carbohidratos, lipídico y de energía y metabolismo de aminoácidos. De la misma manera, a nivel taxonómico de género tampoco se observaron similitudes entre los grupos de comparación, pero sí se mostraron taxones enriquecidos en las tres comparaciones para los niveles taxonómicos superiores a los que pertenecían estos géneros: filo, clase y familia. Esto podría indicar que existe un efecto materno y que aunque el KEGG o el género difieren dependiendo de la genética del individuo (línea alta o línea baja) la función que este KEGG o género realizan es similar.

A nivel taxonómico de filo, no es sorprendente que la mayoría de géneros responsables de la discriminación entre grupos perteneciesen al filo Firmicutes, ya que el filo Firmicutes es el más abundante en conejo y en otras especies de lagomorfos (Li et al., 2017). Además de en conejo, el filo Firmicutes ha sido descrito como el más abundante en la microbiota gastrointestinal de la mayoría de los mamíferos adultos sanos incluidos los humanos (Combes et al., 2011; Davenport et al., 2014), en los que se ha relacionado con la salud gastrointestinal (Kylie et al., 2018). En nuestras muestras, alrededor del 75% de las bacterias pertenecieron a este filo abarcando una gran cantidad de grupos taxonómicos inferiores.

Dentro del filo Firmicutes, la clase identificada más abundante en las tres comparaciones fue Clostridia. Las bacterias pertenecientes a esta clase son anaerobias gram positivas formadoras de esporas. La capacidad de formar esporas las hace altamente estables en el medio ambiente (Yutin and Galperin, 2014). Dentro de esta clase las familias más abundantes fueron Ruminococcaceae y Lachnospiraceae. Estas familias participan en el metabolismo de polisacáridos y aminoácidos no digeribles en la dieta (Biddle et al., 2013) y son vitales para la fermentación de las fibras (Ze et al., 2015).

Las muestras de microbioma de este trabajo provienen del ciego que es el principal órgano fermentador del tubo digestivo en conejos (Abdel-Khhalek et al., 2011). El proceso de fermentación empieza con una conversión de los carbohidratos (glucosa, manosa, fructosa...) a piruvato, el producto final de la primera etapa de la fermentación (glicolisis). El siguiente paso consiste en la formación de acetil-coA, el cual se puede obtener a través del complejo piruvato-deshidrogenasa o del metabolismo de los ácidos-grasos. El acetil-coA entra en el ciclo del ácido cítrico donde a través de diversas

reacciones es catalizado por bacterias fermentativas a alcoholes y ácidos grasos de cadena corta, teniendo como resultado de este proceso, hidrógeno y dióxido de carbono. Por tanto, la presencia de estas familias (Ruminococcaceae y Lachnospiraceae) llevaría a cabo la descomposición de los sustratos que el huésped no puede digerir fermentándolos en ácidos grasos de cadena corta que el huésped puede absorber y utilizar para obtener energía. Esto podría explicar la presencia de la subcategoría de metabolismo de carbohidratos, lípidos y de energía como una de las más representadas en las tres comparaciones para la base de datos de anotación funcional. Además, el ciego se encarga de la síntesis de vitaminas y de aminoácidos esenciales (Cummlings and Macfarlane, 1997) lo que podría explicar la presencia de la subcategoría “metabolismo de aminoácidos” como otra de las más representadas en las tres comparaciones. Estudios previos en porcino, encontraron que la familia Ruminococcaceae estaba positivamente asociadas con el peso a los 60 días de edad y la ganancia media diaria (Ramayo-caldas et al., 2016) y que géneros no identificados de la familia Ruminococcaceae se encontraron en más abundancia en lechones lactantes que no lactantes y que los lechones lactantes presentaron mayores tasas de crecimiento durante la lactación (Mach et al., 2015). Además, Fang et al. (2017) encontraron que la presencia de estas subcategorías en el ciego (Lachnospiraceae) y en las heces (Ruminococcaceae) del cerdo asociadas a funciones relacionadas con la fermentación y a caracteres como la eficiencia alimentaria. Estos estudios reforzarían los resultados obtenidos en nuestro estudio, ya que, aunque en nuestro trabajo solo se utilizaron muestras de ciego, ambas familias están presentes tanto en las muestras de heces como de ciego en conejo (Velasco-galilea et al., 2018).

Dentro de la familia Ruminococcaceae, la familia más abundante, encontramos el género *Faecalibacterium* como el único género en común en la comparación del efecto materno cuando el individuo fue de la línea A y cuando el individuo fue de la línea B. Este género podría ser utilizado como un indicador de la salud intestinal en humanos al ser el más abundante en las heces de individuos sanos (Laurson et al., 2017). Este tipo de bacterias se ha encontrado también en el ciego de terneros (Malmuthuge et al., 2014) relacionado con el mantenimiento del peso corporal adecuado y reduciendo las infecciones entéricas en estadios tempranos. Dentro de la segunda familia más abundante, Lachnospiraceae, encontramos entre otros el género *Coprococcus*, en la comparación del efecto materno cuando el individuo fue de la línea B y en la comparación del efecto materno-genético. *Coprococcus* es una bacteria grampositiva anaeróbica que fermenta activamente los carbohidratos, produciendo butírico (Holdeman and Moore, 1974), un ácido graso de cadena corta que fortalece la barrera de defensa epitelial y reduce la inflamación intestinal (Louis et al., 2014). Por lo tanto, el butirato desempeña un papel clave en el mantenimiento de la salud intestinal y es una fuente principal de energía para la mucosa intestinal (Louis and Flint, 2009).

La presencia del filo Proteobacteria, el segundo filo más representado en las tres comparaciones, ha sido descrito como un posible signo de disbiosis en la microbiota intestinal (Shin et al., 2015). No obstante, no todas las Proteobacterias son patogénicas y el intestino sano de los mamíferos contiene varios miembros de especies bacterianas comensales que pertenecen a este filo como su flora intestinal natural (Caporaso et al., 2011). Además, en este estudio la presencia de este filo no fue muy abundante por lo que no se puede afirmar que en este caso esté relacionado con un desequilibrio de la microbiota en el intestino.

En cuanto al tercer filo más representado, Bacteroidetes, es un filo propio de roedores (Li et al., 2017). No obstante, lo encontramos en baja abundancia en las muestras del ciego, este tipo de bacterias se caracterizan por ser gramnegativas que fermentan carbohidratos no digeribles y producen ácidos grasos de cadena corta que tienen muchos efectos beneficiosos en el intestino (Birg et al., 2019).

Por último, además de las subcategorías de “metabolismo de carbohidratos, lípidos y de energía” y “metabolismo de aminoácidos” que se han nombrado anteriormente, las variables KEGG relacionadas con los procesos celulares y señalización y procesamiento de la información genética también estuvieron enriquecidas en las tres comparaciones. La presencia de estas subcategorías, aunque no esté relacionada con la grasa intramuscular, no es sorprendente ya que a esta subcategoría pertenecen variables relacionadas con funciones propias de las células de división celular, transcripción, reparación del ADN, entre otras.

Este es el primer estudio que aborda la influencia del efecto materno en el microbioma gastrointestinal del conejo. La creación de dos líneas de conejos seleccionadas divergentemente para el carácter de grasa intramuscular durante 10 generaciones nos ha permitido disponer de un material genético único para estudiar el efecto materno en el microbioma. Este efecto materno se pudo estudiar gracias a la posibilidad de realizar transferencias embrionarias aleatorias entre las dos líneas divergentes de la décima generación y la posterior secuenciación del microbioma gastrointestinal de estos individuos. Los resultados mostraron funciones y taxones responsables de la discriminación entre grupos donde está separado el efecto materno del genético. Esto podría explicar diferencias en la grasa intramuscular en función de la línea genética de la madre o del huésped. Conocer el efecto que tiene la madre en la microbiota podría ayudar en un futuro a modificar la microbiota para producir beneficios tanto en la salud como en los caracteres productivos de interés. Sin embargo, más estudios son necesarios para entender mejor la influencia del efecto materno y el genético en el microbioma y cómo aplicar estos conocimientos para modificar el microbioma de los individuos.

6 CONCLUSIÓN

- El filtrado de variables previo a la normalización de los datos y a los análisis estadísticos es crucial para una correcta selección de variables responsables de la discriminación entre grupos.
- El análisis estadístico con técnicas multivariantes ha permitido identificar variables que discriminen entre grupos de efecto materno y efecto materno-genético.
- Los KEGG y taxa género responsables de las discriminaciones entre grupos difieren dependiendo de la comparación (efecto materno cuando el individuo es de la línea A o B), pero no difieren las funciones que realizan.
- Las funciones y taxones responsables del efecto materno en las dos comparaciones (efecto materno cuando el individuo fue de la línea A, efecto materno cuando el individuo fue de la línea B) y en la comparación del efecto materno-genético estaban relacionadas con los procesos de fermentación que tienen lugar en el ciego del intestino del conejo.
- Las funciones más importantes responsables de la separación entre grupos fueron metabolismo de carbohidratos, lipídico y de energía y metabolismo de aminoácidos.
- Las familias más importantes responsables de la separación entre grupos fueron Ruminococcaceae y Lachnospiraceae.

7 BIBLIOGRAFÍA

- Abdel-Khhalek, A. ., Kalaba, Z. ., and El-Gogary, M. . (2011). Functional, Anatomical and Histological Development of Rabbits. doi:10.3923/crps.2011.54.65.
- Aitchison, J. (1982). The Statistical Analysis of Compositional Data. 139–160. doi:10.1111/j.2517-6161.1982.tb01195.x.
- Amann, R. I., Ludwig, W., and Schleifer, K. (1995). Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. 59, 143–169.
- Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data.
- Arumugam, M., Harrington, E. D., Foerstner, K. U., Raes, J., and Bork, P. (2010). SmashCommunity: a metagenomic annotation and analysis tool. 26, 2977–2978. doi:10.1093/bioinformatics/btq536.
- Awany, D., Allali, I., Dalvie, S., Hemmings, S., Mwaikono, K. S., and Pulit, S. L. (2019). Host and Microbiome Genome-Wide Association Studies: Current State and Challenges. 9, 1–12. doi:10.3389/fgene.2018.00637.
- Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Chari, R. S., et al. (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Cmaj* 185, 373–374. doi:10.1503/cmaj.130147.
- Bäumler, A. J., and Sperandio, V. (2016). Interactions between the microbiota and pathogenic bacteria in the gut. 535, 85–93. doi:10.1038/nature18849.Interactions.
- Biddle, A., Stewart, L., Blanchard, J., and Leschine, S. (2013). Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. 627–640. doi:10.3390/d5030627.
- Birg, A., Ritz, N. L., and Lin, H. C. (2019). *The Unknown Effect of Antibiotic- Induced Dysbiosis on the Gut Microbiota*. Elsevier Inc. doi:10.1016/B978-0-12-815249-2.00020-8.
- Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.*, 1–12. doi:10.1186/s13059-015-0759-1.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. 30, 2114–2120. doi:10.1093/bioinformatics/btu170.
- Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., et al. (2016). The effect of host genetics on the gut microbiome. 48. doi:10.1038/ng.3663.
- Breitwieser, F. P., Lu, J., and Salzberg, S. L. (2017). A review of methods and databases for metagenomic classification and assembly. 20, 1125–1139. doi:10.1093/bib/bbx120.
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176.

- Caballero, S., and Pamer, E. G. (2015). Microbiota-Mediated Inflammation and Antimicrobial Defense in the Intestine. 227–256. doi:10.1146/annurev-immunol-032713-120238. Microbiota-Mediated.
- Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-lyons, D., Gonzalez, A., Stombaugh, J., et al. (2011). Moving pictures of the human microbiome.
- Carr, R., and Borenstein, E. (2014). Comparative Analysis of Functional Metagenomic Annotation and the Mappability of Short Reads. 9. doi:10.1371/journal.pone.0105776.
- Castaner, O., Goday, A., Park, Y., Lee, S., Magkos, F., Shio, S. T. E., et al. (2018). The Gut Microbiome Profile in Obesity: A Systematic Review. 2018. doi:10.1155/2018/4095789.
- Chen, C., Huang, X., Fang, S., Yang, H., and He, M. (2018). Contribution of Host Genetics to the Variation of Microbial Composition of Cecum Lumen and Feces in Pigs. 9, 1–13. doi:10.3389/fmicb.2018.02626.
- Cheng, G., Hao, H., Xie, S., Wang, X., Dai, M., and Huang, L. (2014). Antibiotic alternatives: the substitution of antibiotics in animal husbandry? 5, 1–15. doi:10.3389/fmicb.2014.00217.
- Chong, I. G., and Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* 78, 103–112. doi:10.1016/j.chemolab.2004.12.011.
- Combes, S., Michelland, R. J., and Cauquil, L. (2011). Postnatal development of the rabbit caecal microbiota composition and activity. 77, 680–689. doi:10.1111/j.1574-6941.2011.01148.x.
- Crespo-piazuelo, D., Estellé, J., Revilla, M., M, L. C.-, Ramayo-caldas, Y., Óvilo, C., et al. (2018). Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. 1–12. doi:10.1038/s41598-018-30932-6.
- Csige, I., Ujvárosy, D., Szabó, Z., Lorincz, I., Paragh, G., Harangi, M., et al. (2018). The Impact of Obesity on the Cardiovascular System. *J. Diabetes Res.* 2018. doi:10.1155/2018/3407306.
- Cummings, J. H., and Macfarlane, G. T. (1997). Collaborative JPEN-Clinical Nutrition Scientific Publications Role of intestinal bacteria in nutrient metabolism *. 357–365. doi:10.1177/0148607197021006357.
- Danzeisen, J. L., Clayton, J. B., Huang, H., Knights, D., McComb, B., Hayer, S. S., et al. (2015). Temporal relationships exist between cecum, ileum, and litter bacterial microbiomes in a commercial Turkey flock, and subtherapeutic penicillin treatment impacts ileum bacterial community establishment. *Front. Vet. Sci.* 2. doi:10.3389/fvets.2015.00056.
- Davenport, E. R., Mizrahi-man, O., Michelini, K., Barreiro, L. B., Ober, C., and Gilad, Y. (2014). Seasonal Variation in Human Gut Microbiome Composition. 9. doi:10.1371/journal.pone.0090731.
- Dominguez-bello, M. G., Costello, E. K., Contreras, M., Magris, M., and Hidalgo, G. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota

- across multiple body habitats in newborns. 107, 11971–11975. doi:10.1073/pnas.1002601107.
- Dunn, A. B., Jordan, S., Baker, B. J., and Carlson, N. S. (2017). The Maternal Infant Microbiome: Considerations for Labor and Birth. 1–7.
- Elsen, L. W. J. Van Den, Garssen, J., Burcelin, R., and Verhasselt, V. (2019). Shaping the Gut Microbiota by Breastfeeding : The Gateway to Allergy Prevention ? 7, 1–10. doi:10.3389/fped.2019.00047.
- Eren, A. M., Esen, C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi ' o : an advanced analysis and visualization platform for ' omics data. 1–29. doi:10.7717/peerj.1319.
- Fang, S., Xiong, X., Su, Y., Huang, L., and Chen, C. (2017). 16S rRNA gene-based association study identified microbial taxa associated with pork intramuscular fat content in feces and cecum lumen. 1–9. doi:10.1186/s12866-017-1055-x.
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, Y., Eddy, S. R., et al. (2014). Pfam : the protein families database. 42, 222–230. doi:10.1093/nar/gkt1223.
- Fleming, T., Ng, M., Robinson, M., Thomson, B., Graetz, N., Margono, C., et al. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 384, 766–781. doi:10.1016/S0140-6736(14)60460-8.
- Floch, M. H. (2014). Probiotics and Prebiotics. *Nutrition* 10, 680–681.
- Galindo-Prieto, B., Eriksson, L., and Trygg, J. (2014). Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *J. Chemom.* 28, 623–632. doi:10.1002/cem.2627.
- Gevers, D., Pop, M., Schloss, P. D., and Huttenhower, C. (2012). Bioinformatics for the Human Microbiome Project. 8. doi:10.1371/journal.pcbi.1002779.
- 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, 1–6. doi:10.3389/fmicb.2017.02224.
- Greenacre, M. (2019). *Compositional Data Analysis in Practice*. CRC press.
- Handelsman, J. (2004). Metagenomics : Application of Genomics to Uncultured Microorganisms. 68, 669–685. doi:10.1128/MBR.68.4.669.
- Hocquette, J. F., Gondret, F., Baza, E., Mdale, F., Jurie, C., and Pethick, D. W. (2010). Intramuscular fat content in meat-producing animals: Development, genetic and nutritional control, and identification of putative markers. *Animal* 4, 303–319. doi:10.1017/S1751731109991091.
- Holdeman, L. V., and Moore, W. E. . (1974). New Genus, Coprococcus, Twelve New Species, and Emended Descriptions of Four Previously Described Species of Bact. *Int. J. Syst. Bacteriol.* 24, 260–277. doi:10.1099/00207713-24-2-260.
- Hyatt, D., Chen, G., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal : prokaryotic gene recognition and translation initiation site identification.

- Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., et al. (2014). Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section. *Gut* 63, 559–566. doi:10.1136/gutjnl-2012-303249.
- Jovel, J., Patterson, J., Wang, W., Hotte, N., Keefe, S. O., Mitchel, T., et al. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. 7, 1–17. doi:10.3389/fmicb.2016.00459.
- Kamada, N., Seo, S. U., Chen, G. Y., and Núñez, G. (2013). Role of the gut microbiota in immunity and inflammatory disease. *Nat. Rev. Immunol.* 13, 321–335. doi:10.1038/nri3430.
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi:10.1093/nar/27.1.29.
- Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. 30, 42–46.
- Kanehisa, M., Goto, S., Kawashima, S., and Okuno, Y. (2004). The KEGG resource for deciphering the genome. 32. doi:10.1093/nar/gkh063.
- Kaul, A., Mandal, S., Davidov, O., and Peddada, S. D. (2017). Analysis of microbiome data in the presence of excess zeros. *Front. Microbiol.* 8, 1–10. doi:10.3389/fmicb.2017.02114.
- Kim, G., and Kim, H. (2017). Effects of Carcass Weight and Back-fat Thickness on Carcass Properties of Korean Native Pigs. 385–391.
- Kim, J., Kim, M. S., Koh, A. Y., Xie, Y., and Zhan, X. (2016). FMAP : Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinformatics*, 1–8. doi:10.1186/s12859-016-1278-0.
- Kim, M., Park, T., Jeong, J. Y., Baek, Y., and Lee, H. Association between Rumen Microbiota and Marbling Score in Korean Native Beef Cattle. 1–13.
- Kirby, T. O., and Ochoa-rep, J. (2018). The Gut Microbiome in Multiple Sclerosis : A Potential Therapeutic Avenue. doi:10.3390/medsci6030069.
- Kröber, M., Bekel, T., Diaz, N. N., Goesmann, A., Jaenicke, S., Krause, L., et al. (2009). Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. 142, 38–49. doi:10.1016/j.jbiotec.2009.02.010.
- Kumar, A., and Chordia, N. (2017). Role of Microbes in Human Health Applied Microbiology : Open Access. 3, 2–4. doi:10.4172/2471-9315.1000131.
- Kylie, J., Weese, J. S., and Turner, P. V (2018). Comparison of the fecal microbiota of domestic commercial meat , laboratory , companion , and shelter rabbits (*Oryctolagus cuniculi*). 1–15.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923.
- Laursen, M. F., Laursen, P., Larnkjær, A., Mølgaard, C., Michaelsen, K. F., Frøkiær, H., et al. (2017). Faecalibacterium Gut Colonization Is Accelerated by Presence of

Older Siblings. 2, 2–7.

- Lee, L. C., Liong, C. Y., and Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps. *Analyst* 143, 3526–3539. doi:10.1039/c8an00599k.
- Li, D., Liu, C., Luo, R., Sadakane, K., and Lam, T. (2015). MEGAHIT : an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. 31, 1674–1676. doi:10.1093/bioinformatics/btv033.
- Li, F., Li, C., Chen, Y., Liu, J., Zhang, C., Irving, B., et al. (2019). Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. 1–17.
- Li, H., Qu, J., Li, T., Yao, M., Li, J., and Li, X. (2017). Gut microbiota may predict host divergence time during Glires evolution. *FEMS Microbiol. Ecol.* 93, 1–9. doi:10.1093/femsec/fix009.
- Littman, D. R., and Pamer, E. G. (2012). Role of the commensal microbiota in normal and pathogenic host immune responses. 10, 311–323. doi:10.1016/j.chom.2011.10.004.Role.
- Louis, P., and Flint, H. J. (2009). Diversity , metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. doi:10.1111/j.1574-6968.2009.01514.x.
- Louis, P., Hold, G. L., and Flint, H. J. (2014). The gut microbiota , bacterial metabolites and colorectal cancer. *Nat. Publ. Gr.* doi:10.1038/nrmicro3344.
- Ma, J., Prince, A., and Aagaard, K. M. (2014). Use of Whole Genome Shotgun Metagenomics : A Practical Guide for the Microbiome-Minded Physician Scientist. 1.
- 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.
- Malmuthuge, N., Griebel, P. J., and Guan, L. L. (2014). Taxonomic Identification of Commensal Bacteria Associated with the Mucosa and Digesta throughout the Gastrointestinal Tracts of Preweaned Calves. 80, 2021–2028. doi:10.1128/AEM.03864-13.
- Mao, S., Zhang, M., Liu, J., and Zhu, W. (2015). Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle : membership and potential function. *Nat. Publ. Gr.*, 1–14. doi:10.1038/srep16116.
- Mapara, M., Thomas, B., and Bhat, K. . (2012). Rabbit as an animal model for experimental research e. *Dent. Res. J. (Isfahan)*.
- Marchesi, J. R., and Ravel, J. (2015). The vocabulary of microbiome research : a proposal. *Microbiome*, 1–3. doi:10.1186/s40168-015-0094-5.
- Martín-Fernández, J. A., Hron, K., Templ, M., Filzmoser, P., and Palarea-Albaladejo, J. (2015). Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat. Modelling* 15, 134–158. doi:10.1177/1471082X14535524.

- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet*, 5–7.
- Martinez-Álvarez, M., Hernández, P., and Blasco, A. (2016). Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters. doi:10.2527/jas.2016-0590.
- Martinez Álvarez, M., Hernández, P., and Blasco, A. (2016). Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters 1. doi:10.2527/jas2016-0590.
- McDonald, B., and McCoy, K. D. (2019). Maternal microbiota in pregnancy and early life. *Science (80-.)*. 365, 984–985. doi:10.1126/science.aay0618.
- Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. doi:10.1038/ncomms11257.
- Moisá, S. J., Shike, D. W., Faulkner, D. B., Meteer, W. T., Keisler, D., and Loo, J. J. (2014). Central Role of the PPAR γ Gene Network in Coordinating Beef Cattle Intramuscular Adipogenesis in Response to Weaning Age and Nutrition. *Gene Regul. Syst. Bio.*, 17–32. doi:10.4137/GRSB.S11782.Received.
- Muller, J., Szklarczyk, D., Julien, P., Letunic, I., Roth, A., Kuhn, M., et al. (2010). eggNOG v2 . 0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups , species and functional annotations. 38, 190–195. doi:10.1093/nar/gkp951.
- Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., et al. (2012). Host-Gut Microbiota Metabolic Interactions. *Science (80-.)*. 108, 1262–1268.
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes : a new versatile metagenomic assembler. 824–834. doi:10.1101/gr.213959.116.4.
- Osman, M., Neoh, H., Mutalib, N. A., Chin, S., and Jamal, R. (2018). 16S rRNA Gene Sequencing for Deciphering the Colorectal Cancer Gut Microbiome: Current Protocols and Workflows. 9. doi:10.3389/fmicb.2018.00767.
- Palarea-Albaladejo, J., and Martín-Fernández, J. A. (2015). ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. *Chemom. Intell. Lab. Syst.* 143, 85–96. doi:10.1016/j.chemolab.2015.02.019.
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools : a flexible suite of utilities for comparing genomic features. 26, 841–842. doi:10.1093/bioinformatics/btq033.
- Quinn, T. P., Erb, I., Richardson, M. F., and Crowley, T. M. (2018). Understanding sequencing data as compositions: an outlook and review. 34, 2870–2878. doi:10.1093/bioinformatics/bty175.
- 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. 2973–2977. doi:10.1038/ismej.2016.77.
- Rastelli, M., Knauf, C., and Cani, P. D. (2018). Gut Microbes and Health : A Focus on the Mechanisms Linking Microbes , Obesity , and Related Disorders Brief Overview

- of Data Suggesting Link. 26, 792–800. doi:10.1002/oby.22175.
- Ravin, N. V, Mardanov, A. V, and Skryabin, K. G. (2015). Metagenomics as a Tool for the Investigation. 51, 431–439. doi:10.1134/S1022795415050063.
- Reuter, J., Spacek, D., and Synder, M. (2015). High-Throughput Sequencing Technologies. 58, 586–597. doi:10.1016/j.molcel.2015.05.004.High-Throughput.
- Rho, M., Tang, H., and Ye, Y. (2014). FragGeneScan : Predicting Genes in Short and Error-prone Reads FragGeneScan: predicting genes in short and error-prone reads. doi:10.1093/nar/gkq747.
- Roesch, L. F. W., Dobbler, P. T., Pylro, V. S., Kolaczowski, B., Drew, J. C., and Triplett, E. W. (2019). pime: A package for discovery of novel differences among microbial communities. *Mol. Ecol. Resour.* doi:10.1111/1755-0998.13116.
- Sapp, R. L., Bertrand, J. K., Pringle, T. D., and Wilson, D. E. (2002). Effects of selection for ultrasound intramuscular fat percentage in Angus bulls on carcass traits of progeny. *J. Anim. Sci.* 80, 2017–2022. doi:10.2527/2002.8082017x.
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., and Karsch-Mizrachi, I. (2018). GenBank. *Nucleic Acids Res.* 47, D94–D99. doi:10.1093/nar/gky989.
- Schwab, C. R., Baas, T. J., Stalder, K. J., and Nettleton, D. (2009). Results from six generations of selection for intramuscular fat in Duroc swine using real-time ultrasound. I. Direct and correlated phenotypic responses to selection. *J. Anim. Sci.* 87, 2774–2780. doi:10.2527/jas.2008-1335.
- Selengut, J. D., Haft, D. H., Davidsen, T., Ganapathy, A., Gwinn-giglio, M., Nelson, W. C., et al. (2007). TIGRFAMs and Genome Properties : tools for the assignment of molecular function and biological process in prokaryotic genomes. doi:10.1093/nar/gkl1043.
- Senghor, B., Sokhna, C., Ruimy, R., and Lagier, J. (2018). Gut microbiota diversity according to dietary habits and geographical provenance. *Hum. Microbiome J.* 7–8, 1–9. doi:10.1016/j.humic.2018.01.001.
- Shahid, M., Rajoka, R., Shi, J., Mahreen, H., Zhu, J., and Li, Q. (2017). Interaction between diet composition and gut microbiota and its impact on gastrointestinal tract health. 6, 121–130. doi:10.1016/j.fshw.2017.07.003.
- Shin, N., Whon, T. W., and Bae, J. (2015). Proteobacteria : microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.*, 1–8. doi:10.1016/j.tibtech.2015.06.011.
- Singh, B., Crippen, T. L., Tomberlin, J. K., and Plains, S. (2017). An introduction to metagenomic data generation , analysis , visualization , and interpretation.
- Tamames, J., and Puente-sánchez, F. (2019). SqueezeMeta , A Highly Portable , Fully Automatic Metagenomic Analysis Pipeline. 9, 1–10. doi:10.3389/fmicb.2018.03349.
- Tan, Z., Wang, Y., Yang, T., Ao, H., Chen, S., Xing, K., et al. (2018). Differences in gut microbiota composition in finishing Landrace pigs with low and high feed conversion ratios. *Antonie Van Leeuwenhoek* 111, 1673–1685. doi:10.1007/s10482-018-1057-1.

- Tanaka, M., and Nakayama, J. (2017). Development of the gut microbiota in infancy and its impact on health in later life. *Allergol. Int.* 66, 515–522. doi:10.1016/j.alit.2017.07.010.
- Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* 2, 3. doi:10.1186/2042-5783-2-3.
- Treichel, N. S., Prevoršek, Z., Mrak, V., Kostrić, M., Vestergaard, G., Foesel, B., et al. (2019). Effect of the Nursing Mother on the Gut Microbiome of the Offspring During Early Mouse Development. *Microb. Ecol.* 78, 517–527. doi:10.1007/s00248-019-01317-7.
- Turner, P. V (2018). The role of the gut microbiota on animal model reproducibility. 109–115. doi:10.1002/ame2.12022.
- Uritskiy, G. V, Diruggiero, J., and Taylor, J. (2018). MetaWRAP — a flexible pipeline for genome-resolved metagenomic data analysis. 1–13.
- Vanderhaeghen, W., and Dewulf, J. (2017). Antimicrobial use and resistance in animals and human beings. *Lancet Planet. Heal.* 1, e307–e308. doi:10.1016/S2542-5196(17)30142-0.
- Varmuza, K., and Filzmoser, P. (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press.
- Velasco-galilea, M., Piles, M., Viñas, M., Rafel, O., Waite, D. W., and Mcewan, N. R. (2018). Rabbit Microbiota Changes Throughout the Intestinal Tract. 9, 1–14. doi:10.3389/fmicb.2018.02144.
- Vicente, J. S., and García-Ximénez, F. (1996). Direct transfer of vitrified rabbit embryos. *Theriogenology* 45, 811–815. doi:https://doi.org/10.1016/0093-691X(96)00010-6.
- Wang, H., Ji, Y., Yin, C., Deng, M., Tang, T., Deng, B., et al. (2018). Differential Analysis of Gut Microbiota Correlated With Oxidative Stress in Sows With High or Low Litter Performance During Lactation. 9, 1–11. doi:10.3389/fmicb.2018.01665.
- Wang, W., Zhang, P., and Liu, X. (2009). Short read DNA fragment anchoring algorithm. 11, 1–11. doi:10.1186/1471-2105-10-S1-S17.
- Wen, L., and Duffy, A. (2017). Factors Influencing the Gut Microbiota, Inflammation, and Type 2 Diabetes. *J. Nutr.* 147, 1468S–1475S. doi:10.3945/jn.116.240754.
- Wold, S., Sjostrom, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. 109–130.
- Wommack, K. E., Bhavsar, J., and Ravel, J. (2008). Metagenomics: Read Length Matters □ †. 74, 1453–1463. doi:10.1128/AEM.02181-07.
- Xu, Z., and Knight, R. (2015). Dietary effects on human gut microbiome diversity q. 1–5. doi:10.1017/S0007114514004127.
- Yan, Q., Gu, Y., Li, X., Yang, W., Jia, L., and Chen, C. (2016). Alterations of the Gut Microbiome in Hypertension. 7, 1–9. doi:10.3389/fcimb.2017.00381.
- Yang, X., Noyes, N. R., Doster, E., Martin, J. N., Linke, L. M., Magnuson, R. J., et al. (2016). Use of Metagenomic Shotgun Sequencing Technology To Detect

- Foodborne Pathogens within the Microbiome of the Beef Production. 82, 2433–2443. doi:10.1128/AEM.00078-16.Editor.
- Ye, Y., Choi, J., and Tang, H. (2011). RAPSearch : a fast protein similarity search tool for short reads. *BMC Bioinformatics* 12, 159. doi:10.1186/1471-2105-12-159.
- Yok, N. G., and Rosen, G. L. (2011). Combining gene prediction methods to improve metagenomic gene annotation.
- Yutin, N., and Galperin, M. (2014). A genomic update on clostridial phylogeny: Gram-negative spore-formers and other misplaced clostridia. 15, 2631–2641. doi:10.1111/1462-2920.12173.A.
- Ze, X., David, B., Laverde-gomez, J. A., Dassa, B., Sheridan, P. O., Duncan, S. H., et al. (2015). Unique Organization of Extracellular Amylases into Amylosomes in the Resistant Starch-Utilizing Human Colonic Firmicutes Bacterium *Ruminococcus bromii*. 6, 1–11. doi:10.1128/mBio.01058-15.Editor.
- Zhang, S., and Chen, D. (2019). Facing a new challenge: the adverse effects of antibiotics on gut microbiota and host immunity. 0, 1–2. doi:10.1097/CM9.0000000000000245.
- Zhou, Q., Su, X., and Ning, K. (2014). Assessment of quality control approaches for metagenomic data analysis. 1–11. doi:10.1038/srep06957.
- Zhu, B., Wang, X., and Li, L. (2010). Human gut microbiome : the second genome of human body. 1, 718–725. doi:10.1007/s13238-010-0093-z.
- Zomeño, C., Hernández, P., and Blasco, A. (2013). Divergent selection for intramuscular fat content in rabbits . I . Direct response to selection. 4526–4531. doi:10.2527/jas2013-6361.