UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas



Study of the metagenome of the genital tract in sheep and its role in reproductive efficiency

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Edgar Leonardo Reinoso Peláez
Master of Science in Animal Breeding and Reproductive Biotechnology

Madrid, 2025

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas

Doctoral Degree in Agro-environmental Technology for Sustainable Agriculture

Study of the metagenome of the genital tract in sheep and its role in reproductive efficiency

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Edgar Leonardo Reinoso Peláez
Master of Science in Animal Breeding and Reproductive Biotechnology

Under the supervision of:
Dr. Magdalena Serrano Noreña
Dr. María Saura Álvarez

Madrid, 2025

Title: Study of the metagenome of the genital tract in sheep and its role in reproductive efficiency

Author: Edgar Leonardo Reinoso Peláez

Doctoral Programme: Agro-environmental Technology for Sustainable Agriculture

Thesis Supervision:

Magdalena Serrano Noreña, Research Scientist (Investigadora Científica), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria – Consejo Superior de Investigaciones Científicas (Supervisor)

María Saura Álvarez, Tenured Scientist (Científica Titular), Instituto de Investigaciones Marinas – Consejo Superior de Investigaciones Científicas (Supervisor)

External Reviewers:

Thesis Defence Committee:

Thesis Defence Date:

This thesis has been supported by the FPI grant PRE2019-090087 associated to the project RTI-2018-096487-R-C33, from the Ministerio de Ciencia, Innovación y Universidades

To my parents, Marianita and Edgar,

They have given me everything without ever asking for anything in return, teaching me through their example the true meaning of strength, generosity, and unconditional love. To them, I owe it all; for them, I have everything.

Acknowledgements

First of all, to my supervisors, Magdalena Serrano and María Saura. They have been exemplary leaders and have provided me with all the academic support and attention necessary to grow as a researcher. They trusted and invested in my development, facilitating courses, conferences, and research stays. They are outstanding professionals and, above all, people with great empathy. They understood and valued all my ups and downs throughout the process and made my path truly comfortable and motivating. To me, they are not just my thesis supervisors—they are part of my family.

To Malena Serrano, who has always been present and attentive: from every morning when I arrived at the institute, every time I approached her office to discuss results, to the last hours of the day. She always welcomed me with open arms, with a smile, and with unwavering energy and motivation.

To María Saura, who, from the beginning, moved heaven and earth to ensure that bureaucratic procedures would not become an obstacle at the start of my PhD. Without her tireless fighting spirit, I would not be writing these lines today. Even after moving to a different institute, she has never stopped guiding, teaching, and supporting me, always finding solutions when things didn't go well. I've always known I could count on her.

To Fernando Puente Sánchez and Professor Stefan Bertilsson from the Institutionen för vatten och miljö at the Swedish University of Agricultural Sciences (Uppsala, Sweden), for their generous support, for welcoming me into their lab, and for sharing their expertise during my international stay. Their guidance and that of their research group were fundamental in enriching this phase of the project.

To Javier Tamames and the CNB, who welcomed me during a research stay in their department, from whom I learned valuable tools and approaches that enhanced my training as a researcher, and with whom I shared a very professional and welcoming environment.

To Carmen González, lab technician at INIA, whose collaboration has been constant, generous, and precise. Her experience and commitment have been pillars throughout many stages of the experimental work.

To Jorge Hugo Calvo and Manuel Ramón, who have collaborated with the project and our group, and especially for their warmth and closeness. I shared quality moments with them at conferences, during visits to INIA, and in the day-to-day. Their company made me feel at home.

To CESGA (Centro de Supercomputación de Galicia), for providing access to their computational resources, essential for processing the data in this study. Their support was key to carrying out demanding bioinformatics analyses. I also thank the Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), as well as the Asociación de Criadores de Ovejas Manchegas (AGRAMA), the Asociación de Criadores de Ovejas Latxas (CONFELAC), and the Asociación de Criadores de ovino de raza Assaf (ASSAF.E) for their collaboration and willingness to provide access to samples and crucial data. Their involvement and commitment to applied science have been truly valuable.

To the Ministerio de Ciencia, Innovación y Universidades for the support received through the FPI fellowship (PRE2019-090087), which allowed me to develop this doctoral thesis under optimal conditions, as well as to participate in research stays, conferences, and training activities that enriched my academic journey. I also acknowledge the funding for the project in which this work is framed, provided through the RTI-2018-096487-R-C33 grant from the Ministerio, CSIC (2022AEP004), and FEDER funds, whose support was essential for the development of this research.

To Oxford Nanopore Technologies for their support in enabling my attendance at scientific conferences during this thesis. Their contribution was a great help in sharing my work with the scientific community and continuing to learn in high-level academic settings.

To the researchers, technicians, and department staff: Carmen, Almudena, Beatriz, María Jesús, Clara, Óscar, María Muñoz, Cristina, José María, Jesús, Ramón, Rita, Yolanda, Fabián, Teresa, Malena, Luis Alberto… all of whom were always willing to help and fostered a friendly and collaborative environment.

To my colleagues and friends from the office and department: Leo, Hernán, and Pedro. To Paola, Gloria, Mateo, Ilaria, and Chiara, who, although present for short-term stays, helped create a wonderful atmosphere and friendship. To Adrián, Ester, Carlos, Elizabeth, Silvia, Mónica, and all the colleagues from the institute with whom we shared pleasant moments during coffee breaks.

To my family, who have always been with me from far, and who are an infinite source of unconditional love and support. They are the most beautiful and valuable part of my life. To my friends—those in Ecuador, those in Madrid, and those spread around the world—thank you for always making life easier and more meaningful. And to María Manoli, who has been my companion in this final stretch and who supported and motivated me through both the good and the difficult moments.

Agradecimientos

En primer lugar, a mis directoras, Magdalena Serrano y María Saura. Han sido líderes ejemplares y me han brindado todo el soporte académico y la atención necesaria para que pudiera desarrollarme como investigador. Confiaron e invirtieron en mi conocimiento, facilitándome cursos, congresos y estancias. Son grandes profesionales y, además, personas con mucha empatía. Supieron entender y valorar todos mis altibajos durante el proceso y han hecho que mi camino sea totalmente confortable y motivador. Para mí, no son solo mis directoras de tesis, son parte de mi familia.

A Malena Serrano, que ha estado presente y pendiente en todo momento: desde cada mañana cuando llegaba al instituto, cada ocasión en la que me acercaba a su despacho y charlábamos sobre los resultados, hasta las últimas horas del día. Siempre me ha recibido con los brazos abiertos, con una sonrisa, y siempre tan motivada y llena de energía.

A María Saura, quien, para empezar, movió cielo y tierra para que los trámites burocráticos no fuesen un obstáculo al inicio del doctorado. Sin su espíritu de lucha incansable, no estaría escribiendo estas líneas. A pesar de que se mudó de instituto, nunca ha dejado de estar pendiente de mí, de guiarme, enseñarme y hacer que lo imposible sea posible. Siempre he tenido la certeza de que, si algo no iba bien, María encontraría una solución.

A Fernando Puente Sánchez y al profesor Stefan Bertilsson, del Institutionen för vatten och miljö en la Swedish University of Agricultural Sciences (Uppsala, Suecia), por su generosa ayuda, por abrirme las puertas de su laboratorio y por compartir su experiencia durante mi estancia internacional. Su apoyo y el de su grupo de investigación fueron fundamentales para enriquecer esta etapa del trabajo.

A Javier Tamames y al CNB, quienes me acogieron durante una estancia en su departamento, de quienes aprendí valiosas herramientas y enfoques que enriquecieron mi formación como investigador, y con quienes compartí un ambiente muy profesional y acogedor.

A Carmen González, técnica de laboratorio INIA, cuya colaboración ha sido constante, generosa y precisa. Su experiencia y compromiso han sido pilares en muchos momentos del trabajo experimental.

A Jorge Hugo Calvo y Manuel Ramón, que han colaborado con el proyecto y con nuestro grupo, y sobre todo por su calidez y cercanía. Con ellos he compartido momentos de calidad en congresos, visitas al INIA o en el día a día. Su compañía ha hecho que me sienta en casa.

Al CESGA (Centro de Supercomputación de Galicia), por brindarnos acceso a sus recursos computacionales, fundamentales para el procesamiento de datos en este trabajo. Su apoyo ha sido clave para avanzar en los análisis bioinformáticos más exigentes. Mi agradecimiento también al Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), así como a la Asociación de Criadores de Ovejas Manchegas (AGRAMA), a la Asociación de Criadores de Ovejas Latxas (CONFELAC) y a la Asociación de Criadores de Ovino de raza Assaf (ASSAF.E) por su colaboración y disposición para facilitar el acceso a muestras y datos esenciales para esta investigación. Su implicación y compromiso con la ciencia aplicada han sido verdaderamente valiosos.

Al Ministerio de Ciencia, Innovación y Universidades por el apoyo recibido a través de la beca FPI (PRE2019-090087), que me ha permitido desarrollar esta tesis doctoral en condiciones óptimas, así como participar en estancias, congresos y actividades formativas que enriquecieron mi formación. Asimismo, agradezco la financiación del proyecto en el que se enmarca este trabajo, a través de la subvención RTI-2018-096487-R-C33 del propio Ministerio, del CSIC (2022AEP004) y de los Fondos FEDER, cuyo respaldo ha sido fundamental para el desarrollo de la investigación.

A Oxford Nanopore Technologies por el apoyo brindado para asistir a congresos durante el desarrollo de esta tesis. Su contribución fue de gran ayuda para compartir mi trabajo con la comunidad científica y seguir aprendiendo en espacios de alto nivel académico.

A los investigadores, técnicos, y personal del departamento: Carmen, Almudena, Beatriz, María Jesús, Clara, Óscar, María Muñoz, Cristina, José María, Jesús, Ramón, Rita, Yolanda, Fabián, Teresa, Malena, Luis Alberto… quienes siempre han estado dispuestos a colaborar y han generado un ambiente familiar.

A mis compañeros y amigos de despacho y departamento: Leo, Hernán y Pedro. A Paola, Gloria, Mateo, Ilaria y Chiara, que, aunque estuvieron en estancias temporales, generamos un gran ambiente y amistad. A Adrián, Ester, Carlos, Elizabeth, Silvia, Mónica y todos los compañeros del instituto, con quienes compartimos gratos momentos en el café.

A mi familia, que siempre ha estado conmigo desde la distancia, y en quienes encuentro una fuente de amor y apoyo incondicional infinita. Son lo más bonito y valioso que tengo en la vida. A mis amigos, los que están en Ecuador, los que están en Madrid y los que están por el mundo, porque siempre han hecho que mi vida sea más fácil y tenga aún más sentido. A María Manoli, que ha sido mi compañera en este último tramo, y quien me ha motivado y apoyado en los buenos y malos momentos.

Abstract

Sheep farming plays a critical role in global agriculture, particularly in Mediterranean countries such as Spain. Despite advances in genetic selection and reproductive technologies, fertility rates following artificial insemination (AI) remain suboptimal, primarily due to anatomical and physiological constraints. Increasing evidence points to the reproductive tract microbiota as a potential modulator of fertility. However, its composition, functional relevance, and interaction with host genetics remain insufficiently characterised in ovine species.

In this context, this thesis investigates the relationship between vaginal microbiota and reproductive success in sheep, focusing on AI outcomes through four complementary studies. **Study 1** characterised bacterial communities via 16S rRNA metabarcoding to establish a reference microbiota and assess its association with pregnancy. **Study 2** expanded this analysis using Oxford Nanopore Technologies (ONT) based metagenomics to capture broader microbial diversity, including functional pathways. **Study 3** examined the impact of oestrus synchronisation devices (PRID) and treatments on vaginal microbiota and AI success. **Study 4** employed Microbiota Genome-Wide Association Studies (MGWAS) to explore host genetic factors influencing microbial composition.

The work involved multiparous ewes from four Spanish breeds (Latxa, Manchega, Rasa Aragonesa, and Assaf). Vaginal swabs were collected prior to AI and analysed using advanced sequencing techniques, including 16S rRNA metabarcoding and ONT-based metagenomics. Specific bioinformatic tools enabled microbial classification and functional profiling. To explore host genetic influences on microbiota, ewes were genotyped using a high-density SNP chip. Classical statistical inference and machine learning techniques were used to assess microbial diversity, identify differentially abundant taxa, evaluate predictive associations, and explore the host–microbiota interactions associated to fertility.

A consistent core microbiota was identified, dominated by the genera *Streptobacillus, Histophilus, Fusobacterium, Oceanivirga,* and *Parvimonas*. Herd and breed were important drivers of microbial variation, suggesting that environmental and host genetic factors shape microbiome composition and abundance. PRID devices and associated antibiotic treatments significantly modified composition, reducing potentially harmful taxa.

Pregnancy status showed moderate association with vaginal microbiota composition. Taxa were differentially abundant between pregnant and non-pregnant ewes. *Fusobacterium, Histophilus, Leptotrichia, Campylobacter*, and *Treponema* were enriched in non-pregnant ewes, along with functional pathways related to

inflammation and metabolism. In contrast, pregnant ewes more frequently harboured *Pseudomonas, Acinetobacter, Brevundimonas,* and *Glutamicibacter*, suggesting a microbial environment more conducive to successful reproduction.

MGWAS identified host genetic variants associated with microbial genera known to influence fertility. Variants in genes linked to immune regulation, epithelial integrity, and metabolism influenced microbial composition. SNPs associated with beneficial genera such as *Lactobacillus* and *Bifidobacterium breve* were predicted to enhance mucosal immune tolerance and nutrient availability, while alleles linked to potentially pathogenic genera such as *Actinobacillus, Gardnerella*, and *Fusobacterium nucleatum* suggested compromised host defence mechanisms.

In conclusion, this thesis provides an integrative perspective on vaginal microbiota and its relationship to sheep reproductive outcomes. Microbiota structure was considerably shaped by herd, breed, and PRID treatments, while specific microbial taxa and functions were associated with pregnancy outcomes. Moreover, host genomic variation contributed to microbial composition, underscoring the potential of combining microbial and genetic analyses to enhance reproductive efficiency in sheep.

Resumen

La producción ovina desempeña un rol clave en la agricultura global, especialmente en países mediterráneos como España. A pesar de los avances en mejora genética y tecnologías reproductivas, la tasa de fertilidad tras la inseminación artificial (IA) en ovejas sigue siendo subóptima por limitaciones anatómicas y fisiológicas. Estudios recientes sugieren que la microbiota del tracto reproductivo podría modular la fertilidad; sin embargo, su composición, función e interacción con el hospedador está poco caracterizada.

Esta tesis investiga la relación entre la microbiota vaginal y el éxito reproductivo por IA en ovejas mediante cuatro estudios. El **Estudio 1** caracterizó la microbiota mediante secuenciación parcial del gen del ARNr 16S, estableciendo una microbiota de referencia y evaluando su asociación con la tasa de preñez. El **Estudio 2** amplió este análisis utilizando metagenómica basada en secuenciación por nanoporos, lo que permitió analizar genes y funciones microbianas. El **Estudio 3** evaluó el impacto de los dispositivos de sincronización del estro (PRID) y tratamientos asociados sobre la microbiota. El **Estudio 4** investigó la influencia genética del hospedador sobre la abundancia de la microbiota mediante estudios de asociación de genoma completo (MGWAS).

Se recogieron muestras de exudado vaginal antes de la IA en ovejas multíparas de cuatro razas españolas, Latxa, Manchega, Rasa Aragonesa y Assaf, y se secuenció el ADN mediante metabarcoding del ARNr 16S y metagenómica por nanoporos. La caracterización taxonómica y funcional se realizó con herramientas bioinformáticas específicas. Para explorar la influencia de la genética del hospedador sobre la microbiota, las ovejas se genotiparon con un chip de SNPs de alta densidad. Se aplicaron métodos clásicos de inferencia estadística y de aprendizaje automático para evaluar la diversidad microbiana, identificar taxones diferencialmente abundantes, y explorar la interacción hospedador-microbiota y su relación con la preñez.

Se identificó una microbiota de referencia dominada por *Streptobacillus, Histophilus, Fusobacterium, Oceanivirga* y *Parvimonas*. El rebaño, la raza, y el uso de PRID influyeron notablemente en la composición y abundancia de la microbiota, resaltando la importancia de factores ambientales y genéticos. El uso del tratamiento antibiótico asociado al PRID contribuyó a reducir la presencia de taxones potencialmente perjudiciales.

La preñez mostró una asociación moderada con la microbiota vaginal. *Fusobacterium, Histophilus, Leptotrichia, Campylobacter* y *Treponema,* fueron más abundantes en ovejas no preñadas, y se asociaron con rutas relacionadas con

inflamación y metabolismo. Por el contrario, *Pseudomonas, Acinetobacter, Brevundimonas y Glutamicibacter* fueron más comunes en ovejas gestantes, asociándose a un entorno reproductivo más favorable.

El MGWAS identificó variantes genéticas del hospedador asociadas a la abundancia de taxones vinculados a la fertilidad. Estas variantes están relacionadas con la inmunidad, integridad epitelial y metabolismo. Variantes vinculadas a *Lactobacillus y Bifidobacterium breve* se asociaron a una mucosa vaginal más tolerante y a una mayor eficiencia del hospedador para suministrar nutrientes. Mientras que variantes relacionadas con mayor abundancia de *Actinobacillus, Gardnerella y Fusobacterium nucleatum* sugirieron un efecto de debilitamiento en los mecanismos de defensa del hospedador.

En conclusión, esta tesis ofrece una visión integradora de la microbiota vaginal ovina y su vínculo con la reproducción. La composición microbiana estuvo considerablemente determinada por el rebaño, la raza y el uso de PRID. Aunque la asociación con la preñez fue menos marcada, se identificaron taxones y funciones específicos significativamente relacionados. La variación genética del hospedador también influyó en la composición microbiana, destacando el potencial de integrar perfiles microbianos y genéticos para mejorar el rendimiento reproductivo.

Table of Contents

1. Introduction 1

1.1. A brief history of sheep domestication 1

1.2. Current situation of sheep farming in Spain 2

1.3. Fundamentals of sheep farming management 3

1.4. The challenge of reproduction in sheep farming 5

2. Objectives 9

3. State of the art 11

3.1. The role of genital microbiota in mammals' reproduction 11

3.2. Microbiota dynamics with host 12

3.3. Evolution of microbiota studies 13

3.3.1. Current methods for studying microbiota 16

3.3.2. Microbiome data and the problem of compositionality 21

4. Materials and methods 25

4.1. Summary of study designs 26

4.2. Animal samples 26

4.3. DNA extraction 29

4.3.1. Microbial DNA extraction 29

4.3.2. Host DNA extraction 30

4.4. Sequencing 30

4.4.1. Amplicon-based 16S rRNA metabarcoding 30

4.4.2. ONT-based metagenomics sequencing 30

4.5. Bioinformatics analysis 31

4.5.1. 16S rRNA metabarcoding analysis 31

4.5.2. ONT-based metagenomic analysis 32

4.5.3. Host-genotype data processing 32

4.6. Statistical analysis 33

4.6.1. Microbial composition 33

4.6.2. Alpha-diversity 33

4.6.3. Beta-diversity 34

4.6.4. Clustering and prediction 38
4.6.5. Differential abundance analysis 40
4.6.6. MGWAS and functional analysis 44

5. Results 47

5.1. Study 1: Characterisation of vaginal bacterial communities and their association with pregnancy outcomes using 16S rRNA metabarcoding 47
5.1.1. Bacterial composition 47
5.1.2. Alpha-diversity 48
5.1.3. Beta-diversity 50
5.1.4. Clustering and Random Forest prediction 51
5.1.5. Differential abundance analysis 53

5.2. Study 2: Characterisation of vaginal microbiota and microbial genes associated with pregnancy outcomes using metagenomics 56
5.2.1. Microbiota composition 56
5.2.2. Alpha-diversity 58
5.2.3. Beta-diversity 60
5.2.4. Random Forest prediction 62
5.2.5. Differential abundance analysis 63
5.2.6. Integrative gene-taxa associations 65

5.3. Study 3: Impact of oestrus synchronisation devices on vaginal microbiota and artificial insemination success 68
5.3.1. Microbiota composition 68
5.3.2. Alpha-diversity 70
5.3.3. Beta-diversity 72
5.3.4. Differential abundance analysis 75

5.4. Study 4: Host genomic regions associated with vaginal bacterial communities 77
5.4.1. Microbiota genome-wide association analysis 77
5.4.2. SNP associations with key fertility taxa 80
5.4.3. Variant Effect Predictor analysis 81
5.4.4. Functional analysis of non-intergenic SNP-associated genes 83

6. Discussion 89

6.1. Recapitulation 89

6.2. Microbial composition 90

6.3. Drivers of microbial diversity 92
6.3.1. Breed and herd 92

6.3.2. Age and parity 94
6.3.3. PRID effect 94
6.3.4. Pregnancy status 95
6.4. Differential abundance analysis for pregnancy 96
6.4.1. Taxa with increased abundance in non-pregnant ewes 96
6.4.2. Taxa with increased abundance in pregnant ewes 99
6.4.3. Methods and platform for differential abundance analysis 99
6.5. Microbiota genome-wide association studies 101
6.5.1. Host-associated taxa linked to fertility failure 102
6.5.2. Host-associated taxa linked to fertility success 105
6.5.3. Opportunities, limitations, and future directions 108
7. Conclusions 111
References 115
Appendix 143
Appendix for Study 1: Characterisation of vaginal bacterial communities and their association with pregnancy outcomes using 16S rRNA metabarcoding 143
Appendix for Study 2: Characterisation of vaginal microbiota and microbial genes associated with pregnancy outcomes using metagenomics 163
Appendix for Study 3: Impact of oestrus synchronisation devices on vaginal microbiota and artificial insemination success under four treatments 183
Appendix for Study 4: Host genomic regions associated with vaginal bacterial communities 192
Peer-reviewed publications and scientific contributions originating from the doctoral thesis 194
Communications at congresses related to this thesis 195
Research stays during the doctoral process 198

List of Figures

The numbering of figures and tables is designed to facilitate their location within the document, with the first number indicating the corresponding chapter.

Figures for chapter 1: Introduction

Figure 1.1. Historical timeline of sheep farming 1

Figure 1.2. Representative sheep breeds in Spain 3

Figure 1.3. Diagram of foundational pillars of sheep farming 4

Figure 1.4. Anatomy of the ewe's reproductive system. 7

Figures for chapter 3: State of the art

Figure 3.1. Historical timeline of microbiology and sequencing innovations 14

Figure 3.2. The 2D-3D 16S rRNA gene structure 18

Figure 3.3. Compositional nature of sequencing data and its impact on abundance interpretation..... 22

Figures for chapter 4: Material and methods

Figure 4.1. Workflow of study designs and analytical approaches based on the thesis objectives 25

Figure 4.2. Farm locations of the four herds..... 28

Figures for chapter 5: Results

Figure 5.1.1. Core bacterial community composition with a prevalence ≥ 90% across all samples, at the genus (A) and phylum levels (B), for the four herds analysed..... 48

Figure 5.1.2. Alpha-diversity for herd at ASV (A), genus (B), and phylum (C) level for the different coefficients analysed: Observed richness, Chao1, Shannon, and Inverse Simpson index 49

Figure 5.1.3. Principal Component Analysis for breed and herd given the bacterial composition at the ASV level, represented in a 2D (A) and 3D (B) plots..... 50

Figure 5.1.4. Analysis of sample similarities and clustering in bacterial community studies 52

Figure 5.1.5. Heatmap showing significant results (FDR at 5%) for differential abundance analysis for pregnancy success at phylum, genus, species and ASV levels..... 55

Figure 5.2.1. Microbial community composition at genus (A) and phylum (B) level... 58

Figure 5.2.2. Alpha-diversity measures for herd at genus (A) and phylum (B) level for the different coefficients analysed: Observed richness, Chao1, Shannon, and Inverse Simpson index 59

Figure 5.2.3. Principal Component Analysis of microbiota composition by breed, herd, and pregnancy status..... 61

Figure 5.2.4. Heatmap showing significant results (FDR at 5%) for differential abundance analysis for pregnancy success 65

Figure 5.2.5. Heatmap of the relationship between genes and taxa differentially expressed in relation to pregnancy..... 66

Figure 5.2.6. Network diagram of pathways associated with genes differentially expressed in pregnancy 67

Figure 5.3.1. Composition of the microbiota for S1 (before PRID placement) and S2 (after PRID removal) sampling point groups at the phylum (A) and genus (B) levels. 70

Figure 5.3.2. Alpha-diversity measures (Observed, Chao1, Shannon, and Inverse Simpson indices) for sampling point groups at the genus (A) and phylum (B) levels . 71

Figure 5.3.3. Principal Component Analysis for S1 (Before PRID placement) and S2 (After PRID removal) sampling points (A) groups and pregnancy (B) groups at genus, phylum, COG, and KO level 73

Figure 5.3.4. Stacked bar chart showing significant results (FDR at 5%) for differential abundance analysis for sampling point groups S1 and S2 at phylum (A) and genus (B) levels 76

Figure 5.4.1. Relative abundance of significantly associated taxa identified in the MGWAS across sheep breeds. 78

Figure 5.4.2. Distribution of significantly associated SNPs across chromosomes (Chr, x-axis) and taxa (y-axis) 79

Figure 5.4.3. Manhattan plot for Microbiota Genome-Wide Association Studies. 80

Figure 5.4.4. Distribution of genetic variant consequences across host-associated taxa relevant to sheep fertility 82

Figure 5.4.5. Functional descriptions of host genes associated with fertility-related vaginal microbial taxa 85

Figure 5.4.6. Network of host genes significantly associated with fertility-related vaginal microbial taxa 86

Figure 5.4.7. GO classification heatmap of genes in association with vaginal microbial taxa. Low to high impact variants (excluding synonymous ones) were considered to select the genes 87

List of Tables

Tables for chapter 3: State of the art

Table 3.1. Comparison of DNA sequencing technologies 16

Table 3.2. Key differences between 16S rRNA metabarcoding and metagenomic sequencing approaches. 21

Tables for chapter 4: Materials and methods

Table 4.1. Study-specific sampling details 29

Tables for chapter 5: Results

Table 5.1.1. PERMANOVA results for breed, herd, and pregnancy variables at ASV level ……51

Table 5.1.2. Assessment of variable importance and predictive accuracy in bacterial community clustering using Random Forest analysis 53

Table 5.2.1. Summary statistics of reads abundance across taxonomic and functional levels 57

Table 5.2.2. PERMANOVA results for the effects of herd, breed, and pregnancy on microbiota and gene composition 62

Table 5.2.3. Prediction accuracy results (and their SD) from cross-validation analysis using a Random Forest algorithm 63

Table 5.3.1. Summary statistics of reads' abundance across taxonomic and functional levels 69

Table 5.3.2. PERMANOVA results for pregnancy (P and N) and sampling time groups (S1 and S2) at genus, phylum and KO levels. 74

Table 5.4.1. Summary of MGWAS analysis for sheep vaginal microbiota related to fertility and host genome 83

Abbreviations and Acronyms

AI	Artificial Insemination
ASV	Amplicon Sequence Variant
BLAST	Basic Local Alignment Search Tool
BLASTN	BLAST Nucleotide
BLASTX	BLAST Protein
bp	Base Pairs
Chr	Chromosome
CLR	Centered Log-Ratio
COG	Clusters of Orthologous Groups
CSIC	Consejo Superior de Investigaciones Científicas
CV	Cross-validation
D	Simpson's Diversity Index
DESeq2	Differential Expression Analysis for Sequence Count Data
GO	Gene Ontology
IIM	Instituto de Investigaciones Marinas
ILR	Isometric Log-Ratio
INIA	Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria
IPBLN	Instituto de Parasitología y Biomedicina López-Neyra
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthologues
LCA	Lowest Common Ancestor
Limma	Linear Models for Microarray Data
log2FC	Log base-2 Fold Change
MAF	Minor Allele Frequency
MGWAS	Microbiota Genome-Wide Association Study
NCBI-nr	National Center for Biotechnology Information – non-redundant
NGS	Next-Generation Sequencing
NMDS	Non-Metric Multidimensional Scaling
ONT	Oxford Nanopore Technologies

OTU	Operational Taxonomic Unit
PCA	Principal Component Analysis
PERMANOVA	Permutational Multivariate Analysis of Variance
PRID	Progesterone Releasing Intravaginal Device
RA	Relative Abundance
RF	Random Forest
rRNA	ribosomal Ribonucleic Acid
RMSE	Root Mean Squared Error
rpm	Revolutions Per Minute
SNP	Single Nucleotide Polymorphism
SMRT	Single Molecule Real-Time
UPM	Universidad Politécnica de Madrid
VEP	Variant Effect Predictor
VL / RN	Manchega VL / Manchega RN (Herd Codes)

1. INTRODUCTION

1.1. A brief history of sheep domestication

Sheep (*Ovis aries*) farming has been one of the most significant livestock practices due to its influence on the development of early civilizations and its crucial role in the economy of many regions for millennia. Sheep farming dates back approximately 10,500 years, when domestication began from wild ancestors such as the Asiatic mouflon in the Fertile Crescent (Alberto et al., 2018; Ciani et al., 2020). During the Bronze Age, sheep farming practices spread across Europe and Asia, becoming vital to the economies of civilizations such as the Greeks and Romans, where wool production played a significant role (Sabatini et al., 2019). Around 6,000 years ago, shepherds made the first attempts at selective breeding, focusing on enhancing traits like wool quality. This early focus on wool laid the foundation for the development of specialised breeds, and, consequently, distinct lines of milk and meat production were developed in later centuries to meet the growing demand for these resources (Ciani et al., 2020). In the Middle Ages, the wool industry thrived, particularly in Europe, with countries like England or Spanish regions like Castilla, standing out in wool production and trade, supported by institutions such as the Mesta in Spain. The development of specialised breeds for wool, dairy, and meat production eventually spread worldwide, turning sheep farming into one of the most widespread and economically vital agricultural practices (Figure 1.1).

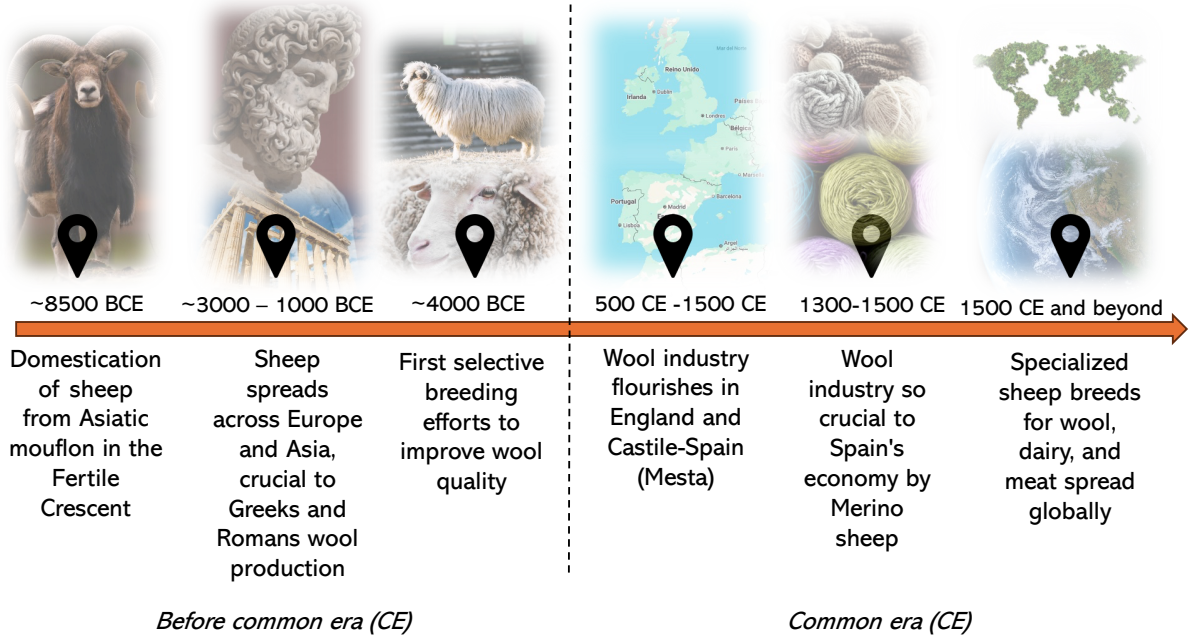


Figure 1.1. Historical timeline of sheep farming.

1.2. Current situation of sheep farming in Spain

Sheep production is one of the most widespread and economically significant agricultural practices worldwide. Sheep are highly adaptable animals, thriving in diverse environments, from the cold highlands of Northern Europe to the arid regions of Africa and Asia. In Spain, sheep farming plays a crucial role in the agricultural economy, with a national flock of over 15 million sheep, representing approximately 25% of the total number of sheep in the European Union. Economically, Spain is a significant producer of lamb meat and milk, generating over 114,000 tonnes of meat and 502 million kg of milk per year (Eurostat, 2023). While wool once held a pivotal position, the focus of the industry has shifted towards meat and dairy production.

In Spain, there are 39 autochthonous and 8 integrated sheep breeds, closely tied to the production systems in which they are raised (Ministerio de Agricultura, 2019). Most sheep farming in this country operates under extensive grazing practices, particularly in regions like Castilla y León, Extremadura, and Aragón, where large expanses of land support meat-producing breeds such as the Merino, native to Spain and historically renowned for its fine wool (Ciani et al., 2015, 2020), and the Rasa Aragonesa, developed in Aragón and well adapted to arid environments (Calvo et al., 2020). In the southeastern regions, breeds like the Segureña, known for their adaptation to local conditions and their high-quality lamb recognised within European markets, are particularly relevant (Marín-Bernal and Navarro-Ríos, 2014). In contrast, dairy production is concentrated in semi-intensive or intensive systems, particularly in Castilla-La Mancha, Castilla y León, and the Basque Country, where breeds like the Manchega and the Latxa thrive. The Manchega, native to Castilla-La Mancha (https://agrama.es), is essential for producing Manchego cheese, while the Latxa (https://aslana.es), from the Basque Country and Navarra, is key to Idiazábal cheese production. Additionally, the Assaf, an imported synthetic breed from Israel, has become significant in Northern Spain for its high milk yields (https://assafe.es).



Figure 1.2. Representative sheep breeds in Spain. The images sourced from the Official Catalogue of Livestock Breeds, provided by the Ministerio de Agricultura, Pesca y Alimentación from Spain (mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-raz).

During the last decade, a significant decline in ovine production has been observed. In particular, Spanish sheep production has dropped from approximately 20 million animals in 2009 to around 14 million in 2022, representing a 30% reduction (Ministerio de Agricultura, 2022). This decline is driven by multiple factors, including rising production costs, an aging farming population with limited generational renewal, environmental challenges such as drought and reduced forage availability, and shifting consumer preferences towards alternative food options and dietary patterns. Additionally, European production is not sufficient for its own consumption (European Parliament, 2018). Therefore, addressing these challenges requires focused efforts on improving the efficiency and sustainability of sheep farming, which will help ensure the sector's resilience and long-term viability.

1.3. Fundamentals of sheep farming management

Sheep management is critical in animal production systems for several key reasons, all of which contribute to ensuring both the welfare of the animals and the overall efficiency and profitability of the farming operation. Effective sheep farming relies on four fundamental pillars: health, nutrition, genetics, and reproduction (National Research Council, 2007; Radostits and Done, 2010; McDonald et al., 2011; Gootwine, 2021; Abbott, 2024).

A robust health program is essential to safeguard flock welfare and productivity. Common preventative measures include targeted vaccination schedules —e.g., against *Clostridium perfringens* types C and D, tetanus, foot rot, bluetongue, and Contagious Ecthyma (Sargison, 2008; Radostits and Done, 2010; Nandi et al., 2011; Uzal et al., 2014)— quarantine of new animals, and monitoring for notifiable diseases such as Maedi-Visna, Foot-Rot, Bluetongue, among others (Beveridge, 1941; Cutlip et al., 1988; Mellor et al., 2008; Radostits and Done, 2010). Strategic parasite control (internal and external) typically involves rotational grazing, targeted deworming, and responsible use of chemical or organic treatments to avoid resistance (Wall and Shearer, 2008; Taylor and Wall, 2015).

Nutrition is a key aspect of sheep management and reinforces efficient growth, reproduction, and overall flock health (McDonald et al., 2011). Grazing forms the basis of many systems, supplemented with hay, silage, or concentrates during periods of high nutrient demand (gestation, lactation, or fattening). Proper mineral and vitamin supplementation (e.g., calcium, phosphorus, copper, selenium) and adequate water intake (3–4 L per adult sheep per day) are indispensable (National Research Council, 2001). Careful ration management helps prevent metabolic issues (e.g., ruminal acidosis) and ensures robust body condition and optimal milk yield.

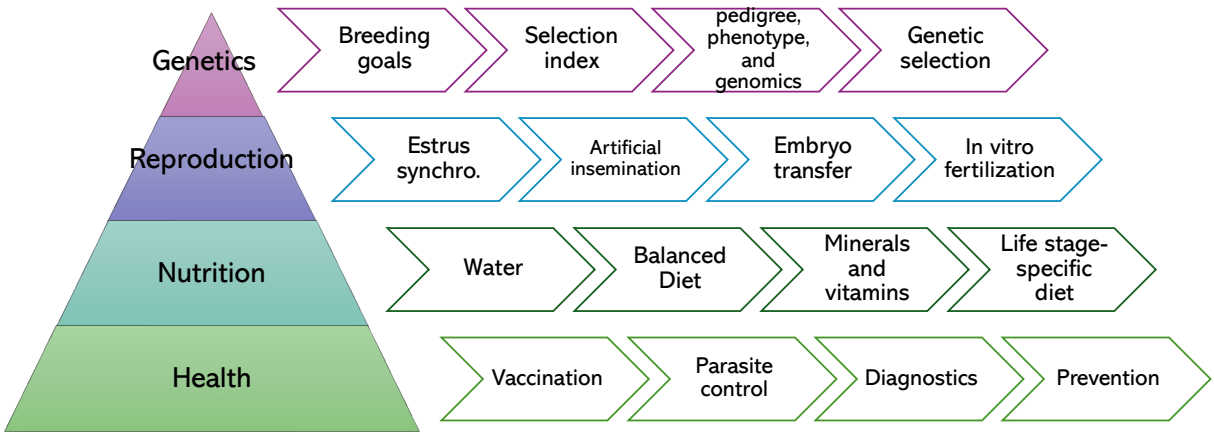


Figure 1.3. Diagram of foundational pillars of sheep farming.

Genetics has played a crucial role in sheep production, significantly influencing the improvement of productive characteristics —primarily meat and milk production— disease resistance, and the adaptability of flocks to various environments and management conditions (Simões et al., 2021). Breeding programs often involve clear objectives —such as meat or dairy yield, wool traits, or health resilience— and rely on

production and pedigree data recording, genetic evaluations, and, increasingly, genomic tools (Kalds et al., 2022). This systematic selection ensures that high-value traits proliferate, strengthening flock performance over time and contributing to both economic viability and sustainable farming practices (Simões et al., 2021).

Artificial insemination plays a pivotal role in modern genetic improvement programs by facilitating the connection between flocks, enabling genetic evaluations to be compared across the entire breeding population. Through the widespread dissemination of superior genetics, AI enhances the accuracy and comparability of genetic assessments, ensuring that selection decisions are based on a broader, more representative dataset. Additionally, AI is essential for the progeny testing of rams for traits that can only be measured in females, such as milk production. By allowing the systematic testing of sires across multiple flocks, AI improves the reliability of estimated breeding values, accelerates genetic progress, and ensures that selection decisions are based on robust, large-scale performance data (Ajmone-Marsan et al., 2023).

Reproductive management in sheep production is a very important element, since the efficiency in the production of herd replacement animals, rests on it (Lianou et al., 2022). Reproductive management not only includes the use of assisted reproduction techniques —such as heat synchronisation devices, AI, embryo transfer, and sperm sorting, among others— but also the reproductive rhythm imposed on the animals by the farmer (Donovan et al., 2004; Tadesse et al., 2022; Gonçalves et al., 2024). Thus, factors such as the age at first mating, the interval between matings, and the body condition of the ewes are also crucial elements for achieving an adequate and efficient reproductive rhythm on livestock farms (Gootwine, 2021).

1.4. The challenge of reproduction in sheep farming

Reproduction practices in sheep farming range from traditional natural mating to advanced biotechnologies. Artificial insemination is a key assisted reproductive technology in sheep breeding. It involves collecting semen from rams and depositing it in the ewe's reproductive tract. However, the success rate of AI in small ruminants, particularly in sheep, tends to be lower compared to other livestock species, ranging between 30% and 70% (Cseh et al., 2012; Vilariño et al., 2013; Granleese et al., 2015). For instance, AI success rates are 60-80% in horses (Aurich, 2012), 50-85% in cattle (Tadesse et al., 2022), and 70-90% in pigs (Roca et al., 2006).

Several factors contribute to the lower efficiency in sheep. One major challenge is their cervical anatomy. The ewe's cervix is highly variable across individuals, with its tightly coiled structure and eccentric folds —irregularly aligned structures within the cervical canal that further complicate transcervical procedures (Figure 1.4)

(Granleese et al., 2015; Gonçalves et al., 2024). Other limiting factors include the difficulty of detecting oestrus —due to its short and often silent expression— requiring synchronisation protocols that may not always align precisely with ovulation (Mohan and Kumar, 2023). Moreover, the reliance on fresh semen limits the flexibility of AI programs, as it demands close coordination, donor-recipient proximity, and presents logistical and biosecurity challenges, all of which may contribute to lower conception rates in ewes (Halaweh et al., 2025).

Several strategies have been developed to improve the success rate of AI in sheep. One highly effective method is laparoscopic AI, which bypasses the cervix to directly deposit semen into the uterine horns. This method is especially beneficial for using frozen semen, although it is invasive, expensive, and r*equi*res a specialised professional to perform (Sathe, 2018). Another critical factor is timing and oestrus synchronisation. It is challenging to precisely determine when the ewe is at the optimal point in her ovulatory cycle. The use of Progesterone Releasing Intravaginal Devices (PRID) plays a crucial role in improving AI success rates in sheep by effectively synchronising oestrus and optimising the timing of ovulation. PRIDs release controlled amounts of progesterone, temporarily suppressing the oestrus cycle and allowing for precise scheduling of AI procedures. Upon removal of the device, a rapid decline in progesterone levels triggers follicular development and ovulation, facilitating better synchronisation within a flock (Gatti et al., 2011; Martinez-Ros et al., 2018; Mohan and Kumar, 2023).

Despite these advancements, the effectiveness of AI in sheep remains highly variable, r*equi*ring tailored management approaches to optimise results. An emerging area of research that may further enhance AI outcomes is the study of the reproductive tract microbiota. The composition of microbial communities within the vaginal and uterine environments has been suggested to influence sperm survival, fertilization success, and early embryo development (Wang et al., 2024a). Understanding the interactions between host genetics, microbial balance, and reproductive performance could lead to novel microbiome-based interventions aimed at improving AI efficiency in sheep. Further investigation in this field may offer new strategies for enhancing reproductive outcomes and increasing the overall success of genetic improvement programs.

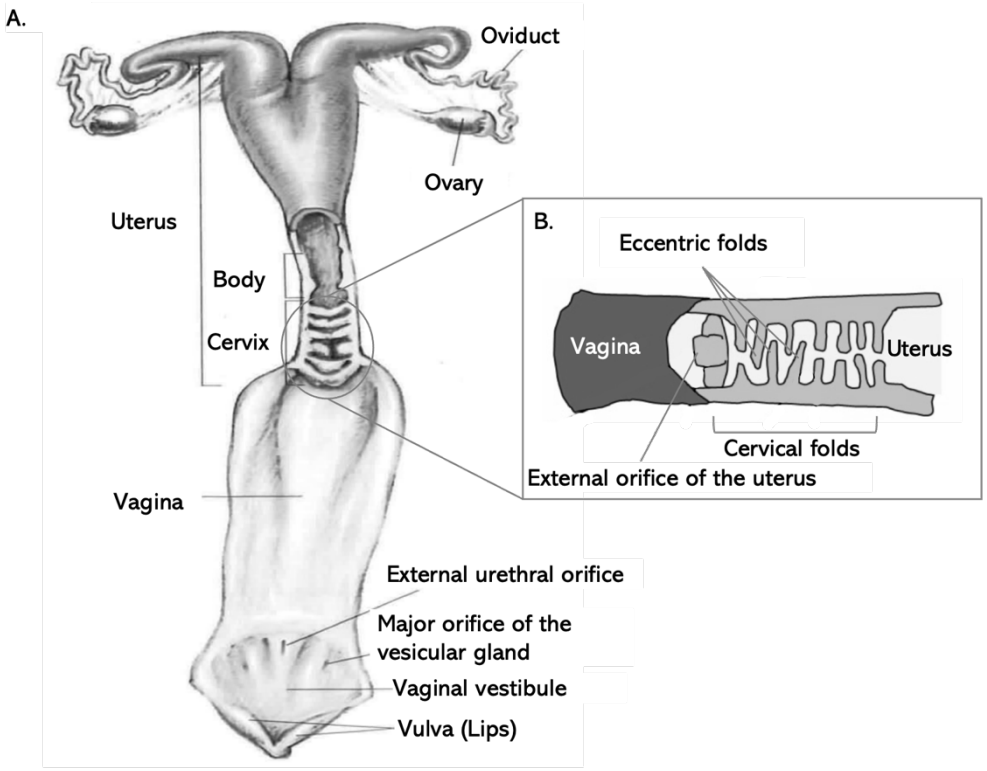


Figure 1.4. Anatomy of the ewe's reproductive system (**A**) (adapted from Spurgeon's Colour Atlas of Large Animal Anatomy, McCracken et al. (2013)). **B.** Detail of the cervical lumen (adapted from Transcervical uterine flushing and embryo transfer in sheep, Gonçalves et al., 2024).

2. OBJECTIVES

Based on the above, the following research questions arise, which constitute the focus of this doctoral thesis:

i. Wha are the reference microbial populations that colonise the sheep reproductive tract?

ii. To what extent do variations in the abundance of these populations influence the reproductive success of this species?

iii. Does the variability of the reproductive tract microbiota in sheep depend on the genetics of the host?

The general aim is to investigate the influence of the reproductive tract microbiota on the reproductive efficiency by AI of four sheep breeds, Assaf, Latxa, Manchega and Rasa Aragonesa, reared in different regions of Spain.

This objective includes the establishment of a reference microbiota (core microbiota), the identification of microbial imbalances (dysbiosis) associated with reproductive failure based on AI outcomes, the assessment of the impact of synchronisation treatments on microbial communities, and the evaluation of host genetic factors influencing the composition and abundance of key taxa associated to fertility.

The specific objectives have been addressed in four studies:

- **Objective 1 (Study 1):** Characterise the composition and abundance of vaginal bacterial communities in different sheep breeds and production systems and evaluate its relationship with AI pregnancy outcome using 16S rRNA metabarcoding.
- **Objective 2 (Study 2):** Characterise the composition and abundance of vaginal microbiota in different sheep breeds and production systems and evaluate their relationship with AI pregnancy outcome using a ONT-based metagenomic approach. This analysis also explores microbial functional pathways associated with fertility by AI.
- **Objective 3 (Study 3):** Determine the impact of oestrus synchronisation devices with different additives on the vaginal microbiota of sheep and their effect on AI pregnancy outcome.
- **Objective 4 (Study 4):** Investigate the relationship between the host genetics and the composition and abundance of the vaginal microbiota, focusing on taxa potentially associated with AI pregnancy outcome.

3. STATE OF THE ART

3.1. The role of genital microbiota in mammals' reproduction

Over the past decade, a growing body of research has highlighted the critical role of the genital tract microbiota in both fertility and the success of assisted reproductive technologies (Tomaiuolo et al., 2020; Adnane and Chapwanya, 2022; Ault-Seay et al., 2023; Bhattacharya et al., 2023; Zangirolamo et al., 2024). Understanding these microbial communities has been enhanced by advances in DNA sequencing technologies, providing new insights for managing fertility and improving livestock sustainability. The reproductive microbiota is dynamic and influenced by several factors such as hormonal fluctuations (Gajer et al., 2012; Cassas et al., 2024), host genetics (Mehta et al., 2020; Greenwood et al., 2022), environment and animal management practices such as reproductive rhythm, sampling season, parity, synchronisation protocols and antibiotic treatment, among others (Martinez-Ros et al., 2018; Ojeda-Hernández et al., 2019; Quadros et al., 2020; Brulin et al., 2023, 2024; Yagisawa et al., 2023). Additionally, the use of probiotics is emerging as a promising alternative for managing reproductive microbiota without adversely affecting fertility (Cribby et al., 2008; Quereda et al., 2020; Gohil et al., 2023). These findings emphasise the importance of understanding reproductive microbiota within each species to optimise reproductive performance and fertility outcomes.

While the role of reproductive tract microbiota in fertility and reproductive health is well-established in humans, it remains poorly understood in livestock. In women, the vaginal microbiota is predominantly composed of *Lactobacilli,* which produce lactic acid that maintains a protective pH (3.0–4.5) and supports reproductive health (Sirota et al., 2014; Witkin and Linhares, 2017). This dominance, often exceeding 70% in humans, is uncommon in other mammals, where *Lactobacilli* typically represent less than 1% of the microbiota (Miller et al., 2016). Conversely, specific bacterial taxa like *Ureaplasma* and *Gardnerella* have been linked to reproductive disorders (Wee et al., 2018). In men, certain bacterial genera, such as *Lactobacillus* and *Gardnerella*, are linked to normal semen quality, whereas *Prevotella* is associated to lower-quality semen (Hou et al., 2013; Weng et al., 2014). However, the unique characteristics of the genital human microbiota limits the direct application of these findings to other species.

In livestock, research on reproductive microbiota and its impact on fertility is still limited. In cattle, the reproductive microbiome, predominantly consists of Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria, with significant variations

depending on anatomical location and breed, affecting fertility and reproductive health (Swartz et al., 2014; Murga Valderrama et al., 2023; Wang et al., 2024c). Unlike in humans, *Lactobacilli* are present at much lower levels in cattle, likely due to near-neutral pH of their reproductive tract (Swartz, 2014). Specific bacterial genera, such as *Fusobacterium*, *Helcococcus*, *Escherichia-Shigella*, and *Histophilus* have been associated with uterine diseases (Kudo et al., 2021), while others like the Lachnospiraceae and Rikenellaceae families, *Lactobacillus*, or *Arcobacter* genera, have been associated with healthier reproductive states in pregnant ewes (Gonzalez Moreno et al., 2016).

Research on the microbial dynamics of the sheep reproductive tract across different reproductive stages remains scarce. Recent studies have identified Firmicutes, Proteobacteria, Bacteroidetes, and Fusobacteria as the dominant phyla in the vaginal microbiota of ewes (Swartz et al., 2014; Barba et al., 2024). Bacteria like Oscillospiraceae, *Mannheimia*, and *Alistipes* have been observed in greater abundance in pregnant ewes, suggesting their role in maintaining favourable conditions for pregnancy (Koester et al., 2021). Conversely, specific genera such as *Mageebacillus*, *Histophilus*, *Actinobacillus*, *Sneathia*, and *Ureaplasma* were found to be less abundant in pregnant ewes (Serrano et al., 2020; Barba et al., 2024). Additionally, there is emerging evidence that the reproductive microbiota of rams may influence the microbiota of the female reproductive tract, potentially impacting pregnancy outcomes. Thus, certain bacteria associated with male reproductive disorders, such as *Corynebacterium*, *Actinobacillus*, and *Histophilus*, can influence the vaginal microbiota of ewes during natural mating and AI practices, impacting reproductive outcomes (Serrano et al., 2018; Barba et al., 2024). These findings highlight the significant role of microbiota in reproduction and the need to explore its impact on fertility in sheep further.

Despite these advances, further investigations are crucial to deepen our understanding of the intricate relationships between reproductive microbiota and fertility in sheep, highlighting the need for more comprehensive research in this area.

3.2. Microbiota dynamics with host

Understanding the role of reproductive microbiota within the broader context of the host's physiology and genetics is crucial to assess its impact on fertility and overall health. However, studies of this nature in sheep are very limited, and the available research often focuses on other species or on different microbial systems such as the rumen (Ravel et al., 2011; Li et al., 2019; Abbas et al., 2020; Martínez-Álvaro et al., 2024).

Heritability of microbiota, or microbiability, refers to the extent to which host genetics, rather than environmental factors, influence microbial community composition (Martinez Boggio et al., 2023; Morris and Bohannan, 2024). Recent research has demonstrated that specific microbial taxa are consistently inherited across generations within a species, suggesting that host genetic variation plays a role in shaping microbial communities (Morris and Bohannan, 2024; Ryu and Davenport, 2025). This, in turn, affects key physiological functions, including metabolism, immune response, and reproduction. Studies on the ruminal microbiota have provided strong evidence of heritability of specific microbial taxa (Li et al., 2019; Abbas et al., 2020). Host genetic effect on microbiota has been reported in across human ethnic groups (Zhou et al., 2007; Ravel et al., 2011), as well as in cattle breeds (Teng et al., 2024) and sheep (Greenwood et al., 2022). Microbiota Genome-Wide Association Studies (MGWAS) have identified host genetic loci associated with key reproductive taxa in humans (Mehta et al., 2020) and pigs (Sanglard et al., 2020). Comparable research on ruminal microbiota has demonstrated a strong genetic influence in cattle (Li et al., 2019; Zhang et al., 2020; Liu et al., 2021; Martínez-Álvaro et al., 2024) and sheep for dairy (Martinez Boggio et al., 2022, 2023) and morphological traits (Wang et al., 2023b). These findings collectively highlight the significant role of host genetics in shaping microbial communities across species. However, to date, studies specifically aiming on the influence of host genetics on sheep reproductive microbiota are absent.

The availability of single nucleotide polymorphism (SNP) information following the completion of the sheep genome allows now a deeper study of the genetic basis of complex traits (Brito et al., 2017). These tools also enable the exploration of links between host genome variations and microbiome diversity and abundance through MGWAS (Martínez-Álvaro et al., 2024). Identifying host genes linked to vaginal microbiota diversity and abundance that support fertility could provide new opportunities for selection breeding programs in sheep. Similarly, linking microbial taxa to fertility may open the door for probiotic supplementation to enhance reproductive success in ewes and rams (Quereda et al., 2020).

3.3. Evolution of microbiota studies

The study of microbiota has evolved from early microscopic observations to advanced sequencing-based approaches, allowing for a deeper understanding of microbial communities and their interactions with the host. Initially, microbiological research relied on culture-based techniques, which were limited by the fact that most microbes are unculturable under standard laboratory conditions (Staley and Konopka, 1985; Amann et al., 1995; Pace, 1997), with only about 1% being cultivable (Handelsman, 2004). The advent of molecular biology, particularly DNA sequencing,

revolutionised microbiological studies by enabling the identification of microorganisms independent of culture methods.

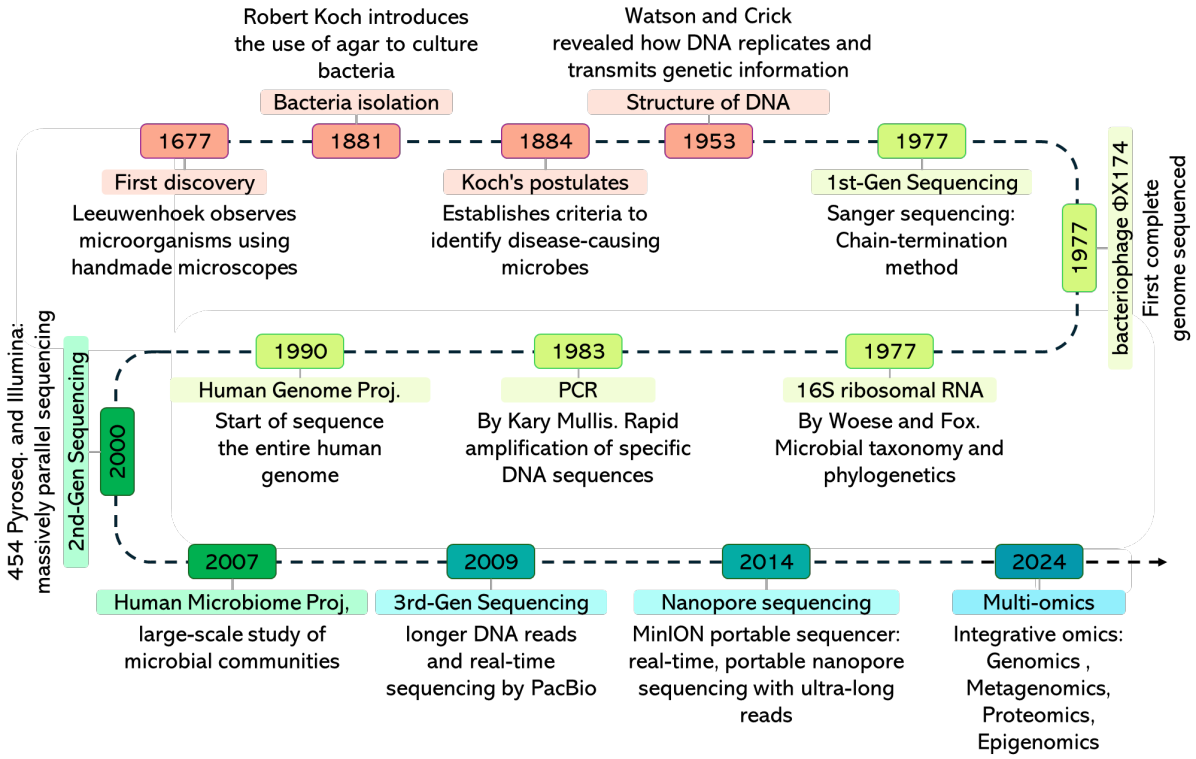


Figure 3.1. Historical timeline of microbiology and sequencing innovations. Hits prior to sequencing technology are marked in orange. First-generation sequencing events are highlighted in yellow-green, second-generation in green, and third generation in blue green.

The first breakthrough in sequencing was Sanger sequencing method developed in the 1970s, which allowed for the characterisation of individual genes, such as 16S ribosomal RNA (rRNA), a widely used marker for bacterial classification (Sanger et al., 1977; Woese and Fox, 1977). While this technique revolutionised microbial taxonomy, it was slow, expensive, and impractical for large-scale studies. To overcome these limitations, next-generation sequencing (NGS), also known as second-generation sequencing, emerged in the early 2000s. This technology introduced massively parallel sequencing, allowing millions of DNA fragments to be sequenced simultaneously, significantly reducing costs and increasing throughput (Wooley et al., 2010). Among NGS platforms, Illumina sequencing became the dominant method due to its high accuracy, scalability, and cost-effectiveness (Bentley et al., 2008; Quail et al., 2008). Based on sequencing-by-synthesis, Illumina produces short DNA reads (50–300 bp), making it the preferred approach for large-scale studies of microbial communities.

Despite the success of NGS, its reliance on short-read sequencing posed challenges for reconstructing complex genomes and identifying structural variations. This led to

the development of third-generation sequencing technologies, which offer long-read sequencing and real-time analysis, providing a more accurate representation of the original genetic material. These technologies can produce reads of thousands or even hundreds of thousands of base pairs in a single sequence. Pacific Biosciences (PacBio) introduced Single-Molecule Real-Time (SMRT) sequencing, while Oxford Nanopore Technologies developed a nanopore-based method capable of generating ultra-long reads (Eid et al., 2009; Jain et al., 2018). Under this method, a DNA molecule passes through a tiny pore in a membrane, causing changes in ion flow that generate unique electrical signals for each base. These advancements have improved metagenomic assemblies, facilitated the detection of epigenetic modifications, and expanded our ability to study microbial ecosystems.

Each method has its advantages and disadvantages, making them more suitable depending on the objectives of the study. Table 3.1 details the differences between various sequencing methods. The evolution of sequencing technologies has revolutionised microbiota research, enabling a shift from culture-based studies to metagenomic and metabarcoding approaches. These advancements have been particularly valuable for studying reproductive microbiota, where sequencing-based techniques provide new insights into the microbial contributions to fertility and reproductive health.

Table 3.1. Comparison of DNA sequencing technologies

	Sanger	Illumina	PacBio (SMRT)	Oxford ONT
Read Length	~800-1000 bp	50-300 bp	>10,000 bp	20 bp to >4 Mb, ultra-long reads possible
Throughput	Low (one read at a time)	High	Moderate	Moderate to High
Accuracy per Read	High	Very High	Very High	Moderate
Error Rate	<1%	<0.1%	<0.1%	~0.4-5%
Cost per Base	High	Low	Moderate	Moderate to Low
Speed	Slow	Fast	Moderate	Moderate to Fast
PCR Requirement	Yes	Yes	No	No
Complex Genome Assembly	Challenging	Challenging	Easier (due to long reads)	Easier (due to long reads)
Primary Technologies	Chain-termination method	Sequencing by synthesis	Single molecule real time sequencing	Real time Nanopore technology
Computing Requirement	Low	High	High	Very High
Portability	No	No	No	Yes

3.3.1. Current methods for studying microbiota

The development of high-throughput sequencing technologies has significantly improved the study of microbiota composition and abundance. Metagenomics (*sensu laxo*) has become the leading method for analysing microbial genomes in environmental samples, addressing the limitations of traditional culture techniques. While partial metagenomics (metabarcoding) targets genes like 16S ribosomal RNA (rRNA) to assess microbial diversity, its resolution is generally limited to the genus level rather than species-level identification (Poole et al., 2023), complete metagenomics allows to characterise all type of microorganisms and provides insights into gene functions (Quince et al., 2017). These methods have proven valuable for studying the microbiota of reproductive tracts and their potential impact on the success of AI in sheep.

3.3.1.1. Amplicon-based sequencing

Amplicon-based sequencing targets highly conserved and hypervariable regions of the microbial rRNA for molecular-level identification and classification. For Bacteria

and Archaea, this typically involves the 16S rRNA gene, which contains nine hypervariable regions (Regueira-Iglesias et al., 2023), while the internal transcribed spacer (ITS) region is used for Fungi (Antil et al., 2023). This methodology is widely employed due to its cost-effectiveness, efficiency, and high accuracy (Bravo-Egana et al., 2021).

The process involves PCR amplification of hypervariable regions with universal primers, and sequencing them on NGS platforms, with the most widely used being Illumina MiSeq, although Ion Torrent is also used to a lesser extent (Bravo-Egana et al., 2021). Bioinformatics pipelines —such as Mothur (Schloss et al., 2009) and QIIME2 (Bokulich et al., 2018), are employed for quality filtering, clustering operational taxonomic units (OTUs) or identifying amplicon sequence variants (ASVs), respectively, and taxonomic assignment. These analyses rely on databases like SILVA (www.arb-silva.de), Greengenes (rnacentral.org/expert-database/greengenes), or the Ribosomal Database Project (bio.tools/rdp).

Amplicon-based sequencing has certain limitations. For instance, it cannot detect viruses or eukaryotic microorganisms (Hugenholtz et al., 1998; Janda and Abbott, 2007). However, since most microorganisms in microbiota studies are bacteria (Pan and Yu, 2014; Laguardia-Nascimento et al., 2015; Lyman et al., 2019; Rowe et al., 2020; Ong et al., 2021; Pyles et al., 2021; Poole et al., 2023), this limitation has less impact. Additionally, it does not provide functional information about the microbial community, limiting insights into the metabolic capabilities of the organisms compared to whole genome sequencing (Regueira-Iglesias et al., 2023). Despite these constraints, a key advantage of 16S rRNA metabarcoding is its effectiveness in studies with low concentrations of microbial DNA. Studies employing amplicon-based sequencing have greatly enhanced our understanding of microbial diversity in multiple habitats (e.g. Zhang et al., 2023).

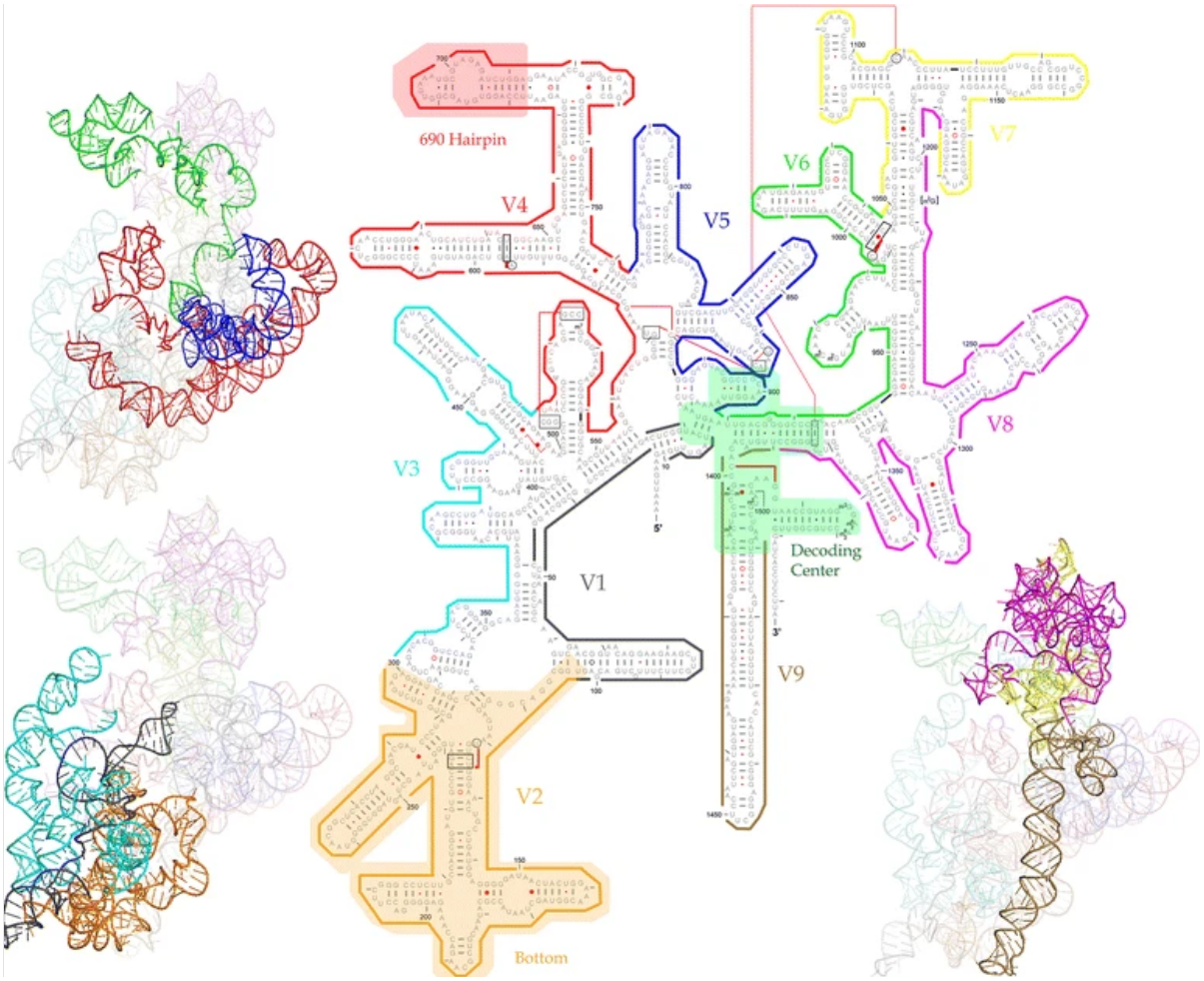


Figure 3.2. The 2D-3D 16S rRNA gene structure. Colour-coded regions illustrate variable sections (V1-V9) and key functional sites, including the 690 Hairpin (red), the Decoding Centre (green), and the structural base "Bottom" (yellow). This figure was reproduced from Yang et al. (2016).

3.3.1.2. Metagenomic sequencing

Metagenomic sequencing is a high-throughput approach used to analyse the collective genomes of microorganisms within a sample —Bacteria, Archaea, Fungi, Eucaryota and viruses.

Metagenomics typically use high-throughput platforms like Illumina NovaSeq, PacBio, and ONT. By sequencing the entire genetic content, metagenomics offers insights into metabolic pathways, antibiotic resistance genes, and the identification of novel microorganisms and previously uncharacterised genetic functions (Venter et al., 2004; Thomas et al., 2012; Parks et al., 2017; Quince et al., 2017).

Several pipelines have been developed to process metagenomic data and identify the diversity and abundance of microbial taxa. The first tool was the Basic Local Alignment Search Tool (BLAST). Because of its ability to align and compare DNA sequences (BLASTN) and proteins (BLASTX) against reference databases, researchers

could identify organisms and genes in complex samples (Altschul et al., 1990). Later, specialised tools like Kraken2 (Wood et al., 2019) and DIAMOND (Buchfink et al., 2014) were developed to optimise and speed up taxonomic and functional classification in large datasets.

Kraken2 is a taxonomic sequence classification tool that assigns taxonomic labels to DNA sequences using the lowest common ancestor (LCA) algorithm. It works by breaking each input sequence into smaller, fixed-length fragments called k-mers. These k-mers are compared against a pre-constructed nucleotide database, such as PlusPF-8, which maps each k-mer to the taxonomic group shared by all genomes containing that specific k-mer. By combining information from multiple k-mers in a query sequence, Kraken2 efficiently determines the most likely taxonomic classification, even for complex datasets. This approach allows for rapid and accurate taxonomic assignment of DNA sequences (Wood and Salzberg, 2014; Wood et al., 2019). This method is used in the EPI2ME workflow developed by ONT.

DIAMOND, on the other hand, is a highly efficient alignment software that translates nucleotide sequences into proteins and compares them against a protein database. It uses substitution matrices like BLOcks Substitution Matrix to calculate similarities between sequences and improve classification accuracy. Common databases include non-redundant protein database (NCBI nr, ncbi.nlm.nih.gov/protein/) for taxonomy, while *Kyoto Encyclopedia of Genes and Genomes* (KEGG, genome.jp/kegg/) and *Clusters of Orthologous Groups* (COG, ncbi.nlm.nih.gov/research/cog-project/) are used for functional annotation. DIAMOND is widely used in the scientific community for its high efficiency and low memory usage, achieving speeds up to 20,000 times faster than BLASTX (Buchfink et al., 2014). It integrates into metagenomic analysis workflows like SqueezeMeta (Tamames and Puente-Sánchez, 2019), which processes data from preparation to classification. With this approach, both read-based and assembly-based methods can be applied.

Despite its promising advantages, metagenomic sequencing faces significant challenges. It is more expensive and computationally intensive than amplicon-based methods due to the large volumes of data generated (Scholz et al., 2012; Thomas et al., 2012). Analysing metagenomic data r*equir*es advanced bioinformatics expertise and substantial computational power (Sharpton, 2014; Sczyrba et al., 2017). The presence of host DNA in samples can complicate data analysis; in some cases, host DNA represents more than 95% of the DNA in the sample (Schmieder and Edwards, 2011), and assembling genomes from complex communities with high diversity can be challenging (Ayling et al., 2020). Additionally, interpreting functional data can be difficult due to the presence of hypothetical proteins and genes with unknown

functions (Yandell and Ence, 2012). The quality of metagenomic databases is lower than that of specialised 16S rRNA databases, especially protein databases. 16S rRNA databases like SILVA and Greengenes are highly curated and focused on precise taxonomic identification of Bacteria and Archaea (Werner et al., 2012). In contrast, k-mer databases and protein databases cover a broader spectrum of organisms and functions but tend to be less curated, with many sequences derived from automatic annotations (Steinegger and Söding, 2018).

Table 3.2. Key differences between 16S rRNA metabarcoding and metagenomic sequencing approaches

	16S rRNA metabarcoding	**Metagenomics**
Target	Specific regions of the 16S rRNA gene	All genetic material
Sequencing Depth Required	Moderate	High
Organisms Detected	Bacteria and archaea	Bacteria, archaea, viruses, fungi, and other eukaryotes
Resolution	Phylum to genus level; OTUs or ASVs as operational units	Kingdom to species level; strain level if high coverage and quality
Functional Information	Taxonomic classification only	Taxonomy and gene functions
Database Requirement	SILVA, Greengenes (Highly curated)	NCBI nr (Less curated for taxonomic assignment), KEGG, COG
Bioinformatics Tools	Mothur, QIIME2, DADA2	BLAST, Kraken2, DIAMOND, SqueezeMeta
Cost	Lower	Higher
Computational Requirement	Moderate	High
PCR Requirement	Yes	No
Bias	PCR bias due to primer selection	Lower, as there is no targeted amplification
Data Analysis Complexity	Moderate	High (assembly-based or read-based approaches)
Host Contamination	Low to Moderate (mitochondrial 16S rRNA gene)	Low to very high (depends on the sample region)
Primary Sequencing Platforms	Illumina MiSeq, Ion Torrent	Illumina NovaSeq, PacBio, Oxford Nanopore

3.3.2. Microbiome data and the problem of compositionality

In microbiome studies, sequencing data are processed to assign reads to specific features (e.g. ASVs, OTUs, taxa, or genes), producing a count matrix. Rows represent these features, columns represent samples, and each value reflects the number of reads associated with a specific feature in a sample. However, these raw counts are influenced by sequencing depth, which can vary between samples and make direct comparisons unreliable. To allow for meaningful comparisons between samples, data

are often normalised. One common approach is to transform counts into relative abundances (RA) by dividing each taxon's count by the total count for that sample. In this way, the data become compositional, as the values are constrained to sum to one (or 100%), making them interdependent. As a result, an increase in the relative abundance of one taxon will necessarily reduce the relative abundances of others, even if their absolute values remain unchanged. This interdependence can alter the original data, potentially distorting the true relationships between taxa and complicating statistical analysis (Figure 3.3). To address this, other type of normalisation, such as log-ratio transformations are applied. These methods adjust for the compositional nature of microbiome data, maintaining the relative information among the different components of the sample, making them more suitable for robust statistical analyses (Gloor et al., 2017).

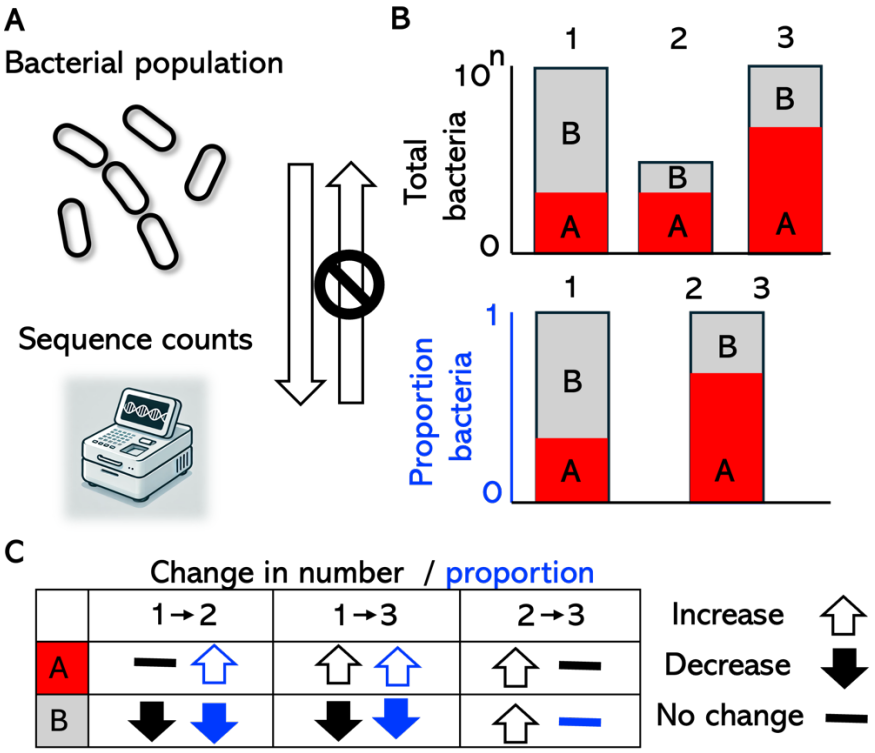


Figure 3.3. Compositional nature of sequencing data and its impact on abundance interpretation. **A.** Illustration that sequencing cannot provide absolute abundances, only proportions based on sequencing depth. **B.** Absolute counts (top) compared to relative abundances (bottom), illustrating that relative proportions can appear identical across samples despite different absolute counts. **C.** Table illustrating that compositional nature can misrepresent changes when transitioning between samples. Figure adapted from Gloor et al. (2017).

The most widely used transformations are Centered Log-Ratio (CLR), Additive Log-Ratio (ALR), and Isometric Log-Ratio (ILR). CLR transformation adjusts each value in relation to the geometric mean of all taxa within a sample, providing a balanced view of the data. This approach maintains symmetry, making data suitable for many statistical analyses. However, this transformation can still be challenging to interpret biologically and r*equi*res handling zero counts through imputation, as zeros are incompatible with logarithmic calculations. ALR transformation, in the other hand, expresses each taxon in relation to a chosen reference taxon, simplifying comparisons in some cases. It allows flexibility but also introduces potential biases, as the choice of the reference taxon can influence the results. Careful selection of the reference taxon is essential to avoid distortions (Greenacre et al., 2021). ILR transformation uses orthogonal balances to create statistically independent comparisons between groups of taxa, resulting in transformed data that are ideal for advanced multivariate analyses. It is mathematically rigorous but more complex to interpret, as the transformed values do not correspond directly to individual taxa (Aitchison, 1982).

4. MATERIALS AND METHODS

Four independent studies were conducted to address the specific objectives outlined previously. Some studies shared animal cohorts and general procedures, but each one used different samples, DNA extraction protocols, and/or statistical analyses, depending on the study design, the animals available, and the resources at that time. Figure 4.1 shows the workflow, which is described in more detail below.

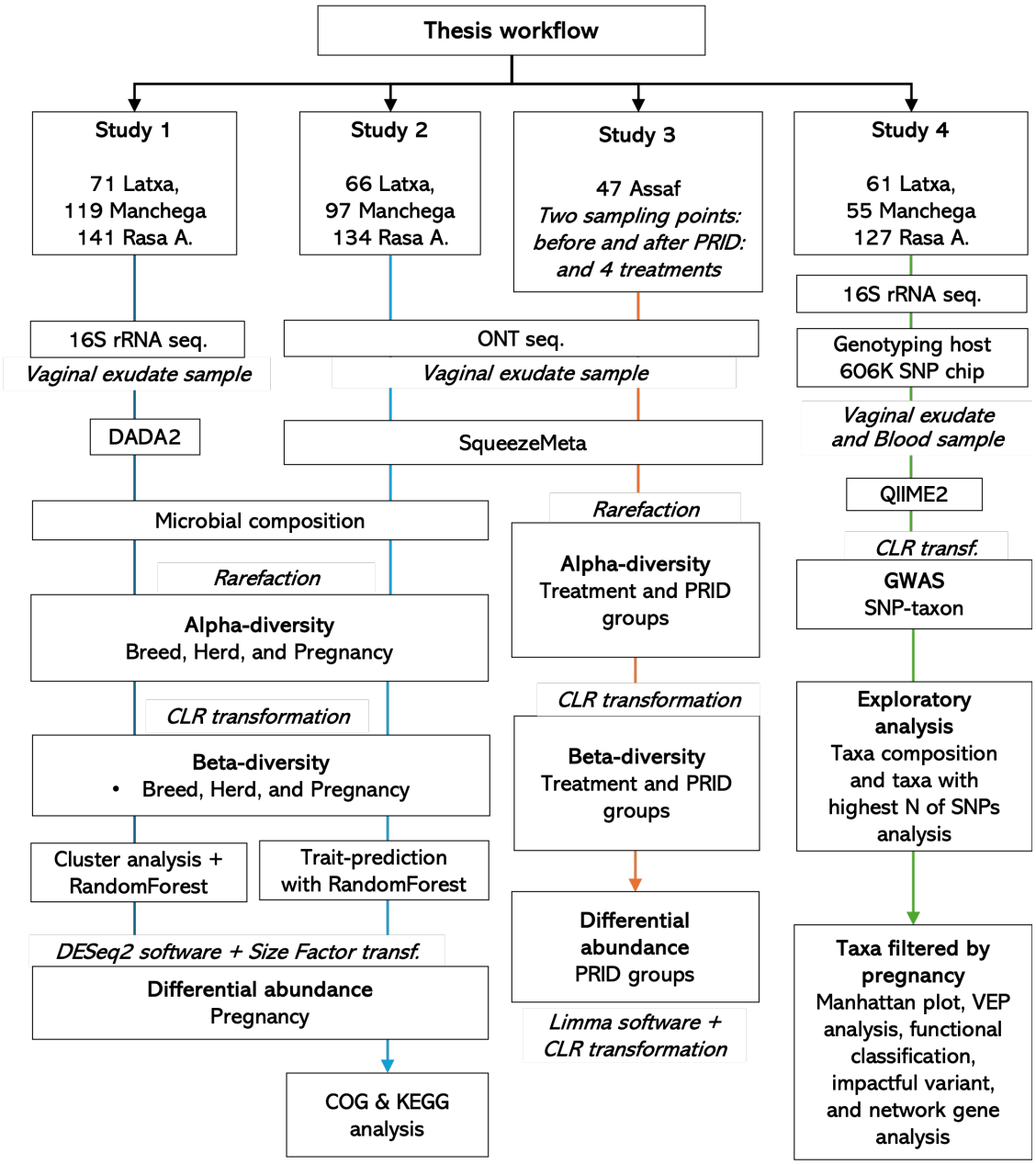


Figure 4.1. Workflow of study designs and analytical approaches based on the thesis objectives. The figure illustrates the methodological workflow used in this thesis.

4.1. Summary of study designs

Study 1 examined the vaginal microbiota of 331 ewes from three breeds, Latxa, Manchega, and Rasa Aragonesa, coming from four herds. Vaginal exudate samples were collected prior to AI and analysed using 16S rRNA metabarcoding to characterise bacterial diversity at multiple taxonomic levels (ASV, genus, species, and phylum). Core bacterial identification, alpha- and beta-diversity, and clustering approaches were conducted. Differential abundance analysis using DESeq2 was performed to identify taxa associated with pregnancy outcomes.

Study 2 focused on a subset of 297 ewes from **Study 1**, using ONT-based metagenomics to expand the analysis to functional genes such as COG and KEGG orthologues (KO). Core microbiota, alpha- and beta-diversity analyses were conducted, and differential abundance analysis identified taxa and genes associated with pregnancy. Network models were applied to link functional pathways with key taxa. Sequencing was conducted on a GridION platform, applying adaptive sampling to exclude host DNA, and the SqueezeMeta pipeline was used for taxonomic and functional annotation.

Study 3 evaluated the impact of intravaginal devices with different additives (probiotic, antibiotic, maltodextrin, and control) on the vaginal microbiota of 47 Assaf ewes sampled at two time points: before PRID insertion (S1) and after PRID removal (S2). Taxonomic profiling was performed using ONT-based metagenomics on a MinION platform, and alpha- and beta-diversity indices were compared across treatments and time points. Differential abundance analysis using Limma identified taxa with significant changes between S1 and S2.

Study 4 investigated the relationship between host genomic variants and vaginal microbiota composition in 243 ewes from **Study 1**. Genotype data were obtained using the Illumina 606K SNP chip and combined with MGWAS analysis to identify host SNPs associated with microbial abundance at species, genus, and phylum levels. Functional relevance of significant SNPs was evaluated using Gene Ontology (GO) annotations, focusing on their effects on taxa linked to pregnancy.

4.2. Animal samples

All animal manipulations complied with Spanish Policy for Animal Protection RD 53/2013 and the European Union Directive 2010/63/EU on the protection of animals used for scientific purposes.

A total of 382 multiparous ewes were included in this study. These animals were from four different breeds with distinct production purposes: Latxa and Assaf (dairy breeds), Rasa Aragonesa (meat breed), and Manchega (dual-purpose breed for milk and

meat). Each group was raised under specific production systems and environmental conditions across various regions of Spain (Figure 4.2). Specifically:

- 71 Latxa ewes were sampled from a herd in Vitoria, Basque Country (Herd L). They were raised in a semi-extensive dairy system, with frequent access to pasture and a relatively mild, humid climate. The average age for this group was 3.01 $\pm$ 1.63 years.
- 141 Rasa Aragonesa ewes were sampled from a herd in Zaragoza, Aragón (Herd R). These animals were kept under a semi-extensive meat production system, with a mix of grazing and supplementary feeding. The average age for this group was 5.71 $\pm$ 2.7 years.
- The Manchega group included 119 ewes from two herds based in Valdepeñas, Castilla-La Mancha. Of these, 60 animals belonged to the RN herd, a non-commercial flock managed by a breeders' association, while the remaining 59 came from the VL herd, a commercially operated flock (where RN and VL correspond to specific farm designations). Ewes were managed under an intensive system, with limited access to pasture. The average age for RN herd was 2.58 $\pm$ 0.94 years, whereas for VL herd was 3.46 $\pm$ 1.41 years.
- 47 Assaf ewes from a farm located in Zamora, Castilla y León. Ewes were managed under an intensive dairy system, with no access to pasture. Feeding was based on formulated rations. The average age for this group was 2.7 $\pm$ 1.15 years.

Age and parity were categorised into five groups to investigate potential non-linear effects. Age category included: <2 years, 2–3, 3–4, 4–5, and >5 years, while parity was categorised as: <2, 2–3, 3–4, 4–5, and >5 parturitions.

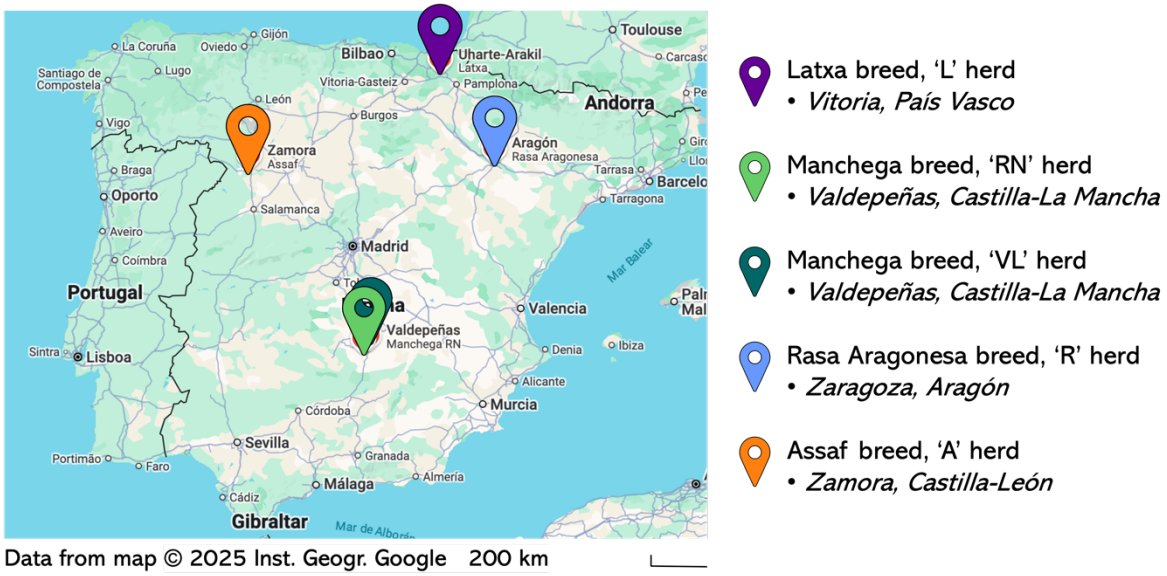


Figure 4.2. Farms locations of the four herds.

All ewes underwent oestrus synchronisation using PRIDs containing 20 mg of flurogestone acetate (Chronogest, MSD Animal Health, Kenilworth, NJ, USA). For Latxa ewes, the PRIDs also included 0.6 g of powdered framycetin (neomycin sulphate, Framicas, Laboratorios Ovejero, Spain). In the Assaf breed, animals were divided into four groups, and in each group, the PRID received a different treatment (see Table 4.1 for further details). After 14 days, PRIDs were removed, and ewes were injected with 300–500 mg of pregnant mare's serum gonadotropin to stimulate ovulation. Cervical AI was performed 53–55 hours post-PRID removal. Vaginal exudate samples were collected using sterile swabs (Real Vaginal Microbiome DNA Kit, Durviz S.L., Valencia, Spain) before PRID insertion and before insemination. Samples were immediately frozen on dry ice and stored at –80°C. All ewes had lambed at least once, were in good body condition, and showed no signs of illness or infection at the time of insemination. AI for Assaf took place on august 2019, whereas for Laxta, Manchega, and Rasa Aragonesa took place along August–September 2020.

Semen was collected from 39 rams across breeds: 13 Rasa Aragonesa, 12 Latxa, 10 Manchega, and 4 Assaf, aged 4–7 years. Rams were managed under standardised conditions at AI centres and underwent quality control, including minimum motility scores (>3.5) and sperm concentrations of 300–400 million/mL. For Rasa Aragonesa, Manchega, and Assaf, semen was diluted in INRA96® (IMV Technologies, L'Aigle, France) supplemented with antibiotics penicillin, gentamicin and amphotericin B and packaged in 0.25 mL straws. For Latxa, powdered skimmed milk plus streptomycin, penicillin, and sodium sulphanilamide was used as the diluent.

Pregnancy diagnosis was assessed by transabdominal ultrasound performed between 37-55 days post insemination. Additionally, for Manchega and Rasa breeds, a progesterone test was also performed.

Blood samples for host genotyping were drawn from the jugular vein into vacutainer® EDTA tubes. Samples were stored at 4°C during transport and stored at –20°C until DNA extraction.

Table 4.1. Study-specific sampling details.

Study	N	Breed	PRID-treatment	Microbiome data	Genotype
1	331	71 Latxa, 119 Manchega (60 RN and 59 VL), and 141 Rasa Aragonesa.	No	16S rRNA metabarcoding	No
2	297	66 Latxa, 97 Manchega (56 RN and 41 VL), and 134 Rasa Aragonesa.	No	ONT-based metagenomics	No
3	47	47 Assaf sampled at two time points: S1 (before PRID insertion) and S2 (after PRID removal, just prior to AI).	Probiotic group (n = 13); Maltodextrin group (n = 10); Antibiotic group (n = 13); and Control group (n = 11) *	ONT-based metagenomics	No
4	243	61 Latxa, 55 Manchega RN, and 127 Rasa Aragonesa	No	16S rRNA Metabarcoding	Yes

* Treatments were established as follows: Probiotic group: PRIDs with 200 mg lyophilised *Lactobacillus* rhamnosus used as probiotic plus maltodextrin as excipient. Maltodextrin group: PRIDs with 200 mg of maltodextrin (the common excipient used in the probiotic preparation). Antibiotic group: PRIDs with 0.6 g powdered framycetin. Control group: PRIDs without additives. At insemination, a complete factorial design was implemented to balance the number of ewes inseminated with semen from the four available rams.

4.3. DNA extraction

4.3.1. Microbial DNA extraction

Two different protocols were used for vaginal microbiota DNA extraction:

The **Real Vaginal Microbiome DNA Kit** (Durviz S.L., Valencia, Spain) (**Studies 1, 2,** and **4**), following the manufacturer's protocol. Swabs were placed in 2 mL microtubes with 900 µL of CTAB extraction buffer and 25 µL of proteinase K, incubated at 70 °C for 10

minutes, bead beaten at maximum speed for 10 minutes, and centrifuged at 14,000 rpm for 5 minutes. The lysate (900 µL) was mixed with 250 µL of binding buffer, loaded onto silica-membrane spin columns, washed with Disinhibition and Wash Buffers (500 µL and 700 µL), and DNA was eluted in 100 µL of pre-heated Elution Buffer (70 °C).

The **QIAamp DNA Microbiome Extraction Kit** (Qiagen Inc., Valencia, CA, USA) (**Study 3**), also following the manufacturer's protocol. Swabs were placed in 1 mL of phosphate buffered saline medium, and host DNA was depleted with buffer AHL and benzonase. Lysis was carried out in Pathogen Lysis Tubes with buffer Animal Tissue Lysis, a detergent-based buffer designed to disrupt cellular and nuclear membranes, facilitating the release of nucleic acids. Mechanical lysis was achieved using a FastPrep®-24 instrument, followed by proteinase K treatment. The lysate was mixed with buffer APL2 and ethanol, applied to QIAamp UCP mini spin columns, washed with buffers AW1 and AW2, and DNA was eluted in 50 µL of buffer AVE.

DNA concentration and purity in all samples were measured using a Qubit 4 fluorometer and a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, DE, USA).

4.3.2. Host DNA extraction

Host DNA was extracted from the whole blood samples using the MasterPure DNA Purification Kit for Blood Version II (Epicentre®), following the manufacturer's instructions. DNA samples were resuspended in water and were measured (concentration and quality) using Nanodrop 2000® spectrophotometer. A total of 1500 ng in 30 µL per DNA sample were then used for genotyping.

4.4. Sequencing

4.4.1. Amplicon-based 16S rRNA metabarcoding

DNA samples were used to generate libraries targeting the V3-V4 regions of the bacterial 16S rRNA gene. The sequencing was conducted on an Illumina MiSeq platform using a 2 × 300 bp paired-end run, performed by the Instituto de Parasitología y Biomedicina López-Neyra (IPBLN-CSIC, Granada, Spain) (**Studies 1** and **4**).

4.4.2. ONT-based metagenomic sequencing

Microbial DNA sequencing was performed using nanopore technology with two distinct platforms:

GridION™ Mk1 DNA/RNA Sequencing Device. DNA (400 ng in 11 µL) was prepared following the Ligation Sequencing gDNA Native Barcoding Kit 24 V14 (SQK-NBD114.24) protocol. Multiplexing was achieved with 24 barcodes per run across 13 runs, including negative controls. Barcoded samples were pooled into 1.5 mL Eppendorf LoBind tubes for adapter ligation and sequenced on R10.4.1 flow cells. Adaptive sampling —a real-time selective sequencing approach that enriches target DNA by dynamically rejecting molecules aligning to a reference genome (Marquet et al., 2022)— was implemented to exclude sequences aligning with the *Ovis aries* reference genome (Oar_rambouillet_v1.0, GCA_002742125.1) (**Study 2**).

MinION™ Mk1C DNA/RNA Sequencing Device. DNA (1 µg) was processed using the SQK-LSK109 ligation sequencing kit and multiplexed with the 1D Native Barcoding genomic DNA kit (EXP-NBD104 and EXP-NBD114). Barcoded samples (700 ng total DNA) were pooled in 1.5 mL Eppendorf LoBind tubes and sequenced on R9.4.1 flow cells across eight runs. Unlike the GridION workflow, adaptive sampling was not applied for these samples because it was not available for MinION at the time the study was performed. Both platforms leveraged ONT technology to provide high-resolution insights into the microbial communities (**Study 3**).

4.5. Bioinformatic analysis

4.5.1. 16S rRNA metabarcoding analysis

In **Study 1**, the V3-V4 regions of the 16S rRNA were processed using the DADA2 package (Callahan et al., 2016) in R (R Core Team, 2023). Primers 341F (5'-CCTACGGGNGGCWGCAG-3') and 785R (5'-GACTACHVGGGTATCTAATCC-3') were removed with Cutadapt software (Martin, 2011) to minimise biases in taxonomic classification. Reads were trimmed to 240 bp for forward and 200 bp for reverse sequences, retaining only bases with a quality score >2, and paired-end reads were merged with a minimum overlap of 12 bp. The DADA2 algorithm identified ASVs, which were classified with a Naive Bayesian classifier trained on the SILVA v132 database (silva_nr_v132_train_set.fa.gz) specific to the V3-V4 regions. In **Study 4**, the same processing parameters were applied, with some subtle differences: DADA2 was used within the QIIME2 platform, and taxonomic assignment was performed using an actualised SILVA nr99_v138.1 database (silva_nr_v138_train_set.fa.gz), in line with the version available at the time of analysis.

For 16S rRNA data, ASVs unclassified at the family level or mapped to mitochondrial 16S rRNA sequences from non-bacterial organisms (e.g., mitochondrial DNA), were excluded.

4.5.2. ONT-based metagenomic analysis

As previously described, metagenomic data were generated using two Nanopore platforms: GridION (**Study 2**) and MinION (**Study 3**). While both datasets were processed following a similar pipeline, platform-specific adaptations were implemented. The following sections describe the data processing steps applied to each platform in detail.

For the data obtained from GridION, basecalling was performed using Guppy (version 6.4.6), converting electrical signals into DNA sequences. The Super Accuracy algorithm was used to ensure a minimum sequence quality of Q10 (Wick et al., 2019), and barcode trimming was applied. Reads shorter than 200 bp were discarded using Cutadapt, and sequencing quality was assessed with NanoPlot. Host DNA was removed by mapping sequences against the *Ovis aries* reference genome using Minimap2. A custom Python pipeline developed by the authors was employed to filter out host-target sequences. Retained reads were processed with the SqueezeMeta v1.6.2 pipeline, including taxonomic assignment via Diamond Blastx against the NCBI-nr, COG, and KO databases. Taxonomic consensus was determined using a LCA algorithm, and gene functions were annotated based on default score thresholds: 60 (genus), 55 (family), 50 (order), 46 (class), and 42 (phylum). Reads below these thresholds were classified as unclassified taxa.

For the data obtained with MinION, basecalling was performed with Guppy (version 4.2.2) using the High Accuracy mode. Reads shorter than 200 bp were discarded, and quality assessment was performed with NanoPlot. Host DNA removal was carried out by mapping sequences to the same *Ovis aries* reference genome but using BBMap software for this step. A notable difference was that retained reads were trimmed to a standard length of 6000 bp to meet software *requi*rements. Subsequent analyses were conducted with the SqueezeMeta v1.3.0 pipeline.

In both datasets, reads from the Bacteria and Archaea domains were retained, along with sequences from Protista (Apicomplexa, Euglenozoa, and Evosea), Fungi (Ascomycota), Animalia (Nematoda), and viral entities.

4.5.3. Host genotype data processing

High density genotype data were obtained (**Study 4**), using the Illumina 606K SNP chip. Quality control measures included a call rate of 95% and a minor allele frequency (MAF) threshold of 5%. The genome assembly was also updated from the Oar_v3.1 (GCF_000298735.1) to the ARS-UI_Ramb_v2.0 (GCF_016772045.1) to improve mapping accuracy.

4.6. Statistical analysis

4.6.1. Microbial composition

Regardless the sequencing technology, microbial composition was analysed at different taxonomic levels, focusing the graphic representations on phylum and genus levels. Core taxa were defined as those present in at least 90% of the samples (**Studies 1** and **2**).

The microbiota composition was visualised using the ggplot2 and base packages in R (R Core Team, 2023).

4.6.2. Alpha-diversity

Alpha-diversity is a term originally described within the field of ecology, used to describe the "within-sample" diversity (Cassol et al., 2025). In microbiome studies, it is a measure of how diverse a single sample is, usually taking into account the number of different microbial taxa observed. Alpha-diversity metrics are often weighted by the abundances at which the individual microbes are observed. There are a number of commonly used metrics to measure alpha-diversity in the literature (Finn, 2024; Cassol et al., 2025), with different assumptions and interpretation. A selection of the most used metrics in microbiome studies include:

Observed richness. The simplest alpha-diversity metric. It measures the total number of unique taxa within a sample (S_{obs}). This metric can take values from 0 to an infinite number of taxa.

Chao1. A non-parametric estimator based on observed richness with an added term that infers the number of taxa not observed –due to sequence technique constraints– but present in the sample according to a rate between the number of singletons (species observed once, F_1) and doubletons (species observed twice, F_2) (Chao, 1984),

$$S_{Chao1} = S_{obs} + \frac{F_1^2}{2F_2}$$

Shannon's Diversity Index. A measure based on the Information theory that considers both taxa richness and evenness, indicating how evenly microorganisms are distributed among the different species (Shannon, 1948). This metric quantifies the entropy, or uncertainty, of predicting which species will be encountered next, based on the number of taxa present and how evenly their abundances are distributed,

$$H' = -\sum \ln (p_i) p_i$$

where p_i is the proportional abundance of the i^{th} taxon. The primary challenge lies in interpreting the value, as it is difficult to determine whether a given variance is due to species richness, dominance, or a combination of both in the sample.

Inverse Simpson index. The Simpson's index (*D*) is based on the evenness of taxa within a sample and measures the probability that two microorganisms randomly selected from a sample belong to the same species (Simpson, 1949),

$$D = \sum p_i^2$$

where p_i is the proportional abundance for each species in the sample. In this way, 0 represents infinite diversity and 1, no diversity. That is, the bigger the Simpson metric, the lower the diversity. As this is unintuitive, the value is usually expressed as its inverse value (1/D).

Before alpha-diversity was estimated, rarefaction was performed to ensure a consistent sequencing depth across all samples. Rarefaction is a commonly used method in microbiome studies that involves randomly subsampling the number of sequencing reads in each sample down to a fixed threshold. This helps to control for differences in sequencing effort, which could otherwise bias diversity estimates (Callahan et al., 2016; Bokulich et al., 2018). In this thesis, thresholds of 20,000 reads (**Study 1**) and 6,000 reads (**Study 2**) were chosen. These values were selected because rarefaction curves showed that most samples reached a plateau at these depths, suggesting that the sequencing coverage was sufficient to capture the majority of microbial diversity present in the samples (Appendix – Figure S1.5 and Figure S2.5).

Alpha-diversity measures were compared among breeds, herds, and pregnancy groups (**Studies 1** and **2**), and between sampling points (S1–S2), treatment, and pregnancy groups (**Study 3**).

Alpha-diversity indices were calculated using the Phyloseq package (McMurdie and Holmes, 2013).

4.6.3. Beta-diversity

Beta-diversity describes the variation in microbial community composition between samples and provides insight into how communities change across environments, treatments, or conditions (van der Plas et al., 2023). Unlike alpha-diversity, which measures diversity within a sample, beta-diversity focuses on the (dis)similarity between pairs of samples, taking into account the presence, absence, and/or abundance of microbial taxa (Baselga and Gómez, 2019; Kers and Saccenti, 2022). An overview of the methods applied in this thesis is presented below.

Principal Component Analysis (PCA). A statistical method used to reduce the complexity of multidimensional data while preserving as much variation as possible. Originally introduced by Pearson (1901) and later formalised by Hotelling (1933), PCA transforms the original dataset into a new set of variables —called principal components— which are uncorrelated and ordered by the amount of variance they capture. This method is commonly applied in microbiome studies to summarise community-level differences and identify sample groupings or gradients related to biological or environmental variables. The analysis begins by centring each variable (e.g., microbial taxon abundance) by subtracting its mean from all observations (e.g., individual samples), such that the empirical mean of each variable is equal to zero. This step produces a matrix $\mathbf{X}$, where rows represent samples and columns correspond to the newly centred variables. From $\mathbf{X}$, the covariance matrix is then computed as:

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$$

where n is the number of samples and $\mathbf{X}^{\mathrm{T}}$ denotes the transpose of the centred data matrix. The eigenvectors of $\mathbf{S}$ define the directions of greatest variance (the principal components), and the original data are projected onto these new axes to obtain the transformed dataset:

$$\mathbf{Z} = \mathbf{XW}$$

Where $\mathbf{W}$ is the matrix of eigenvectors, and $\mathbf{Z}$ contains the coordinates of the samples in the principal component space.

Non-Metric Multidimensional Scaling (NMDS). Originally introduced by (Kruskal, 1964a; b), is an ordination technique used to represent complex patterns of dissimilarity between samples in a low-dimensional space, based on the rank order of distances rather than their absolute values. Unlike PCA, NMDS makes no assumptions about linearity, making it particularly suitable for ecological and microbiome data. The first step is to estimate the dissimilarity matrix, which quantifies how different each pair of samples is, based on their microbial composition. These dissimilarities are calculated mainly through two distances measures, Euclidean distance and Bray-Curtis.

Euclidean distance calculates the straight-line difference between two samples by comparing the abundance of each taxon (Ricotta, 2021). It is defined as:

$$d_{ij}^{Eu} = \sqrt{\sum_{k=1}^{p}(x_{ik} - x_{jk})^2}$$

where x_{ik} and x_{jk} are the abundances of taxon k in samples i and j, respectively, and p is the total number of taxa.

Bray-Curtis, by contrast, considers both the presence and abundance of taxa. It focuses on the shared counts between samples relative to their total abundance (Ricotta and Podani, 2017) and is defined as:

$$d_{ij}^{Bray} = 1 - \frac{2\sum_{k=1}^{p} min(x_{ik}, x_{jk})}{\sum_{k=1}^{p}(x_{ik} + x_{jk})}$$

where the numerator represents the shared abundance between samples, while the denominator represents the total abundance. A Bray–Curtis value of 0 means the two samples are identical, while a value of 1 indicates they share no taxa.

Once the dissimilarity matrix is constructed, NMDS uses an iterative algorithm to arrange the samples in a reduced-dimensional space, typically two dimensions. The quality of this configuration is evaluated using a stress function, which measures the disagreement between the distances in the reduced space ($\hat{d}_{ij}$) and the original dissimilarities (d_{ij})

$$\text{Stress} = \sqrt{\frac{\sum_{i<j}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{i<j} d_{ij}^2}}$$

Lower stress values indicate a better representation of the data, with values below 0.1 generally considered to reflect a good fit.

Permutational Multivariate Analysis of Variance (PERMANOVA). A non-parametric statistical method used to test for differences in multivariate data across predefined groups (Garrido-Martín et al., 2023). Unlike traditional ANOVA, which relies on raw data and assumptions of normality, PERMANOVA operates on a dissimilarity matrix (e.g., Bray–Curtis or Euclidean), making it well suited for ecological and microbiome datasets (Gail et al., 2021).

The method partitions the total variation in the data into within-group and between-group components and evaluates whether the centroids of different groups differ significantly in multivariate space. This is quantified using a pseudo-F statistic, calculated as:

$$F = 1 - \frac{SS_{between}/(k-1)}{SS_{within}/(N-k)}$$

Where $SS_{between}$ and SS_{within} are the sums of squares between and within groups, K is the number of groups, and N is the total number of observations. Statistical significance is assessed by repeatedly permuting the group labels and recalculating the F-statistic to generate a null distribution. The proportion of permutations with an F-value greater than or equal to the observed value determines the p-value. This procedure allows PERMANOVA to test for group differences without relying on parametric assumptions.

Data normalisation. To ensure comparability across samples and appropriately address the compositional nature of microbiota data, normalisation is a crucial pre-processing step. Among the available approaches, CLR is particularly suitable as it respects the relative structure of the data, reduces the influence of scale differences, and improves comparability across samples by transforming values relative to their geometric mean (see section State of the Art). The CLR transformation of a composition vector $\mathrm{x} = [x_1; x_2; \ldots, x_D]$ is defined as:

$$\mathrm{CLR}(\mathrm{x}) = \left[\log\left(\frac{x_1}{g(\mathrm{x})}\right), \log\left(\frac{x_2}{g(\mathrm{x})}\right), \ldots, \log\left(\frac{x_D}{g(\mathrm{x})}\right)\right]$$

Where $g(\mathrm{x}) = \left(\prod_{i=1}^{D} x_i\right)^{1/D}$ is the geometric mean of the components in x.

Since the logarithm is undefined for zero, CLR requires strictly positive values. Therefore, zero counts —common in microbiome datasets— must first be addressed. This is typically achieved by either adding a small pseudo-count or employing more robust approaches such as Bayesian multiplicative replacement, which estimates plausible values for zero entries without overly distorting the data distribution.

Building on the previously described statistical methods, beta-diversity analyses were conducted using PCA and PERMANOVA to explore differences in microbial communities across breeds, herds, age, parity, and pregnancy groups (**Studies 1** and **2**), as well as across PRID treatments and sampling points (**Study 3**). Pregnancy-related differences were examined both collectively across all samples and within individual herds. Pairwise comparisons were included in all PERMANOVA analyses to assess differences between specific group pairs. CLR transformation was applied to all beta-diversity analyses. Prior to this, zero values were imputed using a Bayesian multiplicative replacement method.

Prior to beta-diversity analyses, no abundance filtering was applied before compositional transformation in **Studies 1** and **2**, as their larger sample sizes and higher sequencing depths helped mitigate the effects of zero inflation. In contrast, a 0.05% relative abundance filter was applied in **Study 3** to minimise noise from low-abundance features.

Principal Component Analysis was conducted using the stats and FactoMineR R packages, PERMANOVA was performed with the vegan package (Oksanen et al., 2022), and the RVAideMemoire package (Hervé and Hervé, 2020) was used for pairwise comparisons. CLR transformation was applied using the microbiome and easyCODA R packages (Lahti and Shetty, 2018; Maintainer and Boutros, 2018).

4.6.4. Clustering and prediction

Prediction refers to the process of using existing data to estimate or classify unknown outcomes. In microbiome research, predictive models can be applied to assess whether microbial community patterns are associated with specific phenotypic traits. In this study, two complementary approaches were implemented to explore and predict phenotypic traits using vaginal microbiota composition. Although both relied on the same dataset, each employed distinct methodologies: the first (in **Study 1**) focused on clustering microbial profiles derived from 16S rRNA metabarcoding, whereas the second (in **Study 2**) utilised ONT-based metagenomic data to predict phenotypic variables. In both approaches, the Random Forest (RF) algorithm was used as the predictive model due to its robustness and flexibility in handling high-dimensional data. For model training and more robust results, cross-validation (CV) was also implemented to assess model performance and generalisation. The following sections first describe the principles of the RF algorithm and the CV procedure, followed by the specific methodological steps applied in each predictive approach.

Random Forest. A supervised machine learning algorithm introduced by Breiman (2001), which makes predictions by aggregating the outputs of multiple decision trees. A decision tree is a predictive model that splits the data through a sequence of decision rules based on variable thresholds. Its structure resembles a branching tree, where each internal node represents a condition, and each terminal node corresponds to a prediction. In a RF, each tree is trained on a different random subset of the data and variables, resulting in a diverse set of trees. Together, these trees form the "forest", where their combined output introduces variability and reduces the risk of overfitting. The final prediction is made by majority vote in classification tasks or by averaging the predictions in regression (Reinoso-Peláez et al., 2022). The model can be built as follows:

$$\hat{y} = \mu + \sum_{t=1}^{T} c_t h_t(\mathrm{x}_n; V_{m,t})$$

Where $\hat{y}$ is the predicted outcome for observation n; μ is the overall intercept; T is the total number of trees in the forest; c_t is a shrinkage factor applied to tree t (commonly $c = 1/T$); $h_t(\mathrm{x}_n; V_{m,t})$ is the prediction from the t-th decision tree, where x_n is the vector of predictor values for observation n, and $V_{m,t}$ is a random subset of m predictor variables. Typically, $m = \sqrt{p}$ for classification tasks or $m = p/3$ for regression, where p is the total number of predictors. Each decision tree is trained using bootstrap aggregation (bagging), where random subsets of the training data and predictor variables are used to increase model diversity and reduce overfitting. The final model aggregates the outputs of all trees by majority vote (for classification tasks) or by

averaging (for regression tasks). Model performance is typically evaluated using out-of-bag samples, which refer to observations that are not included in the bootstrap sample for a given tree.

Cross-validation. A method used to evaluate how well a predictive model performs on new, unseen data. It works by splitting the dataset into subsets: some are used to train the model, while the remaining are used to test it. This process is repeated several times to obtain a more reliable estimate of the model's predictive performance (Sweet et al., 2023). One of the most widely used and robust approaches is 5-fold cross-validation, performed as follows: the dataset is randomly divided into five subsets of equal size. In each fold, four subsets are used for training and the remaining one for testing. The performance of the RF model is then assessed by calculating the overall classification accuracy for categorical outcomes, or the Pearson correlation coefficient and root mean squared error (RMSE) for continuous outcomes, comparing observed and predicted values (Wilimitis and Walsh, 2023).

Microbial clustering (Study 1). This analysis first grouped samples based on the similarity of their microbial composition (microbial clusters) to identify natural groupings within the microbial communities and then evaluated whether these microbial clusters could be predicted using variables under study (such as breed, herd, age category, parity category, or pregnancy status). For this purpose, the ASV table with CLR transformation was reduced to two-dimension NMDS based on Euclidean distances. Next, the position of each sample was recalculated using the radial theta algorithm, which measures the angle of each point relative to the centre of the dataset in this two-dimensional space (Lahti and Shetty, 2018). The angle was computed as:

$$\theta = atan2(y - \overline{y}, x - \overline{x}),$$

Where *atan2* is the two-parameter arctangent function, x and y represent the coordinates of a point, and $\overline{x}$ and $\overline{y}$ are the mean coordinates of the NMDS dataset.

K-means clustering was then applied to group samples based on these angles. To determine the optimal number of clusters, the Silhouette method was applied. This method assigns a Silhouette score to each sample, which measures how well the sample fits within its own cluster compared to the nearest cluster. The score is calculated as the difference between the average distance to samples in the same cluster and the average distance to samples in the closest different cluster, divided by the larger of the two. Scores range from −1 to 1, with higher values indicating better-defined clusters. Several clustering solutions were tested, and the one balancing a high Silhouette score and biological interpretability was selected.

To assess whether phenotypic traits (breed, herd, age, parity, or pregnancy status) could predict microbial cluster, a RF model was implemented following the general RF

framework described above, where the predicted variable ($\hat{y}$) was the microbial cluster assignment, while the predictors were the phenotypic traits under study. The model was built using 10,000 trees ($T = 10{,}000$) and evaluated through 5-fold-CV. A separate model was trained for each phenotypic variable. The aim was to determine whether host phenotype could explain variation in microbial composition and to identify which trait yielded the highest classification accuracy, as reflected by cluster structure. Model performance was assessed by calculating the percentage of correctly classified samples, comparing predicted and actual cluster assignments across the cross-validation folds.

Phenotype prediction with microbiota as proxy (Study 2). In contrast to the clustering-based approach, where phenotypic traits were used as predictors of microbial clusters, this second analysis reversed the modelling direction by using directly microbiota composition to predict phenotypic traits.

The model followed the general RF framework described above, but with a different configuration. In this case, the predicted variable ($\hat{y}$) was the phenotypic trait, while the predictors were the CLR-microbial abundance. The model was built using 10,000 trees, with each tree constructed using random sampling of 63.3% of the animals. At each split, a random subset of predictor variables ($m = \sqrt{p}$) was selected, where p is the total number of predictors (e.g., for genus level, p = 2,305 genera). The model was evaluated using 5-fold-CV. Performance was assessed by calculating overall classification accuracy for categorical traits (pregnancy, herd, breed, age category, or parity category), while *r* and RMSE for continuous traits (age and parity as numerical values). An independent model was established for each trait at the phylum, genus, COG, and KO levels.

In all clustering and prediction approaches, a prevalence filter of 1% was applied to exclude very rare taxa, and CLR transformation was applied to the data for compositional normalisation (**Studies 1** and **2**).

Silhouette scores were calculated using the cluster R package (Maechler et al., 2022). The randomForest R package was used to perform the RF analysis (Liaw and Wiener, 2022), while base R was utilised for conducting the K-fold-CV (R Core Team, 2023).

4.6.5. Differential abundance analysis

Differential abundance analysis was performed to identify and quantify variations in the abundance of microbial taxa across pregnancy status and PRID groups. To ensure clarity and methodological transparency, the statistical frameworks employed for differential analysis are first described, followed by a detailed description of the specific modelling strategies implemented in each individual study.

Negative binomial generalised linear model. A robust statistical framework for analysing over-dispersed count data, as commonly encountered in high-throughput or microbiome sequencing studies. The model structure follows the implementation provided by the DESeq2 R package, developed by Love et al. (2014). Let K_{ij} represent the observed number of reads assigned to microbial taxon i in sample j. The model assumes:

$$K_{ij} \sim NB(\mu_{ij}, \alpha_i)$$

Where μ_{ij} denotes the expected abundance of taxon i in sample j, and α_i is the dispersion parameter that accounts for biological and technical variability beyond Poisson noise (Love et al., 2014). The variance structure is defined as:

$$\mathrm{Var}(K_{ij}) = \mu_{ij} + \alpha_i \mu_{ij}^2$$

To incorporate experimental covariates —such as pregnancy status— the expected counts are linked to a linear predictor through a log function:

$$\log(\mu_{ij}) = \log(s_j) + \mathrm{x}_j^T \beta_i$$

Here, s_j is a size factor that corrects for sample-specific sequencing depth, such as differences in the total number of reads obtained per sample. x_j encodes the experimental design (e.g., pregnancy vs. non-pregnancy), and β_i represents the effect of those variables on the abundance of taxon i. Click or tap here to enter text.

To account for differences in sequencing depth, a specific normalisation procedure, hereafter referred to as "SizeFactor" (SF) normalisation, was applied. This approach, based on the poscount method in DESeq2 (Love et al., 2014), estimates a size factor for each sample by comparing its non-zero counts to a pseudo-reference derived from the geometric mean across features. It is particularly suitable for microbiome data, where zero inflation is common, as it avoids the need to introduce artificial pseudo-counts or apply logarithmic transformations (McMurdie and Holmes, 2013). The resulting size factors scale the counts across samples, enabling fair comparisons while maintaining compatibility with the assumptions of negative binomial modelling. SF normalisation is explicitly designed to preserve these assumptions, allowing accurate estimation of dispersion and fold changes without violating key statistical r*equi*rements (i.e., untransformed integer counts, negative binomial-distributed variance, and multiplicative covariate effects on the count scale), which would not be met by log-ratio transformed data.

Dispersion estimates α_i are fitted by modelling a global mean–dispersion trend, followed by empirical Bayes shrinkage of individual dispersions towards this trend. This improves the stability of estimates, particularly for rare taxa or in designs with limited replication (Love et al., 2014). Statistical testing is performed via Wald tests on the

regression coefficients β_i, and significance is assessed using Benjamini–Hochberg correction to control the False Discovery Rate (Benjamini and Hochberg, 1995).

This modelling strategy is well suited for microbiome studies, where overdispersion, zero inflation, and high biological variability are prevalent. It enables robust differential abundance analysis across conditions while accommodating the complex structure of metataxonomic data.

Linear modelling with empirical Bayes moderation. This statistical framework, implemented in the Limma R package (Ritchie et al., 2015), models differential abundance in high-dimensional RNA-seq or microbiome data using linear models combined with empirical Bayes variance moderation. This approach is well suited to settings with limited sample sizes, where borrowing information across features enhances the stability and power of statistical inference.

Let Y_{ij} represent the CLR-transformed abundance of microbial feature i in sample j. The model assumes:

$$Y_{ij} = \mathrm{x}_j^T \beta_i + e_{ij}$$

Where x_j encodes the experimental design (e.g., time, treatment, or pregnancy status); β_i are the regression coefficients for feature i (e.g. Taxa); and e_{ij} is the residual error, assumed to follow a normal distribution with constant variance.

The CLR transformation is applied prior to modelling to address the compositional nature of microbiome data and to approximate the assumptions of linearity, normality, and homoscedasticity *requi*red by the linear model. After model fitting, empirical Bayes moderation is applied to the feature-wise variances, shrinking them toward a common value. This results in moderated t-statistics, which improve inference accuracy, particularly in small-sample or heterogeneous datasets. P-values are adjusted using the Benjamini–Hochberg procedure to control the false discovery rate (Benjamini and Hochberg, 1995).

4.6.5.1. Study 1. Vaginal microbiota association with pregnancy using 16S rRNA metabarcoding

The DESeq2 model was used at ASV, species, genus or phylum level, employing the poscount method and SF normalisation. Taxa with prevalence lower than 1% of the samples were excluded to improve model stability. Differential abundance analysis in relation to pregnancy was conducted using two complementary approaches: (i) a global model, which included all samples, and (ii) herd-specific models, performed separately within each herd.

To minimise overparameterization, only pregnancy and herd were included as fixed effects in the global model. The herd variable accounted for breed-related structure and population-level differences. Age and parity were not incorporated, as preliminary analyses suggested that their inclusion did not significantly improve model performance. This approach allowed for a more stable and interpretable estimation of the primary effects of interest. The final model was defined as follows:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{e}$$

Where, $\boldsymbol{y}$ is the vector of abundance of a given taxon; $\boldsymbol{b}$ is a vector that includes the fixed effects of pregnancy and herd in the case of the global model, while pregnancy in the case of the herd-specific model; $\boldsymbol{e}$ is the vector of residual errors and $\boldsymbol{X}$ is incidence matrix for the fixed effects (pregnancy and herd in global model, pregnancy for herd-specific model). False Discovery Rate multitest correction was applied to adjust p-values at 5% level.

4.6.5.2. Study 2. Association of vaginal microbiota with pregnancy using ONT-based metagenomics

Comparable modelling framework was implemented for the ONT dataset, extending the analysis to both taxonomic levels (phylum and genus) and functional features (COG and KO). DESeq2 was again used, with the poscount method and SF normalisation. As in **Study 1**, taxa with prevalence lower than 1 were excluded and then differential abundance was assessed using (i) a global model and (ii) herd-specific models. Pregnancy and herd were included as fixed effects in the global model, while only pregnancy was used in the herd-specific analyses. The same model structure was applied, where $\boldsymbol{y}$ represents the abundance of the feature (taxon, COG, or KO), and the remaining elements are defined as above. Multiple testing correction was performed using the FDR method at the 5% level.

4.6.5.3. Study 3. Impact of PRID on vaginal microbiota composition and artificial insemination outcome

The Limma model was applied to ONT sequencing data from the Assaf-bred group, focusing on both taxonomic levels (phylum and genus) and functional features (COG and KO). Differential abundance was assessed between sampling points and, within the S2 subset, between pregnancy statuses. To minimise noise from low-abundance features, a RA filter of 0.05% was applied prior to analysis, followed by CLR transformation.

Three modelling approaches were developed: (i) a pregnancy status model using all S2 animals, corrected for treatment; (ii) a sampling point model including all samples,

also corrected for treatment; and (iii) a sampling point model run separately within each treatment group. The model structure was similar to that used in **Studies 1** and **2**, although with a different configuration. In these models, $\mathbf{y}$ represents the abundance of a given taxa, and $\boldsymbol{b}$ encodes the fixed effects: pregnancy and treatment in model (i); sampling point and treatment in model (ii); and sampling point in model (iii). P-values were adjusted using FDR correction at a 5% significance threshold.

4.6.6. MGWAS and functional analysis

Microbiota Genome-Wide Association Studies were performed in **Study 4** to examine the relationship between the host genome and the abundance of specific microbial taxa within the vaginal microbiota. The objective was to identify SNPs significantly associated with variations in microbial abundance across multiple taxonomic levels and explore their potential association with fertility.

Separate MGWAS analyses were conducted for each taxon at the phylum, family, genus, and species levels. The model was defined as:

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

Where $\mathbf{y}$ is the vector of phenotypes, representing the CLR-transformed RA of a specific taxon; $\mathbf{b}$ is the vector of fixed effects, including breed, the first two eigenvectors from the PCA and SNP genotypes; $\boldsymbol{u}$ denotes the vector of random additive genetic effects; and $\mathbf{e}$ represents the vector of random residual errors. $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating the observations with the fixed and random effects, respectively. Vector $\mathbf{u}$ is assumed to follow a multivariate normal distribution $\mathbf{u} \sim N(0, \mathbf{G}\sigma_u^2)$, where **G** is the genomic relationship matrix and σ_u^2 is the additive genetic variance. The residuals are also assumed to be normally distributed as $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is an identity matrix and σ_e^2 is the residual variance.

Results from the analyses were represented in a combined Manhattan plot. Significant SNPs associated to each taxon's abundance were identified after applying multitest FDR correction at 10%.

A detailed review of the literature was conducted to identify microbial taxa previously linked to fertility. The complete list, along with relevant bibliographic references, is provided in Appendix – Table S4.1. From this list, specific attention was given to taxa

that also exhibited one or more SNPs significantly associated with their abundance in the MGWAS analyses.

Next, the Variant Effect Predictor (VEP) tool was applied to predict the functional consequences of those SNPs identified with a significant association to reproductive taxa, using the *Ovis aries* reference genome ARS-UI_Ramb_v2.0 (GCF_016772045.1), as the reference assembly (Durinck et al., 2009). Subsequently, variants located within intergenic regions were excluded, with priority given to those predicted to have functional relevance. Gene functions were explored further through GO annotations. For consistency, standardised gene names were used; however, when a variant lacked an assigned name, its Ensembl gene identifier was reported instead.

An interaction network analysis was also performed to uncover host genes shared among multiple taxa. This network focused on identifying genes associated with at least two distinct taxa within the vaginal microbiota.

Finally, the predicted impacts of SNPs were classified into four categories —"Modifier," "Low," "Moderate," and "High"— using the VEP framework. Only SNPs with "Low," "Moderate," or "High" impacts were included in downstream analyses, as these are more likely to affect gene or protein function. To understand the SNP effect on the abundance of the microbial taxa, the limitation in interpretation derived from the use of CLR-transformed abundance tables was addressed, which do not allow a direct evaluation of the effects in absolute or relative abundance. In this line, we approximated a relative effect of each SNP as follows:

$$\boldsymbol{b_t} = \boldsymbol{e^b}$$

Where b_t represents the proportional change in the RA of a taxon associated with one allele compared to the other, expressed as a multiplicative factor relative to the total microbiome. This back-transformation allows for interpreting $\boldsymbol{b}$ as an approximated log2FC, consistent with compositional data principles, where differences in log-transformed values correspond to proportional differences in the original scale (Mandal et al., 2015). Based on the principle that a difference in log-abundance translates into a multiplicative factor in the original scale (Mandal et al., 2015), the exponentiated coefficient $\boldsymbol{b_t}$ represents an approximate multiplicative change in a taxon's RA with respect to the geometric mean of the community (or all other taxa), rather than an absolute shift. According to compositionality characteristics (Gloor et

al., 2017), a value of $\boldsymbol{b_t}$ greater than 1 indicates a relative increase for the alternative allele, whereas a value below 1 suggests a relative decrease.

Statistical analyses and data visualisation were conducted using R, with packages such as ggplot2 (Wickham, 2016) for visualisation and dplyr for data manipulation. VEP analysis was carried out using tools provided by the Ensembl Genome browser.

5. RESULTS

5.1. Study 1: Characterisation of vaginal bacterial communities and their association with pregnancy outcomes using 16S rRNA metabarcoding

This first study aimed to characterise the composition and abundance of the vaginal bacterial communities and to investigate its potential association with reproductive success following AI, by amplifying and sequencing the V3–V4 hypervariable regions of the 16S rRNA gene. To achieve this, vaginal samples from 331 ewes were collected in four herds from three Spanish sheep breeds (Latxa, Manchega, and Rasa Aragonesa). The results presented below comprise the characterisation of the core bacterial composition, assessment of alpha and beta diversities, clustering analyses, cluster prediction using RF with cross-validation, and differential abundance analyses to identify associations between specific taxa and alpha-diversity metrics.

The overall fertility rate of the 331 artificially inseminated ewes was 43%, with differences among herds: 73% in Latxa, 42% in Rasa, 30% in Manchega VL, and 24% in Manchega RN.

5.1.1. Bacterial composition

A total of 12,235 ASVs were identified from 12,742,281 reads across the 331 samples, with an average of 38,496 reads per sample. Taxonomic classification revealed 33 phyla, 67 classes, 161 orders, 351 families, and 906 genera.

The core bacterial community at genus level consisted of *Streptobacillus* (12.80%), *Histophilus* (11.35%), *Fusobacterium* (9.53%), *Oceanivirga* (6.50%), *Parvimonas* (3.77%), *Porphyromonas* (3.25%), *Bacteroides* (1.44%), *Anaerococcus* (1.19%), *Aerococcus* (1.19%), *Streptococcus* (0.89%), and *Trueperella* (0.65%) (Figure 5.1.1 A). The RA of these genera differed among herds, with *Fusobacterium* being more abundant in Latxa and Rasa herds, while *Histophilus* was predominant in Manchega VL and RN herds (Figure 5.1.1 A).

At the phylum level, the core microbiota was composed of Proteobacteria (16%), Fusobacteria (52%), Firmicutes (21%), Bacteroidota (10%), and Actinobacteria (2%). Variations in phylum abundance were observed across herds, with Rasa exhibiting the lowest proportion of Proteobacteria and the highest abundance of Firmicutes (Figure 5.1.1 B).

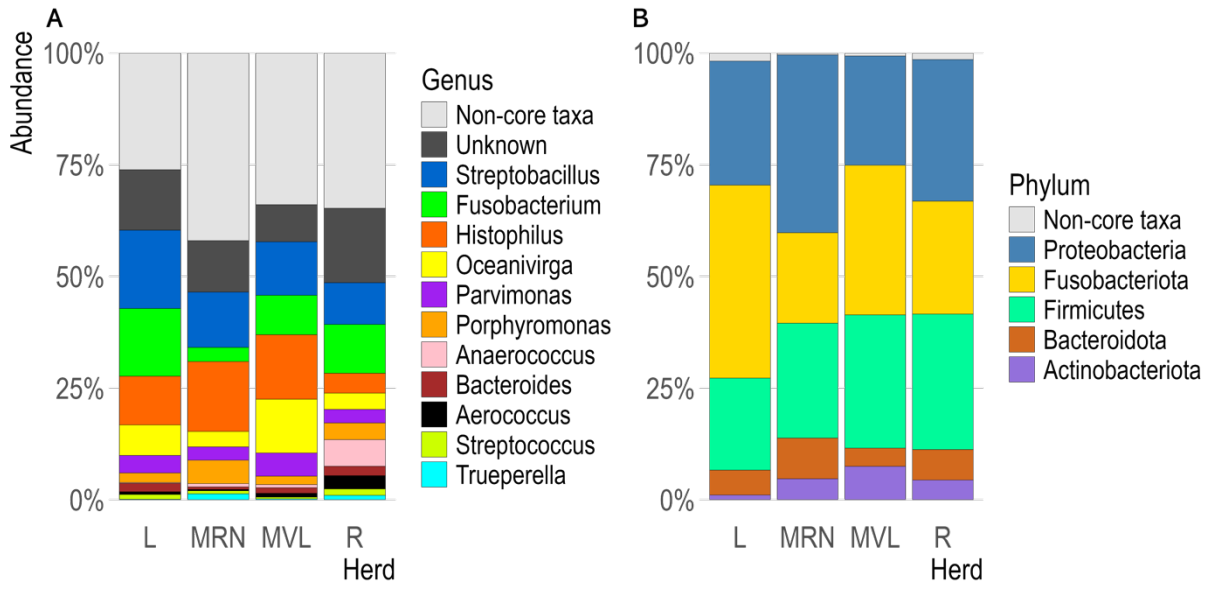


Figure 5.1.1. Core bacterial community composition with a prevalence ≥ 90% across all samples, at the genus (**A**) and phylum levels (**B**), for the four herds analysed. L: Latxa, MRN: Manchega RN, MVL: Manchega VL, R: Rasa.

5.1.2. Alpha-diversity

In general, the data indicated differences in alpha-diversity among the herds, with Manchega VL and Rasa Aragonesa typically exhibiting higher diversity compared to Latxa and Manchega RN, depending on the index and taxonomic level considered (Figure 5.1.2). Alpha-diversity varied across herds, with Latxa consistently showed the lowest alpha-diversity values across most indices and taxonomic levels. Followed by Manchega RN, which presented slightly higher diversity than Latxa, although in many comparisons the differences were not statistically significant. In contrast, Manchega VL exhibited the highest diversity, particularly in indices such as Shannon and Inverse Simpson, which account for both richness and evenness. The differences between MVL and the lower-diversity herds (Latxa and Manchega RN) were frequently significant.

Rasa generally presented a high degree of variability in its diversity values. In several cases, Rasa significantly differed from Latxa and Manchega RN, indicating a generally higher diversity. However, its values do not always differ significantly from those of Manchega VL, suggesting comparable levels of diversity in some contexts, although with data more dispersed.

No statistically significant differences were found between pregnant and non-pregnant ewes at the analysed levels considering the herd or when considering the entire dataset (Appendix – Figure S1.2 and Figure S1.2). Similarly, alpha-diversity

analyses based on categorised age and parity also revealed no significant associations (Appendix – Figure S1.3 and Figure S1.4).

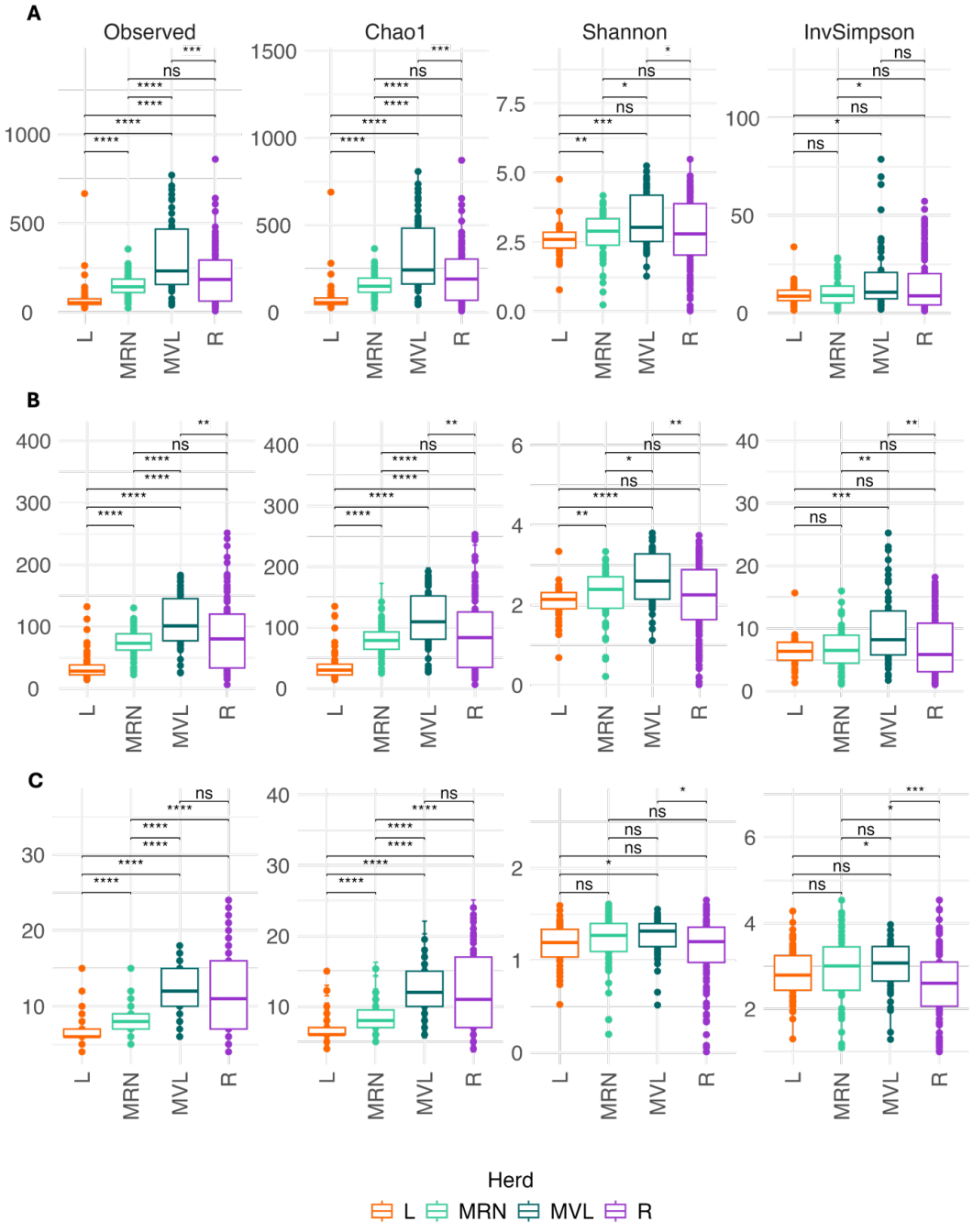


Figure 5.1.2. Alpha-diversity for herd at ASV (**A**), genus (**B**), and phylum (**C**) level for the different coefficients analysed: Observed richness, Chao1, Shannon, and Inverse Simpson index. Data were rarefied. Statistical significance is indicated by asterisks: * ($p < 0.05$), *** ($p < 0.001$). ns: not significant. L: Latxa, MRN: Manchega RN, MVL: Manchega VL, R: Rasa.

5.1.3. Beta-diversity

Principal Component Analysis showed distinct clustering among breeds and herds (Figure 5.1.3), with the first three axes accounting for 12.46%, 10.09%, and 7.09% of the total variation, respectively. Higher dispersion was observed in Manchega RN and Rasa compared to the other groups. No notable differences were detected with respect to pregnancy status, or to the categorised age and parity variables (Appendix – Figure S1.6).

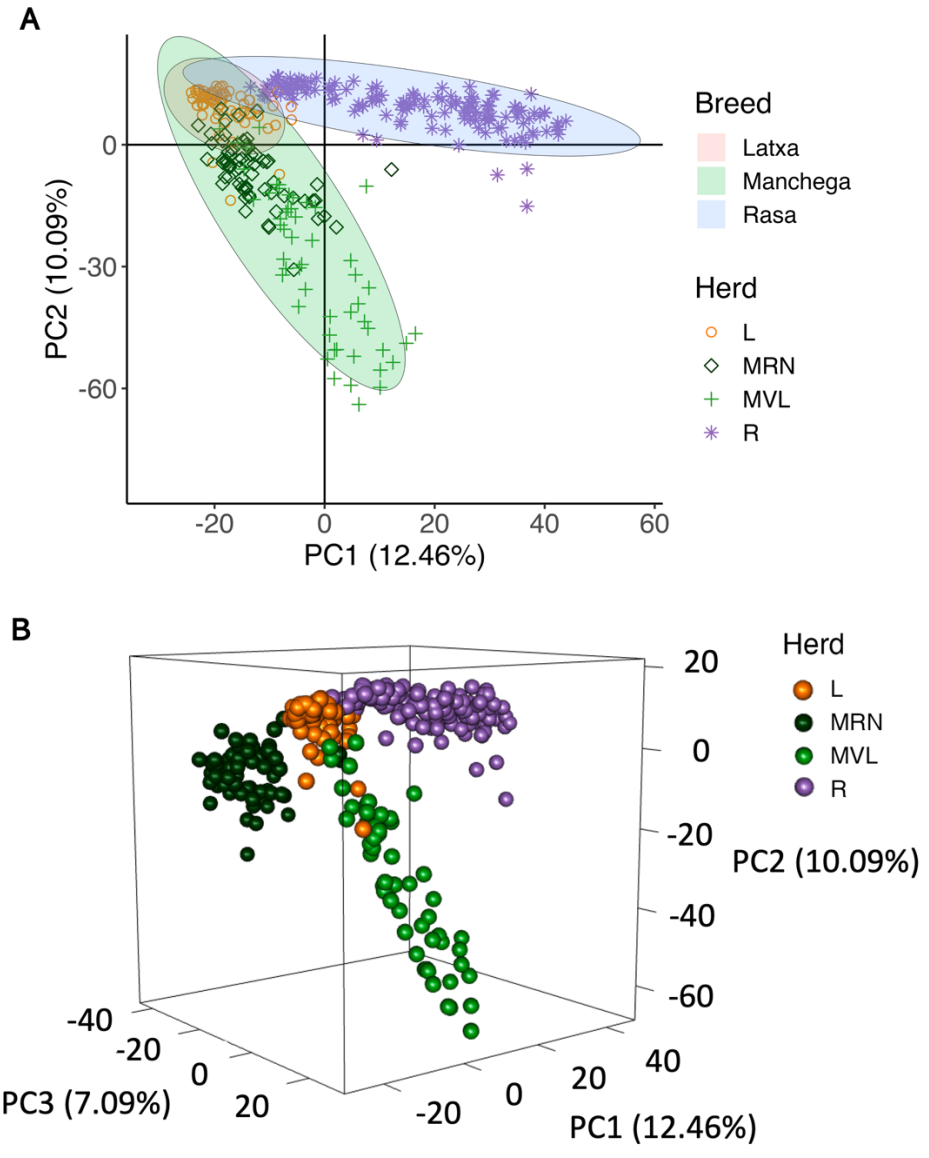


Figure 5.1.3. Principal Component Analysis for breed and herd given the bacterial composition at the ASV level, represented in 2D (**A**) and 3D (**B**) plots. Ellipses in panel (**A**) are calculated using covariance to visualize data variability within each breed group.

The PERMANOVA analysis (with Euclidean distances) revealed significant differences for herd and breed in the global model (considering all samples and correcting by herd), as well as across all pairwise comparisons among herd groups. Pregnancy status, age, and parity categories also displayed significant differences for the global model (Table 5.1.1). No differences were found for pregnancy within individual herds. Regarding pairwise comparisons between age categories, significant differences were observed across all groups, except between <2 and 2–3 years, and between 2–3 and 3–4 years (Appendix – Table S1.2). For parity categories, significant differences were only observed between ewes with <2 and >4 parturitions, and between 2–3 and >4 parturitions (Appendix – Table S1.3).

Table 5.1.1. PERMANOVA results for breed, herd, and pregnancy variables at ASV level

Factor	SumOfSqs	R2	F	Pr(>F)
Herd	207,627	0.2192	30.601	**0.001*** **
Breed	14,4089	0.1521	29.424	**0.001*** **
Pregnancy	5,816	0.0061	2.032	**0.001*** **
Age category	44,048	0.0465	3.975	**0.001*** **
Parity category	29,688	0.0313	2.637	**0.001*** **

Df: Degrees of Freedom; SumOfSqs: Sum of Squares; R^2: Coefficient of Determination; F: F-statistic; Pr(>F): P-value for the F-statistic; MRN: Manchega RN, MVL: Manchega VL.

5.1.4. Clustering and RF prediction

To further explore patterns in microbial community structure, clustering analysis was performed using the Silhouette method to determine the optimal number of clusters. This analysis initially identified three distinct clusters, which largely reflected herd-related patterns. Notably, samples from the Latxa and Manchega RN herds grouped together, suggesting similarities in their microbial profiles. However, when the number of clusters was increased to four (k = 4), the clustering more clearly aligned with individual herd identities, indicating a finer resolution of herd-specific community structure (Figure 5.1.4). No apparent patterns were observed for categorised age or parity in relation to the clustering structure.

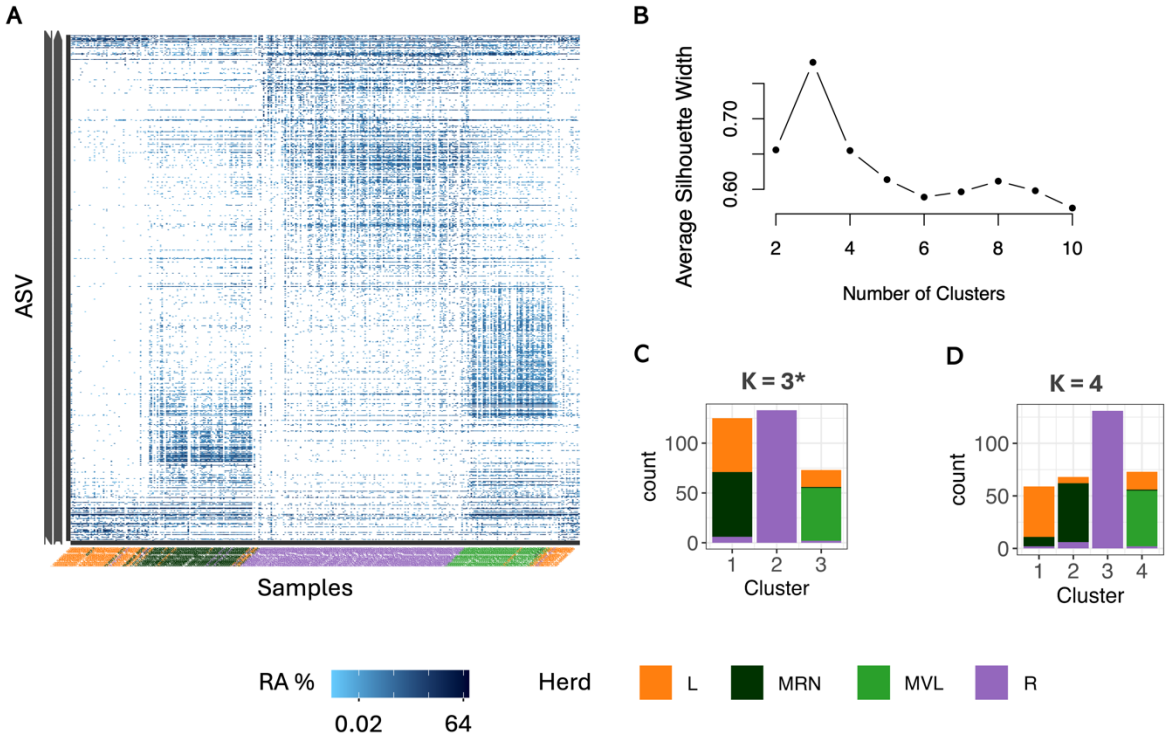


Figure 5.1.4. Analysis of sample similarities and clustering in bacterial community studies. (**A**) Heatmap showing the RA of bacterial community composition, arranged to samples similarity parameters. (**B**) Silhouette width plot for the different numbers of clusters. (**C**) and (**D**) Barplot representing the clusters assuming different k values. Each bar represents the number of samples within each cluster, colour-coded by sample type: L (Latxa), MRN (Manchega RN), MVL (Manchega VL), and R (Rasa). The asterisk (*) highlights the optimal K-value determined by the highest Silhouette scores achieved in the analysis.

To complement the clustering analysis and identify key factors driving microbial community differences, an RF classification was performed. Results indicated that herd was the most influential variable for cluster prediction, closely followed by breed, regardless of whether RA and CLR normalisation methods were used. In contrast, pregnancy status had minimal influence on prediction. Cross-validation further supported these findings, showing that herd was the variable with the highest prediction accuracy (0.87), followed by breed (0.67) (Table 5.1.2).

Table 5.1.3. Assessment of variable importance and predictive accuracy in bacterial community clustering using RF analysis. Impact of herd, breed, and pregnancy, on bacterial cluster categorisation, evaluated by RF importance scores under CLR data transformations. Predictive accuracy was assessed through cross-validation

	[1]Random Forest	
Variable	Mean Decrease Accuracy	Mean Decrease Gini
Herd	0.347	90.82
Breed	0.277	61.27
Age category	0.053	13.77
Parity category	0.053	9.87
Pregnancy	0.011	3.91
	[2]Cross-validation	
	Accuracy	SD
Herd	0.8700[a]	0.0598
Breed	0.6701[ab]	0.0250
Age category	0.4453[bc]	0.0244
Parity category	0.4537[bc]	0.0449
Pregnancy	0.3984[c]	0.0397

[1]**Random Forest** results where the predicted values were the clusters (n = 4) with variables evaluated including herd, breed, and pregnancy. Mean Decrease Accuracy: The average reduction in model accuracy when a variable is omitted. Mean Decrease Gini: The reduction in the Gini coefficient when a variable is omitted, indicating variable importance. [2]**Cross-validation**: Assesses the predictive accuracy of the RF model via k-fold Method (k = 5). Accuracy: The proportion of correct predictions made by the model, estimated by the mean percentage of those which were correctly assigned. Std: Standard deviation of the accuracy, providing a measure of its variation across the cross-validation folds. Significance between accuracy results is denoted by superscript letters: different letters (a, b, c) indicate statistically significant differences between groups as determined by post-hoc Dunn's testing with Benjamini-Hochberg adjustment.

5.1.5. Differential abundance analysis

To gain deeper insight into specific microbial taxa driving differences between groups, differential abundance analysis was performed across multiple taxonomic levels. Figure 5.1.5 provides an overview of the analyses conducted across different taxonomic levels, including ASV, species, genus, and phylum. In the global model, six ASVs were found to have significantly higher abundance in pregnant ewes. These ASVs belonged to the genera *Staphylococcus*, *Porphyromonas*, *Aerococcus*, *Corynebacterium*, *Arcanobacterium*, and *Histophilus*. Conversely, fifteen ASVs showed greater abundance in non-pregnant ewes, including taxa from the genera *Escherichia-Shigella, Leptotrichia, Bacteroides, Fusobacterium, Porphyromonas, Oceanivirga, Campylobacter*, and *Histophilus*. Additionally, two ASVs were identified within the Pasteurellaceae and Weeksellaceae families. At the genus level, *Histophilus, Escherichia-Shigella, and Leptotrichia* were found to be more prevalent in

non-pregnant ewes. Supplementary Boxplots (Appendix – Figure S1.10 to Figure S1.19) visually illustrate these patterns.

Within the herd-specific model, results varied across taxa and groups. For Latxa, two ASVs from *Escherichia-Shigella* and *Fusobacterium* were significantly enriched in pregnant ewes, while no taxa were associated with non-pregnant ewes. In contrast, the Manchega RN herd presented ASVs from the genera *Fusobacterium*, *Histophilus*, *Parvimonas*, and *Bacteroides*, along with the species *Histophilus somni* and the phylum Fusobacteriota, significantly more abundant in non-pregnant ewes. Pregnant ewes in this herd showed higher abundance of the genera *Acinetobacter*, *Pseudomonas*, and *Brevundimonas*. In Manchega VL, two ASVs from the genus *Histophilus* and the Pasteurellaceae family were more abundant in non-pregnant ewes. Finally, the Rasa herd, consistent with the global model, showed a significant increase in the genus *Escherichia-Shigella* among non-pregnant ewes.

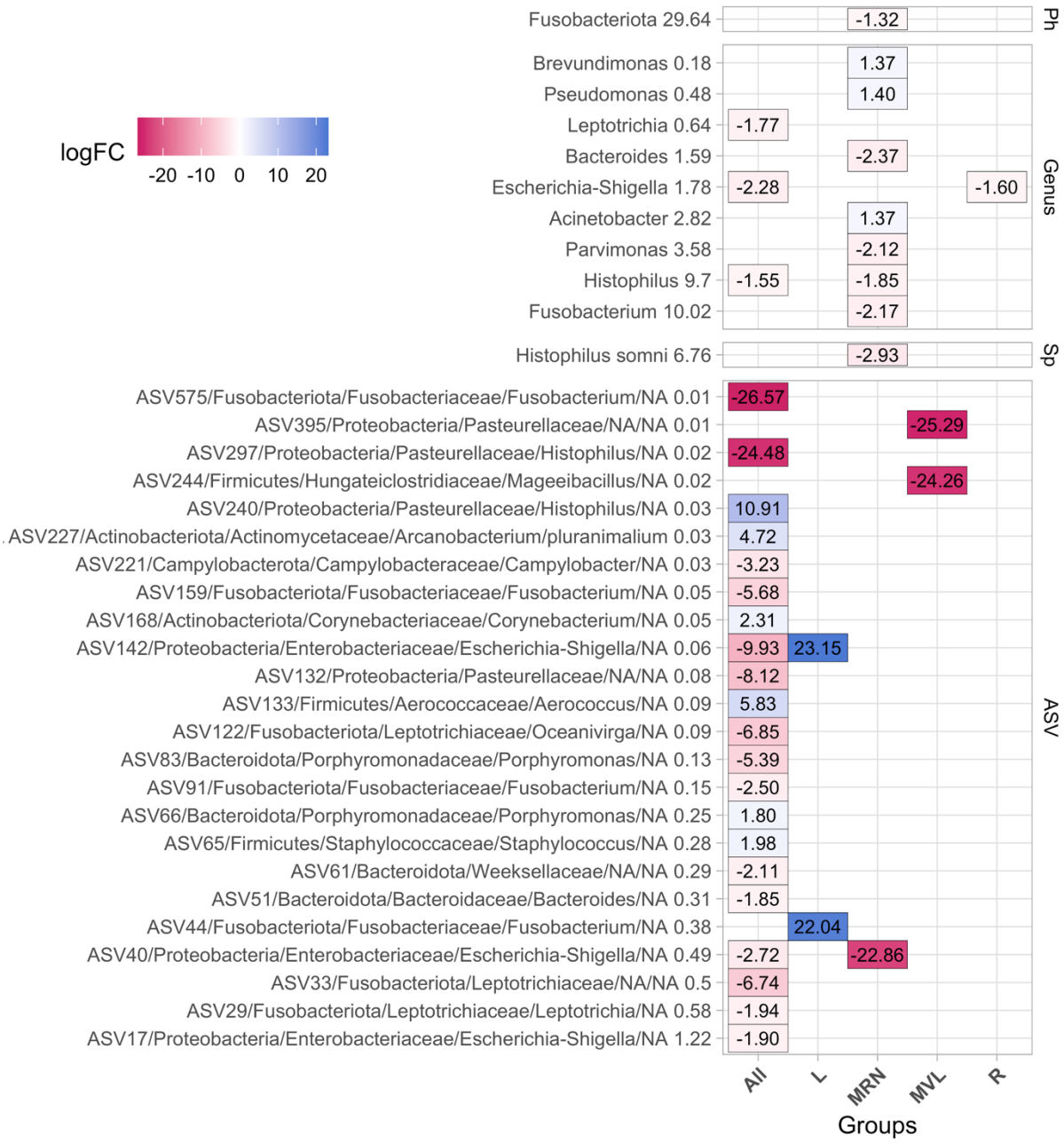


Figure 5.1.5. Heatmap showing significant results (FDR at 5%) for differential abundance analysis for pregnancy success at phylum, genus, species and ASV levels. The x-axis shows the results for the global model (ALL) and herd-specific model (L: Latxa, MRN and MVL: Manchega, R: Rasa). The log2FC is represented by colour gradations, red for negative and blue for positive with respect to pregnancy success. The colour intensity correlates with the log2FC value; values close to zero are represented in white. The log2FC value is also presented in each square of the heatmap. The y-axis categorizes the taxonomic assignments, delineating a hierarchical classification that includes the ASV code, followed by its corresponding phylum, family, genus, and species, and its RA.

5.2. Study 2: Characterisation of vaginal microbiota and microbial genes associated with pregnancy outcomes using metagenomics

This second study also focuses on the characterisation of vaginal bacterial communities and their association with pregnancy outcomes, using the same three Spanish sheep breeds and the same four herds analysed in the first study. However, it includes 297 vaginal samples, 34 samples less than in **Study 1**.

The main difference lies in the methodology: this study employs ONT-based metagenomic sequencing instead of 16S rRNA metabarcoding. Nanopore technology sequences all DNA present in the sample without prior PCR amplification, allowing not only for taxonomic profiling but also for the study of functional genes (such as COG and KO). This section presents the core microbiota composition, assessments of alpha and beta-diversity, predictions of traits (herd, breed, age, parity, and pregnancy) using taxonomic and functional profiles, differential abundance analyses to identify associations between specific taxa/genes and pregnancy outcomes, and interactive gene taxa association.

The overall fertility rate across the 297 ewes subjected to AI was 48.49%. Fertility rates by herd were as follows: 81.82% in Latxa, 46.27% in Rasa Aragonesa, 34.15% in Manchega VL, and 25% in Manchega RN.

5.2.1. Microbiota composition

The sequencing data from the 297 samples comprised a total of 63,430,285 reads. Following quality filtering, 3,122,826 reads were assigned to the Bacteria, Archaea, Protista and Fungi kingdoms, and virus. Taxonomic classification identified 113 phyla, 551 families, 1,389 genera, 9,911 COG entries, and 8,417 KOs. A detailed summary of the microbiota composition is provided in Table 5.2.1.

Table 5.2.2. Summary statistics of reads abundance across taxonomic and functional levels.

Level	Total Reads	Mean Sample	Median Sample	Number of Taxa/Genes
Phylum	3,161,540	10,644.92	6250	113
Genus	1,821,558	6,133.19	3328	1389
COG	3,662,878	12,332.92	6467	9911
KO	3,776,544	12,715.64	6872	8417

Level: Taxonomic or functional classification level; **Mean Sample**: The average number of reads per sample; **Median Sample**: The median number of reads per sample; **Number of Taxa/Genes**: The total number of distinct taxa or genes identified at each classification level. All the unclassified reads were discarded at each taxonomic level, retaining taxa that belong to the Kingdoms Bacteria, Archaea, and virus, as well as to the phyla Apicomplexa, Ascomycota, Euglenozoa, Evosea, and Nematoda.

The core microbiota consisted of 11 phyla and 21 genera (Figure 5.2.1). After applying a 1% prevalence threshold, the most abundant genera included *Staphylococcus* (18.2%), *Escherichia* (3.2%), *Histophilus* (2.3%), *Mycoplasma* (1.2%), *Campylobacter* (0.7%), *Bacteroides* (0.7%), *Fusobacterium* (0.6%), *Actinobacillus* (0.5%), *Parvimonas* (0.5%), and *Anaplasma* (0.3%). Variations in the abundance of these genera were observed across herds; *Staphylococcus* was more prevalent in Rasa Aragonesa and Manchega RN ($p = 4.715e{-}16$; Kruskal-Wallis test), while Latxa had higher levels of *Histophilus* than the other herds ($p = 2.2e{-}16$; Kruskal-Wallis test). At the phylum level, the most abundant groups were Bacillota (64%), Pseudomonadota (20%), Bacteroidota (3.6%), Fusobacteriota (3.2%), and Actinomycetota (3%).

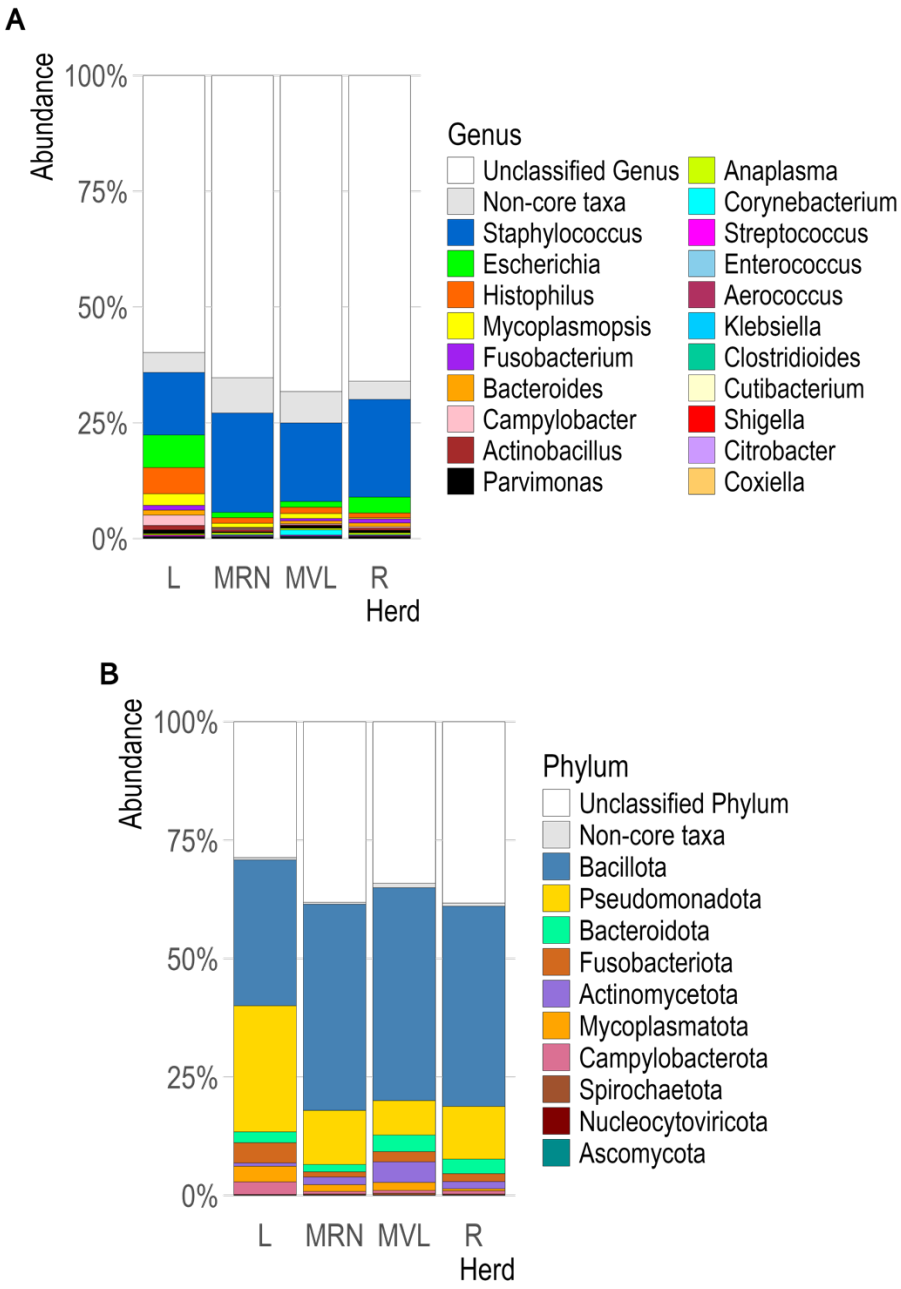


Figure 5.2.1. Microbial community composition at genus (**A**) and phylum (**B**) level. Taxa with a prevalence ≥ 90% were considered as part of the core microbiota. L: Latxa, MRN: Manchega RN, MVL: Manchega VL, R: Rasa. KO: KEGG Orthologous. Non-core taxa: those taxa with a prevalence < 90%.

5.2.2. Alpha-diversity

In general, the data from this study indicated differences in alpha-diversity among herds, following patterns similar to those observed in **Study 1** (Figure 5.2.2). Manchega VL once again exhibited the highest diversity across most indices at both the phylum and genus levels, while Latxa continued to show the lowest diversity, particularly for

the Observed and Chao1 indices. However, in contrast to **Study 1**, Latxa displayed greater variance in the Shannon index.

Consistent with the previous findings, no statistically significant differences in alpha-diversity were found between pregnant and non-pregnant ewes (Appendix – Figure S2.1 and Figure S2.2). Similarly, analyses based on age and parity categories revealed no strong associations with alpha-diversity (Appendix – Figure S2.3 and Figure S2.4).

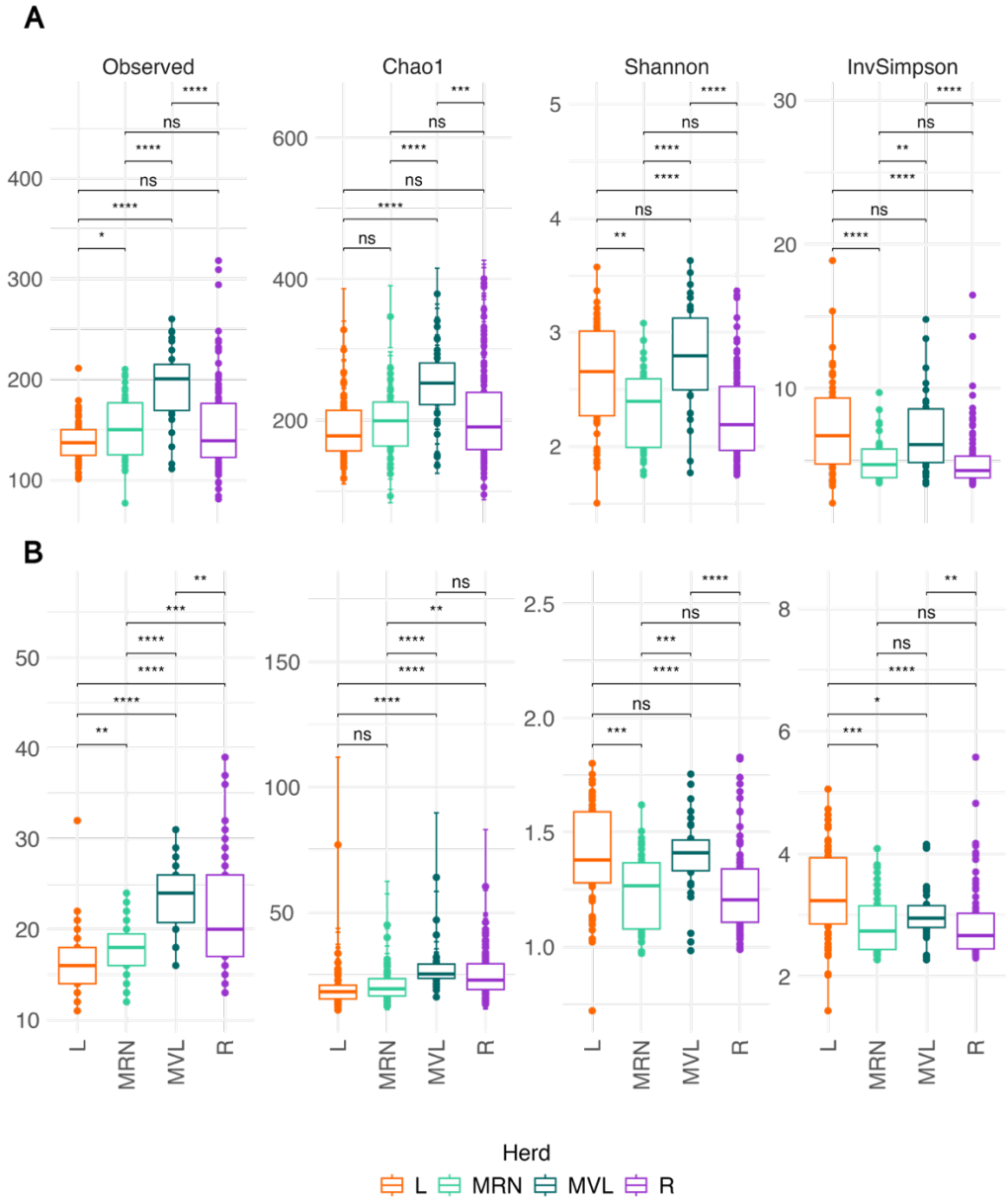


Figure 5.2.2. Alpha-diversity measures for herd at the genus (**A**) and phylum (**B**) levels for the different coefficients analysed: Observed richness, Chao1, Shannon, and Inverse Simpson index. Data were rarefied. Statistical significance is indicated by asterisks: * ($p < 0.05$), *** ($p < 0.001$). ns: not significant.

5.2.3. Beta-diversity

Overall, the beta-diversity results largely align with those reported in **Study 1**. In this study, PCA revealed differentiation among breeds and herds, broadly consistent with the patterns observed previously, although the clustering patterns here were less evident than in **Study 1**. Specifically, some overlap was observed at the genus level between Rasa Aragonesa, Manchega RN, and parts of the Latxa samples (Figure 5.2.3 A). At the phylum level, Latxa samples overlapped with both groups. As previously observed in **Study 1**, the two Manchega herds formed closely related clusters but still retained some degree of separation. Similarly, no distinct clustering was observed based on pregnancy status, age, or parity categories (see Figure 5.2.3 B and Appendix – Figure S2.6). Regarding COG and KO analyses, no clear clustering was identified when using functional classifications.

Concerning PERMANOVA results, as in **Study 1**, the analysis revealed significant differences for both breed and herd in the global model, as well as across all pairwise comparisons (Table 5.2.3). However, in this study, breed differences (F = 21.52 at the phylum level and F = 20.44 at the genus level) were slightly greater than herd differences (F = 18.42 at the phylum level and F = 18.31 at the genus level). Pregnancy status, age, and parity categories showed significant variation in the global model, but no significant differences were found within individual herds (Appendix – Table S2.1), similarly to **Study 1**. For age and parity categories, pairwise comparisons in this study revealed specific differences: ewes older than five years differed significantly from younger groups, and ewes with more than three parturitions differed from those with fewer parturitions (Appendix – Table S2.3 and Table S2.4).

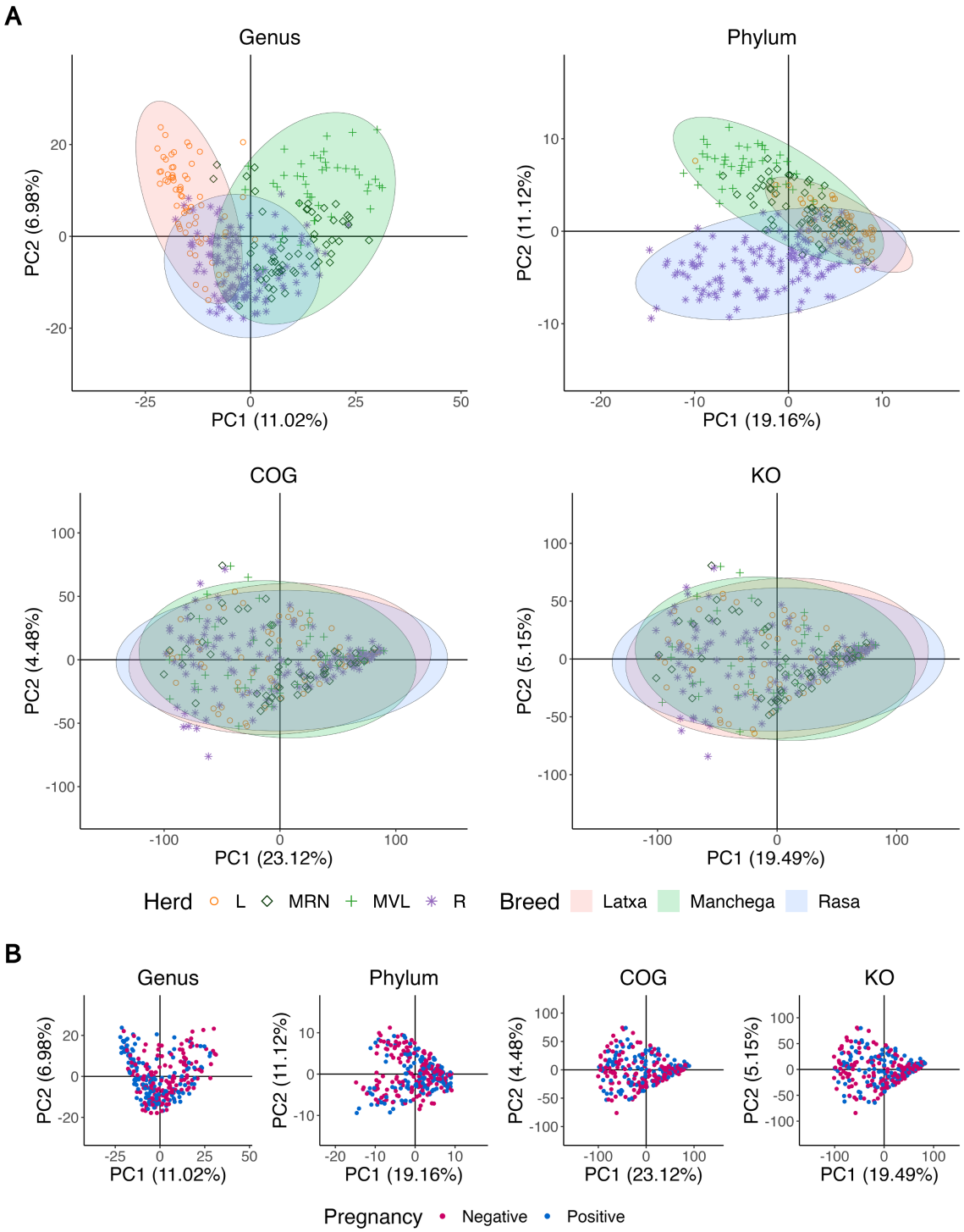


Figure 5.2.3. Principal Component Analysis of microbiota composition by breed, herd, and pregnancy status. Phylum, genus, COG, and KO levels were analysed. Panel **A** shows the analysis for breed and herd, while panel **B** shows the PCA for pregnancy. Ellipses in panel **A** were calculated according to the covariance within each breed group.

Table 5.2.4. PERMANOVA results for the effects of herd, breed, and pregnancy on microbiota and gene composition

Level	Factor	SumOfSqs	R2	F	Pr(>F)
Phylum	Herd	8,092.83	0.159	18.420	**0.001*** **
	Breed	6,513.67	0.128	21.522	**0.001*** **
	Pregnancy	424.10	0.008	2.474	**0.008** **
	Age category	1,303.59	0.028	2.092	**0.001*** **
	Parity category	1,302.72	0.028	2.090	**0.001*** **
Genus	Herd	67,636.83	0.158	18.312	**0.001*** **
	Breed	52,296.28	0.122	20.442	**0.001*** **
	Pregnancy	4,394.10	0.010	3.057	**0.001*** **
	Age category	18,897.07	0.031	2.352	**0.001*** **
	Parity category	16,006.77	0.026	1.983	**0.001*** **
COG	Herd	264,896.48	0.071	7.403	**0.001*** **
	Breed	229,012.52	0.061	9.535	**0.001*** **
	Pregnancy	15,110.55	0.004	1.190	0.184
	Age category	82,987.99	0.022	1.648	**0.008** **
	Parity category	72,909.75	0.019	1.444	**0.020***
KO	Herd	285,831.10	0.068	7.082	**0.001*** **
	Breed	242,043.93	0.057	8.927	**0.001*** **
	Pregnancy	19,077.24	0.005	1.337	0.101
	Age category	87,442.57	0.021	1.542	**0.005** **
	Parity category	78,401.38	0.019	1.379	**0.020***

SumOfSqs: Sum of squares, a measure of variance. **R2**: Coefficient of determination, indicating the proportion of variance explained by the factor. **F**: F-statistic, a ratio used to determine the significance of the factor. **Pr(>F)**: P-value, indicating the statistical significance of the factor. Values in bold indicate significant results ($P < 0.05$).

5.2.4. Random forest prediction

The RF model showed different levels of accuracy depending on the factor and classification level. Among the categorical variables, Breed reached the highest accuracy, with 89.24% (SD = 5.88) at the phylum level and 92.40% (SD = 3.29) at the genus level. Herd also performed well, with 84.18% (SD = 6.45) at phylum and 91.07% (SD = 3.64) at genus level. Pregnancy followed, with 52.18% (SD = 10.57) at phylum and 56.11% (SD = 4.73) at genus. Age category achieved 38.78% (SD = 5.84) and 39.08% (SD = 5.29), while parity category reached 42.76% (SD = 4.26) and 42.88% (SD = 5.45) at the phylum and genus levels, respectively. Prediction accuracy for COG and KO was generally lower. When treating Age and Parity as continuous variables, correlation was higher for age at genus level (r = 0.47, SD = 0.11) than for parity (r =

0.34, SD = 0.15). RMSE values were similar for age (RMSE = 2.07, SD = 1.64) and parity (RMSE = 2.09, SD = 1.50). A full summary of the results is presented in Table 5.2.5.

Table 5.2.6. Prediction accuracy results (and their SD) from cross-validation analysis using a RF algorithm

	Phylum	Genus	COG	KO
		% Correctly classified		
Breed	89.24 (5.88)	92.40 (3.29)	70.20 (3.01)	70.11 (3.40)
Herd	84.18 (6.45)	91.07 (3.64)	67.57 (5.75)	69.27 (6.11)
Pregnancy	52.18 (10.57)	56.11 (4.73)	56.30 (6.30)	55.48 (6.05)
Age category	38.78 (5.84)	39.08 (5.29)	31.99 (9.55)	32.58 (7.50)
Parity category	42.76 (4.26)	42.88 (5.45)	40.12 (10.47)	39.19 (9.89)
		Correlation		
Age	0.43 (0.06)	0.47 (0.11)	0.32 (0.07)	0.31 (0.05)
Parity	0.23 (0.03)	0.34 (0.15)	0.19 (0.10)	0.18 (0.11)
		RMSE		
Age	2.02 (1.71)	2.07 (1.64)	1.88 (1.37)	1.91 (1.39)
Parity	2.08 (1.58)	2.09 (1.50)	1.75 (1.39)	1.77 (1.37)

For categorical variables, including herd, breed, age category, parity category, and pregnancy, accuracy is measured as the percentage of correctly classified instances, which reflects the proportion of predictions where the model's output matches the real category labels. For numerical variables, such as age and parity, accuracy is assessed using the correlation coefficient, which quantifies the linear relationship between predicted values and real values. The values in parentheses represent the standard deviation (SD) of the predictive accuracy across the cross-validation folds (k = 5).

5.2.5. Differential abundance analysis

The differential-abundance analysis in the present study was performed at phylum and genus taxonomic levels, as well as at COG entries and KOs. The results showed a number of concordant patterns —and several noteworthy discrepancies— emerged when compared with the first study.

Under the global model, fourteen genera displayed significantly higher abundance in non-pregnant ewes. Six of these —*Histophilus*, *Fusobacterium*, *Bacteroides*, *Campylobacter*, *Leptotrichia* and *Oceanivirga*— had likewise been enriched in non-pregnant animals in **Study 1**. The remaining eight genera —*Streptobacillus*, *Trueperella*, *Gemella*, *Caviibacter*, *Peptostreptococcus*, *Treponema*, *Sneathia* and *Pseudoestreptobacillus*— were newly identified in this study. In contrast, this study presented the absence of several genera —*Escherichia–Shigella*, *Staphylococcus*, *Porphyromonas*, *Aerococcus*, *Corynebacterium* and *Arcanobacterium*— that had been significant in the previous study. At phylum rank, non-pregnant ewes presented

a higher abundance of Fusobacteriota, Bacteroidota, and Uroviricota. Only one of them (Fusobacteriota) was also reported in **Sudy 1**. Functional profiling further revealed twelve COG entries and two KO that were more abundant in the non-pregnant group (Figure 5.2.4).

In the Manchega RN herd, non-pregnant ewes showed higher abundance of the genera *Histophilus*, *Fusobacterium*, *Actinobacillus* and *Streptobacillus*; at phylum level they also had higher abundance of Fusobacteriota and Uroviricota. Pregnant ewes, by contrast, displayed higher abundance of *Pseudomonas*. Compared with **Study 1**, the enrichment of *Histophilus* and *Fusobacterium* was replicated, whereas *Actinobacillus*, *Streptobacillus* and Uroviricota were detected only in the present study; *Escherichia-Shigella*, observed previously, was not significant here (Figure 5.2.4).

In the Rasa Aragonesa herd, non-pregnant ewes had higher abundance of the genera *Histophilus*, *Bacteroides*, *Parvimonas*, *Pseudomonas*, *Trueperella* and *Gaviibacter*, together with higher abundance of the phyla Bacteroidota and Fusobacteriota. *Histophilus*, *Bacteroides*, *Fusobacterium* and *Pseudomonas* confirm the associations noted in **Study 1**, whereas *Parvimonas*, *Trueperella*, *Gaviibacter* and Bacteroidota are new to the present dataset; again, *Escherichia-Shigella* was not significant (Figure 5.2.4).

No differentially abundant genera or phyla were identified in the Latxa or Manchega VL herds, whereas **Study 1** had reported a few significant taxa in those groups.

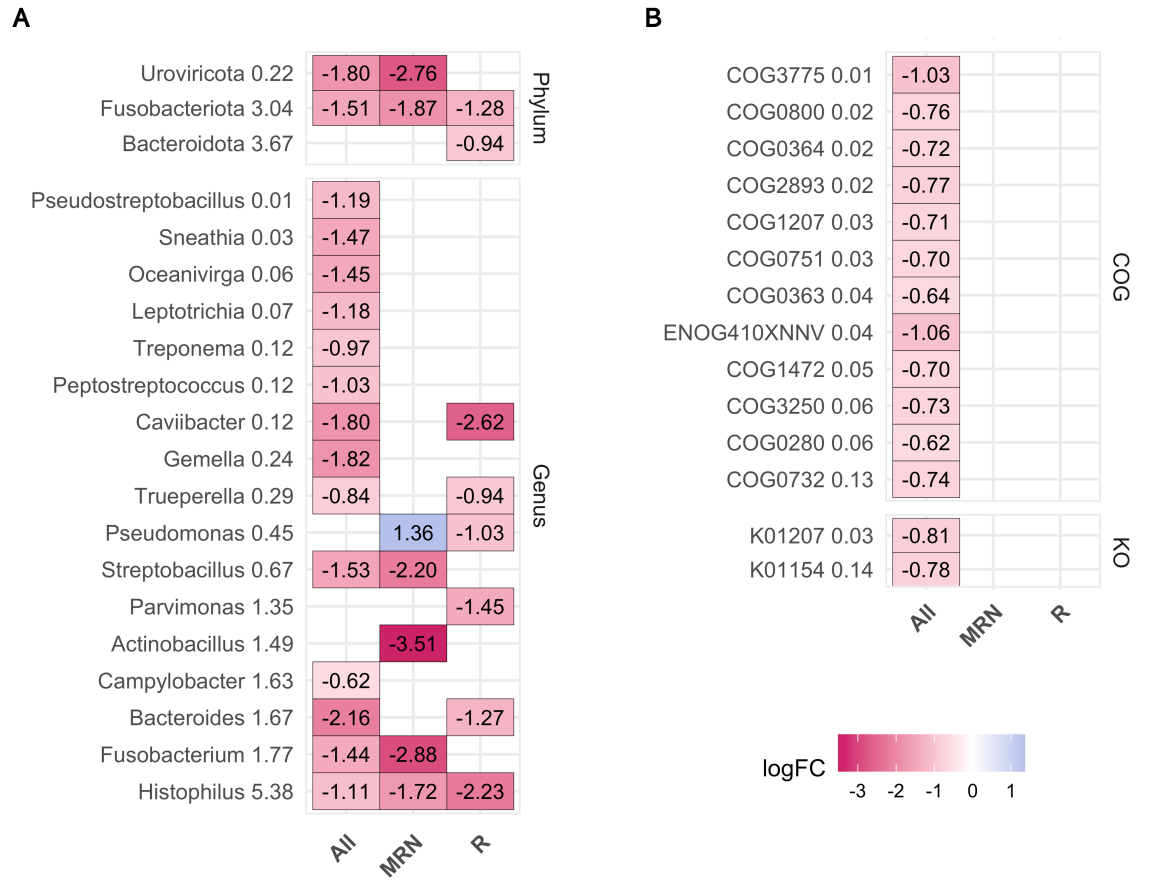


Figure 5.2.4. Heatmap showing significant results (FDR at 5%) for differential abundance analysis for pregnancy success. Panel (A) shows the results at phylum and genus taxonomic levels, while panel (**B**) presents results for COG and KO gene levels. The x-axis represents results for the global model (ALL) as well as within the Manchega RN (MRN) and Rasa (R) herds. No significant taxa were associated with the Latxa (L) or Manchega VL (MVL) herds. The log2FC is represented by colour gradations and its corresponding value, with red indicating higher abundance in non-pregnant ewes and blue indicating higher abundance in pregnant ewes. The colour intensity correlates with the log2FC values, values close to zero are represented in white. The log2FC value is also presented in each square of the heatmap. The y-axis categorizes the taxonomic assignments followed by its RA.

5.2.6. Integrative gene-taxa associations

Sequences aligned with the more significantly abundant COG entries and KOs in non-pregnant ewes corresponded predominantly to *Bacteroides*, *Histophilus*, *Streptobacillus*, *Fusobacterium*, *Actinobacillus*, *Campylobacter*, *Parvimonas*, and *Trueperella*, as well as the phyla Fusobacteriota and Bacteroidota. Figure 5.2.5 illustrates these associations, highlighting the connections between functional pathways and bacterial taxa in non-pregnant ewes.

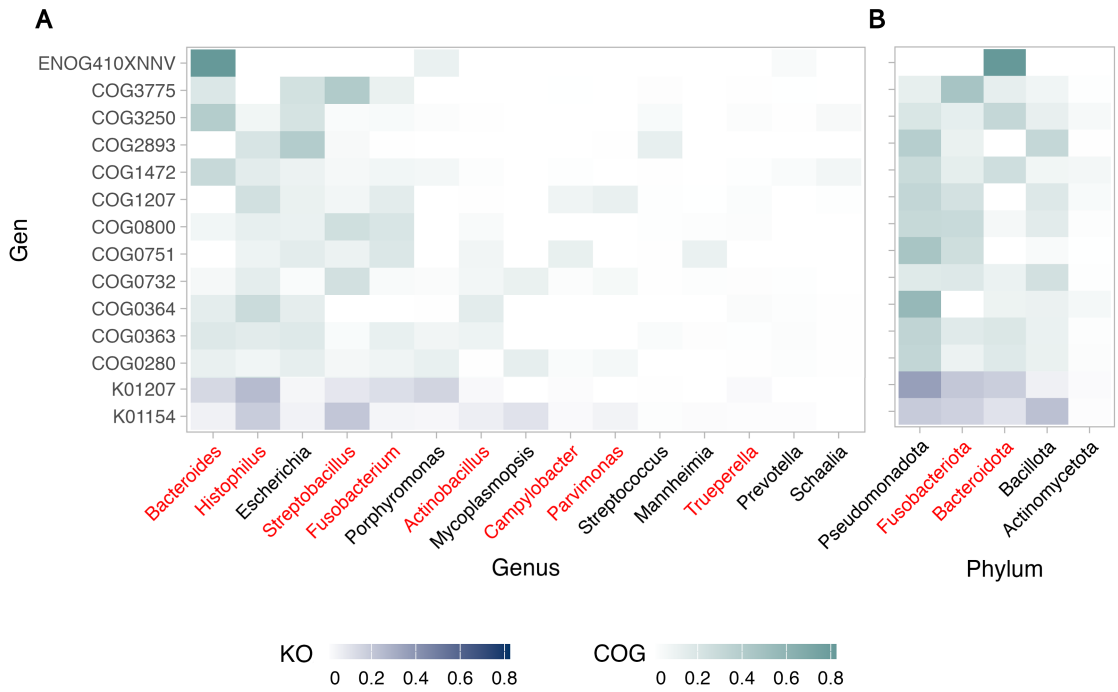


Figure 5.2.5. Heatmap of the relationship between genes and taxa differentially expressed in relation to pregnancy. The y-axis displays the genes (COG in green and KO in blue) that are significantly associated with pregnancy status, along with their corresponding taxa represented in terms of RA. The x-axis outlines the 15 most abundant genera (Panel A) and the 5 most abundant phyla (Panel b) associated with these COGs and KOs. Taxa that were significantly more abundant in non-pregnant ewes are highlighted in red, while those with no significant association are shown in black.

The identified COG entries were linked to essential cellular functions, including carbohydrate transport and metabolism (COG3775, COG2893, COG1472, COG0363, COG0364, COG3250, and COG0800), translation, ribosomal structure, and biogenesis (COG0751), defence and restriction systems (COG0732), energy production (COG0280), cell envelope biogenesis and outer membrane processes (COG1207), and inorganic ion transport and metabolism (ENO4G10XNNV). Similarly, the KO were predominantly associated with glycosaminoglycan degradation, amino sugar and nucleotide sugar metabolism (K01154), and beta-lactam resistance (K01207). These findings illustrate a diverse range of biochemical activities, spanning from fundamental cellular processes and energy production to specialised functions in environmental interactions and defence mechanisms (Figure 5.2.6).

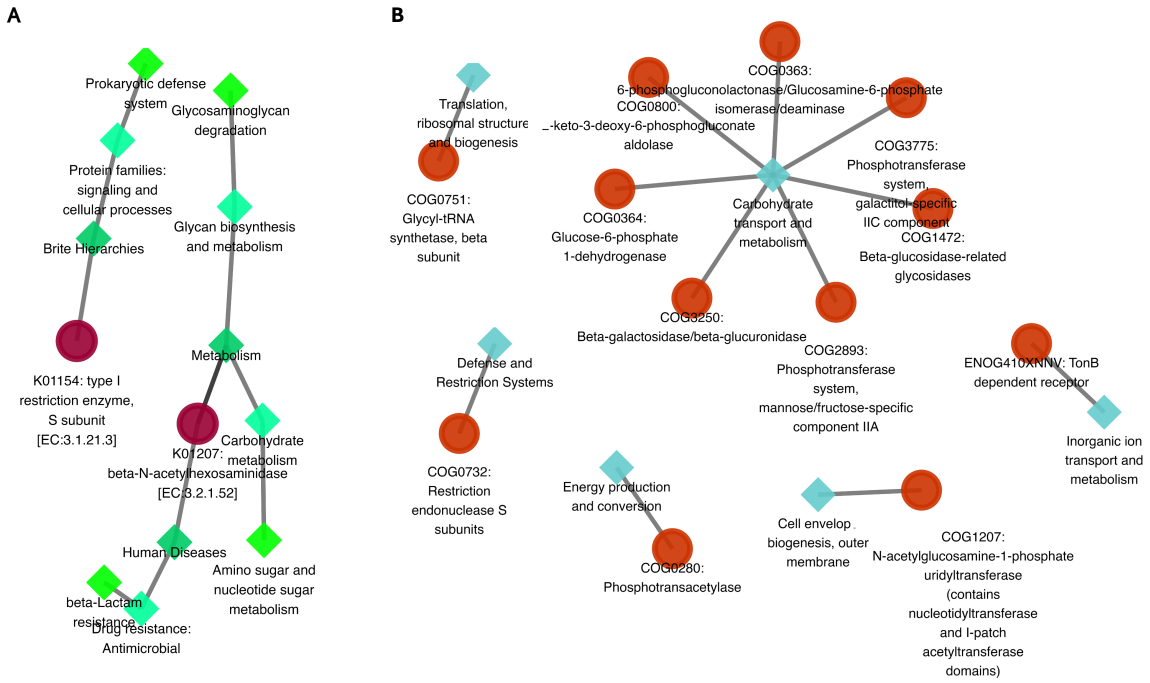


Figure 5.2.6. Network diagram of pathways associated with genes differentially expressed in pregnancy. Panel (**A**) displays the pathways involving KO, while Panel (**B**) illustrates the pathways linked to COG entries.

5.3. Study 3: Impact of oestrus synchronisation devices on vaginal microbiota and artificial insemination success

While the first two studies focused on characterising the microbial communities and their association with pregnancy outcomes, breed, and herd, this study introduces additional variables related to the synchronisation protocols and treatments commonly applied in the context of oestrus synchronisation.

Three main variables were considered: the effect of PRID on the vaginal microbiota; the influence of different treatments (see Table 4.1 for details) added to the PRIDs; and the association between microbiota composition and pregnancy outcomes within this experimental framework. To address these aims, vaginal samples were collected from 47 Assaf ewes —a different breed and dataset from those included in the two previous studies— at two time points: before PRID insertion and after PRID removal.

A ONT-based metagenomic approach was employed, enabling taxonomic characterisation at the phylum and genus levels, as well as functional profiling based on COGs and KOs. The results presented in this section include assessments of alpha and beta diversities, differential abundance analyses to identify microbial changes associated with PRID, and the detection of microbial taxa and functional genes potentially linked to pregnancy outcomes.

The overall fertility rate of the 47 inseminated ewes was 55%. Fertility rates varied among treatments: 31% in the probiotic group, 54% in the antibiotic group, 60% in the maltodextrin group, and 82% in the control group.

5.3.1. Microbiota composition

The sequencing data from 93 samples —comprising two samples per ewe, except for one sample that was discarded due to failing technical and quality control criteria— yielded a total of 1,890,475 reads. After quality filtering, 482,543 reads were taxonomically assigned to Bacteria, Archaea, Protista, Fungi, and Viruses. Taxonomic classification identified 74 phyla and 557 genera, along with 5,048 COG entries and 5,968 KEGG orthologues (Table 5.3.1).

Table 5.3.1. Summary statistics of reads abundance across taxonomic and functional levels.

Level	Total Reads	Mean Sample	Median Sample	Number of Taxa/Genes
Phylum	442,871	4,762	1,695	74
Genus	222,994	2,398	759	557
COG	1,278,862	13,751	2,542	5048
KO	1,218,980	13,107	2,412	5968

Level: Taxonomic or functional classification level; **Mean Sample**: The average number of reads per sample; **Median Sample**: The median number of reads per sample; **Number of Taxa/Genes**: The total number of distinct taxa or genes identified at each classification level. All the unclassified reads were discarded at each taxonomic level, retaining taxa that belong to Bacteria, Archaea, and Viruses, as well as to the phyla Apicomplexa, Ascomycota, Euglenozoa, Evosea, and Nematoda.

Although the main objective of this chapter was not to define a core microbiota — thus, no prevalence threshold was applied as in previous studies— an overview of the microbial composition is provided to highlight the most abundant taxa in this dataset.

The most abundant genera for Assaf ewes included *Staphylococcus* (5.95 %), *Mycoplasma* (4.62 %), *Histophilus* (4.24 %), *Fusobacterium* (3.12 %), *Escherichia* (2.37%), *Porphyromonas* (2.26 %), *Actinobacillus* (2.21 %), *Streptobacillus* (1.95 %), *Bacteroides* (1.82 %), and *Ureaplasma* (0.65 %). At the phylum level, the most abundant taxa were Firmicutes (31.57 %), Proteobacteria (24.17 %), Fusobacteria (10.71%), Bacteroidetes (8.69%), Tenericutes (6.73%), Actinobacteria (5.32%), Spirochaetes (0.62%), Chlamydiae (0.13%), Ascomycota (0.9%), and Euryarchaeota (0.08%). Figure 5.3.1 provides an overview of the microbiota's RA at both the phylum and genus levels in the S1 and S2 time points. A supplementary figure by pregnancy status, as well as at treatment is available in Appendix – Figure S3.1 and Figure S3.2.

The five most abundant COGs identified were ENOG410YQYP (2.79 %), ENOG410Y65V (1.21 %), ENOG410XST2 (1.19 %), COG5295 (1 %), and ENOG4111K87 (0.50 %). Similarly, the five most prevalent KOs were K09228 (5.55 %), K21449 (1.09 %), K08111 (0.97 %), K06147 (0.49 %), and K01154 (0.47 %). Detailed information on the composition of COGs and KOs is presented in Appendix – Figure S3.3.

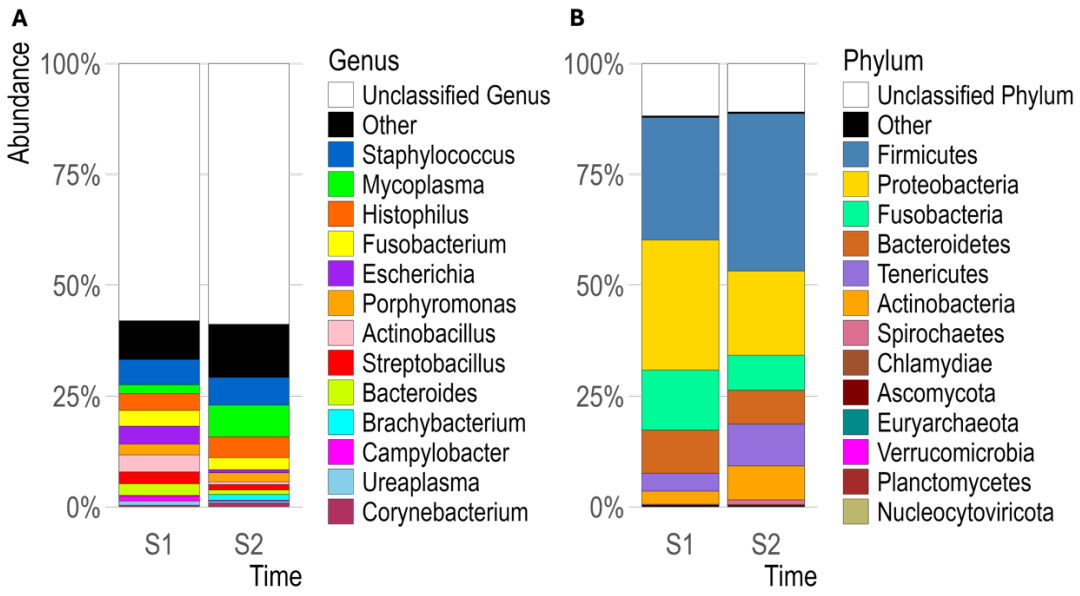


Figure 5.3.1. Composition of the microbiota for S1 (before PRID placement) and S2 (after PRID removal) sampling point groups at the phylum (**A**) and genus (**B**) levels. In these composition analyses, no prevalence or RA filters were applied.

5.3.2. Alpha-diversity

Alpha-diversity was significantly higher in S2 compared to S1 across most metrics (Figure 5.3.2). Within the S2 treatments, the control group showed significantly higher alpha-diversity values compared to the antibiotic group. No significant differences were observed among the other treatments.

Although not statistically significant, pregnant ewes exhibited higher alpha-diversity than non-pregnant ewes across all four metrics (Appendix – Figure S3.4).

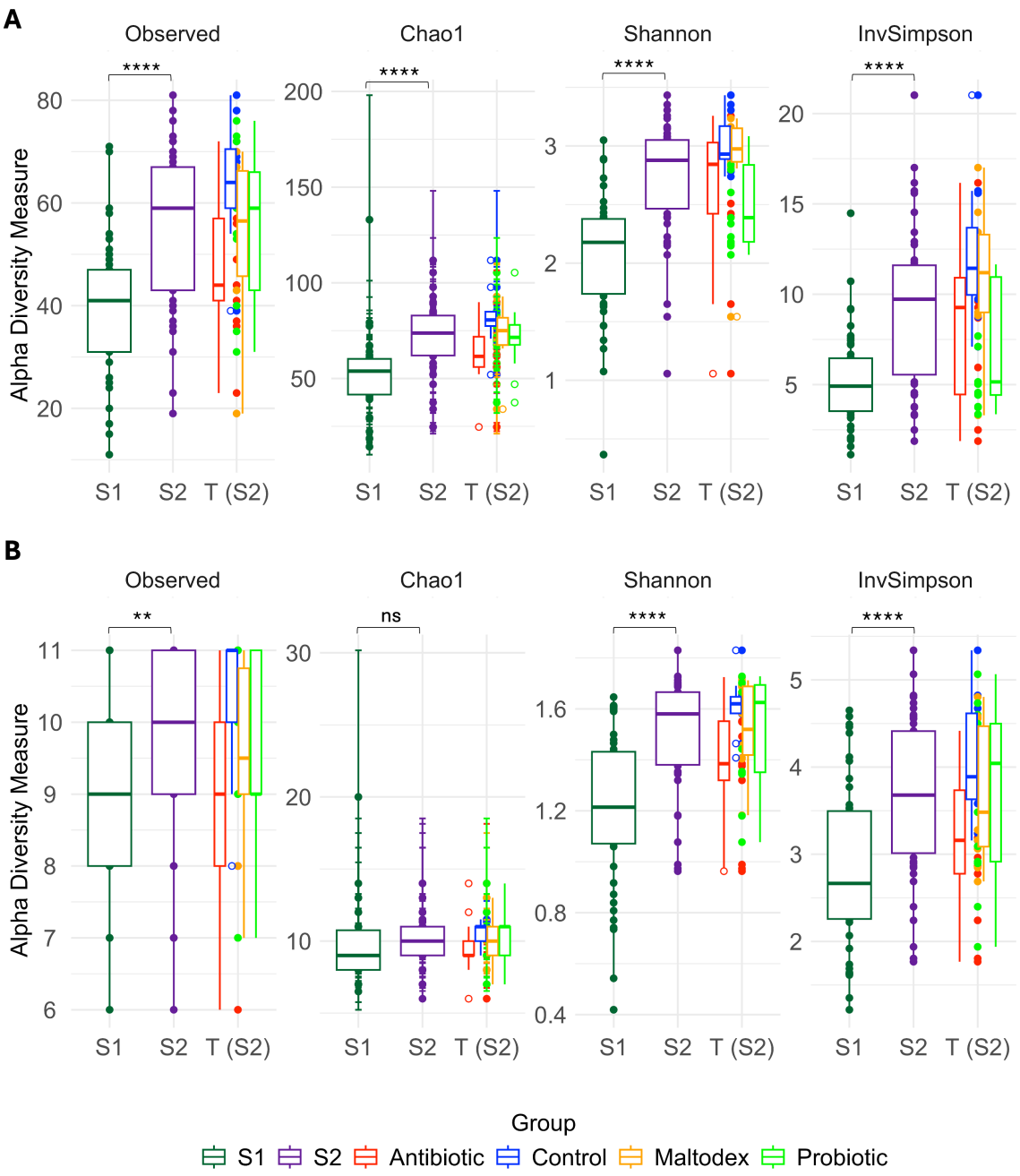


Figure 5.3.2. Alpha-diversity measures (Observed, Chao1, Shannon, and Inverse Simpson indices) for sampling point groups at the genus (**A**) and phylum (**B**) levels. S1: before PRID treatment, S2: after PRID treatment, T(S2): treatment within the S2 group. Treatments within the S1 group were not considered, as they were administered alongside PRID, and their effects were evaluated within the S2 group. Statistical significance is indicated by asterisks: * ($p < 0.05$), *** ($p < 0.001$); ns: not significant. Data were filtered to retain taxa with > 0.05% relative abundance.

5.3.3. Beta-diversity

Principal Component Analysis revealed a clear clustering across S1 and S2 sampling points, particularly at the genus level, accounting for 17.25% (PCA1) and 8.51% (PCA2) of the total variance (Figure 5.3.3 A). While there were relevant differences between time points S1 and S2, explaining 21.3% (PCA1) and 7.9% (PCA2) of the total variance (Figure 5.3.3 B). For COGs and KOs no pattern of differentiation was observed between groups.

PERMANOVA analysis revealed significant differences in overall microbiota community composition between pregnant and non-pregnant ewes only at the genus level, and no significant differences among PRID treatments at any level. However, highly significant differences were found at the phylum, genus and KEGG categories between sampling points (S1 vs. S2). At the phylum and genus levels, the highest changes observed in beta-diversity were for the control and maltodextrin treatments. At the functional level (KO), only for the antibiotic (0.006) and maltodextrin (0.025) treatments, significant changes in beta-diversity between time points S1 and S2 were detected. Table 5.3.2 provides a detailed summary of the PERMANOVA results for pregnant vs. non-pregnant ewes and S1 vs. S2 time points, specified by treatment.

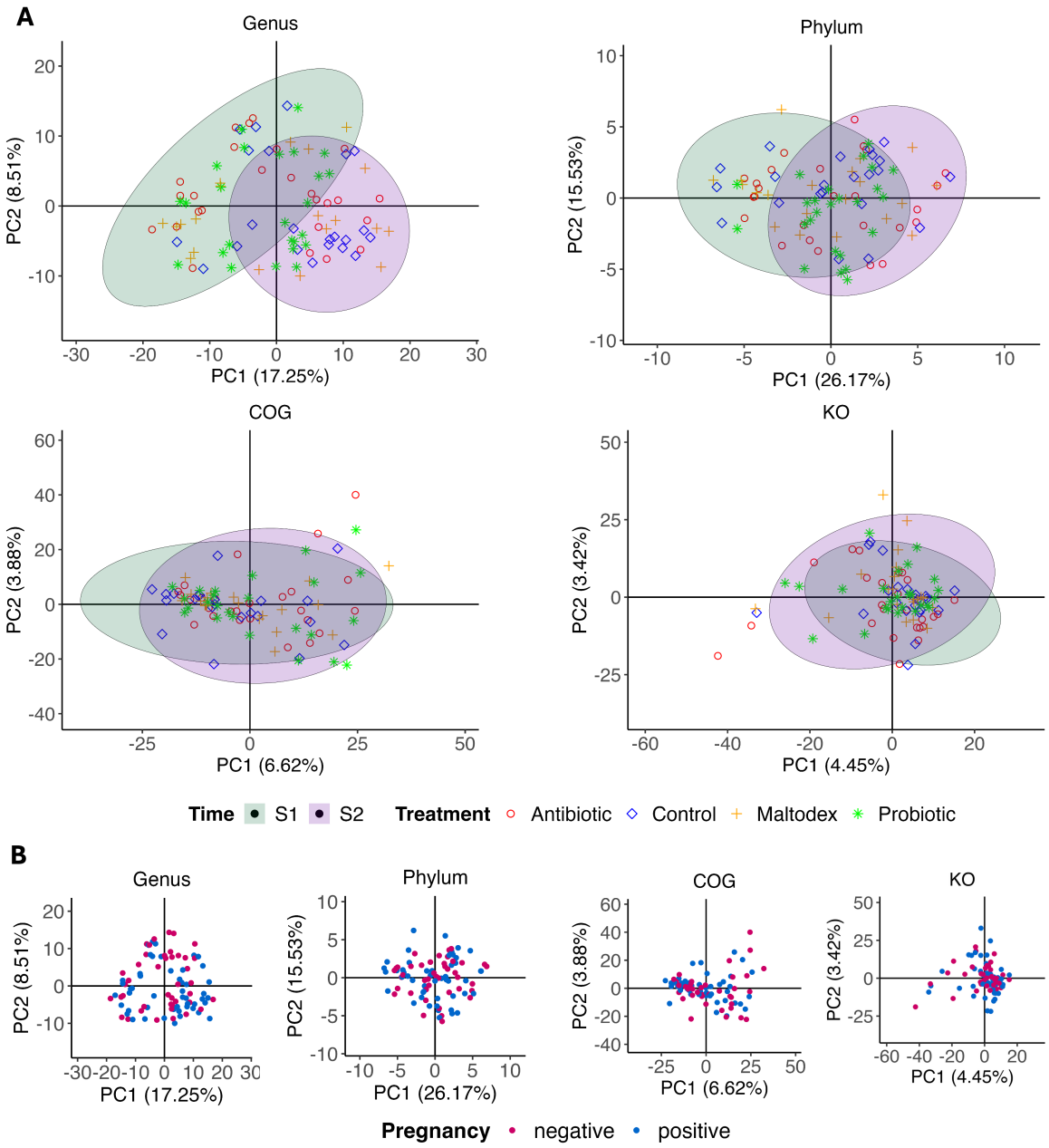


Figure 5.3.3. Principal Component Analysis for S1 (before PRID placement) and S2 (after PRID removal) sampling points (**A**) groups and pregnancy (**B**) groups at genus, phylum, COG, and KO level. Data were filtered to retain taxa with > 0.05% relative abundance.

Table 5.3.3. PERMANOVA results for pregnancy (P and N) and sampling time groups (S1 and S2) at genus, phylum and KO levels

		Genus		Phylum		COG		KO	
	Number (P/N)	SumOfSqs	Pr(>F)	SumOfSqs	Pr(>F)	SumOfSqs	Pr(>F)	SumOfSqs	Pr(>F)
S1 vs S2	93	4962.02	0.001	483.72	0.001	3557.73	0.023	3529.48	0.004
Antibiotic	26	1904.68	0.001	174.58	0.001	3678.81	0.036	3307.45	0.055
Control	22	1536.22	0.001	194.03	0.001	2266.06	0.842	2122.47	0.917
Maltodextrin	20	1812.13	0.001	191.27	0.001	3537.31	0.009	3617.95	0.002
Probiotic	26	1259.65	0.001	64.40	0.044	2724.34	0.669	2715.77	0.283
Treatment	47	679.52	0.036	44.88	0.115	4390.04	0.088	4726.50	0.368
P vs N	47 (26/21)	282.38	0.018	6.18	0.81	1536.72	0.105	1691.26	0.122
Antibiotic	13 (7/6)	271.18	0.095	15.36	0.312	1755.47	0.376	1838.85	0.701
Control	11 (9/2)	212.97	0.081	5.59	0.665	839.30	0.385	1039.63	0.362
Maltodextrin	10 (6/4)	154.49	0.86	7.96	0.824	1438.63	0.976	1766.14	0.742
Probiotic	13 (4/9)	129.65	0.888	2.94	0.958	1316.80	0.319	1335.09	0.565

P: positive for pregnancy, N: negative for pregnancy. Numbers of Pr(>F) in black colour represent those that were significant at an alpha value of 0.05. Data were filtered to retain taxa with > 0.05% relative abundance.

5.3.4. Differential abundance analysis

For the comparison between sampling points (S1 vs S2), six phyla and 24 genera were significantly more abundant in the S1 group. The phyla with the highest log2FC were Ascomycota and Nematoda, while the genera with the highest log2FC included *Actinobacillus*, *Aggregatibacter*, *Sneathia*, and *Oceanivirga*. No COGs or KO pathways showed differential abundance between sampling points. In contrast, the S2 group displayed higher abundance for five phyla (Euryarchaeota, Spirochaetes, Tenericutes, Candidatus Saccharibacteria, and Actinobacteria), 25 genera (the most abundant being *Psychrobacter, Kocuria, Jeotgalicoccus, Brachybacterium, Micrococcus, Mycoplasma, Leucobacter, Salinicoccus, Olsenella*, and *Corynebacterium*), two COGs (COG1196 and ENOG410XQ90), and one KO pathway (K06919). These results are summarised in Figure 5.3.4 A and B for phylum and genus levels, respectively.

Among PRID treatments, many genera in common were observed to decrease and increase between S1 and S2 time points, except for the probiotic-treated samples, where only one genus (*Campylobacter*) decreased, and five genera increased from S1 to S2. The antibiotic and maltodextrin treatments promoted the shift of a greater number of genera between S1 and S2, many of them common and others specific to each treatment. Detailed results of the differential abundance analyses for individual treatments are provided in Appendix – Figure S3.5 (antibiotic), Figure S3.6 (control), Figure S3.7 (maltodextrin), and Figure S3.8 (probiotic).

Regarding pregnancy, significant differences between pregnant and non-pregnant ewes were detected exclusively at the genus level. Specifically, the genera *Oenococcus* and *Neisseria* were more abundant in non-pregnant ewes than in pregnant ones.

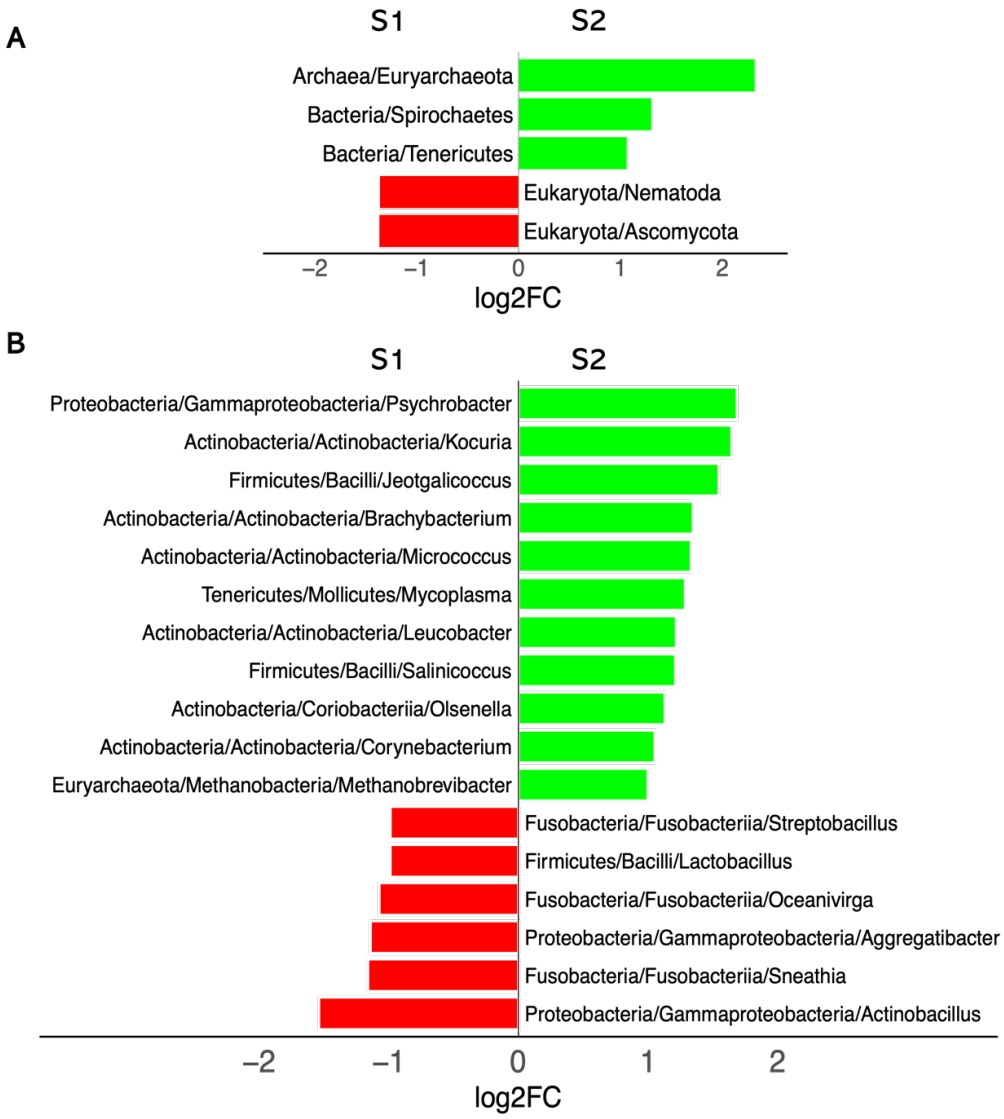


Figure 5.3.4. Stacked bar chart showing significant results (FDR at 5%) for differential abundance analysis for sampling point groups S1 and S2 at phylum (**A**) and genus (**B**) levels. A threshold for $|\log_2 FC| > 1$ was imposed for representing the figure. Red represents a higher abundance for S1 compared to S2, while green represents a higher abundance for S2 compared to S1. Data were filtered to retain taxa with > 0.05% relative abundance.

In order to evaluate whether the samples were a random selection from the population and the potential occurrence of bias in the distribution of samples across groups, comparative analyses were performed within the S1 group (before PRID). Results from these analyses showed non-significant differences between pregnant and non-pregnant groups within S1 (Table S3.1), thus validating the absence of bias due to sampling when ewes were assigned to each PRID treatment. The distribution of taxa across treatments within groups is shown in Appendix – Figure S3.1.

5.4. Study 4: host genomic regions associated with vaginal bacterial communities

This fourth study explores how host genetics may influence microbial abundance across different taxonomic levels. Unlike the previous three studies, which primarily examined the microbiota in relation to environmental and physiological factors, this analysis focuses on the host–microbiota relationship from a genomic perspective, aiming to identify genetic variants in the host that are statistically associated with variations in microbial abundance.

As a reminder, host genotype data were obtained from 243 ewes belonging to the set of samples analysed in **Studies 1** and **2**, and microbiota characterisation was based on 16S rRNA metabarcoding. These datasets were integrated within an MGWAS framework to identify SNPs significantly associated with the abundance of microbial taxa at the phylum, genus, and species levels, with particular emphasis on taxa potentially linked to reproductive function.

From that point onward, the associated genetic variants were subsequently examined through functional analyses to elucidate the biological roles of the involved genes. Additionally, network analysis was performed to further explore their functional relationships.

Microbiota genome-wide association analysis

A separate MGWAS was performed for the relative abundance (CLR transformation) of each taxon analysed (35 phyla, 353 families, 885 genera, and 349 species), following the taxonomic assignment as detailed in Material and methods-DNA extraction section.

From the initial number of SNPs, a total of 24,641 SNPs were identified as significantly associated with the vaginal microbiome in ewes after FDR correction. Specifically, 17 SNPs were linked to 9 phyla, 11,741 SNPs to 508 genera, and 19,884 SNPs to 260 species.

Among the most abundant microbial taxa, several exhibited significant associations with host-SNPs, including the phyla Fusobacteriota and Actinobacteriota; the genera *Rothia*, *Moraxella*, *Alisiella*, and *Flavobacterium*; and the species *Streptococcus equi*, *Spingobacterium alimentarum*, *Moraxella boboculi*, and *Parvimonas micra* (Figure 5.4.1).

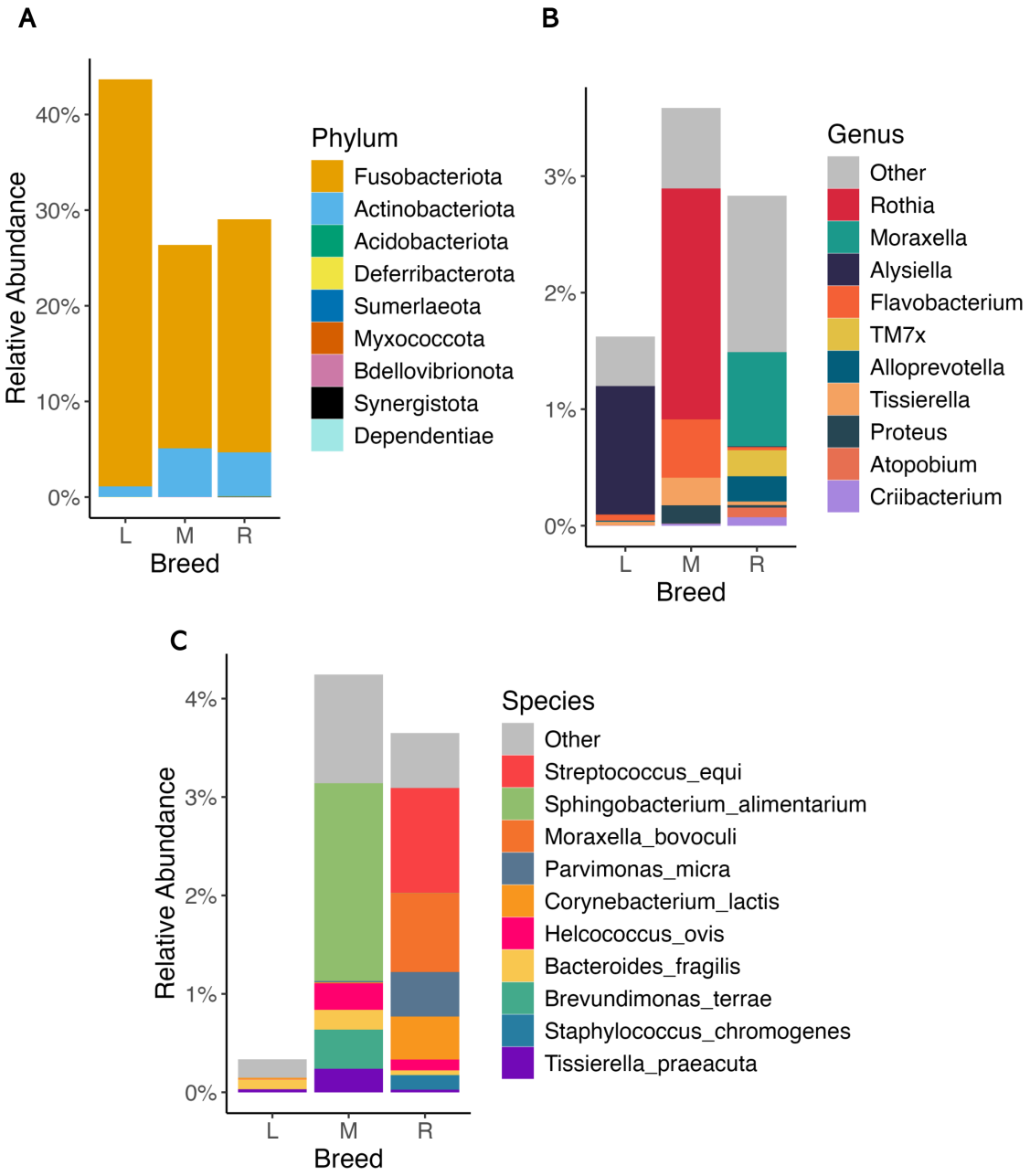


Figure 5.4.1. Relative abundance of significantly associated taxa identified in the MGWAS across sheep breeds. The figure shows the RA of taxa with at least one SNP significantly associated, categorised at three taxonomic levels: (**A**) phylum, (**B**) family, and (**C**) species, across three sheep breeds (L: Latxa, M: Manchega, and R: Rasa). The y-axis represents the RA of each taxon within the whole vaginal microbiota.

When considering taxa with the highest number of associated SNPs (irrespective of their abundance), Figure 5.4.2 summarises the top 5 phyla, 15 genera, and 15 species based on the total number of significant associations. These included: the phyla Synergistota (5 SNPs) and Sumerlaeota (4 SNPs); the genera *Weeksella* (577 SNPs),

Diaphorobacter (443 SNPs), *Ottowia* (428 SNPs), *Aquabacterium* (404 SNPs), and *Acidovorax* (361 SNPs); and the species *Kocuria rosea* (682 SNPs), *Acinetobacter gandensis* (599 SNPs), *Pseudomonas oryzihabitans* (529 SNPs), *Novosphingobium lindaniclasticum* (526 SNPs), and *Thiopseudomonas denitrificans* (500 SNPs).

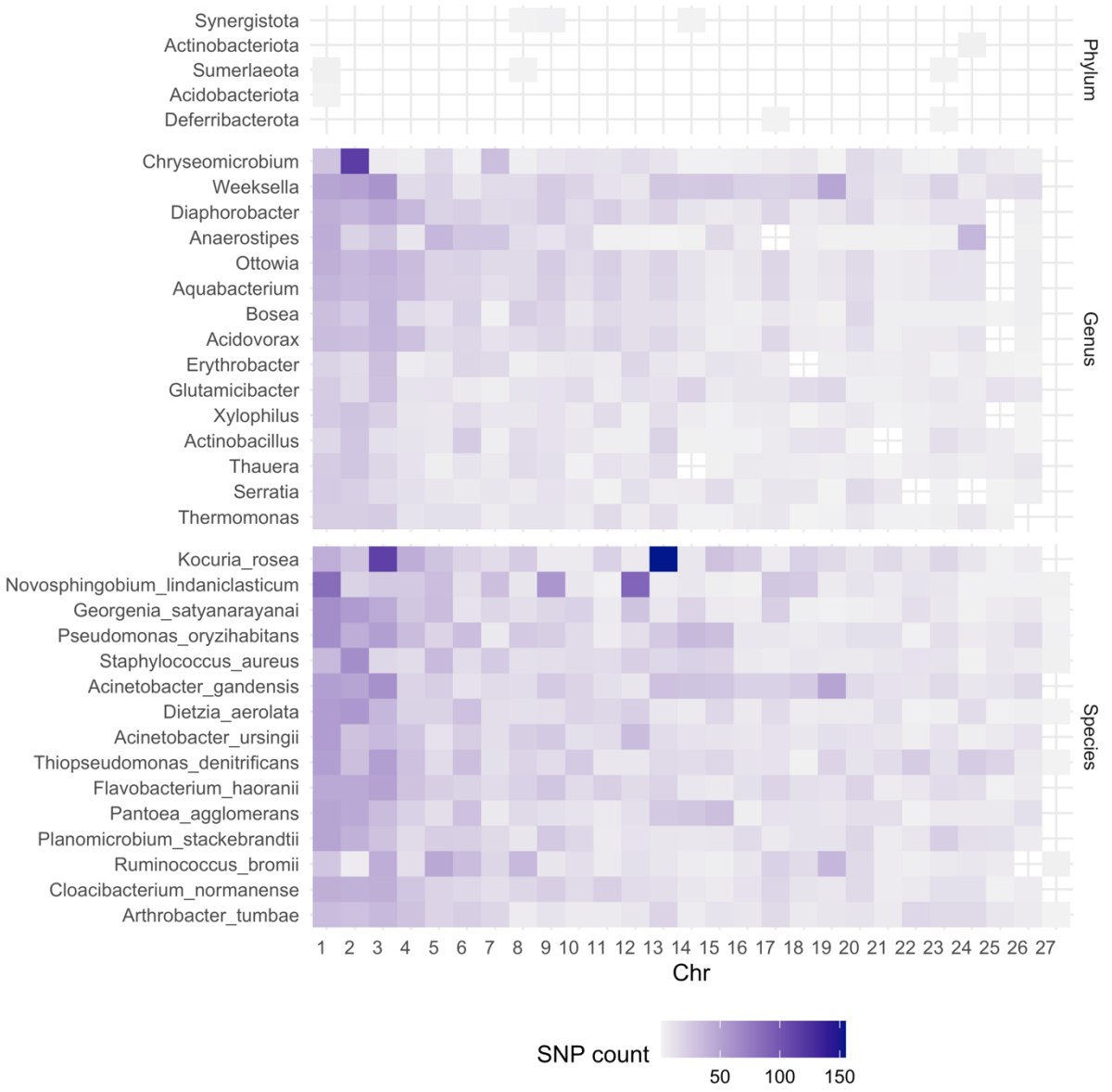


Figure 5.4.2. Distribution of significantly associated SNPs across chromosomes (Chr, x-axis) and taxa (y-axis). Taxa are grouped by taxonomic level, ordered by the number of associated SNPs, and include the 5 phyla, 15 genera, and 15 species with the highest SNP counts; other taxa are not shown in this plot. The x-axis represents the 27 sheep chromosomes, while the colour intensity reflects the number of associated SNPs, with darker shades indicating higher counts.

5.4.1. SNP associations with key fertility taxa

This section focuses on microbial taxa that have been linked to fertility in previous research, including findings from both this thesis and the scientific literature (see Appendix – Table S4.1). A total of 25 of these fertility-related taxa were found to have significant associations with 3,790 SNPs in the host genome. These taxa included the phylum Fusobacteriota; the genera *Actinobacillus*, *Anaplasma*, *Bergeyella*, *Coxiella*, *Erysipelothrix*, *Falsiporphyromonas*, *Fluviicola*, *Glutamicibacter*, *Lactobacillus*, *Mannheimia*, *Methanosphaera*, *Mycobacterium*, *Mycoplasma*, *Neisseria*, and *Streptomyces*; and the species *Bacteroides fragilis*, *Bifidobacterium breve*, *Fusobacterium nucleatum*, *Gardnerella vaginalis*, *Helcococcus ovis*, *Peptostreptococcus anaerobius*, *Porphyromonas levii*, *Prevotella nigrescens*, and *Staphylococcus aureus*. Figure 5.4.3 presents a Manhattan plot showing the chromosomal distribution of SNPs associated with these taxa.

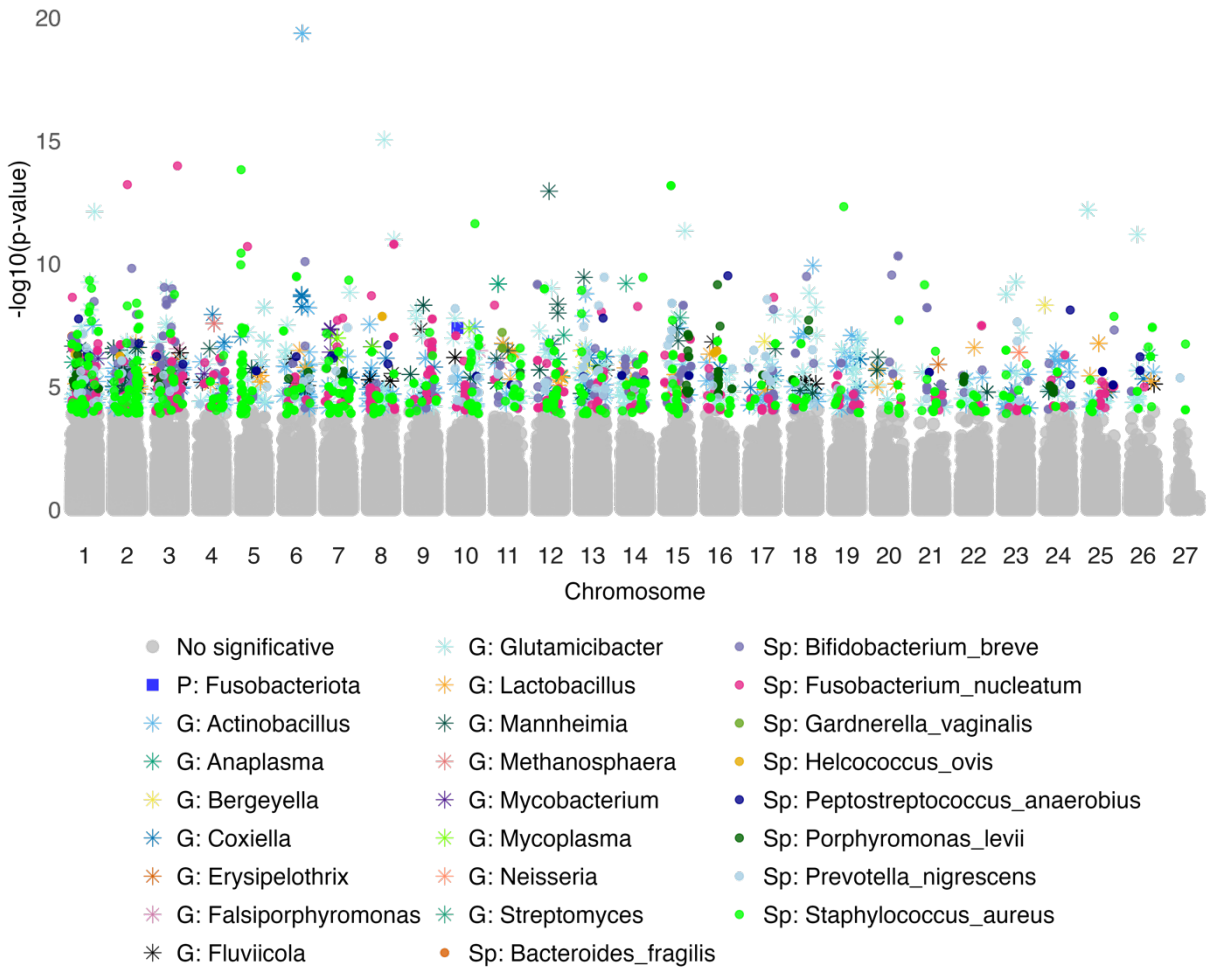


Figure 5.4.3. Manhattan plot for the Microbiota Genome-Wide Association Studies. Each point represents the -log(p-value) associated with the regression of the phenotype on the SNP

genotype, with the corresponding position in the chromosome represented on the x-axis. Different colours denote distinct taxa with significant associations at a FDR-adjusted p-value < 0.10, while grey points indicate non-significant associations. P: phylum; G: genus; Sp: species.

Fertility-related taxa with significant host-SNP associations were selected for downstream analyses to investigate the functional relevance of the host variants involved.

5.4.2. Variant Effect Predictor analysis

To evaluate the potential biological relevance of the SNPs associated with fertility-related microbial taxa, a Variant Effect Predictor (VEP) analysis was conducted. This analysis identifies the genomic position of each SNP and categorizes its predicted impact on gene function. A total of 3,790 variants were examined, with the majority linked to taxa at the species level (2,302 variants), followed by the genus level (1,483 variants), and a smaller number at the phylum level (5 variants; Table 5.4.1). Across all taxonomic levels, intronic variants were the most common (58.4%), followed by intergenic (23.4%), upstream (6.9%), and downstream gene variants (5.3%) (Figure 5.4.4).

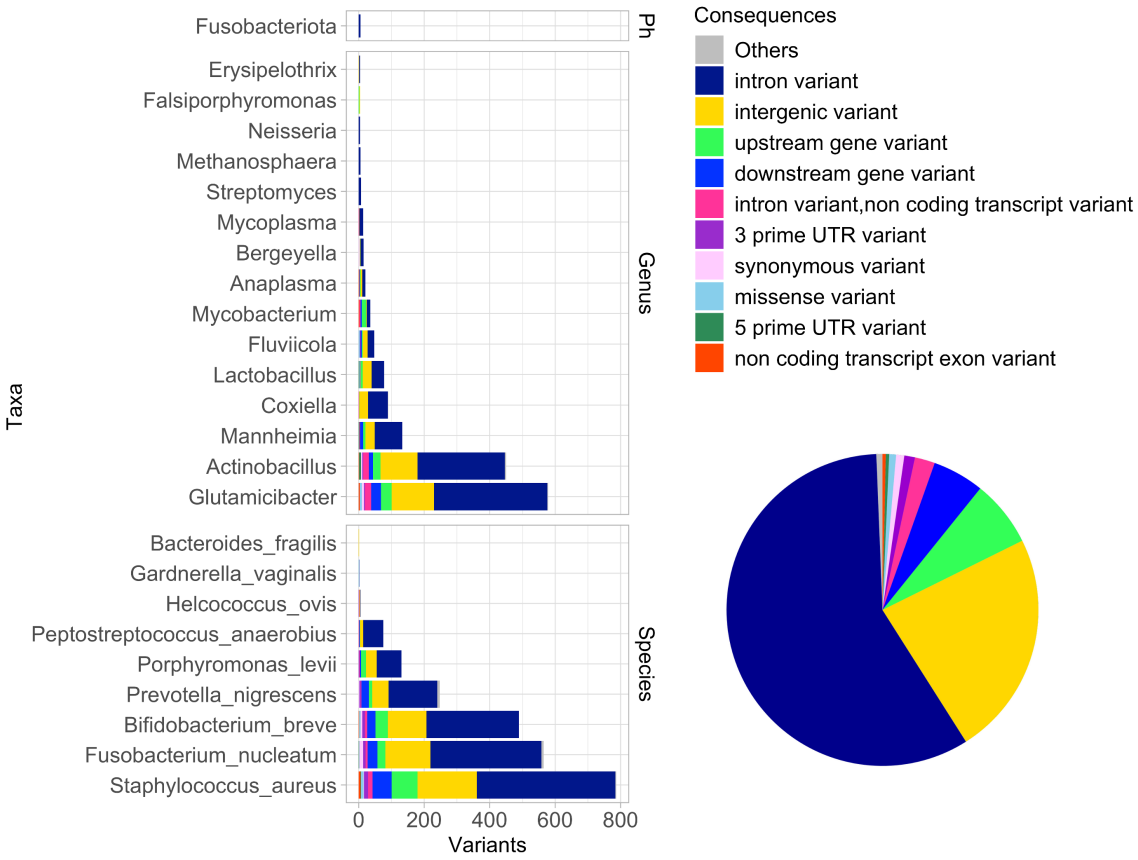


Figure 5.4.4. Distribution of genetic variant consequences across host-associated taxa relevant to sheep fertility. The bar chart displays the number of variants categorised by their predicted consequences, grouped by taxonomic levels (phylum, family, genus, and species) along the y-axis. The pie chart provides a summary of the overall distribution of variant types across all taxa.

Table 5.4.1. Summary of microbial taxa significantly associated with the host genomic regions, including those previously reported in the literature as linked to fertility.

Taxa	*SNPs*	*Genes*
Phylum		
Fusobacteriota	1	1
Genus		
Glutamicibacter	287	140
Actinobacillus	231	114
Mannheimia	66	36
Lactobacillus	47	17
Coxiella	41	15
Fluviicola	26	16
Mycobacterium	11	7
Anaplasma	10	5
Bergeyella	6	5
Mycoplasma	5	2
Falsiporphyromonas	3	1
Erysipelothrix	2	1
Methanosphaera	1	1
Neisseria	1	2
Streptomyces	1	1
Species		
Staphylococcus_aureus	419	230
Fusobacterium_nucleatum	291	145
Bifidobacterium_breve	242	109
Prevotella_nigrescens	112	59
Porphyromonas_levii	63	22
Peptostreptococcus_anaerobius	31	20
Helcococcus_ovis	5	3
Gardnerella_vaginalis	2	1
Bacteroides_fragilis	1	0

The table shows the number of significant SNPs identified along with the number of genes within ±5 kb of each significant SNP.

5.4.3. Functional analysis of non-intergenic SNP-associated genes

For this analysis, only variants located at gene regions were considered, excluding intergenic SNPs. A total of 512 host-associated genes were identified. Among these, 322 were detected at the species level, 209 at the genus level, and one at the phylum level. The taxa with the largest number of associated genes included the species *Staphylococcus aureus* (123 genes), *Fusobacterium nucleatum* (85 genes), and

Bifidobacterium breve (65 genes), as well as the genera *Glutamicibacter* (85 genes), *Actinobacillus* (60 genes), and *Mannheimia* (28 genes). These genes are involved in a variety of molecular processes, including protein complex assembly and interactions, enzymatic activity and regulation, signal transduction, protein synthesis and processing, and metabolic and energy-related pathways. Additionally, their roles extend to cellular processes such as cellular structure organization and cytoskeletal dynamics, as well as biological processes involving gene expression and regulation, development and differentiation, and nervous system function and neurotransmission. At the phylum level (Fusobacteriota), the single gene identified, SETDB2, is associated with functions related to cell division, cell cycle regulation, cellular components, and epigenetic modifications (Figure 5.4.5).

A total of 32 host genes were shared across multiple taxa within the vaginal microbiota. *Fusobacterium nucleatum* shared three genes with *Glutaminibacter*, three with *Bifidobacterium breve*, and two with *Lactobacillus*. The genus *Glutaminibacter* also shared two genes with *Staphylococcus aureus*. Additionally, *Actinobacillus* shared two genes with *Mannheimia*, while *Bifidobacterium breve* shared two genes with *Actinobacillus*. Figure 5.4.6 presents a comprehensive visualisation of all shared genes and their associated taxa.

Finally, each variant was classified according to how strongly it might affect gene function. Only variants predicted to have a low, moderate, or high impact on the gene were included in this analysis. Variants labelled as "modifiers", which are unlikely to have a relevant effect, were excluded (Figure 5.4.7). This analysis identified 28 host genes associated with significant SNPs across 11 taxa. Among the key taxa, the genus *Glutaminibacter*, was linked to genes with roles in cellular signalling, gene regulation, and protein degradation, processes relevant to reproductive function (LRP2, CLK3, ASCC2, *MYMK, KAT6B, EXOSC10*, and *ENSOARG00020003095*). *Actinobacillus* showed associations with genes related to metabolic regulation and chromatin modification (*UBR5*, *HLCS*, and *BLTP3A*), while *Coxiella* was associated with *MOV10L1*, a gene a gene directly implicated in germ cell development. At the species level, *Staphylococcus aureus* and *Bifidobacterium breve* were connected to genes involved in hormone regulation, transcriptional control, and early developmental pathways (*PAPPA2*, *MLLT6*, *ZBTB17*, *MVD* and *TRIM46*, *EIF2D*, *IFRD2,* respectively). *Prevotella nigrensis* was associated with *SDHD,* a gene essential for mitochondrial function and energy metabolism, which are critical in reproductive tissues, along with two additional genes with potential regulatory roles, as well as *HADHB*, and *PITX3*. Figure 5.4.7 provides a detailed summary, including GO annotations for these genes.



Figure 5.4.5. Functional descriptions of host genes associated with fertility-related vaginal microbial taxa. Each circle represents the number of genes within a functional category associated for each specific taxon, with the size of the circle proportional to the number of genes. The colours indicate different functional categories. Taxa are grouped by taxonomic level.

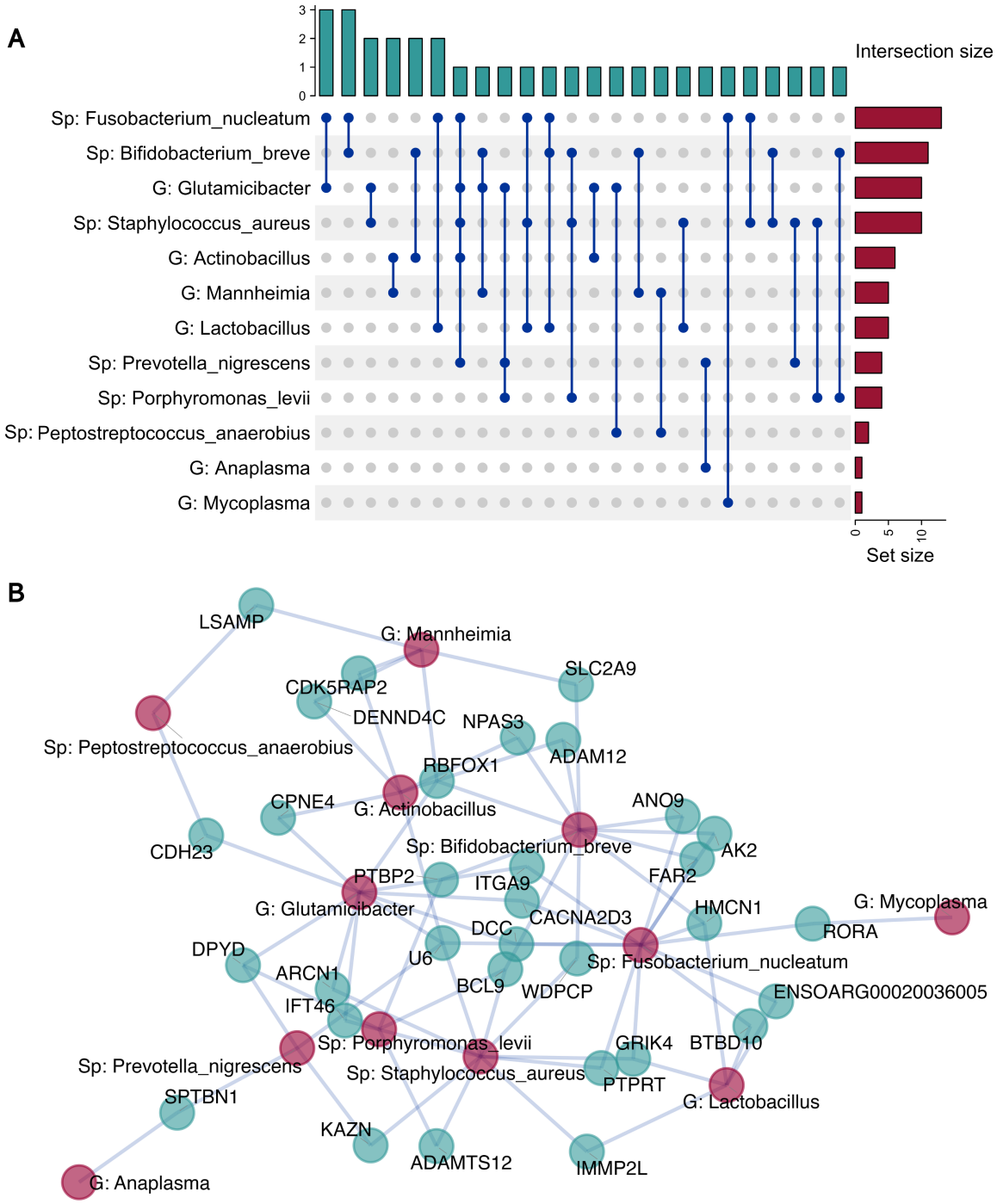


Figure 5.4.6. Network of host genes significantly associated with fertility-related vaginal microbial taxa. The upset plot (**A**) shows the number of genes associated with each taxon (set size) and the intersections between taxa, indicating how many genes are shared among multiple taxa (intersection size). The gene network diagram (**B**) visualises the connections between individual genes and their associated taxa.



Figure 5.4.7. GO classification heatmap of genes with consequential SNPs in association with vaginal microbial taxa. Low to high impact variants (excluding synonymous ones) were considered to select the genes. The y-axis lists the specific host genes, while the x-axis categorizes their broader functional roles based on GO classifications. Each square represents a SNP, with the size indicating the variant's impact and the colour intensity reflecting the SNP effect on the associated taxon. Colours indicate the biological category (Biological Processes, Cellular Components, or Molecular Functions). The right-hand side of the y-axis lists the vaginal microbiota taxa associated with these genes, denoted as G: Genus and Sp: Species.

6. DISCUSSION

6.1. Recapitulation

Sheep breeding programmes face, among others, a significant challenge, the low success rate of AI. Although several factors are known to influence ovine fertility by AI —such as the complexity of the ewe's reproductive tract, the requirement for fresh semen, or the difficulty in accurately detecting the ovulation stage— the role of the ewe's reproductive microbiota remains not fully understood. Advances in sequencing technologies offer promising opportunities to explore these microbial communities. In this context, the aim of this thesis was to achieve a comprehensive understanding of the ewe's vaginal microbiota through four complementary studies, which collectively aimed to: (i) characterise the vaginal microbiota and identify a core microbiota common to the sheep species; (ii) evaluate the influence of key factors —including breed, herd, age, synchronisation treatments, and pregnancy status— on microbial composition; (iii) detect specific bacterial taxa and functional genes associated with AI fertility outcome; and (iv) investigate the host's genetic contribution to microbial composition, aiming to identify genes linked to the abundance of key taxa.

The structure of this thesis follows a conceptual, rather than chronological, order. Hence, some methodological decisions were shaped by timing and sample availability. For instance, two DNA extraction protocols were employed; samples in **Study 3** were processed prior to the start of this thesis using the QIAamp DNA Microbiome Kit, while **Studies 1**, **2**, and **4** used the Real Vaginal Microbiome DNA Kit, selected for its greater efficiency in lysing diverse microbial cells and reducing host DNA contamination —particularly relevant in low-biomass vaginal samples. Furthermore, differences in sequencing platforms may have influenced sequencing depth and microbial resolution. **Study 2** used the GridION device with adaptive sampling, allowing real-time depletion of host DNA and resulting in higher microbial read yields. In contrast, **Study 3** used the MinION device without adaptive sampling, which likely contributed to lower microbial recovery. These methodological variations should be taken into account when comparing microbial profiles across studies.

In this section, we discuss the key findings across all four studies, addressing microbiota composition, drivers of microbial variation, differentially abundant taxa, and host genetic associations revealed through MGWAS.

6.2. Microbial composition

Our assessment of the compositional core microbiota of the ewes' vaginal tract (determined in **Studies 1** and **2**) identified several common phyla with both 16S rRNA metabarcoding and ONT-based metagenomic methods, including Fusobacteriota, Bacillota (Firmicutes), Proteobacteria (also referred to as Pseudomonadota and Campylobacterota), Bacteroidota, and Actinomycetota (Actinobacteriota). These phyla were also identified in the vaginal tract of ovine species by Swartz et al. (2014), Serrano et al. (2020), and Greenwood et al. (2022), as well as in cattle by Swartz et al. (2014), Ong et al. (2021), and Amat et al. (2022). At the genus level, *Histophilus*, *Fusobacterium*, *Parvimonas*, *Aerococcus*, *Bacteroides*, and *Streptococcus* were consistently detected by both sequencing methods, as also reported by Swartz et al. (2014) and Serrano et al. (2020), reinforcing their role as stable members of the ewe's core vaginal microbiota.

Several genera, such as *Streptobacillus*, *Oceanivirga*, *Porphyromonas*, *Anaerococcus*, and *Trueperella,* were exclusively detected through 16S rRNA metabarcoding. This is likely due to the amplification step inherent to the method, which enhances the detection of low-abundance taxa, along with a more comprehensive representation of these genera in 16S rRNA-specific databases and the typically higher number of reads generated compared to the ONT-based metagenomic approach. In contrast, ONT-based metagenomics revealed broader diversity, identifying taxa not captured by 16S rRNA metabarcoding. These included phyla such as Mycoplasmatota, Spirochaetota, Nucleocytoviricota (virus), and Ascomycota (Fungi), as well as genera such as *Staphylococcus*, *Escherichia*, *Mycoplasmopsis*, *Campylobacter*, *Actinobacillus*, *Anaplasma*, *Corynebacterium*, *Enterococcus*, *Klebsiella*, *Clostridioides*, *Cutibacterium*, *Shigella*, *Citrobacter*, and *Coxiella*. This broader detection is likely due to multiple factors. Firstly, 16S rRNA metabarcoding relies on the amplification of specific regions of the 16S rRNA gene, limiting its ability to detect taxa with mismatches to commonly used primers. Genera such as *Mycoplasma* and *Campylobacter* are known to be underrepresented in 16S rRNA studies for this reason (Winand et al., 2019). In contrast, although ONT sequencing does not rely on amplification, it captures all DNA present in the sample. Moreover, differences in reference databases and taxonomic assignment algorithms may influence classification accuracy. These findings highlight the added value of ONT-based metagenomics for capturing a more comprehensive landscape of the vaginal microbial ecosystem.

It is important to note that core microbiota was assessed using a 90% prevalence, a conventional threshold imposed to accurately detect all those microorganisms that systematically colonise the vaginal tract of ewes, regardless of their breed and

management system, and are therefore necessary for their proper functioning. Therefore, taxa absent from each method's core list may still have been detected in the dataset, but did not meet the prevalence r*equi*red to be considered part of the core. This was the case for several taxa that were not part of the core microbiota in 16S rRNA or ONT analyses but were among the most abundant genera in **Study 3**, which did not apply a prevalence filter (e.g. *Escherichia*, *Actinobacillus*, and *Streptobacillus)*. These findings illustrate that core microbiota defined by prevalence capture strictly the most consistently shared taxa across individuals.

Staphylococcus was highly abundant when using metagenome data (**Studies 2** and **3**), but not 16S rRNA data. These technical differences may reflect both biological and technical factors. ONT sequencing samples faced high host contamination (The Human Microbiome Project Consortium, 2012), yielding only 2-5% microbial reads, of which 64% remained unclassified. In low-biomass vaginal samples, this limited sequencing depth may allow a few background taxa, such as *Staphylococcus*, to dominate the output. Additionally, microbial recovery is influenced by DNA extraction and preservation methods. Effective extraction approaches —incorporating mechanical and enzymatic lysis and, where feasible, host DNA depletion— play a critical role in maximising microbial DNA yield. Recent technological advances, particularly the adaptive sampling implemented in **Study 2**, have further enhanced microbial DNA enrichment by dynamically rejecting host DNA during sequencing on the GridION platform (Loose et al., 2016). Adaptive sampling more than doubles the raw sequencing output compared to **Study 3**, which utilised MinION without this capability (average yield: 10,644.92 vs. 4,762 reads). Despite ongoing challenges posed by host contamination, continuous refinement of selective sequencing techniques and advanced flow cell chemistries (e.g., R10.4.1), it is expected that ONT significantly improves accuracy and facilitates deeper characterisation of complex microbial communities, particularly in host-rich environments such as the ewe's reproductive tract. Other recent laboratory advances, including ultra-high molecular weight extraction using Nanobind disks (Ay¸se et al., 2025) and host DNA depletion via functionalised magnetic beads (Yuan et al., 2025), also represent promising strategies to enhance microbial DNA recovery in challenging samples.

Interestingly, several of the consistently identified genera, such as *Histophilus* and *Mycoplasma,* are recognised inhabitants of the ruminant respiratory tract (Corbeil, 2007; Nicholas et al., 2008), while *Bacteroides* and *Fusobacterium* typically dominate the rumen or hindgut (Tapio et al., 2017; Schwarz et al., 2023). Although our genus-level resolution does not allow for precise species or strain identification, the presence of overlapping genera across different mucosal environments suggests that host immunity, environmental stressors, and management practices may influence their

distribution (Amabebe and Anumba, 2020; Rojas et al., 2022). This observation raises the possibility that certain microbial taxa may transit between, or be simultaneously shaped by, multiple mucosal environments. Recent studies have highlighted the existence of functional mucosal axes, such as the gut–testis and gut–placenta axes through which microbial metabolites and immune signaling mediate systemic effects on reproductive tissues (Baker et al., 2017; Oh et al., 2023; Wang et al., 2023a; Zhang et al., 2024). Such evidence supports the broader concept that microbial overlap across mucosal sites may reflect integrated host–microbiota interactions rather than direct colonisation alone.

These results demonstrate that the consistent detection of certain phyla and genera across methods and studies reinforces the existence of a vaginal microbial core, while the detection of method-specific or context-dependent taxa highlights the influence of technical, biological, and environmental factors. Notably, this study provides one of the first attempts to define a core vaginal microbiota in ewes based on cross-methodological evidence. Integrating multiple perspectives —prevalence-based, abundance-driven, and taxonomically-diverse— enhances our ability to characterise microbiomes more accurately and meaningfully, laying the foundation for future investigations into their functional roles and potential impact on reproductive outcomes.

6.3. Drivers of microbial diversity

Understanding the factors that shape the vaginal microbiota is essential for identifying its composition. This section explores how factors such as breed, herd, age, parity, pregnancy status, synchronisation devices, and their alternative additives, influence microbial composition, using various analytical approaches, including alpha and beta-diversity, clustering, and predictive modelling. By integrating these perspectives, we aim to unravel the complex interactions between host physiology and environmental factors that govern microbial diversity. Results revealed notable differences for breed, herd, age, parity, and PRID effect, which are discussed below.

6.3.1. Breed and herd

In general, breed and herd emerged as the most influential factors shaping the vaginal microbiota, with consistent patterns observed across alpha-diversity indices, PCA, PERMANOVA, clustering, and prediction analyses. While the breed effect would be related to genetic differences in the host, the herd effect would be due to different management (feeding, reproductive rate, climatic and geographic conditions) which reflects ultimately environmental effects. These findings were consistent across both 16S rRNA metabarcoding and ONT-based metagenomics approaches and may

reflect the combined influence of genetic and environmental factors. However, disentangling the relative contributions of breed and herd to the variation in bacterial composition remains challenging, primarily due to the confounding structure of the experimental design. Each breed in our study is represented by a single herd, with the exception of one breed, which includes two herds. As a result, breed and herd effects are largely confounded making it difficult to determine whether observed differences in microbiota profiles are driven by genetic background, environmental factors specific to each herd, or an interaction between the two. This overlap limits the ability to attribute microbial variation exclusively to host genetics or local environmental conditions

The observed patterns can also vary depending on the normalisation method applied. For example, clustering based on CLR-transformed data highlighted herd-level differences (Figure 5.1.4), while relative abundance data suggested stronger breed-level separation (Appendix – Figure S1.7). Although direct evidence linking host genetics to reproductive tract microbiota is still limited, Li et al. (2019) and Wang et al. (2023b) have reported associations between host genetics and microbiota in other body sites, such as the rumen in cattle and sheep, respectively. In our case, only two herds were sampled within the same breed, which introduces design limitations. However, the Manchega herds (MNR and MVL) differ markedly in their management systems: MNR is a non-commercial herd primarily dedicated to research, while MVL is a commercial operation optimised for production. The pattern observed in the PCA results for Manchega MNR and MVL herds (Figure 5.1.3 and Figure 5.2.3), may therefore reflect closer genetic relationships among animals of the same breed, suggesting that host genetic background could also be contributing to differences in microbiota composition.

A particularly illustrative case of these complexities can be observed in Latxa. Regarding alpha-diversity, Latxa consistently exhibited the lowest richness (Observed and Chao1 indices), potentially due to the antibiotic-containing intravaginal sponges used for oestrus synchronisation. This result is in line with findings from **Study 3**, where reduced diversity was also observed in association with the antibiotic use. Notably, despite this reduction in richness, Latxa displayed relatively higher evenness in **Study 2** (ONT), as indicated by Shannon and Inverse Simpson indices, suggesting a stable yet taxonomically limited microbial community. The discrepancy between sequencing platforms likely reflects methodological differences, as previously mentioned. While 16S rRNA metabarcoding targets a specific gene region and has therefore limited taxonomic resolution, particularly at the species level,, ONT-based metagenomic sequencing analyses the entire genomic content of the sample, allowing for more precise species-level identification and a more comprehensive representation of the

microbial community structure. Interestingly, the Latxa herd also achieved the highest pregnancy rate, which may suggest that reduced richness does not necessarily reflect a compromised reproductive environment but could instead indicate a reduction in potentially pathogenic taxa, due to the antibiotic treatment. These observations highlight the need to interpret alpha-diversity alongside microbial composition and functional potential, rather than relying on it as a standalone predictor of fertility outcomes.

6.3.2. Age and parity

Age and parity showed similar results in both **Studies 1** and **2**. They were found to be significantly different across their categories by PERMANOVA analysis. These results are evident especially in the predictive analysis of **Study 2**, where both age and parity exhibited moderate RF classification accuracy (~40%). However, given that this is a predictive classification with five categories, this accuracy is notably higher than the 20% expected if predictions were random (i.e., 100%/5 levels). Additionally, when treated as continuous variables, the model achieved a RMSE of approximately 2 years, indicating a moderate ability to distinguish younger from older animals based on microbial composition. This aligns with pairwise comparisons (**Study 2**, Appendix – Table S2.3), where older animals (greater than 5 years and more than 3 parturitions) exhibited distinct microbiota compositions, suggesting an important difference in microbiota composition between young and old animals. Given the high correlation between age and parity ($r = 0.94$), it remains challenging to disentangle their individual contributions. Moreover, parity is more susceptible to errors in farm records, particularly when integrating new animals with incomplete reproductive histories.

6.3.3. PRID effect

Regarding the PRID effect (**Study 3**), we found a significant shift in the structure of the vaginal microbiota when comparing the two sampling points, i.e. before placing the PRID, and after removing it. This effect was showed by both PCA and PERMANOVA analyses, indicating that the use of the PRID alone exerts a consistent influence on microbial composition. This observation is in line with previous studies. Laguardia-Nascimento et al. (2015) reported that increased progesterone concentrations were associated with larger microbial populations in the vaginal tract. Similarly, Kaur et al. (2020) observed in humans that elevated levels of progesterone and oestrogen at the start of the menstrual cycle coincided with higher vaginal bacterial diversity. This has been attributed to an increase in glycogen production, which serves as a primary nutrient source for vaginal microbes (Mirmonsef et al., 2011).

We also observed that the progestogen included in the synchronisation sponges significantly increased the alpha-diversity. This result was expected, since under higher progesterone concentrations, the populations of microbes in the vagina have been described to be relatively larger (Laguardia-Nascimento et al., 2015). Linking with the previously mentioned effect of alpha-diversity on pregnancy, it may suggest a beneficial effect of the progestogen treatment on the pregnancy outcome. However, Martinez-Ros et al. (2018) found that the effect of progesterone varies depending on the duration of exposure, indicating that treatment timing may critically modulate its impact on microbial communities and reproductive outcomes.

Consistently, group S2 (samples collected after PRID removal) showed a notable decrease in Gammaproteobacteria and Fusobacteria, along with a concurrent increase in Actinobacteria and Firmicutes. These trends are particularly relevant, as Serrano et al. (2020) identified species from the Gammaproteobacteria and Fusobacteria classes —such as *Actinobacillus seminis*, *Sneathia vaginalis*, and *Streptobacillus notomytis*— as potentially associated with reduced fertility in ewes. In the human context, species such as *Pasteurella multocida* and *Leptotrichia amnionii*, also belonging to these classes, have been linked to reproductive disorders (Gundi et al., 2004). These results support the hypothesis that PRID-use modulates the vaginal microbiota in a potentially beneficial fashion. Further studies are needed to clarify the mechanistic links between these microbial shifts and fertility outcomes. Additionally, no significant differences in microbial composition were detected between PRID-treatment groups overall, which may be related to the limited sample size of each treatment batch.

6.3.4. Pregnancy status

Significant differences related to pregnancy status were consistently detected in PERMANOVA in the global model, for **Studies 1**, **2**, and **3**, confirming that microbiota composition may influence to some extent reproductive success. Random Forest also showed moderate predictive accuracy for pregnancy (**Study 2**, 56.11% at the genus level) but given that this is a binary classification (pregnant vs. non-pregnant), this accuracy is only slightly above the 50% expected by chance. These results suggest a differential effect of microbiota on pregnancy, but the predictive signal might reside in more specific taxonomic groups rather than broad community shifts. Notably, this points toward the potential role of particular genera in exerting deleterious effects on reproductive outcomes, possibly through inflammation, tissue damage, failure of fertilisation, defective development of the embryo, or interference with local immune tolerance, mechanisms commonly described in pathogen-driven reproductive failures. While this does not exclude the contribution of broader community alterations,

it suggests that the effects observed may be driven by the presence or relative abundance of specific taxa with potential pathogenic properties.

6.4. Differential abundance analysis for pregnancy

Differential abundance analyses were implemented for both the data from 16S rRNA metabarcoding (**Study 1**) and ONT-based metagenomics (**Study 2**). While some microorganisms were identified exclusively by one method —such as *Acinetobacter, Escherichia-Shigella, Staphylococcus, Corynebacterium,* and *Aerococcus* in the metabarcoding dataset, or *Actinobacillus, Caviibacter, Campylobacter, Gemella, Peptostreptococcus, Pseudostreptobacillus, Sneathia, Streptobacillus, Treponema*, and *Trueperella* in the metagenomic dataset— other taxa were consistently found in both. Notably, *Bacteroides*, *Fusobacterium*, *Histophilus*, *Leptotrichia*, *Oceanivirga*, *Parvimonas*, and *Pseudomonas* emerged across both platforms, and these shared detections underscore their potential relevance for reproductive outcome in ewes. In the following section, we present an integrated discussion encompassing the taxa identified through both studies, highlighting their possible roles and implications for sheep fertility.

6.4.1. Taxa with increased abundance in non-pregnant ewes

Treponema, *Peptostreptococcus*, *Caviibacter*, *Gemella*, and *Trueperella* genera, which were found in higher proportions in non-pregnant ewes, are rarely addressed in the literature on ovine reproductive microbiota. Nevertheless, certain species, such as *Treponema pallidum*, are known to cause serious human diseases like syphilis (Radolf et al., 2016). *Peptostreptococcus* has been associated with bacterial vaginosis in humans (Puapermpoonsiri et al., 1996). *Caviibacter abscessus* is implicated in purulent vaginal discharge in cattle (Ault-Seay et al., 2023) and cervical abscesses in Guinea pigs (Bemis et al., 2016). *Gemella asaccharolytica* has been linked to a heightened likelihood of HIV acquisition in humans (McClelland et al., 2018), while *Trueperella pyogenes* is recognised as an opportunistic pathogen causing metritis and mastitis in various livestock species (Rzewuska et al., 2019). The higher abundance of these genera in non-pregnant ewes may indicate a dysbiotic state or an environment less conducive to successful pregnancy, potentially contributing to poorer reproductive outcomes. Whether reproductive failure is due to fertilization impairment or failed embryo development should be studied in future work.

Streptobacillus, *Leptotrichia*, *Oceanivirga*, *Sneathia*, and *Treponema* were also more abundant genera in non-pregnant ewes. Although these taxa are not considered part of the core microbiota, their substantial prevalence, 38–88%, (Appendix – Figure S1.9

and Figure S2.8), implies that they may play transient or condition-specific roles within the vaginal microbiome.

Focusing on *Streptobacillus* genus, which exceeded 80% prevalence, Swartz et al. (2014), identified it among the most frequent genera in the ovine vaginal microbiota, indicating that it is both common and relatively stable. Despite this prevalence, no clear evidence currently links *Streptobacillus* to positive reproductive outcomes in ewes. However, some species of this genus may negatively affect pregnancy; *Streptobacillus notomytis*, for instance, was reported to be 19 times more abundant (adjusted p = 0.05) in non-pregnant than in pregnant ewes (Serrano et al., 2020). In our investigation (**Study 2**), numerous genes were substantially more prevalent in non-pregnant ewes, some of which were primarily derived from *Streptobacillus* (Figure 5.2.5). These genes are crucial to carbohydrate metabolism, transport, and defence mechanisms. For example, COG0800 is vital to the Entner-Doudoroff pathway of glucose metabolism (Figure 5.2.6), thus potentially influencing energy balance and reproductive health. COG3775, which relates to the phosphotransferase system, specifically its galactitol-specific IIC component (Figure 5.2.6), may suggest an adaptation favouring environments where galactitol is a key nutrient (López et al., 2019). COG0732, involved in the restriction endonuclease S subunit, points to a defence-oriented microbial community that may foster a hostile environment unfavourable to pregnancy, aligning with Wang et al. (2022) and Merk (2006), who emphasised the importance of such genes in microbial adaptation and competition across various prokaryotic and human contexts. Additionally, the gene K01154, encoding the Type I restriction enzyme S subunit (Figure 5.2.6), is notable for its role in microbial defence mechanisms, especially restricting foreign DNA. Freese et al. (2017) identified comparable systems in *Phaeobacter gallaeciensis* from the Pacific Ocean, suggesting that the presence of K01154 could create an antagonistic environment potentially affecting fertility.

Leptotrichia was also found to be detrimental to pregnancy in ewes by Koester et al. (2021). Although it is frequently observed in the female genital tract (Eribe and Olsen, 2017; Zhang et al., 2022), certain species of *Leptotrichia* in humans have been linked to spontaneous abortion and chorioamnionitis (Shukla et al., 2002; Boennelycke et al., 2007; Fontanals et al., 2018). *Oceanivirga* and *Sneathia* were identified in sheep by Serrano et al. (2020). Pathogenic species such as *Sneathia sanguinegens* have been strongly associated with reproductive disorders in humans (Eisenberg et al., 2021), implying that these genera could likewise impact fertility outcomes in ewes.

Finally, the genera *Histophilus*, *Fusobacterium*, *Bacteroides*, *Campylobacter*, *Actinobacillus*, and *Parvimonas* were markedly more abundant in non-pregnant ewes across various herds (**Study 1** – Figure 5.1.5; **Study 2** – Figure 5.2.4), despite being part

of the core microbiota (Figure 5.1.1 and Figure 5.2.1). These findings are in line with previous research by Kirkbride (1993), Serrano et al. (2020), Koester et al. (2021), and Yaeger et al. (2021). They suggest that an overrepresentation of certain core taxa in non-pregnant ewes could be due to a dysbiosis that would produce the reproductive failures observed. While some of these genera might contain both pathogenic and beneficial species, excessive levels may prove harmful. Environmental and genetic factors, including hormonal fluctuations and immune responses, likely contribute to this imbalance, ultimately disturbing the microbial ecosystem necessary for a successful pregnancy (Quereda et al., 2020; Serrano et al., 2020; Koester et al., 2021; Ault-Seay et al., 2023). Notable species include *Histophilus somni* (also reported in **Study 1**, Figure 5.2.4), *Campylobacter fetus, Campylobacter jejuni, Campylobacter coli*, *Actinobacillus seminis*, and *Parvimonas micra*, which have been associated with abortion, vaginitis, polyarthritis/tenosynovitis, and epididymitis in sheep (Foster et al., 1999; De La Puente-Redondo et al., 2000; Gonzalez Moreno et al., 2016; Serrano et al., 2020; Yaeger et al., 2021; Arif, 2022). Although typically commensal, certain *Bacteroides* strains can adopt an opportunistic behaviour and have been linked to gynaecological infections in humans (Wexler, 2007). In particular, *Bacteroides heparinolyticus* and *Bacteroides pyogenes*, along with *Fusobacterium necrophorum*, have been implicated in metritis and clinical endometritis in dairy cattle (Bicalho et al., 2012; Jeon et al., 2016; Gibiino et al., 2018; Çömlekcioğlu et al., 2024), underlining their capacity to contribute to reproductive impairment.

These genera were additionally linked to multiple genes significantly more prevalent in non-pregnant ewes. For instance, COG0280, predominantly found in *Fusobacterium* and *Bacteroides*, is connected to acetate kinase, suggesting that these taxa may utilise acetate as a primary energy source, favouring alterations in microbial metabolism (Mariño-Ramírez et al., 2004). ENOG410XNNV, mainly derived from *Bacteroides* (Bacteroidota), facilitates ion transport, suggesting that these microbes contribute to ionic balance that may influence the reproductive environment (Qi et al., 2019). COG3250, abundant in *Bacteroides*, underpins carbohydrate metabolism, implying a reliance on complex carbohydrate breakdown consistent with observations by Li et al. (2011) and Wu et al. (2017) in microbial communities related to *Microcystis aeruginosa* and *Bifidobacterium pseudocatenulatum*, respectively. The presence of COG1207, a gene related to membrane transport, was primarily found in *Histophilus* and *Fusobacterium*, indicating its possible involvement in nutrient acquisition or environmental adaptation. COG1472, contained in *Bacteroides* and *Histophilus*, is crucial for complex carbohydrate catabolism (Lee et al., 2015), whereas COG0364 and COP0363 (core subunits of bacterial RNA), likewise detected in *Histophilus* and *Bacteroides*, are essential components of the enzyme complex responsible for gene expression. Finally, COG0751, associated with *Fusobacterium*, is

central to protein synthesis (Pragya et al., 2017). Concerning KO, K01207 in *Histophilus*, *Bacteroides*, and *Fusobacterium,* encodes β-N-acetylhexosaminidase, underscoring its role in carbohydrate use and glycosaminoglycan degradation. Similar gene functions have been reported in the gut microbiota of laying hens, marine ecosystems, and probiotic kefir (Borrelli et al., 2017; Wright et al., 2019; Zeng et al., 2022), suggesting that these metabolic pathways are widespread and may likewise influence the ovine reproductive tract.

6.4.2. Taxa with increased abundance in pregnant ewes

Acinetobacter, *Brevundimonas*, and *Pseudomonas* displayed notably higher abundance in pregnant ewes (Figure 5.1.5), suggesting a potential association with a reproductive microbiota compatible with successful pregnancy. Supporting evidence in females is limited, but *Pseudomonas* has been detected in the vaginal microbiota of women across reproductive states and, in some contexts, associated with eubiotic profiles (Pyles et al., 2021). Nevertheless, our ONT data showed two-way associations (Figure 5.2.4), implying that *Pseudomonas* may exhibit context-dependent impacts on reproductive success. For *Acinetobacter* and *Brevundimonas*, positive associations have been reported in males, where their presence correlates with favourable semen parameters (Garcia-Segura et al., 2022; Koort et al., 2023). While these findings derive from a different reproductive context, they may indicate a broader role for these genera in reproductive health.

Complementing these observations, our GWAS also identified host genetic associations involving *Fluviicola, Lactobacillus, Glutamicibacter,* and *Bifidobacterium breve*. These taxa were not differentially abundant in our dataset but previously reported as enriched in pregnant females (Table 5.4.1). Their potential relevance is considered in subsequent sections.

6.4.3. Methods and platform for differential abundance analysis

As previously mentioned, both similarities and differences emerged when comparing 16S rRNA metabarcoding and ONT-based metagenomic methodologies. Metabarcoding targets a conserved gene to identify bacteria and archaea but provides limited resolution at the species level (Zhang et al., 2023). In contrast, ONT-based metagenomic sequencing encompasses a broader scope by detecting other microorganisms in addition to bacterial taxa, and the longer reads facilitate functional profiling (COG and KO), offering richer insights into microbial capabilities (Magi et al., 2018). Notably, this functional dimension allows not only taxonomic characterisation but also the estimation of other measures, such as the diversity based on microbial gene functions, providing a complementary perspective on the ecological potential

of the microbial community (Mellor et al., 2008; Sargison, 2008; Radostits and Done, 2010; Nandi et al., 2011; Uzal et al., 2014; Finn, 2024). Despite these advantages, ONT is characterised by higher sequencing error rates (although it has decreased to < 1% with the last advances in flow-cells R10.4.1 and chemical Kit 14) and is more prone to host DNA contamination, especially challenging in vaginal samples with substantial host DNA (Rup, 2012; Magi et al., 2018). This complexity, alongside potential underrepresentation of vaginal microbes in reference databases, complicates species-level classification, necessitating thorough filtering and advanced bioinformatics pipelines. However, both beta-diversity and differential abundance analyses consistently identified similar taxa associated with pregnancy, thus reinforcing the validity of our findings across multiple sequencing platforms. Integrating 16S rRNA metabarcoding and ONT-based metagenomic techniques enhances our overall understanding of microbial communities, capitalising on the higher accuracy of the 16S rRNA metabarcoding and the broader scope provided by metagenomics.

As stated in previous sections, determining the appropriate method for data normalisation is a challenging yet crucial step in microbiome data. Normalisation aims to correct for technical biases such as differences in sequencing depth, library size, and compositional constraints inherent to microbial datasets, all of which significantly influence downstream results (Gloor et al., 2017). Several widely adopted approaches, notably DESeq2 and Limma, implement distinct strategies for normalisation. DESeq2 employs size-factor normalisation combined with the 'poscount' approach. This method modifies the geometric mean calculation by taking the n-th root of the product of non-zero counts, effectively addressing the challenge of zero-inflation commonly encountered in microbiota datasets (Love et al., 2014). Additionally, DESeq2's negative binomial model, inherently accommodates data overdispersion — typical in microbial community analyses— thereby facilitating robust and reliable statistical inference in differential abundance testing (Love et al., 2014). On the other hand, Limma offers considerable flexibility, particularly through alternative transformations such as CLR, explicitly addressing the compositional nature of microbiota data. Incorporating CLR transformation within Limma analysis represents a significant strength, as it directly manages compositional constraints and relative abundances among microbial taxa, allowing for robust statistical inference (Ritchie et al., 2015). Limma's linear modelling approach assumes normally distributed residuals, which supports valid statistical inference on fold changes while also providing computational efficiency (Ritchie et al., 2015). Throughout this thesis, DESeq2 was prioritised in **Studies 1** and **2** primarily because of its robust management of low-count and highly dispersed data, its well-established validation within microbiome research, and its tailored statistical assumptions. Nevertheless, the flexibility, computational efficiency, and compositional sensitivity offered by Limma (used in **Study 3**) were also

highly valued, highlighting its complementary strengths within distinct analytical contexts. Limma with CLR transformation was used in **Study 3** due to its suitability for modelling compositional data and adjusting for treatment effects in a structured, balanced design. In contrast, DESeq2 was preferred in **Studies 1** and **2** given its robustness to overdispersion, zero inflation, and greater inter-sample variability. Method choice was thus aligned with the statistical characteristics and analytical focus of each dataset.

6.5. Microbiota genome-wide association studies

Study 4 aimed to investigate the link between host genetics and vaginal microbiota composition in sheep. Our results evidenced genomic regions significantly involved in modulating the abundance of specific microbial taxa, several of which are known to influence fertility.

Recent research has shown that host genetics can influence microbial communities across various animal species (Wright et al., 2019; Mehta et al., 2020; Sanglard et al., 2020). However, to date, this relationship has not been explored in sheep. In humans, Mehta et al. (2020) conducted one of the limited MGWAS addressing the vaginal microbiome, identifying host genetic variants —mostly linked to immune functions— related to bacterial vaginosis and variations in *Lactobacillus* predominance. Similarly, Wright et al. (2019) reported moderate heritability (~35%) for *Lactobacillus crispatus* abundance in twin studies, suggesting that host genetics contribute meaningfully to shaping vaginal microbial communities. In livestock species, Sanglard et al. (2020) found several vaginal bacterial genera in pigs with relatively high heritability (40–60%), influenced by specific genomic regions associated with immune functions. Although a similar MGWAS has not yet been performed in sheep reproductive microbiota, comparative breed studies (e.g., Merino vs. Suffolk) indicate that genetic factors likely influence microbiota composition. This aligns with findings from other ruminants, such as cattle, where heritable microbiota taxa and genetic loci related to immune functions, metabolism, and mucosal health have been documented (Roehe et al., 2016; Mani et al., 2022). Altogether, this evidence supports the concept that host genetics substantially shapes reproductive microbiota, highlighting the value of using MGWAS to identify genetic variants linked to vaginal microbiota composition in sheep that can improve the reproductive outcome. While numerous SNPs were significantly associated with different microbial taxa (Figure 5.4.1 and Figure 5.4.2), our discussion focuses specifically on those taxa previously implicated in fertility-related contexts in sheep and other mammals (Appendix – Figure S4.1).

Genetic associations at the phylum level appeared less robust compared to lower ranks, as expected, reflecting greater ecological variability within these broad

taxonomic groups. For instance, Fusobacteriota —previously associated with non-pregnant ewes in **Studies 1** and **2**, as well as by Koester et al. (2021) and Brulin et al. (2023)— was linked to SNPs located in intronic or intergenic regions, which can indicate regulatory roles or be in linkage disequilibrium with causal variants.

At the genus and species levels, 1,875 SNPs were significantly associated with 25 microbial taxa abundance, suggesting that more precise host-microbe interactions occur at these detailed taxonomic resolutions. Certain genes were relevant due to predicted functional impacts based on reference allele changes, as determined by the VEP analysis (excluding synonymous variants, Figure 5.4.7). These gene impacts are addressed in greater detail in subsequent sections.

6.5.1. Host-associated taxa linked to fertility failure

Several genetic variants were associated with the abundance variability of microbial taxa generally associated with unfavourable reproductive outcomes. At the genus level, *Actinobacillus*, *Bergeyella*, and *Coxiella* were associated with six host genes, while five species —*Staphylococcus aureus*, *Fusobacterium nucleatum*, *Prevotella nigrescens*, and *Gardnerella vaginalis*— were linked to fourteen.

Actinobacillus is a Gram-negative bacillus previously reported as an opportunistic pathogen in association with reproductive disorders in sheep (Serrano et al., 2020). In the current study, three host genes —*UBR5*, *HLCS*, and *BLTP3A*— were identified as significantly correlated with its abundance. The SNP oar3_OAR9_67669602 (A/C) located at the chromosome 9, is a synonymous variant of the exon 45 of the *UBR5* gene. The allele C showed a frequency of 12% and produces and increase of Actinobacillus abundance. *UBR5* encodes an E3 ubiquitin ligase involved in cell cycle control, metabolic regulation, and antiviral defence (Wang et al., 2024b), with demonstrated immunomodulatory functions in murine models (Yang et al., 2024). Genetic variation in this gene could disturb immune homeostasis, potentially impairing the host's capacity to manage opportunistic bacteria. *The SNP oar3_OAR20_7378588 (G/A) located at the chromosome 20 of the ovis aries genome, is a synonymous variant at the exon 8 of the BLTP3A gene, which is highly expressed in lymphoid and reproductive tissues* (GeneCards, 2024a), participates in antigen processing, immune signalling, and reproductive development (Payne et al., 2018; Wen et al., 2020). It also regulates lipid trafficking and lysosomal integrity, potentially influencing nutrient availability and bacterial adhesion within the vaginal microenvironment (Hanna et al., 2024). Finally, oar3_OAR1_245046296 (A/G) variant at chromosome 1, is a splice region variant, splice polypyrimidine tract variant, intron variant, located at the intron 7 of the *HLCS* gene. The alternative allele G of this SNP has a frequency of 4.5% and produces an increase of Actinobacillus abundance. *HLCS*, essential for biotin metabolism,

modulates cytokine pathways through biotinylating of heat-shock proteins and metabolic enzymes (Xue, 2013). As biotin is a critical cofactor in bacterial growth, variants in *HLCS* may affect biotin availability and could promote *Actinobacillus* proliferation in a competitive microbial setting.

Bergeyella, another Gram-negative genus commonly found on mucosal surfaces, was previously reported as less prevalent in non-pregnant ewes (Koester et al., 2021). In our dataset, we found one SNP, oar3_OAR13_23093023 (A/G), associated to this genus. This SNP is a synonymous variant located at the exon 17 of the *ARHGAP21* gene. It showed a frequency of 7% of the G allele, and an increasing effect in the Bergeyella abundance. This gene encodes a Rho GTPase-activating protein expressed in tissues such as the placenta, brain, and heart (Wang et al., 2023c), and is essential for maintaining epithelial integrity and modulating immune responses through epithelial–mesenchymal transition (Barcellos et al., 2013). Variants that impair *ARHGAP21* function may compromise epithelial cohesion or pathogen internalisation, increasing susceptibility to colonisation by opportunistic bacteria.

Coxiella, particularly *Coxiella burnetii*, an intracellular pathogen responsible for Q fever, is known to contribute to abortion events in ruminants (Ruiz-Fons et al., 2014; Smith et al., 2023). The SNP oar3_OAR3_208609769 (G/A), located on chromosome 3 within the *MOV10L1* gene, is associated with increased C. burnetii abundance (A allele frequency = 4.2%). This variant causes a high-impact stop-gained mutation in exon 21, generating a premature stop codon that results in a truncated polypeptide. *MOV10L1*, which encodes an RNA helicase essential for gametogenesis and transposon silencing during meiosis (NCBI, 2025a). Dysfunctional alleles could elevate retrotransposon activity in placental tissue, potentially triggering localised inflammation and diminishing fertility. As *Coxiella* frequently resides in placental tissue, a compromised host genetics profile may facilitate its persistence, especially in non-pregnant animals due to the inability to clear the infection.

Staphylococcus aureus is a Gram-positive opportunistic microbe whose increased presence in the vaginal environment correlates with inflammation and reduced reproductive efficiency. Four genes —*ZBTB17, PAPPA2, MVD,* and *MLLT6*— were associated with its differential abundance. The SNP oar3_OAR12_45402829 (G/A) located in chromosome 2 is a missense variant with moderate effect in the exon 2 of the *ZBTB17* gene. The A allele presented a frequency of 9.2% and was associated with an increase of the *S. aureus* abundance, showing a predicted deleterious effect by producing an amino acid shift from Glutamine to Lysine. *ZBTB17* encodes a zinc-finger transcription factor critical for early development and immune cell maturation (Cheng et al., 2021; Piskor et al., 2022). Impaired expression of this gene may compromise adaptive immune function. The SNP oar3_OAR12_51413790 (A/G) is a synonymous

variant located at chromosome 12 in the exon 11 of the *PAPPA2* gene. The mutated allele G showed a frequency of 6.6% and was associated with an increase of *S. aureus* abundance. *PAPPA2*, expressed predominantly in placental and reproductive tissues, enhances IGF bioavailability, and its polymorphisms may impair endometrial support during gestation (Andrew et al., 2018). The SNP oar3_OAR14_12203359 (A/G), located on chromosome 14, is a missense variant in exon 8 of the *MVD* gene. The alternative allele G showed a frequency of 7.2% and results in an amino acid substitution from arginine to serine (R/S), with a predicted moderate impact on protein function. This SNP is associated with an increase in the abundance of *S. aureus*. *MVD* is involved in sterol biosynthesis and immune cell function (NCBI, 2025b), and deficiencies in this pathway could weaken innate immune defence. The SNP oar3_OAR11_23909163 (A/G), located on chromosome 11, is a missense variant in the *MLLT6* gene, causing an amino acid substitution from aspartic acid to glycine, predicted to be deleterious. The alternative allele G has a frequency of 5.3% and is associated with an increase in the *S. aureus* abundance. *MLLT6*, linked to chromatin remodelling, regulates *PD-L1* expression and may suppress immune responses in the reproductive tract, favouring pathogenic bacterial persistence (Sreevalsan et al., 2020).

Fusobacterium nucleatum, a Gram-negative anaerobic species, has been implicated in adverse reproductive conditions such as preterm labour, chorioamnionitis, and stillbirth (Vander Haar et al., 2018). In this study, its abundance was associated with variants at the *SORL1, RBPJ,* and *LRP6* genes. The SNP oar3_OAR15_28610319 (A/G), located on chromosome 15, is a splice region and intronic variant in the *SORL1* gene. The alternative allele G has a frequency of 2.8% and is associated with an increase in the abundance of the taxon. *SORL1* is involved in endocytic recycling and innate immunity via CD14-LPS signalling (Talbot et al., 2019; Ovesen et al., 2025). A reduced-function allele may delay immune recognition. The SNP oar3_OAR6_41617009 (G/A), located on chromosome 6, is a synonymous variant in exon 4 of the *RBPJ* gene. The alternative allele A has a frequency of 7.0% and is associated with an increase in the abundance of the taxon. *RBPJ* gene is crucial for cytokine regulation and uterine tissue repair (Strug et al., 2018; Huang et al., 2020). Finally, The SNP oar3_OAR3_189334388 (G/A), located on chromosome 3, is a synonymous variant in exon 13 of the *LRP6* gene. The alternative allele A has a frequency of 10.4% and is associated with an increase in taxon abundance. *LRP6* regulates Wnt/β-catenin signalling, contributing to epithelial barrier stability. Disruption in this pathway may promote microbial overgrowth (Auton et al., 2015).

Prevotella nigrescens, typically linked to bacterial vaginosis, was associated with *SDHD, PITX3,* and *HADHB* genes. The SNP oar3_OAR15_19568260 (G/A), located on chromosome 15, is a stop-gained variant in exon 2 of the *SDHD* gene. The alternative

allele A has a frequency of 7.0% and is associated with an increase in the abundance of *P. nigrescens*. This variant introduces a premature stop codon, likely resulting in a truncated and nonfunctional SDHD protein. Mutations on this gene can lead to succinate accumulation and proinflammatory responses (Nazar et al., 2019), which might serve as a favourable environment for *Prevotella* species. The SNP oar3_OAR22_19175882 (A/G), located on chromosome 22, is a synonymous variant in exon 5 of the *PITX3* gene. The alternative allele G has a frequency of 6.4% and is associated with an increase in the abundance of this taxon. *PITX3* is mainly known for its role in lens formation and is expressed in the genitourinary system during development, potentially influencing glandular structures that secrete mucus and antimicrobials. Variants impairing these glands may alter mucus composition, favouring *Prevotella* proliferation. The SNP oar3_OAR3_31806623 (A/C), located on chromosome 3, is a splice polypyrimidine tract and intron variant in the *HADHB* gene. The alternative allele C is associated with an increase in the abundance of this taxon. *HADHB*, critical for fatty acid metabolism, supports macrophage function and anti-inflammatory pathways (Margulies et al., 2005; Wang et al., 2022). Alterations in this gene may favour inflammatory conditions conductive to anaerobic colonisation.

Gardnerella vaginalis, a Gram-variable organism often linked to bacterial vaginosis and reproductive complications, was associated with the SNP oar3_OAR11_45030002 (C/A). This missense variant in exon 13 of the PSMD11 gene results in an amino acid substitution from lysine to methionine. The A allele has a frequency of 4.3% and is associated with increased abundance of the corresponding taxon. The variant is predicted to be deleterious, suggesting a potential functional impact on proteasome activity. *PSMD11*, which encodes a regulatory subunit of the proteasome. Reduced *PSMD11* expression may compromise antigen presentation and T-cell function, facilitating persistence of potentially pathogenic taxa (Huang et al., 2024).

6.5.2. Host-associated taxa linked to fertility success

In contrast, we identified the genera *Glutamicibacter*, *Fluviicola*, and *Lactobacillus*, along with the species *Bifidobacterium breve*, that have previously been linked to higher prevalence in pregnant ewes, as being associated with twelve host genes.

Koester et al. (2021) reported an OTU belonging to *Glutamicibacter* that appeared more frequently in pregnant animals. In our analysis, polymorphisms at seven host genes —*MYMK, LRP2, KAT6B, EXOSC10, ENSOARG00020003095, CLK3,* and *ASCC2*— were associated with overabundance of *Glutamicibacter*. For instance, the SNP oar3_OAR3_2382704 (A/G), a synonymous variant in exon 5 of the *MYMK* gene, has an alternative G allele with a frequency of 5.9% and was linked to higher *Glutamicibacter* abundance. *MYMK* plays a key role in muscle cell fusion during embryogenesis

(GeneCards, 2024b), and its early expression suggests potential involvement in uterine or placental development. Genetic variations may thus contribute to maternal conditions favourable to commensal colonisation. The SNP oar3_OAR2_128555884 (A/G) is a missense variant in exon 39 of the *LRP2* gene. The alternative allele G, present at a frequency of 4.2%, causes an amino acid substitution from phenylalanine to valine. This variant is associated with an increase in the abundance of *Glutamicibacter*. *LRP2* encodes the receptor megalin, essential for the uptake of vitamins A and D, immune molecules, and lipids (Dufour et al., 2024), as well as for ciliary function and mucosal balance through planar polarity pathways (Bunatyan et al., 2023). Its upregulation by progesterone in uterine tissue (Oh et al., 2013) involves hormonal regulation of microbiota composition. Changes in *LRP2* may therefore impact nutrient dynamics and immune interactions, enabling persistence of *Glutamicibacter*. The SNP oar3_OAR25_28788821 (G/A) is a synonymous variant located in exon 8 of *KAT6B*. The A allele (frequency 4.5%) is associated with increased abundance of the taxon. *KAT6B* encodes a histone acetyltransferase essential for transcriptional activation via histone H3 acetylation (Sun et al., 2023), contributing to early development and epithelial differentiation (National Library of Medicine, 2023). Disruption in *KAT6B* could compromise epithelial integrity and favour bacterial adhesion (Bergamasco et al., 2024). The SNP oar3_OAR12_36656119 (A/G) is a synonymous variant in exon 7 of *EXOSC10*. The G allele (3.4%) is linked to increased abundance of *Glutamicibacter*. *EXOSC10*, part of the nuclear RNA exosome, supports RNA turnover, genomic maintenance, and oocyte development (Jamin et al., 2017; Zhou et al., 2024), and is recognised as an autoantigen in some immune disorders (NCBI, 2025c). Dysfunction of *EXOSC10* gene may lead to RNA accumulation, promoting immune dysregulation and microbiome shifts during pregnancy. The SNP oar3_OAR18_32765946 (C/G) is a 3′ UTR variant in *CLK3*. This gene, a dual-function kinase expressed in mature spermatozoa (Menegay et al., 1999), may influence the uterovaginal environment through metabolic and immune pathways. The SNP oar3_OAR17_59545379 (C/A) is a missense variant in exon 17 of ASCC2, leading to an amino acid change from glutamine to histidine. The A allele (4.0%) is predicted to have moderate impact and is associated with increased abundance of the taxon. *ASCC2*, which connects genome surveillance with stress signalling (Jia et al., 2020), could regulate inflammation and thereby shape microbiota composition via allelic variation.

Regarding *Fluviicola*, prior work by Serrano et al. (2020) found it more abundant in pregnant ewes. In our dataset, we identified a genetic association involving the SNP oar3_OAR18_62811847 (A/G), a missense variant in exon 3 of *WDR20* gene, resulting in a histidine-to-aspartic acid substitution. The G allele (frequency of 11.5%) is linked to increased *Fluviicola* abundance and is annotated with moderate impact and deleterious potential. *WDR20* encodes an adaptor protein involved in regulating

deubiquitinating enzymes. Since ubiquitination plays a central role in innate receptor signalling (Jiang et al., 2024), allelic variants of *WDR20* might support a more stable mucosal environment conducive to colonisation by beneficial taxa such as *Fluviicola*.

Lactobacillus, a key genus in maintaining vaginal health in women, is known for reducing pH and providing defence against pathogens (Kwon and Lee, 2022). In our study, it was linked to a missense variant, oar3_OAR12_60229149 (A/G), located in exon 54 of *HMCN1*, which causes an aspartic acid-to-glycine substitution via a gAc to gGc codon change. The G allele (frequency of 19.8%) was associated with increased abundance of *Lactobacillus*. *HMCN1* encodes a structural glycoprotein found in basement membranes and cell–matrix adhesion zones (Lin et al., 2020). In mammals, this gene contributes to epithelial cohesion in the reproductive tract. High estrogen levels during gestation enhance glycogen storage in the vaginal epithelium, favouring lactic acid–producing *Lactobacillus*, which may maintain an acidic environment that suppresses pathogen growth.

Bifidobacterium was previously identified in the vaginal microbiome of dairy cattle by Brulin et al. (2023), who suggested a possible association with increased milk fat yield, indicating a role in productive performance. At the species level, *Bifidobacterium breve* has been detected in the human vagina (Bottacini et al., 2014), where it may contribute to maternal-neonatal microbial transmission and stability of the vaginal ecosystem through mechanisms such as pili synthesis, exopolysaccharide production, and CRISPR-related defence. In our analysis, *B. breve* abundance was associated with variants in three genes: *TRIM46*, *IFRD2*, and *EIF2D*. The SNP oar3_OAR1_98604666 (C/A), a synonymous variant in exon 2 of *TRIM46*, is classified as low impact; the A allele (frequency of 10.6%) was associated with increased *B. breve* abundance. *TRIM46* organises neuronal microtubules and modulates antiviral responses by targeting kinase 1 for degradation. Downregulation of *TRIM46* intensifies interferon signalling and suppresses viral replication (Su et al., 2022), suggesting a role in fine-tuning immune responses. Similarly, the SNP oar3_OAR19_48122704 (G/A), a missense variant in exon 11 of *IFRD2*, causes an aspartic acid-to-glutamic acid substitution. The A allele (8.3%) was associated with increased abundance of *B. breve*. *IFRD2* acts as a translational repressor by binding to ribosomes and interfering with translation RNA (tRNA) positioning (Brown et al., 2018), and it has also been implicated in cell differentiation and interferon regulation. Its paralogue *IFRD1* has been connected to neutrophil ROS production (Hector et al., 2013), supporting a broader immunomodulatory function. By downregulating inflammatory protein synthesis, *IFRD2* may protect mucosal tissues and commensal microbes. The third SNP, oar3_OAR12_3852718 (A/G), is a missense variant in exon 10 of *EIF2D*, leading to a tyrosine-to-serine substitution. The G allele (4.7%) is associated with increased *B. breve* abundance and is predicted to have

moderate impact. *EIF2D* is a non-canonical translation initiator, delivering aminoacyl-tRNA to the ribosome independently of GTP, and plays a role in regulating immune-related protein production. It is also essential for ATF4 translation during the integrated stress response, ensuring cellular adaptation and endoplasmic reticulum function under stress (Vasudevan et al., 2020). Together, *IFRD2* and *EIF2D* may modulate inflammatory tone, maintaining a balanced environment that supports *Bifidobacterium breve* colonisation enhancing reproductive success. Regulatory genes that prevent excessive immune activation —such as *TRIM46* and *IFRD2*— likely contribute to this equilibrium.

Overall, while no single microbial taxon can be definitively linked to fertility, our findings highlight associations between beneficial genera and host genes involved in immune regulation, epithelial integrity, nutrient transport, and hormonal response. These associations suggest that host genetic variation may shape the reproductive tract environment in ways that favour the persistence of commensal microbes with potentially supportive roles during pregnancy. In particular, genes related to inflammation control, epithelial cohesion, and metabolic regulation appear to contribute to a mucosal environment conducive to microbial stability. Thus, rather than the presence of specific taxa alone, it is the host's ability to maintain a balanced, non-inflammatory environment that may be key to promoting reproductive success.

6.5.3. Opportunities, limitations, and future directions

Although this approach contributes to interpreting proportional changes in microbial communities, it also underscores the inherent complexity of microbiome studies, where overall bacterial load can be influenced by numerous interacting factors. Despite the insights gained, our study is not exempt from limitations. The sample size (243 ewes) of the MGWAS, while substantial, may limit the statistical power to detect associations involving rare variants or low-abundance taxa, and the possibility of false positives remains a key concern in this type of high-dimensional analyses (Pearson and Manolio, 2008). Furthermore, the Latxa breed was exposed to the antibiotic framycetin added to the PRID, which may have affected the structure of the vaginal microbiota in such a way that cannot be completely separated from host genetic effects. The inclusion of animals from three genetically distinct breeds adds another layer of complexity, as population stratification —despite being accounted for in the MGWAS model— may influence allele frequencies and microbiome patterns. Nonetheless, this variation also provides an opportunity to identify alleles associated with beneficial microbial profiles that could be leveraged in selective breeding programs. Additionally, all samples were collected shortly after the removal of progesterone sponges and immediately prior to AI, a time point that may influence microbial dynamics, as both progesterone-

based treatments and antibiotics have been shown to modify vaginal microbiota composition (as described in **Study 3**). Still, the dataset offers valuable insights, particularly considering this is one of the first efforts to characterise the genetic architecture underlying host–microbiota interactions in the reproductive tract of sheep at a functionally relevant time. Future research should prioritise the functional validation of candidate genes, expand to include larger and more diverse cohorts, and integrate complementary omics approaches —such as meta transcriptomics and metabolomics— to enhance causal interpretation. Ultimately, combining genomic selection with targeted microbiota modulation strategies, whether through probiotics, nutritional interventions, or breeding schemes that integrate host and microbial data, could offer new avenues for improving reproductive efficiency in sheep.

This study focused on SNPs located in gene regions with predicted functional consequences. Although the observed associations are promising, they should not yet be interpreted as causal. Functional validation —whether in vivo or through molecular assays— remains essential to confirm their biological relevance. Overall, while these findings do not yet support direct selection based solely on these SNPs, they highlight a promising direction. A more robust framework would involve integrating host genomic data with microbial and phenotypic information as part of a broader strategy to enhance fertility traits in sheep.

As a general reflection on the development of this doctoral thesis, it is worth asking to what extent the research has successfully addressed the questions originally posed as its objectives

i. What are the reference microbial populations that colonize the reproductive tract of sheep?

   We have been able to characterise a core microbiota of the vaginal tract of sheep, which has shown great taxonomic diversity, and certain differences dependent on the host genome (breed) and the environment and management conditions (herd) in which ewes are reared.

ii. To what extent do variations in the abundance of these taxa influence the reproductive success of this species?

   We have shown that the variation in the abundance of certain microbial taxa influences, to a greater or lesser extent, the reproductive failure or success of ewes subjected to artificial insemination.

iii. Does the variability of the reproductive tract microbiota in sheep depend on the genetics of the host?

Our results suggest that genetic variation in certain genomic regions may influence the abundance of specific microbial taxa, some of which are associated with reproductive traits. These findings open the door to future research on the potential integration of microbiota-related markers in breeding strategies to improve fertility.

7. CONCLUSIONS

This thesis provides an integrative characterisation of the ewe's vaginal microbiota and its relationship with reproductive success, combining high-throughput sequencing technologies, functional microbial profiling, and host genomic analysis. Across four complementary studies, this research reveals that the vaginal microbiome is shaped by both host and environment related factors and that specific microbial communities and functions may influence fertility outcomes. The findings underscore the complexity of the ewe's reproductive microbiota and open new opportunities for the implementation of targeted interventions and genomic strategies aimed at improving AI success in sheep. By integrating microbial, functional, and host-genomic data, this thesis provides a multidimensional view of the vaginal ecosystem in sheep and its influence on reproductive performance. Future work should prioritise functional validation of candidate genes, microbial interventions, and longitudinal designs to assess causality. Combining selective breeding with microbiota-informed strategies holds potential for improving reproductive outcomes in the ovine sector. The main conclusions are:

1. A stable core vaginal microbiota exists in ewes, consistently identified across both 16S rRNA metabarcoding and ONT-based metagenomic approaches. Genera such as *Histophilus, Fusobacterium, Parvimonas, Aerococcus, Bacteroides,* and *Streptococcus* were prevalent in over 90% of samples across methods, highlighting their potential ecological relevance.
2. Herd and breed emerged as the primary drivers of vaginal microbial diversity in ewes, highlighting the combined influence of both environmental conditions and host genetic background; although due to the structure of the experimental design, the ability to accurately distinguish the relative contributions of each factor was limited. Another important aspect was the use of synchronisation devices (PRID), which induced consistent and substantial alterations in the microbial composition. Although with a lower impact, age and parity also contributed to shaping the microbiota, with older and multiparous ewes displaying differential microbial profiles compared to younger animals.
3. Vaginal microbiota composition showed a moderate yet consistent association with pregnancy outcomes. While the overall community structure had limited predictive value, specific taxa were differentially abundant between pregnant and non-pregnant ewes. Notably, *Fusobacterium, Histophilus, Leptotrichia, Campylobacter, Streptobacillus, Trueperella, Treponema,* and *Bacteroides*

were more abundant in non-pregnant ewes and were linked to increased representation of microbial functions (COG and KO) related to carbohydrate metabolism, defence mechanisms, and restriction-modification systems, suggesting a more competitive or inflammatory environment. Conversely, *Pseudomonas, Acinetobacter,* and *Brevundimonas* were more abundant in pregnant ewes. These findings indicate that an imbalance or overrepresentation of certain core and opportunistic taxa may contribute to a suboptimal vaginal environment for successful reproduction. Rather than supporting reproductive success, the microbiota dysbiosis in non-pregnant ewes may reflect or contribute to an environment that is less permissive to pregnancy establishment.

4. Host genomic variation can contribute to shape the composition and abundance of vaginal microbial taxa in sheep, several of which have been linked to fertility outcomes. Although most identified SNPs exert modest individual effects, their cumulative impact —particularly across immune pathways, epithelial integrity, and metabolic regulation— appears to modulate the reproductive tract microbiota. Notably, pathogenic and opportunistic taxa such as *Actinobacillus, Bergeyella, Coxiella, Staphylococcus aureus, Fusobacterium nucleatum, Prevotella nigrescens,* and *Gardnerella vaginalis* were associated with gene variants potentially impairing local immune surveillance, whereas taxa positively associated with pregnancy, such as *Glutamicibacter, Fluviicola, Lactobacillus,* and *Bifidobacterium breve*, were linked to host genes that could foster a more tolerogenic or nutrient-rich environment. These findings offer a promising avenue for incorporating microbiome information into breeding programs aimed at improving AI success. Nevertheless, further validation in larger multi-breed cohorts, functional analyses of candidate genes, and integrative "omics" approaches are warranted to confirm causal mechanisms. By refining our understanding of how host genetics and microbiota interact, future strategies —ranging from selective breeding to targeted microbial interventions— may ultimately enhance reproductive efficiency in sheep.

5. The combined use of 16S rRNA metabarcoding and ONT-based metagenomics provided complementary insights into the ewe's vaginal microbiota. While metabarcoding offered higher sensitivity for low-abundance taxa and produced a greater number of reads, it was limited to bacterial profiles and constrained by primer bias. In contrast, ONT-based metagenomics enabled broader taxonomic coverage —including other microbial groups— and allowed functional gene profiling (COG and KO) but was more affected by

host DNA contamination and lower read depth. Despite these methodological differences, both approaches consistently identified key taxa associated with core microbiota and pregnancy outcomes, reinforcing the robustness of the findings.

REFERENCES

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

Abbott, K.A. (2024). Practice of Sheep Veterinary Medicine. CRC Press.

Adnane, M., and Chapwanya, A. (2022). A review of the diversity of the genital tract microbiome and implications for fertility of cattle. Animals 12:460. doi:10.3390/ani12040460.

Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society 44:139–160. doi:10.1111/j.2517-6161.1982.tb01195.x.

Ajmone-Marsan, P., Boettcher, P.J., Colli, L., Ginja, C., Kantanen, J., and Lenstra, J.A. (2023). Genomic characterization of animal genetic resources: practical guide. FAO Animal Production and Health Guidelines No. 32. doi:10.4060/cc3079en.

Alberto, F.J., Boyer, F., Orozco-Terwengel, P., Streeter, I., Servin, B., De Villemereuil, P., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. Nat Commun 9:1–9. doi:10.1038/s41467-018-03206-y.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol 215:403–410. doi:10.1016/s0022-2836(05)80360-2.

Amabebe, E., and Anumba, D.O.C. (2020). Female gut and genital tract microbiota-induced crosstalk and differential effects of short-chain fatty acids on immune sequelae. Front Immunol 11:553047. doi:10.3389/fimmu.2020.02184.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169. doi:10.1128/mr.59.1.143-169.1995.

Amat, S., Dahlen, C.R., Swanson, K.C., Ward, A.K., Reynolds, L.P., and Caton, J.S. (2022). Bovine animal model for studying the maternal microbiome, in utero microbial colonization and their role in offspring development and fetal programming. Front Microbiol 13:937157. doi:10.3389/fmicb.2022.854453.

Andrew, M., Liao, L., Fujimoto, M., Khoury, J., Hwa, V., and Dauber, A. (2018). PAPPA2 as a therapeutic modulator of IGF-I bioavailability: in vivo and in vitro evidence. J Endocr Soc 2:646–656. doi:10.1210/js.2018-00106.

Antil, S., Abraham, J.S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., et al. (2023). DNA barcoding, an effective tool for species identification: a review. Mol Biol Rep 50:761–775. doi:10.1007/s11033-022-08015-7.

Arif, E.D. (2022). Detection of Campylobacter fetus in aborted ewes in Sulaimani province by PCR. Iraqi Journal of Veterinary Sciences 36:647–651. doi:10.33899/ijvs.2021.131225.1931.

Ault-Seay, T.B., Moorey, S.E., Mathew, D.J., Schrick, F.N., Pohler, K.G., McLean, K.J., et al. (2023). Importance of the female reproductive tract microbiome and its relationship with the uterine environment for health and productivity in cattle: a review. Frontiers in Animal Science 4:1111636. doi:10.3389/fanim.2023.1111636.

Aurich, J.E. (2012). Artificial insemination in horses—more than a century of practice and research. J Equine Vet Sci 32:458–463. doi:10.1016/j.jevs.2012.06.011.

Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A., et al. (2015). A global reference for human genetic variation. Nature 526:68–74. doi:10.1038/nature15393.

Ay¸se, A., Ova, Ö., Joffre, E., Shafagh, R.Z., Assunção, M.F.G., Sidorov, R.Y., et al. (2025). Improved isolation of ultra-high-molecular-weight genomic DNA suitable for third-generation sequencing. Microorganisms 13:534. doi:10.3390/microorganisms13030534.

Ayling, M., Clark, M.D., and Leggett, R.M. (2020). New approaches for metagenome assembly with short reads. Brief Bioinform 21:584–594. doi:10.1093/bib/bbz020.

Baker, J.M., Al-Nakkash, L., and Herbst-Kralovetz, M.M. (2017). Estrogen–gut microbiome axis: physiological and clinical implications. Maturitas 103:45–53. doi:10.1016/j.maturitas.2017.06.025.

Barba, M., Toquet, M., García-Roselló, E., Gomis, J., Quereda, J.J., González-Torres, P., et al. (2024). Description of the vaginal microbiota in nulliparous ewes during natural mating and pregnancy: preliminary signs of the male preputial microbiota modulation. Front Microbiol 14:1224910. doi:10.3389/fmicb.2023.1224910.

Barcellos, K.S.A., Bigarella, C.L., Wagner, M. V., Vieira, K.P., Lazarini, M., Langford, P.R., et al. (2013). ARHGAP21 protein, a new partner of a-tubulin involved in cell-cell adhesion formation and essential for epithelial-mesenchymal transition. J Biol Chem 288:2179–2189. doi:10.1074/jbc.m112.432716.

Baselga, A., and Gómez, C. (2019). Diversidad alfa, beta y gamma: ¿cómo medimos diferencias entre comunidades biológicas?. Nova Acta Científica Compostelana (Bioloxía) 26:39–45.

Bemis, D.A., Johnson, B.H., Bryant, M.J., Jones, R.D., McCleery, B. V., Greenacre, C.B., et al. (2016). Isolation and identification of Caviibacter abscessus from cervical abscesses in a series of pet guinea pigs (Cavia porcellus). Journal of Veterinary Diagnostic Investigation 28:763–769. doi:10.1177/1040638716665660.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57:289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59. doi:10.1038/nature07517.

Bergamasco, M.I., Ranathunga, N., Abeysekera, W., Li-Wai-Suen, C.S.N., Garnham, A.L., Willis, S.N., et al. (2024). The histone acetyltransferase KAT6B is required for hematopoietic stem cell development and function. Stem Cell Reports 19:469–485. doi:10.1016/j.stemcr.2024.02.005.

Beveridge, W.I.B. (1941). Foot-rot in sheep : a transmissible disease due to infection with Fusiformis nodosus (n. sp.): studies on its cause, epidemiology, and control. Council for Scientific and Industrial Research 1–58. doi:10.5555/19422200367.

Bhattacharya, K., Dutta, S., Sengupta, P., and Bagchi, S. (2023). Reproductive tract microbiome and therapeutics of infertility. Middle East Fertil Soc J 28:1–14. doi:10.1186/s43043-023-00136-8.

Bicalho, M.L.S., Machado, V.S., Oikonomou, G., Gilbert, R.O., and Bicalho, R.C. (2012). Association between virulence factors of Escherichia coli, Fusobacterium necrophorum, and Arcanobacterium pyogenes and uterine diseases of dairy cows. Vet Microbiol 157:125–131. doi:10.1016/j.vetmic.2011.11.034.

Boennelycke, M., Christensen, J.J., Arpi, M., and Krause, S. (2007). Leptotrichia amnionii found in septic abortion in Denmark. Scand J Infect Dis 39:382–383. doi:10.1080/00365540601053022.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:1–17. doi:10.1186/s40168-018-0470-z.

Borrelli, L., Coretti, L., Dipineto, L., Bovera, F., Menna, F., Chiariotti, L., et al. (2017). Insect-based diet, a promising nutritional source, modulates gut microbiota composition and SCFAs production in laying hens. Sci Rep 7:1–11. doi:10.1038/s41598-017-16560-6.

Bottacini, F., O'Connell Motherway, M., Kuczynski, J., O'Connell, K.J., Serafini, F., Duranti, S., et al. (2014). Comparative genomics of the Bifidobacterium breve taxon. BMC Genomics 15:1–19. doi:10.1186/1471-2164-15-170.

Bravo-Egana, V., Sanders, H., and Chitnis, N. (2021). New challenges, new opportunities: next generation sequencing and its place in the advancement of HLA typing. Hum Immunol 82:478–487. doi:10.1016/j.humimm.2021.01.010.

Breiman, L. (2001). Random forests. Mach Learn 45:5–32. doi:10.1023/a:1010933404324.

Brito, L.F., Clarke, S.M., McEwan, J.C., Miller, S.P., Pickering, N.K., Bain, W.E., et al. (2017). Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. BMC Genet 18:1–17. doi:10.1186/s12863-017-0476-8.

Brown, A., Baird, M.R., Yip, M.C.J., Murray, J., and Shao, S. (2018). Structures of translationally inactive mammalian ribosomes. Elife 7. doi:10.7554/elife.40486.

Brulin, L., Ducrocq, S., Even, G., Sanchez, M., Martel, S., Merlin, S., et al. (2023). Characterization of bovine vaginal microbiota and its relationship with host fertility, health, and production. bioRxiv 2023.12.13.571514. doi:10.1101/2023.12.13.571514.

Brulin, L., Ducrocq, S., Even, G., Sanchez, M.P., Martel, S., Merlin, S., et al. (2024). Characterization of bovine vaginal microbiota using 16S rRNA sequencing: associations with host fertility, longevity, health, and production. Sci Rep 14:1–18. doi:10.1038/s41598-024-69715-7.

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

Bunatyan, L., Margineanu, A., Boutin, C., Montcouquiol, M., Bachmann, S., Ilsø Christensen, E., et al. (2023). LRP2 contributes to planar cell polarity-dependent coordination of motile cilia function. Cell Tissue Res 392:535–551. doi:10.1007/s00441-023-03757-7.

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

Calvo, J.H., Serrano, M., Tortereau, F., Sarto, P., Iguacel, L.P., Jiménez, M.A., et al. (2020). Development of a SNP parentage assignment panel in some North-Eastern Spanish meat sheep breeds. Spanish Journal of Agricultural Research 18:e0406. doi:10.5424/sjar/2020184-16805.

Cassas, M.S., Jonas, L.C., Anderson, C.J., Schmitz-Esser, S., and Youngs, C.R. (2024). Temporal changes in ewe vaginal microbiota throughout gestation. Front Microbiol 15:1359678. doi:10.3389/fmicb.2024.1359678.

Cassol, I., Ibañez, M., and Bustamante, J.P. (2025). Key features and guidelines for the application of microbial alpha diversity metrics. Sci Rep 15:1–13. doi:10.1038/s41598-024-77864-y.

Catlin, B.W. (1992). Gardnerella vaginalis: characteristics, clinical considerations, and controversies. Clin Microbiol Rev 5:213–237. doi:10.1128/cmr.5.3.213.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. Scandinavian Journal of Statistics 11:265–270.

Cheng, Z.Y., He, T.T., Gao, X.M., Zhao, Y., and Wang, J. (2021). ZBTB transcription factors: key regulators of the development, differentiation and effector function of T cells. Front Immunol 12:713294. doi:10.3389/fimmu.2021.713294.

Ciani, E., Lasagna, E., D'Andrea, M., Alloggio, I., Marroni, F., Ceccobelli, S., et al. (2015). Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. Genetics Selection Evolution 47:1–12. doi:10.1186/s12711-015-0139-z.

Ciani, E., Mastrangelo, S., Da Silva, A., Marroni, F., Ferenčaković, M., Ajmone-Marsan, P., et al. (2020). On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. Genetics Selection Evolution 52:1–14. doi:10.1186/s12711-020-00545-7.

Çömlekcioğlu, U., Jezierska, S., Opsomer, G., and Pascottini, O.B. (2024). Uterine microbial ecology and disease in cattle: a review. Theriogenology 213:66–78. doi:10.1016/j.theriogenology.2023.09.016.

Corbeil, L.B. (2007). Histophilus somni host-parasite relationships. Anim Health Res Rev 8:151–160. doi:10.1017/s1466252307001417.

Cribby, S., Taylor, M., and Reid, G. (2008). Vaginal microbiota and the use of probiotics. Interdiscip Perspect Infect Dis 2008:1–9. doi:10.1155/2008/256490.

Cseh, S., Faigl, V., and Amiridis, G.S. (2012). Semen processing and artificial insemination in health management of small ruminants. Anim Reprod Sci 130:187–192. doi:10.1016/j.anireprosci.2012.01.014.

Cutlip, R.C., Lehmkuhl, H.D., Schmerr, M.J.F., and Brogden, K.A. (1988). Ovine progressive pneumonia (maedi-visna) in sheep. Vet Microbiol 17:237–250. doi:10.1016/0378-1135(88)90068-5.

Donovan, A., Hanrahan, J.P., Kummen, E., Duffy, P., and Boland, M.P. (2004). Fertility in the ewe following cervical insemination with fresh or frozen–thawed semen at a

natural or synchronised oestrus. Anim Reprod Sci 84:359–368. doi:10.1016/j.anireprosci.2003.12.014.

Dufour, A., Kurylo, C., Stöckl, J.B., Laloë, D., Bailly, Y., Manceau, P., et al. (2024). Cell specification and functional interactions in the pig blastocyst inferred from single-cell transcriptomics and uterine fluids proteomics. Genomics 116:110780. doi:10.1016/j.ygeno.2023.110780.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4:1184–1191. doi:10.1038/nprot.2009.97.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. doi:10.1126/science.1162986.

Eisenberg, T., Gronow, S., Falgenhauer, J., Imirzalioglu, C., Mühldorfer, K., Rau, J., et al. (2021). Sneathia vaginalis sp. nov.(Fusobacteriales, Leptotrichiaceae) as a replacement of the species 'Sneathia amnii'Harwich et al. 2012 and 'Leptotrichia amnionii'Shukla et al. 2002, and emended description of Sneathia Collins et al. 2001. Int J Syst Evol Microbiol 71:004663. doi:10.1099/ijsem.0.004663.

Elad, D., Friedgut, O., Alpert, N., Stram, Y., Lahav, D., Tiomkin, D., et al. (2004). Bovine necrotic vulvovaginitis associated with Porphyromonas levii. Emerg Infect Dis 10:505. doi:10.3201/eid1003.020592.

Eribe, E.R.K., and Olsen, I. (2017). Leptotrichia species in human infections II. J Oral Microbiol 9:1368848. doi:10.1080/20002297.2017.1368848.

European Parliament. (2018). Report on the Current Situation and Future Prospects for the Sheep and Goat Sectors in the EU. Accessed June 25, 2024. https://www.europarl.europa.eu/doceo/document/A-8-2018-0064_EN.html.

Eurostat. (2023). Agricultural Production - Livestock and Meat - Statistics Explained. Accessed October 21, 2024. https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=470510#Sheep_and_goat_meat.

Evaldson, G., Malmborg, A.S., Nord, C.E., and Östensson, K. (1983). Bacteroides fragilis, Streptococcus intermedius and group B streptococci in ascending infection of pregnancy: an animal experimental study. Gynecol Obstet Invest 15:230–241. doi:10.1159/000299415.

Finn, D.R. (2024). A metagenomic alpha-diversity index for microbial functional biodiversity. FEMS Microbiol Ecol 100. doi:10.1093/femsec/fiae019.

Fontanals, D., García-Miralles, C., Ballester, R., Cochs, B., Orellana, R., and Capilla, S. (2018). Chorioamnionitis due to Leptotrichia trevisanii. Anaerobe 49:18–20. doi:10.1016/j.anaerobe.2017.11.001.

Foster, G., Collins, M.D., Lawson, P.A., Buxton, D., Murray, F.J., and Sime, A. (1999). Actinobacillus seminis as a cause of abortion in a UK sheep flock.. Vet Rec 144:479–480. doi:10.1136/vr.144.17.479.

Freese, H.M., Methner, A., and Overmann, J. (2017). Adaptation of surface-associated bacteria to the open ocean: a genomically distinct subpopulation of Phaeobacter gallaeciensis colonizes pacific mesozooplankton. Front Microbiol 8:277070. doi:10.3389/fmicb.2017.01659.

Gail, M.H., Wan, Y., and Shi, J. (2021). Power of microbiome beta-diversity analyses based on standard reference samples. Am J Epidemiol 190:439–447. doi:10.1093/aje/kwaa204.

Gajer, P., Brotman, R.M., Bai, G., Sakamoto, J., Schütte, U.M.E., Zhong, X., et al. (2012). Temporal Dynamics of the Human Vaginal Microbiota. Sci Transl Med 4:132ra52. doi:10.1126/scitranslmed.3003605.

Garcia-Segura, S., del Rey, J., Closa, L., Garcia-Martínez, I., Hobeich, C., Castel, A.B., et al. (2022). Seminal microbiota of idiopathic infertile patients and its relationship with sperm DNA integrity. Front Cell Dev Biol 10:937157. doi:10.3389/fcell.2022.937157.

Garrido-Martín, D., Calvo, M., Reverter, F., and Guigó, R. (2023). A fast non-parametric test of association for multiple traits. Genome Biol 24:1–32. doi:10.1186/s13059-023-03076-8.

Gatti, M., Zunino, P., and Ungerfeld, R. (2011). Changes in the aerobic vaginal bacterial mucous load after treatment with intravaginal sponges in anoestrous ewes: effect of medroxiprogesterone acetate and antibiotic treatment use. Reproduction in domestic animals 46:205–208. doi:10.1111/j.1439-0531.2010.01626.x.

GeneCards. (2024a). BLTP3A Gene: Bridge-like Lipid Transfer Protein Family Member 3A. Accessed March 9, 2025. https://www.genecards.org/cgi-bin/carddisp.pl?gene=BLTP3A.

GeneCards. (2024b). MYMK Gene: Myomaker, Myoblast Fusion Factor. Accessed March 9, 2025. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MYMK.

Gibiino, G., Lopetuso, L.R., Scaldaferri, F., Rizzatti, G., Binda, C., and Gasbarrini, A. (2018). Exploring Bacteroidetes: metabolic key points and immunological tricks of our gut commensals. Digestive and Liver Disease 50:635–639. doi:10.1016/j.dld.2018.03.016.

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

Gohil, P., Nanavati, B., Patel, K., Suthar, V., Joshi, M., Patil, D.B., et al. (2023). Assessing the efficacy of probiotics in augmenting bovine reproductive health: an integrated in vitro, in silico, and in vivo study. Front Microbiol 14:1137611. doi:10.3389/fmicb.2023.1137611.

Gonçalves, J.D., Dias, J.H., Machado-Neves, M., Vergani, G.B., Ahmadi, B., Pereira Batista, R.I.T., et al. (2024). Transcervical uterine flushing and embryo transfer in sheep: morphophysiological basis for approaches currently used, major challenges, potential improvements, and new directions (alas, including some old ideas). Reprod Biol 24:100920. doi:10.1016/j.repbio.2024.100920.

Gonzalez Moreno, C., Fontana, C., Cocconcelli, P.S., Callegari, M.L., and Otero, M.C. (2016). Vaginal microbial communities from synchronized heifers and cows with reproductive disorders. J Appl Microbiol 121:1232–1241. doi:10.1111/jam.13239.

Gootwine, E. (2021). Management of dairy animals: sheep- reproductive management. Elsevier.

Granleese, T., Clark, S.A., Swan, A.A., and van der Werf, J.H.J. (2015). Increased genetic gains in sheep, beef and dairy breeding programs from using female reproductive technologies combined with optimal contribution selection and genomic breeding values. Genetics Selection Evolution 47:70. doi:10.1186/s12711-015-0151-3.

Greenacre, M., Martínez-Álvaro, M., and Blasco, A. (2021). Compositional data analysis of microbiome and any-omics datasets: a validation of the additive logratio transformation. Front Microbiol 12:727398. doi:10.3389/fmicb.2021.727398.

Greenwood, E.C., Torok, V.A., and Van Wettere, W.H.E.J. (2022). Characterizing the vaginal microbiota of high and low producing Poll Merino and White Suffolk ewes. Transl Anim Sci 6:txac133. doi:10.1093/tas/txac133.

Gundi, V.A.K.B., Desbriere, R., and La Scola, B. (2004). Leptotrichia amnionii and the female reproductive tract. Emerg Infect Dis 10:2056. doi:10.3201/eid1011.031019.

Vander Haar, E.L., So, J., Gyamfi-Bannerman, C., and Han, Y.W. (2018). Fusobacterium nucleatum and adverse pregnancy outcomes: epidemiological and mechanistic evidence. Anaerobe 50:55–59. doi:10.1016/j.anaerobe.2018.01.008.

Halaweh, W., Khnissi, S., Ben Souf, I., Salman, M., and M'Hamdi, N. (2025). Impact of mating methods and semen preservation on reproductive and growth

performances in palestinian Assaf sheep. Biology (Basel) 14:80. doi:10.3390/biology14010080.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews 68:669–685. doi:10.1128/mmbr.68.4.669-685.2004.

Hanna, M.G., Rodriguez Cruz, H.O., Fujise, K., Wu, Y., Xu, C.S., Pang, S., et al. (2024). Bridge-like lipid transfer protein 3A (BLTP3A) is associated with membranes of the late endocytic pathway and is an effector of CASM. bioRxiv 2024.09.28.615015. doi:10.1101/2024.09.28.615015.

Hector, A., Kormann, M., Kammermeier, J., Burdi, S., Marcos, V., Rieber, N., et al. (2013). Expression and regulation of interferon-related development regulator-1 in cystic fibrosis neutrophils. Am J Respir Cell Mol Biol 48:71–77. doi:10.1165/rcmb.2012-0061oc.

Hervé, M., and Hervé, M.M. (2020). Package 'RVAideMemoire' 0–9.

Hosny, A.E.D.M.S., El-khayat, W., Kashef, M.T., and Fakhry, M.N. (2017). Association between preterm labor and genitourinary tract infections caused by Trichomonas vaginalis, Mycoplasma hominis, Gram-negative bacilli, and coryneforms. Journal of the Chinese Medical Association 80:575–581. doi:10.1016/j.jcma.2016.10.007.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–441. doi:10.1037/h0071325.

Hou, D., Zhou, X., Zhong, X., Settles, M.L., Herring, J., Wang, L., et al. (2013). Microbiota of the seminal fluid from healthy and infertile men. Fertil Steril 100:1261–1269. doi:10.1016/j.fertnstert.2013.07.1991.

Huang, Q., Tian, R., Yu, J., and Du, W. (2024). Identification of PSMD11 as a novel cuproptosis‐ and immune‐related prognostic biomarker promoting lung adenocarcinoma progression. Cancer Med 13:e7379. doi:10.1002/cam4.7379.

Huang, Z., DerGarabedian, B.P., He, L., Sha, Y., Chen, Z., Kang, J., et al. (2020). Impact of periodonto-pathogenic microbiota and sociodemographic variables on periodontal status during pregnancy and postpartum period. Oral Health Prev Dent 18:855–864. doi:10.3290/j.ohpd.a45355.

Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774. doi:10.1128/jb.180.18.4765-4774.1998.

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

Jamin, S.P., Petit, F.G., Kervarrec, C., Smagulova, F., Illner, D., Scherthan, H., et al. (2017). EXOSC10/Rrp6 is post-translationally regulated in male germ cells and controls the onset of spermatogenesis. Sci Rep 7:1–15. doi:10.1038/s41598-017-14643-y.

Janda, J.M., and Abbott, S.L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol 45:2761–2764. doi:10.1128/jcm.01228-07.

Jeon, S.J., Cunha, F., Ma, X., Martinez, N., Vieira-Neto, A., Daetz, R., et al. (2016). Uterine microbiota and immune parameters associated with fever in dairy cows with metritis. PLoS One 11:e0165740. doi:10.1371/journal.pone.0165740.

Jia, J., Absmeier, E., Holton, N., Pietrzyk-Brzezinska, A.J., Hackert, P., Bohnsack, K.E., et al. (2020). The interaction of DNA repair factors ASCC2 and ASCC3 is affected by somatic cancer mutations. bioRxiv 2020.05.22.111591. doi:10.1101/2020.05.22.111591.

Jiang, L., Qi, X., Lai, M., Zhou, J., Yuan, M., You, J., et al. (2024). WDR20 prevents hepatocellular carcinoma senescence by orchestrating the simultaneous USP12/46-mediated deubiquitination of c-Myc. Proceedings of the National Academy of Sciences 121:e2407904121. doi:10.1073/PNAS.2407904121/SUPPL_FILE/PNAS.2407904121.SAPP.PDF.

Kalds, P., Zhou, S., Gao, Y., Cai, B., Huang, S., Chen, Y., et al. (2022). Genetics of the phenotypic evolution in sheep: a molecular look at diversity-driving genes. Genetics Selection Evolution 54:1–27. doi:10.1186/s12711-022-00753-3.

Kaur, H., Merchant, M., Haque, M.M., and Mande, S.S. (2020). Crosstalk between female gonadal hormones and vaginal microbiota across various phases of women's gynecological lifecycle. Front Microbiol 11:503746. doi:10.3389/fmicb.2020.00551.

Kers, J.G., and Saccenti, E. (2022). The power of microbiome studies: some considerations on which alpha and beta metrics to use and how to report results. Front Microbiol 12:796025. doi:10.3389/fmicb.2021.796025.

Kirkbride, C.A. (1993). Diagnoses in 1,784 ovine abortions and stillbirths. J Vet Diagn Invest 5:398–402. doi:10.1177/104063879300500316.

Koester, L.R., Petry, A.L., Youngs, C.R., and Schmitz-Esser, S. (2021). Ewe vaginal microbiota: associations with pregnancy outcome and changes during gestation. Front Microbiol 12:745884. doi:10.3389/fmicb.2021.745884.

Koort, K., Sõsa, K., Türk, S., Lapp, E., Talving, E., Karits, P., et al. (2023). Lactobacillus crispatus-dominated vaginal microbiome and Acinetobacter-dominated seminal microbiome support beneficial ART outcome. Acta Obstet Gynecol Scand 102:921–934. doi:10.1111/aogs.14598.

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27. doi:10.1007/BF02289565.

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:115–129. doi:10.1007/bf02289694.

Kudo, H., Sugiura, T., Higashi, S., Oka, K., Takahashi, M., Kamiya, S., et al. (2021). Characterization of reproductive microbiota of primiparous cows during early postpartum periods in the presence and absence of endometritis. Front Vet Sci 8:736996. doi:10.3389/FVETS.2021.736996.

Kwon, M.S., and Lee, H.K. (2022). Host and microbiome interplay shapes the vaginal microenvironment. Front Immunol 13:919728. doi:10.3389/fimmu.2022.919728.

De La Puente-Redondo, V.A., García Del Blanco, N., Pérez-Martínez, C., González-Rodríguez, M.C., Rodríguez-Ferri, E.F., and Gutiérrez-Martín, C.B. (2000). Isolation of Actinobacillus seminis from the genital tract of rams in spain. J Comp Pathol 122:217–222. doi:10.1053/jcpa.1999.0359.

Laguardia-Nascimento, M., Branco, K.M.G.R., Gasparini, M.R., Giannattasio-Ferraz, S., Leite, L.R., Araujo, F.M.G., et al. (2015). Vaginal microbiome characterization of Nellore cattle using metagenomic analysis. PLoS One 10:e0143294. doi:10.1371/journal.pone.0143294.

Lahti, L., and Shetty, S. (2018). Introduction to the microbiome R package.

Lee, K.C.Y., Morgan, X.C., Power, J.F., Dunfield, P.F., Huttenhower, C., and Stott, M.B. (2015). Complete genome sequence of the thermophilic Acidobacteria, Pyrinomonas methylaliphatogenes type strain K22T. Stand Genomic Sci 10:1–8. doi:10.1186/s40793-015-0099-5.

Legaria, M.C., Nastro, M., Camporro, J., Heger, F., Barberis, C., Stecher, D., et al. (2021). Peptostreptococcus anaerobius: Pathogenicity, identification, and antimicrobial susceptibility. Review of monobacterial infections and addition of a case of urinary tract infection directly identified from a urine sample by MALDI-TOF MS. Anaerobe 72:102461. doi:10.1016/j.anaerobe.2021.102461.

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. Microbiome 7:1–17. doi:10.1186/s40168-019-0699-1.

Li, N., Zhang, L., Li, F., Wang, Y., Zhu, Y., Kang, H., et al. (2011). Metagenome of microorganisms associated with the toxic Cyanobacteria Microcystis aeruginosa analyzed using the 454 sequencing platform. Chinese Journal of Oceanology and Limnology 29:505–513. doi:10.1007/s00343-011-0056-0.

Lianou, D.T., Vasileiou, N.G.C., Michael, C.K., Valasi, I., Mavrogianni, V.S., Caroprese, M., et al. (2022). Patterns of reproductive management in sheep and goat farms in Greece. Animals 12:3455. doi:10.3390/ani12243455.

Liaw, A., and Wiener, M. (2022). randomForest: classification and regression.

Liu, K., Zhang, Y., Yu, Z., Xu, Q., Zheng, N., Zhao, S., et al. (2021). Ruminal microbiota–host interaction and its effect on nutrient metabolism. Animal Nutrition 7:49–55. doi:10.1016/J.ANINU.2020.12.001.

Locatelli, C., Scaccabarozzi, L., Pisoni, G., Bronzo, V., Casula, A., Testa, F., et al. (2013). Helcococcus kunzii and Helcococcus ovis isolated in dairy cows with puerperal metritis. J Gen Appl Microbiol 59:371–374. doi:10.2323/jgam.59.371.

Loose, M., Malla, S., and Stout, M. (2016). Real-time selective sequencing using nanopore technology. Nat Methods 13:751–754. doi:10.1038/nmeth.3930.

López, M.G., Irla, M., Brito, L.F., and Wendisch, V.F. (2019). Characterization of D-arabitol as newly discovered carbon source of Bacillus methanolicus. Front Microbiol 10:1725. doi:10.3389/fmicb.2019.01725.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21. doi:10.1186/s13059-014-0550-8.

Lyman, C.C., Holyoak, G.R., Meinkoth, K., Wieneke, X., Chillemi, K.A., and DeSilva, U. (2019). Canine endometrial and vaginal microbiomes reveal distinct and complex ecosystems. PLoS One 14:e0210157. doi:10.1371/journal.pone.0210157.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2022). Cluster: cluster analysis basics and extensions.

Magi, A., Semeraro, R., Mingrino, A., Giusti, B., and D'Aurizio, R. (2018). Nanopore sequencing data analysis: state of the art, applications and challenges. Brief Bioinform 19:1256–1272. doi:10.1093/bib/bbx062.

Maintainer, C., and Boutros, P. (2018). Package "VennDiagram". R package.

Mandal, S., Van Treuren, W., White, R.A., Eggesbø, M., Knight, R., and Peddada, S.D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 26:27663. doi:10.3402/mehd.v26.27663.

Mani, S., Aiyegoro, O.A., and Adeleke, M.A. (2022). Association between host genetics of sheep and the rumen microbial composition. Trop Anim Health Prod 54. doi:10.1007/s11250-022-03057-2.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. doi:10.1038/nature03959.

Marín-Bernal, A.M., and Navarro-Ríos, M.J. (2014). Análisis y diagnóstico de los sistemas de producción ovina en el sureste español. Arch. Zootec 63:9–13.

Mariño-Ramírez, L., Minor, J.L., Reading, N., and Hu, J.C. (2004). Identification and mapping of self-assembling protein domains encoded by the Escherichia coli K-12 genome by use of λ repressor fusions. J Bacteriol 186:1311–1319. doi:10.1128/jb.186.5.1311-1319.2004.

Marquet, M., Zöllkau, J., Pastuschek, J., Viehweger, A., Schleußner, E., Makarewicz, O., et al. (2022). Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. Sci Rep 12. doi:10.1038/s41598-022-08003-8.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. doi:10.14806/ej.17.1.200.

Martinez Boggio, G., Christensen, O.F., Legarra, A., Meynadier, A., and Marie-Etancelin, C. (2023). Microbiability of milk composition and genetic control of microbiota effects in sheep. J Dairy Sci 106:6288–6298. doi:10.3168/jds.2022-22948.

Martinez Boggio, G., Meynadier, A., Buitenhuis, A.J., and Marie-Etancelin, C. (2022). Host genetic control on rumen microbiota and its impact on dairy traits in sheep. Genetics Selection Evolution 54:1–19. doi:10.1186/s12711-022-00769-9.

Martínez-Álvaro, M., Mattock, J., González-Recio, Ó., Saborío-Montero, A., Weng, Z., Lima, J., et al. (2024). Including microbiome information in a multi-trait genomic evaluation: a case study on longitudinal growth performance in beef cattle. Genetics Selection Evolution 56:1–16. doi:10.1186/s12711-024-00887-6.

Martinez-Ros, P., Lozano, M., Hernandez, F., Tirado, A., Rios-Abellan, A., López-Mendoza, M.C., et al. (2018). Intravaginal device-type and treatment-length for ovine estrus synchronization modify vaginal mucus and microbiota and affect fertility. Animals 8:226. doi:10.3390/ani8120226.

McClelland, R.S., Lingappa, J.R., Srinivasan, S., Kinuthia, J., John-Stewart, G.C., Jaoko, W., et al. (2018). Key vaginal bacteria associated with increased risk of HIV acquisition in African women: a nested case-control study. Lancet Infect Dis 18:554. doi:10.1016/s1473-3099(18)30058-6.

McCracken, T., Kainer, R., and Spurgeon, T. (2013). Spurgeon's Color Atlas of Large Animal Anatomy: The Essentials. John Wiley & Sons, ed.

McDonald, P., Edwards, R.A., Greenhalgh, J.F.D., Morgan, C.A., Sinclair, L.A., Wilkinson, R.G., et al. (2011). Animal Nutrition (7th Edition). Pearson, ed.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. doi:10.1371/journal.pone.0061217.

Mehta, S.D., Nannini, D.R., Otieno, F., Green, S.J., Agingu, W., Landay, A., et al. (2020). Host genetic factors associated with vaginal microbiome composition in Kenyan women. mSystems 5. doi:10.1128/msystems.00502-20.

Mellor, P.S., Carpenter, S., Harrup, L., Baylis, M., and Mertens, P.P.C. (2008). Bluetongue in Europe and the Mediterranean Basin: history of occurrence prior to 2006. Prev Vet Med 87:4–20. doi:10.1016/j.prevetmed.2008.06.002.

Menegay, H., Moeslein, F., and Landreth, G. (1999). The dual specificity protein kinase CLK3 is abundantly expressed in mature mouse spermatozoa. Exp Cell Res 253:463–473. doi:10.1006/excr.1999.4655.

Merk, R. (2006). A comparative categorization of protein function encoded in bacterial or archeal genomic islands. J Mol Evol 62:1–14. doi:10.1007/s00239-004-0311-5.

Miller, E.A., Beasley, D.A.E., Dunn, R.R., and Archie, E.A. (2016). Lactobacilli dominance and vaginal pH: why is the human vaginal microbiome unique?. Front Microbiol 7:224527. doi:10.3389/fmicb.2016.01936.

Ministerio de Agricultura. (2022). Ovino-Caprino. Accessed June 25, 2024. https://www.mapa.gob.es/es/ganaderia/temas/produccion-y-mercados-ganaderos/sectores-ganaderos/ovino-caprino/.

Ministerio de Agricultura, P. y A. (2019). BOE-A-2019-2859 real decreto 45/2019. Boletín Oficial del Estado.

Mirmonsef, P., Gilbert, D., Zariffard, M.R., Hamaker, B.R., Kaur, A., Landay, A.L., et al. (2011). The effects of commensal bacteria on innate immune responses in the female genital tract. American Journal of Reproductive Immunology 65:190–195. doi:10.1111/j.1600-0897.2010.00943.x.

Mohan, K., and Kumar, N. (2023). Comparative evaluation of estrus synchronization protocols on reproductive performance and estrus behavior in Barbados Black Belly sheep. Vet World 16:2244. doi:10.14202/vetworld.2023.2244-2249.

Morris, A.H., and Bohannan, B.J.M. (2024). Estimates of microbiome heritability across hosts. Nat Microbiol 9:3110–3119. doi:10.1038/s41564-024-01865-w.

Murga Valderrama, N.L., Segura Portocarrero, G.T., Romani Vasquez, A.C., Frias Torres, H., Flores Durand, G.J., Cornejo Villanueva, V.G., et al. (2023). Exploring the microbiome of two uterine sites in cows. Sci Rep 13:1–13. doi:10.1038/s41598-023-46093-0.

Nandi, S., De, U.K., and Chowdhury, S. (2011). Current status of contagious ecthyma or orf disease in goat and sheep—A global perspective. Small Ruminant Research 96:73–82. doi:10.1016/j.smallrumres.2010.11.018.

National Library of Medicine. (2023). KAT6B Gene. Accessed March 9, 2025. https://medlineplus.gov/genetics/gene/kat6b/.

National Research Council. (2001). Nutrient Requirements of Dairy Cattle. The National Academies Press, Washintong, USA.

National Research Council. (2007). Nutrient Requirements of Small Ruminants: Sheep, Goats, Cervids, and New World Camelids. National Academies Press.

Nazar, E., Khatami, F., Saffar, H., and Tavangar, S.M. (2019). The emerging role of succinate dehyrogenase genes (SDHx) in tumorigenesis. Int J Hematol Oncol Stem Cell Res 13:72. doi:10.18502/ijhoscr.v13i2.692.

NCBI. (2025a). Mov10l1 Mov10 like RISC Complex RNA Helicase 1 [Mus Musculus (House Mouse)]. Accessed March 10, 2025. https://www.ncbi.nlm.nih.gov/gene/83456.

NCBI. (2025b). MVD Mevalonate Diphosphate Decarboxylase [Homo Sapiens (Human)] . Accessed March 10, 2025. https://www.ncbi.nlm.nih.gov/gene/4597.

NCBI. (2025c). EXOSC10 Exosome Component 10 [Homo Sapiens (Human)]. Accessed March 10, 2025. https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=5394.

Nicholas, R., Ayling, R., and McAuliffe, L. (2008). Mycoplasma Diseases of Ruminants. CABI International.

Oh, S.J., Kim, T.H., Lim, J.M., and Jeong, J.W. (2013). Progesterone induces expression of Lrp2 in the murine uterus. Biochem Biophys Res Commun 441:175–179. doi:10.1016/j.bbrc.2013.10.037.

Oh, Y., Quiroz, E., Wang, T., Medina-Laver, Y., Redecke, S.M., Dominguez, F., et al. (2023). The NR2F2-HAND2 signaling axis regulates progesterone actions in the uterus at early pregnancy. Front Endocrinol (Lausanne) 14:1229033. doi:10.3389/fendo.2023.1229033.

Ojeda-Hernández, F., del Moral-Ventura, S., Capataz-Tafur, J., Peña-Castro, J., Abad-Zavaleta, J., Chay-Canul, A., et al. (2019). Vaginal microbiota in Pelibuey sheep treated with antimicrobials at the removal of intravaginal sponges impregnated with flurogestone acetate. Small Ruminant Research 170:116–119. doi:10.1016/j.smallrumres.2018.11.015.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., et al. (2022). vegan: community ecology package.

Ong, C.T., Turni, C., Blackall, P.J., Boe-Hansen, G., Hayes, B.J., and Tabor, A.E. (2021). Interrogating the bovine reproductive tract metagenomes using culture-independent approaches: a systematic review. Anim Microbiome 3:1–12. doi:10.1186/s42523-021-00106-3.

Ovesen, P.L., Juul-Madsen, K., Telugu, N.S., Schmidt, V., Frahm, S., Radbruch, H., et al. (2025). Alzheimer's disease risk gene SORL1 promotes receptiveness of human microglia to pro-inflammatory stimuli. Glia 73:857–872. doi:10.1002/GLIA.24659.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science 276:734–740. doi:10.1126/science.276.5313.734.

Pan, D., and Yu, Z. (2014). Intestinal microbiome of poultry and its interaction with host and diet. Gut Microbes 5:108–119. doi:10.4161/gmic.26945.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature Microbiology 2017 2:11 2:1533–1542. doi:10.1038/s41564-017-0012-7.

Payne, K.K., Svoronos, N., Chaurio, R.A., Sanz, J.P., Anadon, C., Calmette, J., et al. (2018). Polymorphic UHRF1BP1 drives superior anti-tumor immunity in ovarian cancer. The Journal of Immunology 200:178.15-178.15. doi:10.4049/jimmunol.200.supp.178.15.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space . The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2:559–572. doi:10.1080/14786440109462720.

Pearson, T.A., and Manolio, T.A. (2008). How to interpret a genome-wide association study. JAMA - Journal of the American Medical Association 299:1335–1344. doi:10.1001/jama.299.11.1335.

Piskor, E.M., Ross, J., Möröy, T., and Kosan, C. (2022). Myc-interacting zinc finger protein 1 (Miz-1) is essential to maintain homeostasis and immunocompetence of the B cell lineage. Biology (Basel) 11:504. doi:10.3390/biology11040504.

van der Plas, F., Hennecke, J., Chase, J.M., van Ruijven, J., and Barry, K.E. (2023). Universal beta-diversity–functioning relationships are neither observed nor expected. Trends Ecol Evol 38:532–544. doi:10.1016/j.tree.2023.01.008.

Poole, R.K., Soffa, D.R., McAnally, B.E., Smith, M.S., Hickman-Brown, K.J., and Stockland, E.L. (2023). Reproductive microbiomes in domestic livestock: insights utilizing 16S rRNA gene amplicon community sequencing. Animals 13:485. doi:10.3390/ani13030485.

Pragya, P., Kaur, G., Ali, S.A., Bhatla, S., Rawat, P., Lule, V., et al. (2017). High-resolution mass spectrometry-based global proteomic analysis of probiotic strains Lactobacillus fermentum NCDC 400 and RS2. J Proteomics 152:121–130. doi:10.1016/j.jprot.2016.10.016.

Puapermpoonsiri, S., Kato, N., Watanabe, K., Ueno, K., Chongsomchai, C., and Lumbiganon, P. (1996). Vaginal microflora associated with bacterial vaginosis in Japanese and Thai pregnant women. Clinical Infectious Diseases 23:748–752. doi:10.1093/clinids/23.4.748.

Pybus, V., and Onderdonk, A.B. (1998). A commensal symbiosis between Prevotella bivia and Peptostreptococcus anaerobius involves amino acids: potential significance to the pathogenesis of bacterial vaginosis. FEMS Immunol Med Microbiol 22:317–327. doi:10.1111/j.1574-695x.1998.tb01221.x.

Pyles, R.B., Miller, A.L., Maxwell, C., Dawson, L., Richardson-Harman, N., Swartz, G., et al. (2021). Characterization of the ovine vaginal microbiome and inflammation patterns as an improved testing model of human vaginal irritation. Frontiers in Reproductive Health 3:714829. doi:10.3389/frph.2021.714829.

Qi, Z., Shi, S., Tu, J., and Li, S. (2019). Comparative metagenomic sequencing analysis of cecum microbiotal diversity and function in broilers and layers. 3 Biotech 9:1–10. doi:10.1007/s13205-019-1834-1.

Quadros, D.L., Zanella, R., Bondan, C., Zanella, G.C., Facioli, F.L., da Silva, A.N., et al. (2020). Study of vaginal microbiota of Holstein cows submitted to an estrus synchronization protocol with the use of intravaginal progesterone device. Res Vet Sci 131:1–6. doi:10.1016/j.rvsc.2020.03.027.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., et al. (2008). A large genome center's improvements to the Illumina sequencing system. Nat Methods 5:1005–1010. doi:10.1038/nmeth.1270.

Quereda, J.J., García-Roselló, E., Barba, M., Mocé, M.L., Gomis, J., Jiménez-Trigos, E., et al. (2020). Use of probiotics in intravaginal sponges in sheep: A pilot study. Animals 10:719. doi:10.3390/ani10040719.

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35:833–844. doi:10.1038/nbt.3935.

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Radolf, J.D., Deka, R.K., Anand, A., Šmajs, D., Norgard, M. V., and Yang, X.F. (2016). Treponema pallidum, the syphilis spirochete: making a living as a stealth pathogen. Nat Rev Microbiol 14:744–759. doi:10.1038/nrmicro.2016.141.

Radostits, O.M.., and Done, S.H.. (2010). Veterinary Medicine—a Textbook of the Diseases of Cattle, Horses, Sheep, Pigs and Goats. Canadian Veterinary Medical Association.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S.K., McCulle, S.L., et al. (2011). Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 108:4680–4687. doi:10.1073/pnas.1002611107.

Regueira-Iglesias, A., Balsa-Castro, C., Blanco-Pintos, T., and Tomás, I. (2023). Critical review of 16S rRNA gene sequencing workflow in microbiome studies: from primer selection to advanced data analysis. Mol Oral Microbiol 38:347–399. doi:10.1111/omi.12434.

Reinoso-Peláez, E.L., Gianola, D., and González-Recio, O. (2022). Genome-enabled prediction methods based on machine learning. Springer US, New York.

Reinoso-Peláez, E.L., Puente-Sánchez, F., Serrano, M., Calvo, J.H., Ramon, M., and Saura, M. (2025a). Characterization of bacterial communities of ewe's vaginal tract and its potential impact on reproductive efficiency. Anim Microbiome 7:1–14. doi:10.1186/s42523-025-00383-2.

Reinoso-Peláez, E.L., Saura, M., González, C., Ramón, M., Calvo, J.H., and Serrano, M. (2025b). The influence of vaginal microbiota on ewe fertility: a metagenomic and functional genomic approach. Microbiome. doi:10.1186/s40168-025-02165-z.

Reinoso-Peláez, E.L., Saura, M., González-Recio, Ó., González, C., Fernández, A., Peiro-Pastor, R., et al. (2023). Impact of oestrus synchronization devices on ewes vaginal microbiota and artificial insemination outcome. Front Microbiol 14:1063807. doi:10.3389/fmicb.2023.1063807.

Ricotta, C. (2021). From the euclidean distance to compositional dissimilarity: what is gained and what is lost. Acta Oecologica 111:103732. doi:10.1016/j.actao.2021.103732.

Ricotta, C., and Podani, J. (2017). On some properties of the Bray-Curtis dissimilarity and their ecological meaning. Ecological Complexity 31:201–205. doi:10.1016/j.ecocom.2017.07.003.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. doi:10.1093/nar/gkv007.

Roca, J., Vázquez, J., Gil, M., Cuello, C., Parrilla, I., and Martínez, E. (2006). Challenges in pig artificial insemination. Reproduction in Domestic Animals 41:43–53. doi:10.1111/j.1439-0531.2006.00768.x.

Roehe, R., Dewhurst, R.J., Duthie, C.A., Rooke, J.A., McKain, N., Ross, D.W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. PLoS Genet 12:e1005846. doi:10.1371/journal.pgen.1005846.

Rojas, C., Gálvez-Jirón, F., De Solminihac, J., Padilla, C., Cárcamo, I., Villalón, N., et al. (2022). Crosstalk between body microbiota and the regulation of immunity. J Immunol Res 2022:6274265. doi:10.1155/2022/6274265.

Rowe, M., Veerus, L., Trosvik, P., Buckling, A., and Pizzari, T. (2020). The reproductive microbiome: an emerging driver of sexual selection, sexual conflict, mating systems, and reproductive isolation. Trends Ecol Evol 35:220–234. doi:10.1016/j.tree.2019.11.004.

Ruiz-Fons, F., González-Barrio, D., Aguilar-Ríos, F., Soler, A.J., Garde, J.J., Gortázar, C., et al. (2014). Infectious pathogens potentially transmitted by semen of the black variety of the Manchega sheep breed: health constraints for conservation purposes. Anim Reprod Sci 149:152–157. doi:10.1016/j.anireprosci.2014.07.006.

Rup, L. (2012). The human microbiome project. Indian J Microbiol 52:315. doi:10.1007/s12088-012-0304-9.

Ryu, E.P., and Davenport, E.R. (2025). Host genetic determinants of the microbiome across animals: from Caenorhabditis elegans to cattle. Annu Rev Anim Biosci 10:203–226. doi:10.1146/annurev-animal-020420-032054.

Rzewuska, M., Kwiecień, E., Chrobak-Chmiel, D., Kizerwetter-Świda, M., Stefańska, I., and Gieryńska, M. (2019). Pathogenicity and virulence of Trueperella pyogenes: a review. Int J Mol Sci 20:2737. doi:10.3390/ijms20112737.

Sabatini, S., Bergerbrant, S., Brandt, L., Margaryan, A., and Allentoft, M.E. (2019). Approaching sheep herds origins and the emergence of the wool economy in continental Europe during the Bronze Age. Archaeol Anthropol Sci 11:4909–4925. doi:10.1007/s12520-019-00856-x.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74:5463–5467. doi:10.1073/pnas.74.12.5463.

Sanglard, L.P., Schmitz-Esser, S., Gray, K.A., Linhares, D.C.L., Yeoman, C.J., Dekkers, J.C.M., et al. (2020). Investigating the relationship between vaginal microbiota and host genetics and their impact on immune response and farrowing traits in commercial gilts. Journal of Animal Breeding and Genetics 137:84–102. doi:10.1111/jbg.12456.

Sargison, N. (2008). Sheep Flock Health. John Wiley & Sons.

Sathe, S.R. (2018). Laparoscopic artificial insemination technique in small ruminants-a procedure review. Front Vet Sci 5:390825. doi:10.3389/fvets.2018.00266.

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

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6:e17288. doi:10.1371/journal.pone.0017288.

Scholz, M.B., Lo, C.C., and Chain, P.S.G. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Curr Opin Biotechnol 23:9–15. doi:10.1016/j.copbio.2011.11.013.

Schwarz, C., Mathieu, J., Gomez, J.L., Miller, M.R., Tikhonova, M., Nagaraja, T.G., et al. (2023). Unexpected finding of Fusobacterium varium as the dominant Fusobacterium species in cattle rumen: potential implications for liver abscess etiology and interventions. J Anim Sci 101:skad130. doi:10.1093/jas/skad130.

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., et al. (2017). Critical assessment of metagenome interpretation—a benchmark of metagenomics software. Nat Methods 14:1063–1071. doi:10.1038/nmeth.4458.

Serrano, M., Climent, E., Freire, F., Martínez-Blanch, J.F., González, C., Reyes, L., et al. (2020). Influence of the ovine genital tract microbiota on the species artificial insemination outcome. A pilot study in commercial sheep farms. High Throughput 9:16. doi:10.3390/ht9030016.

Serrano, M., Freire, F., Martinez-Blanch, J.F., González, C., Reyes, L., Solaz-Fuster, M.C., et al. (2018). Deciphering the role of vaginal and sperm metagenome in the success of artificial insemination in sheep. Page 106.

Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal 27:379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Sharpton, T.J. (2014). An introduction to the analysis of shotgun metagenomic data. Front Plant Sci 5:86894. doi:10.3389/fpls.2014.00209.

Shukla, S.K., Meier, P.R., Mitchell, P.D., Frank, D.N., and Reed, K.D. (2002). Leptotrichia amnionii sp. nov., a novel bacterium isolated from the amniotic fluid of a woman after intrauterine fetal demise. J Clin Microbiol 40:3346–3349. doi:10.1128/jcm.40.9.3346-3349.2002.

Simões, J., Abecia, J.A., Cannas, A., Delgadillo, J.A., Lacasta, D., Voigt, K., et al. (2021). Review: managing sheep and goats for sustainable high yield production. Animal 15:100293. doi:10.1016/j.animal.2021.100293.

Simpson, E.H. (1949). Measurement of diversity. Nature 163:688–688. doi:10.1038/163688a0.

Sirota, I., Zarek, S.M., and Segars, J.H. (2014). Potential influence of the microbiome on infertility and assisted reproductive technology. Semin Reprod Med 32:35. doi:10.1055/s-0033-1361821.

Smith, M.S., Hickman-Brown, K.J., Mcanally, B.E., Oliveira Filho, R. V., De Melo, G.D., Pohler, K.G., et al. (2023). Reproductive microbiome and cytokine profiles associated with fertility outcomes of postpartum beef cows. J Anim Sci 101:skad219. doi:10.1093/jas/skad219.

Sreevalsan, S., Döring, M., Paszkowski‐Rogacz, M., Brux, M., Blanck, C., Meyer, M., et al. (2020). MLLT6 maintains PD-L1 expression and mediates tumor immune resistance. EMBO Rep 21. doi:10.15252/embr.202050155.

Staley, J.T., and Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol 39:321–346. doi:10.1146/annurev.mi.39.100185.001541.

Steinegger, M., and Söding, J. (2018). Clustering huge protein sequence sets in linear time. Nat Commun 9:1–8. doi:10.1038/s41467-018-04964-5.

Strug, M.R., Su, R.W., Kim, T.H., Mauriello, A., Ticconi, C., Lessey, B.A., et al. (2018). RBPJ mediates uterine repair in the mouse and is reduced in women with recurrent pregnancy loss. The FASEB Journal 32:2452. doi:10.1096/fj.201701032r.

Su, W., Lin, X.T., Zhao, S., Zheng, X.Q., Zhou, Y.Q., Xiao, L.L., et al. (2022). Tripartite motif-containing protein 46 accelerates influenza A H7N9 virus infection by promoting K48-linked ubiquitination of TBK1. Virol J 19. doi:10.1186/s12985-022-01907-x.

Sun, X., Luo, X., Lin, L., Wang, S., Wang, C., Yuan, F., et al. (2023). Clinical features and underlying mechanisms of KAT6B disease in a Chinese boy. Mol Genet Genomic Med 11:e2202. doi:10.1002/mgg3.2202.

Swartz, J.D. (2014). METABOLIC AND MICROBIAL IMPACTS ON REPRODUCTION IN LIVESTOCK. Montana State University - Bozeman, College of Agriculture.

Swartz, J.D., Lachman, M., Westveer, K., O'Neill, T., Geary, T., Kott, R.W., et al. (2014). Characterization of the vaginal microbiota of ewes and cows reveals a unique microbiota with low levels of lactobacilli and near-neutral pH. Front Vet Sci 1:116383. doi:10.3389/fvets.2014.00019.

Sweet, L., Müller, C., Anand, M., and Zscheischler, J. (2023). Cross-validation strategy impacts the performance and interpretation of machine learning models. Artificial Intelligence for the Earth Systems 2. doi:10.1175/aies-d-23-0026.1.

Tadesse, B., Reda, A.A., Kassaw, N.T., and Tadeg, W. (2022). Success rate of artificial insemination, reproductive performance and economic impact of failure of first service insemination: a retrospective study. BMC Vet Res 18:1–10. doi:10.1186/s12917-022-03325-1.

Talbot, H., Saada, S., Naves, T., Gallet, P.F., Fauchais, A.L., and Jauberteau, M.O. (2019). Regulatory roles of sortilin and SorLA in immune-related processes. Front Pharmacol 9:1507. doi:10.3389/fphar.2018.01507.

Tamames, J., and Puente-Sánchez, F. (2019). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. Front Microbiol 9. doi:10.3389/FMICB.2018.03349.

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

Taylor, M.A., and Wall, R. (2015). Veterinary Parasitology (4th Edition). John Wiley & Sons, ed.

Teng, Y., Feng, S., Gu, Z., Hou, C., Xu, H., Li, Z., et al. (2024). Comparison of microbiota structure in reproductive tract of Yanbian cattle and Yanhuang cattle. Front Microbiol 15:1419914. doi:10.3389/fmicb.2024.1419914.

The Human Microbiome Project Consortium. (2012). A framework for human microbiome research. Nature 486:215–221. doi:10.1038/nature11209.

Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. Microb Inform Exp 2:1–12. doi:10.1186/2042-5783-2-3.

Tomaiuolo, R., Veneruso, I., Cariati, F., and D'argenio, V. (2020). Microbiota and human reproduction: the case of male infertility. High Throughput 9. doi:10.3390/ht9020010.

Uzal, F.A., Freedman, J.C., Shrestha, A., Theoret, J.R., Garcia, J., Awad, M.M., et al. (2014). Towards an understanding of the role of Clostridium perfringens toxins in human and animal disease. Future Microbiol 9:361–377. doi:10.2217/fmb.13.168.

Vasudevan, D., Neuman, S.D., Yang, A., Lough, L., Brown, B., Bashirullah, A., et al. (2020). Translational induction of ATF4 during integrated stress response requires noncanonical initiation factors eIF2D and DENR. Nat Commun 11:4677. doi:10.1038/s41467-020-18453-1.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74. doi:10.1126/science.1093857.

Vilariño, M., Rubianes, E., and Menchaca, A. (2013). Ovarian responses and pregnancy rate with previously used intravaginal progesterone releasing devices for fixed-time artificial insemination in sheep. Theriogenology 79:206–210. doi:10.1016/j.theriogenology.2012.10.007.

Wagener, K., Drillich, M., Baumgardt, S., Kämpfer, P., Busse, H.J., and Ehling-Schulz, M. (2014). Falsiporphyromonas endometrii gen. nov., sp. nov., isolated from the post-partum bovine uterus, and emended description of the genus Porphyromonas Shah and Collins 1988. Int J Syst Evol Microbiol 64:642–649. doi:10.1099/ijs.0.057307-0.

Wall, R., and Shearer, D. (2008). Veterinary Ectoparasites: Biology, Pathology and Control (2nd Ed.). John Wiley & Sons.

Wang, C., Zhang, L., Jiang, X., Ma, W., Geng, H., Wang, X., et al. (2022). Toward efficient and high-fidelity metagenomic data from sub-nanogram DNA: evaluation of library preparation and decontamination methods. BMC Biol 20:1–12. doi:10.1186/s12915-022-01418-9.

Wang, M., Ren, C., Wang, P., Cheng, X., Chen, Y., Huang, Y., et al. (2023a). Microbiome–metabolome reveals the contribution of the gut–testis axis to sperm motility in sheep (Ovis aries). Animals 13:996. doi:10.3390/ani13060996.

Wang, N., Chen, L., Yi, K., Zhang, B., Li, C., and Zhou, X. (2024a). The effects of microbiota on reproductive health: a review. Crit Rev Food Sci Nutr 64:1486–1507. doi:10.1080/10408398.2022.2117784.

Wang, W., Zhang, Y., Zhang, X., Li, C., Yuan, L., Zhang, D., et al. (2023b). Heritability and recursive influence of host genetics on the rumen microbiota drive body weight variance in male Hu sheep lambs. Microbiome 11:1–17. doi:10.1186/s40168-023-01642-7.

Wang, Y., Niu, K., Shi, Y., Zhou, F., Li, X., Li, Y., et al. (2024b). A review: targeting UBR5 domains to mediate emerging roles and mechanisms - chance or necessity?. International journal of surgery 110:4947–4964. doi:10.1097/js9.0000000000001541.

Wang, Y., Wang, Z., Liu, W., Xie, S., Ren, X., Yan, L., et al. (2024c). Genetic background of blood β-hydroxybutyrate acid concentrations in early-lactating holstein dairy cows based on genome-wide association analyses. Genes (Basel) 15:412. doi:10.3390/genes15040412.

Wang, Z., Wu, S., Wang, G., Yang, Z., Zhang, Y., Zhu, C., et al. (2023c). ARHGAP21 is involved in the carcinogenic mechanism of cholangiocarcinoma: a study based on bioinformatic analyses and experimental validation. Medicina (Lithuania) 59:139. doi:10.3390/medicina59010139.

Wee, B.A., Thomas, M., Sweeney, E.L., Frentiu, F.D., Samios, M., Ravel, J., et al. (2018). A retrospective pilot study to determine whether the reproductive tract microbiota differs between women with a history of infertility and fertile women. Aust N Z J Obstet Gynaecol 58:341–348. doi:10.1111/ajo.12754.

Wen, L., Liu, L., Shen, X., Li, H., Zhu, Z., Huang, H., et al. (2020). The association of the UHRF1BP1 gene with systemic lupus erythematosus was replicated in a Han Chinese population from mainland China. Ann Hum Genet 84:221–228. doi:10.1111/ahg.12362.

Weng, S.L., Chiu, C.M., Lin, F.M., Huang, W.C., Liang, C., Yang, T., et al. (2014). Bacterial communities in semen from men of infertile couples: metagenomic sequencing reveals relationships of seminal microbiota to semen quality. PLoS One 9. doi:10.1371/journal.pone.0110152.

Werner, J.J., Koren, O., Hugenholtz, P., Desantis, T.Z., Walters, W.A., Caporaso, J.G., et al. (2012). Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. ISME J 6:94–103. doi:10.1038/ismej.2011.82.

Wexler, H.M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. Clin Microbiol Rev 20:593–621. doi:10.1128/cmr.00008-07.

Wick, R.R., Judd, L.M., and Holt, K.E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:1–10. doi:10.1186/s13059-019-1727-y.

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer-Verlag New York.

Wilimitis, D., and Walsh, C.G. (2023). Practical considerations and applied examples of cross-validation for model development and evaluation in health care: tutorial. JMIR AI 2:e49023. doi:10.2196/49023.

Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Van Braekel, J., et al. (2019). Targeting the 16S rRNA gene for bacterial identification in complex mixed samples: comparative evaluation of second (Illumina) and third (Oxford Nanopore Technologies) generation sequencing technologies. Int J Mol Sci 21:298. doi:10.3390/ijms21010298.

Witkin, S.S., and Linhares, I.M. (2017). Why do lactobacilli dominate the human vaginal microbiota?. BJOG 124:606–611. doi:10.1111/1471-0528.14390.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74:5088–5090. doi:10.1073/pnas.74.11.5088.

Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol 20:1–13. doi:10.1186/s13059-019-1891-0.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:1–12. doi:10.1186/gb-2014-15-3-r46.

Wooley, J.C., Godzik, A., and Friedberg, I. (2010). A primer on metagenomics. PLoS Comput Biol 6:e1000667. doi:10.1371/journal.pcbi.1000667.

Wright, R.J., Gibson, M.I., and Christie-Oleza, J.A. (2019). Understanding microbial community dynamics to improve optimal microbiome selection. Microbiome 7:1–14. doi:10.1186/s40168-019-0702-x.

Wu, G., Zhang, C., Wu, H., Wang, R., Shen, J., Wang, L., et al. (2017). Genomic microdiversity of Bifidobacterium pseudocatenulatum underlying differential strain-level responses to dietary carbohydrate intervention. mBio 8:10–1128. doi:10.1128/mbio.02348-16.

Xue, J. (2013). Holocarboxylase synthetase represses long terminal repeats through epigenetic synergies between biotin and methyl donors, and activates immune response by catalyzing biotinylation of heat shock protein 72. ETD collection for University of Nebraska-Lincoln.

Yaeger, M.J., Sahin, O., Plummer, P.J., Wu, Z., Stasko, J.A., and Zhang, Q. (2021). The pathology of natural and experimentally induced Campylobacter jejuni abortion

in sheep. Journal of Veterinary Diagnostic Investigation 33:1096–1105. doi:10.1177/10406387211033293.

Yagisawa, T., Uchiyama, J., Takemura-Uchiyama, I., Ando, S., Ichii, O., Murakami, H., et al. (2023). Metataxonomic analysis of the uterine microbiota associated with low fertility in dairy cows using endometrial tissues prior to first artificial insemination. Microbiol Spectr 11. doi:10.1128/spectrum.04764-22.

Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13:329–342. doi:10.1038/NRG3174.

Yang, B., Wang, Y., and Qian, P.Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics 17:1–8. doi:10.1186/s12859-016-0992-y.

Yang, D., Geng, T., Harrison, A.G., Cahoon, J.G., Xing, J., Jiao, B., et al. (2024). UBR5 promotes antiviral immunity by disengaging the transcriptional brake on RIG-I-like receptors. Nat Commun 15. doi:10.1038/s41467-024-45141-1.

Yuan, J., Ma, L., Du, J., Sun, H., Li, S., Zhou, G., et al. (2025). Host DNA depletion assisted metagenomic sequencing of bronchoalveolar lavage fluids for diagnosis of pulmonary tuberculosis.. Ann Clin Microbiol Antimicrob 24:13. doi:10.1186/s12941-025-00782-y.

Zangirolamo, A.F., Souza, A.K., Yokomizo, D.N., Miguel, A.K.A., Costa, M.C. da, Alfieri, A.A., et al. (2024). Updates and current challenges in reproductive microbiome: a comparative analysis between cows and women. Animals 14:1971. doi:10.3390/ani14131971.

Zeng, X., Wang, Y., Jia, H., Wang, Z., Gao, Z., Luo, Y., et al. (2022). Metagenomic analysis of microflora structure and functional capacity in probiotic Tibetan kefir grains. Food Research International 151:110849. doi:10.1016/j.foodres.2021.110849.

Zhang, H., Zha, X., Zhang, B., Zheng, Y., Elsabagh, M., Wang, H., et al. (2024). Gut microbiota contributes to bisphenol A-induced maternal intestinal and placental apoptosis, oxidative stress, and fetal growth restriction in pregnant ewe model by regulating gut–placental axis. Microbiome 12. doi:10.1186/s40168-024-01749-5.

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

Zhang, W., Fan, X., Shi, H., Li, J., Zhang, M., Zhao, J., et al. (2023). Comprehensive assessment of 16S rRNA gene amplicon sequencing for microbiome profiling across multiple habitats. Microbiol Spectr 11. doi:10.1128/spectrum.00563-23.

Zhang, Y., Wu, Z., Li, L., Wang, X., Fan, W., and Zhao, J. (2022). Characterizing the supragingival microbiome of healthy pregnant women. Front Cell Infect Microbiol 12. doi:10.3389/fcimb.2022.1016523.

Zhou, M., Yu, J., Xu, Y., Li, H., Feng, Y.-Q., Wang, X., et al. (2024). Exosc10 deficiency in the initial segment is dispensable for sperm maturation and male fertility in mice. Zygote 32:437–445. doi:10.1017/s0967199424000418.

Zhou, X., Brown, C.J., Abdo, Z., Davis, C.C., Hansmann, M.A., Joyce, P., et al. (2007). Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. ISME J 1:121–133. doi:10.1038/ismej.2007.12.

APPENDIX

Appendix for Study 1: characterisation of vaginal bacterial communities and their association with pregnancy outcomes using 16S rRNA metabarcoding

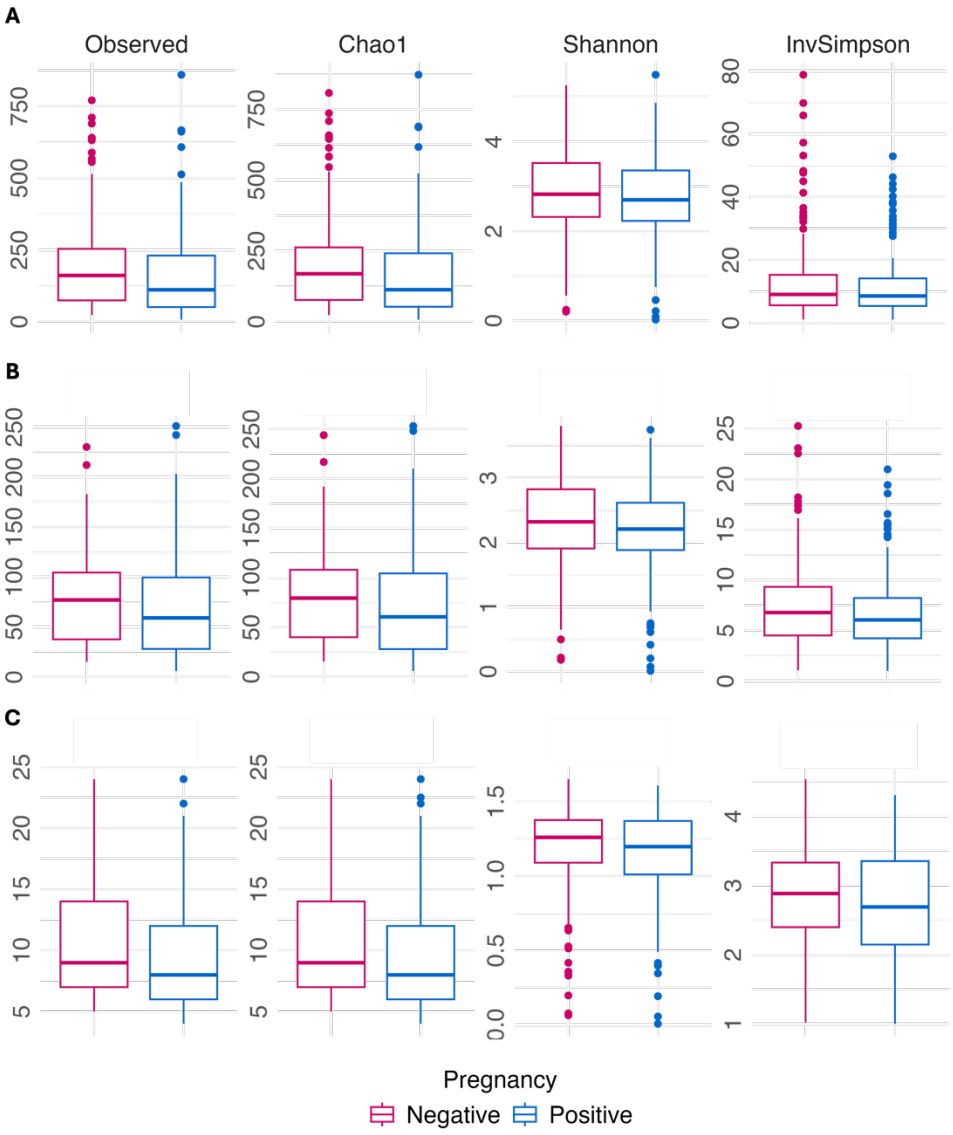


Figure S1.1. Alpha-diversity indices (Observed, Chao1, Shannon, and InvSimpson index) for pregnancy at the ASV (**A**), genus (**B**), and phylum (**C**) level using the 16S rRNA metabarcoding approach. Data were rarefied.

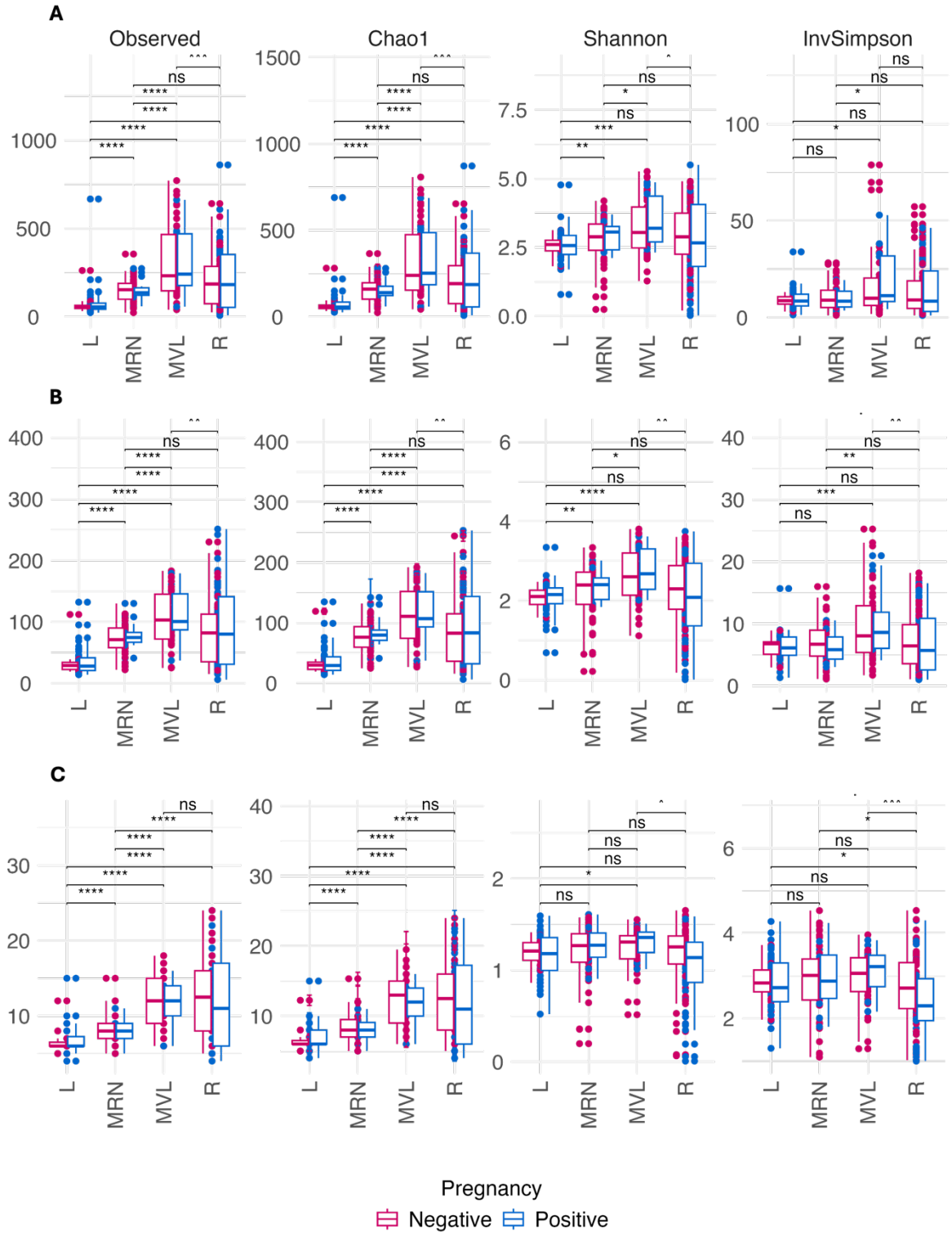


Figure S1.2. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for Pregnancy and herd at the ASV (**A**), genus (**B**) and phylum (**C**) level using the 16S rRNA metabarcoding approach. Data were rarefied. Statistical significance is indicated by asterisks: * ($p < 0.05$), *** ($p < 0.001$). ns: not significant.

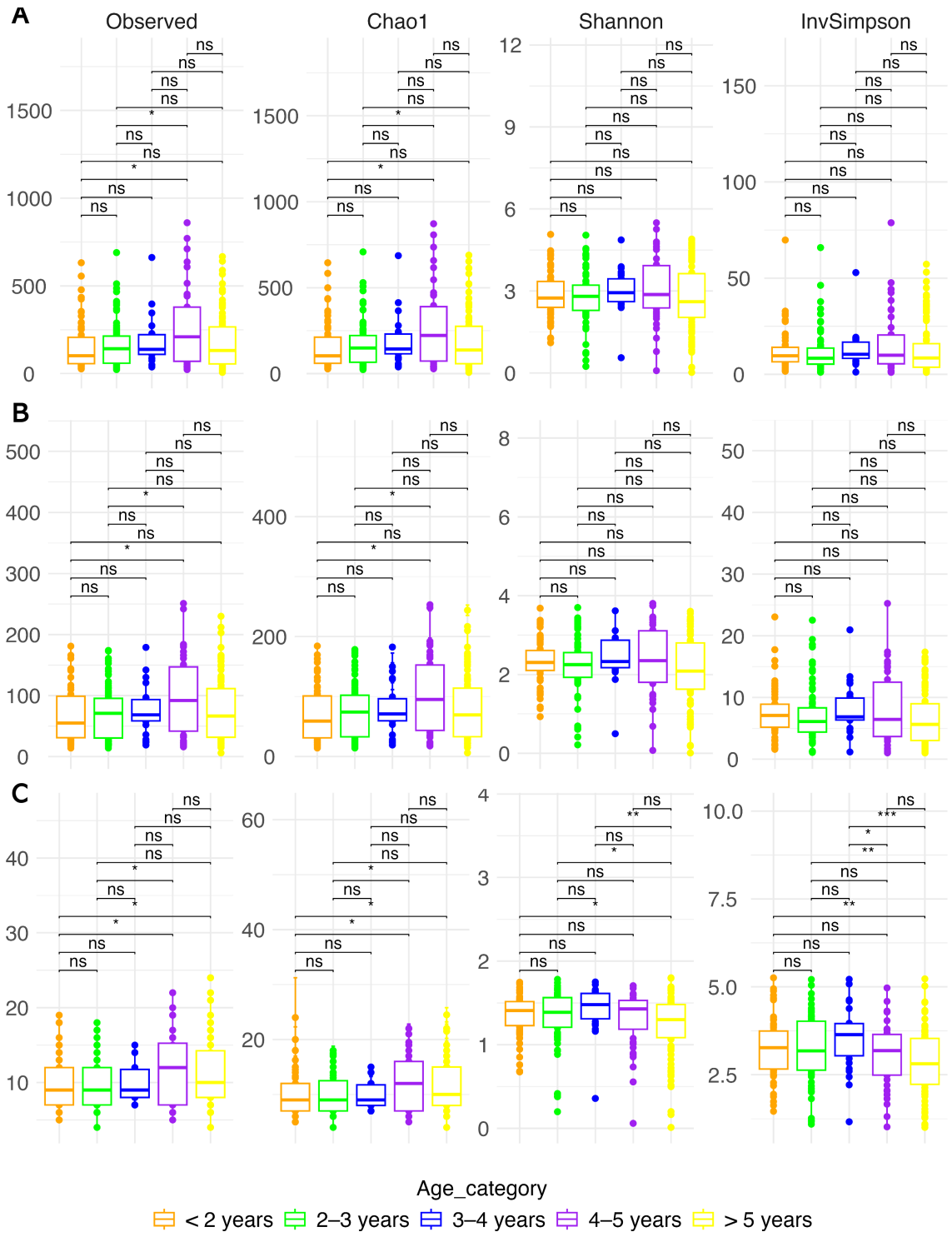


Figure S1.3. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for categorised age at the ASV (**A**), genus (**B**) and phylum (**C**) level using the 16S rRNA metabarcoding approach. Statistical significance is indicated by asterisks: * ($p < 0.05$), *** ($p < 0.001$). ns: not significant.

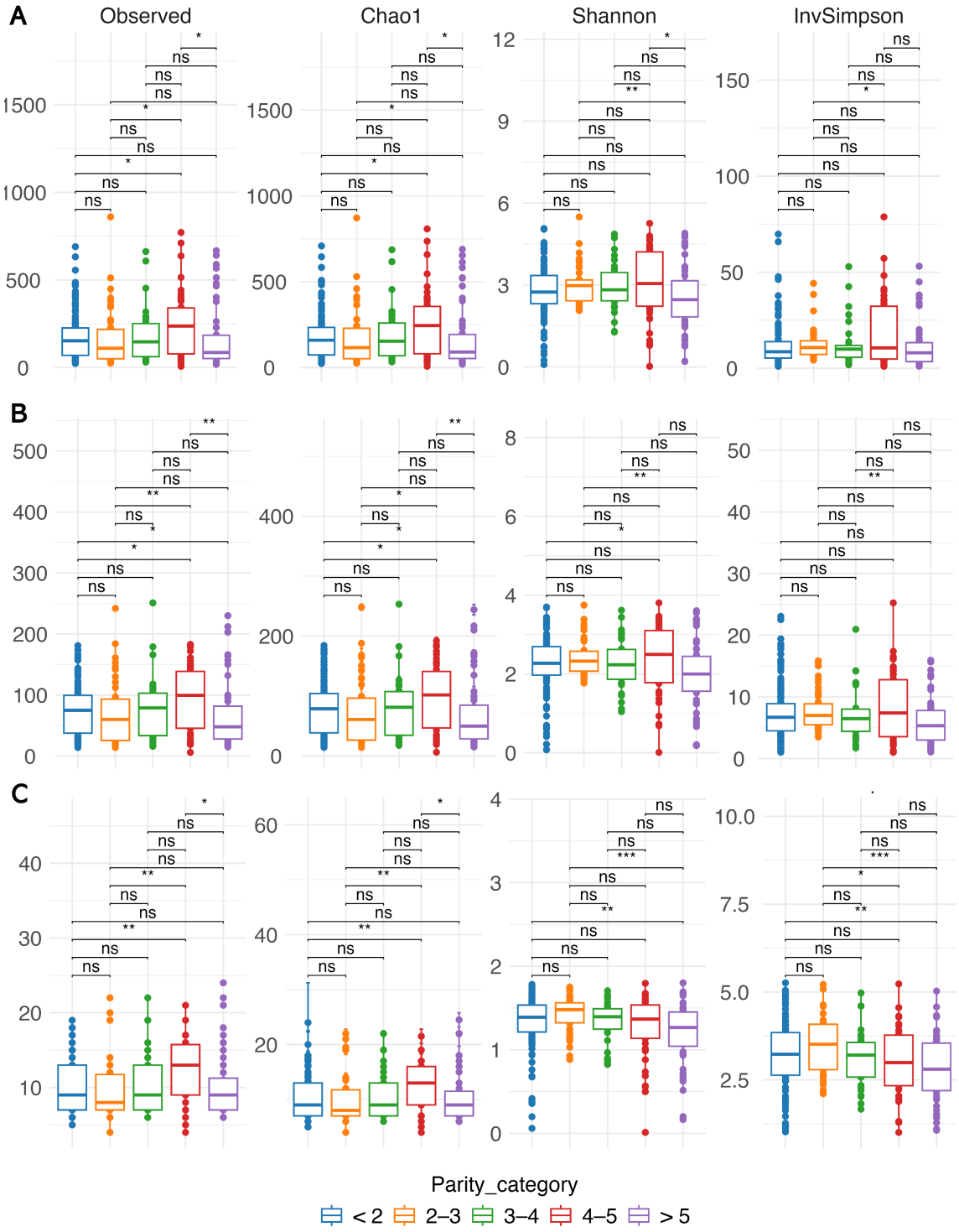


Figure S1.4. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for categorised parity at the ASV (**A**), genus (**B**) and phylum (**C**) level using the 16S rRNA metabarcoding approach. Data were rarefied. Statistical significance is indicated by asterisks: * ($p < 0.05$), *** ($p < 0.001$). ns: not significant.

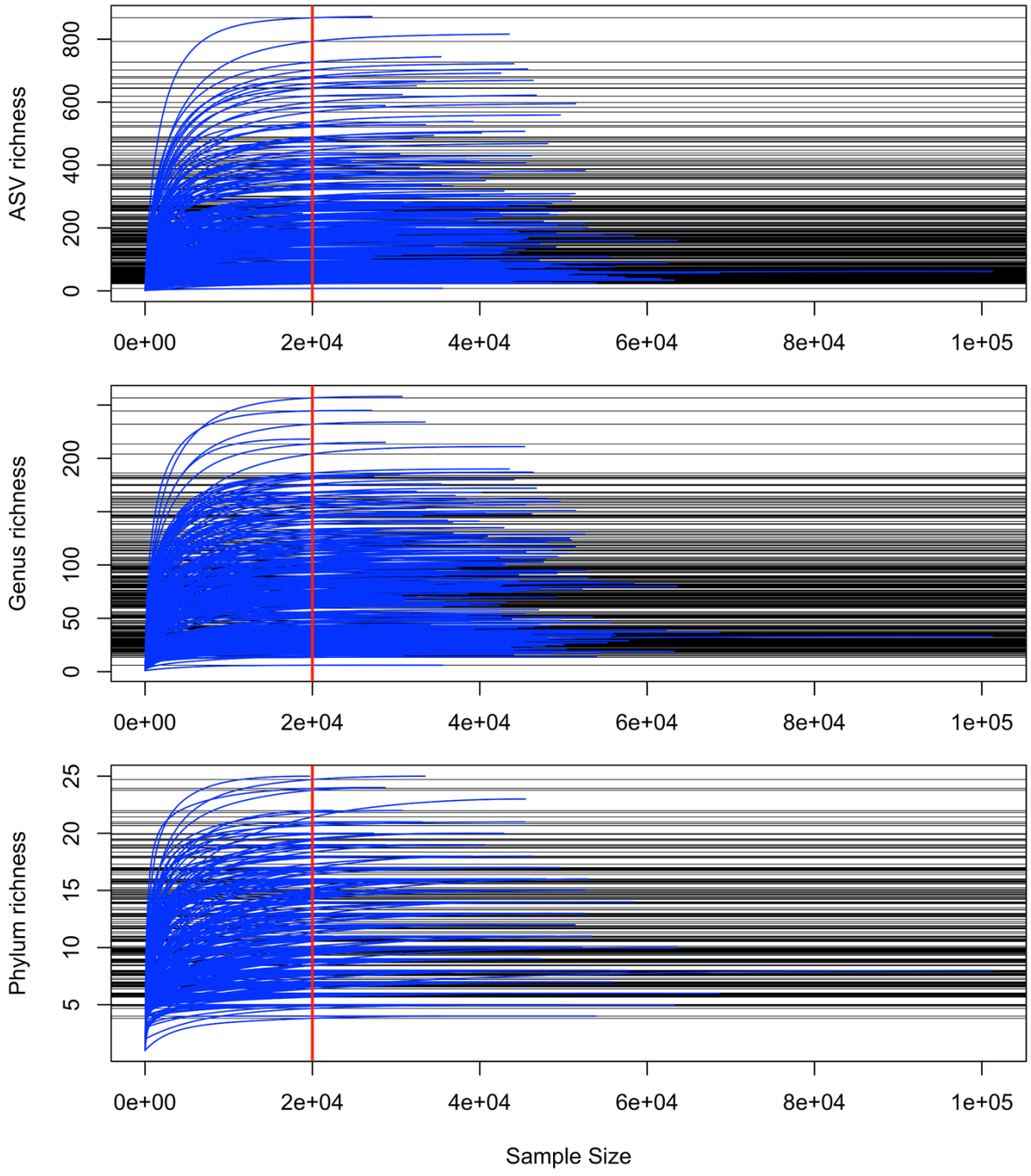


Figure S1.5. Rarefaction curves for all analysed samples for alpha-diversity analysis using the 16S rRNA metabarcoding approach. Sequencing depth in the x-axis and observed taxa richness in the y-axis for ASV, phylum and genus level with 16s rRNA metabarcoding sequencing approach. Blue lines represent individual samples. Red vertical dashed line indicates the threshold value (20,000). Black horizontal lines indicate the total sequencing depth for each sample.

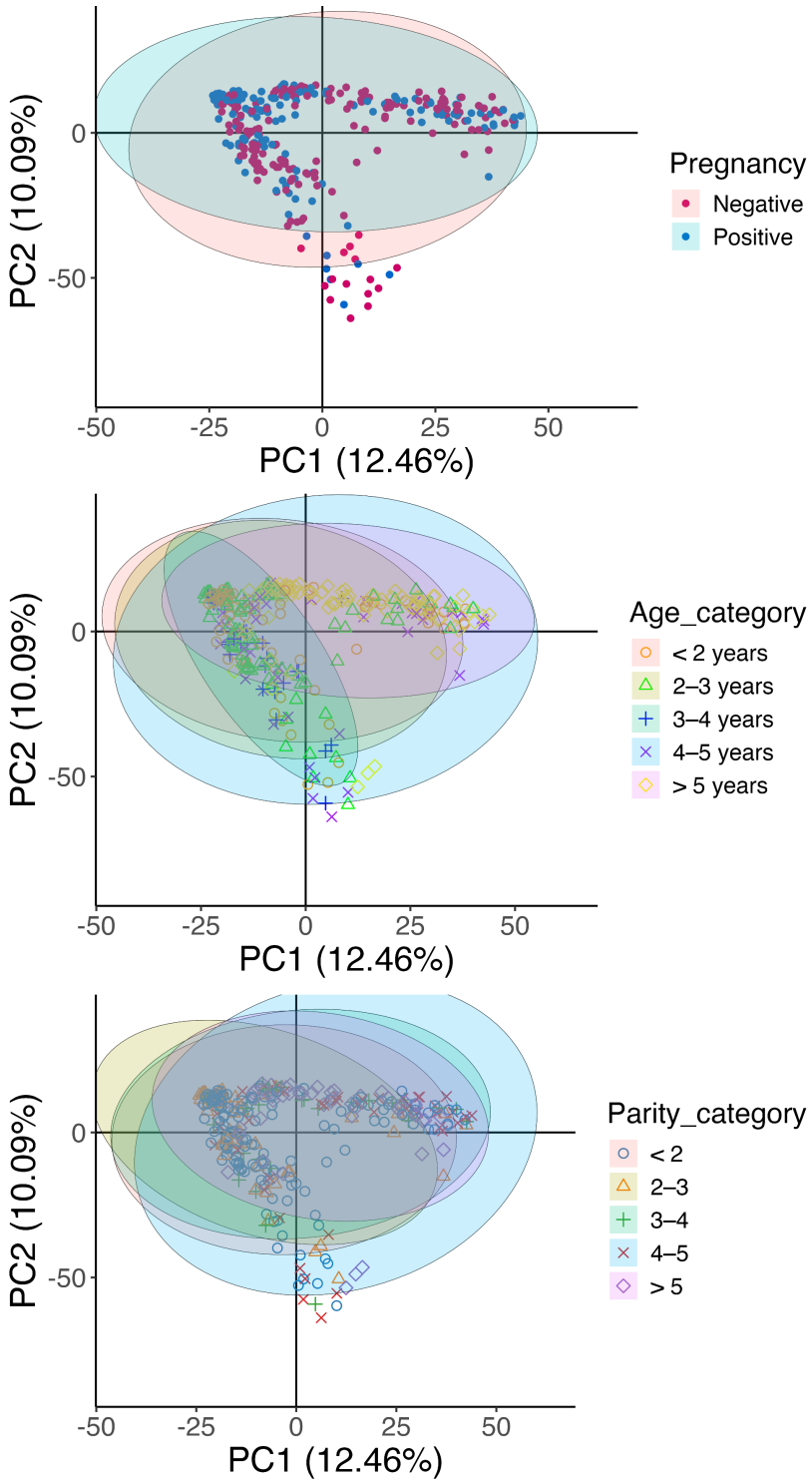


Figure S1.6. Principal component analysis for pregnancy, age and parity categorised given the microbiota composition at the ASV level using the 16S rRNA metabarcoding approach.

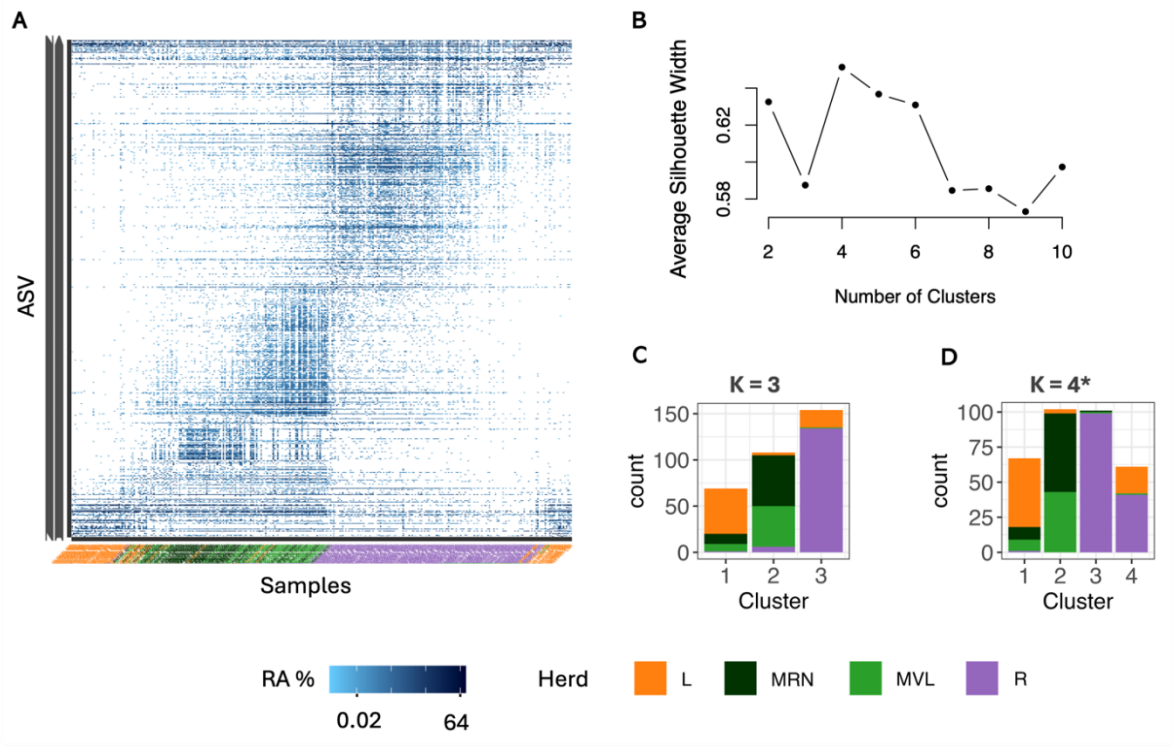


Figure S1.7. Analysis of sample similarities and clustering in bacterial community studies using the 16S rRNA metabarcoding approach. (**A**) Heatmap showing the RA of microbial community composition, arranged according to samples similarity parameters. (**B**) Silhouette width plot for the different numbers of clusters. (**C**) and (**D**) Barplot representing the clusters assuming different number of K. Each bar represents the number of samples within each cluster, colour-coded by sample type: L (Latxa), MRN (Manchega RN), MVL (Manchega VL), and R (Rasa). The asterisk (*) highlights the optimal K-value determined by the highest silhouette scores achieved in the analysis.

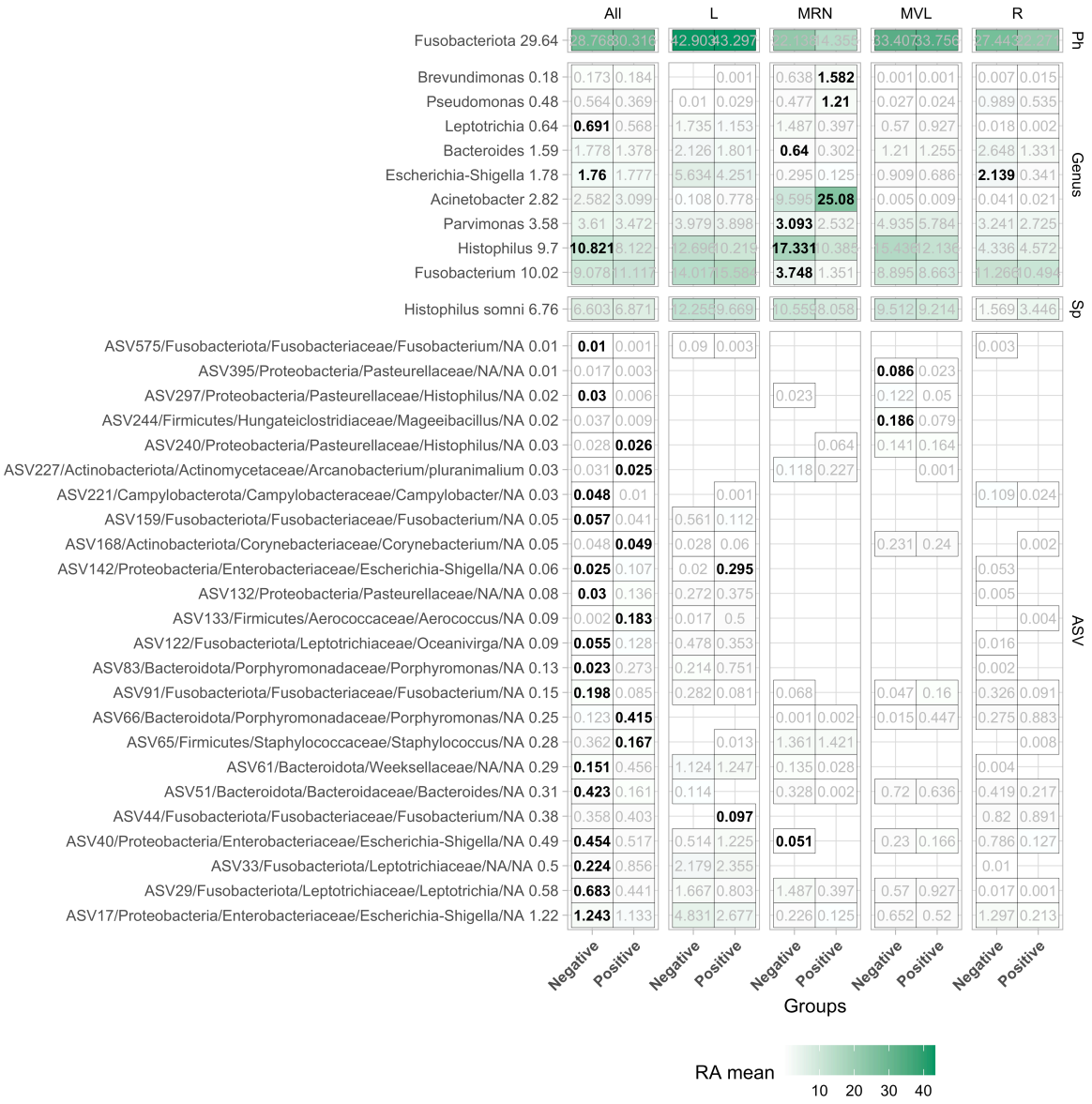


Figure S1.8. Heatmap representing the RA of taxa against group means (on a 100-point scale) from the differential abundance analysis for pregnancy, for the global model (ALL) and within each herd using the 16S rRNA metabarcoding approach. Taxa levels include ASV (ASV/code/Phylum/Family/Genus/Species RA), Species, Genus, and Phylum. In bold, significantly differential abundant taxa between pregnant (positive) and non-pregnant (negative) ewes.



Figure S1.9. Prevalence of pregnancy by herd groups using the 16S rRNA metabarcoding approach. Heatmap representation of prevalence of taxa normalised by group on a 100-point scale. Taxa levels include ASV (ASV/code/Phylum/Family/Genus/Species RA), Species, Genus, and Phylum, with significantly abundant pregnancy groups in differential analysis denoted in bold.

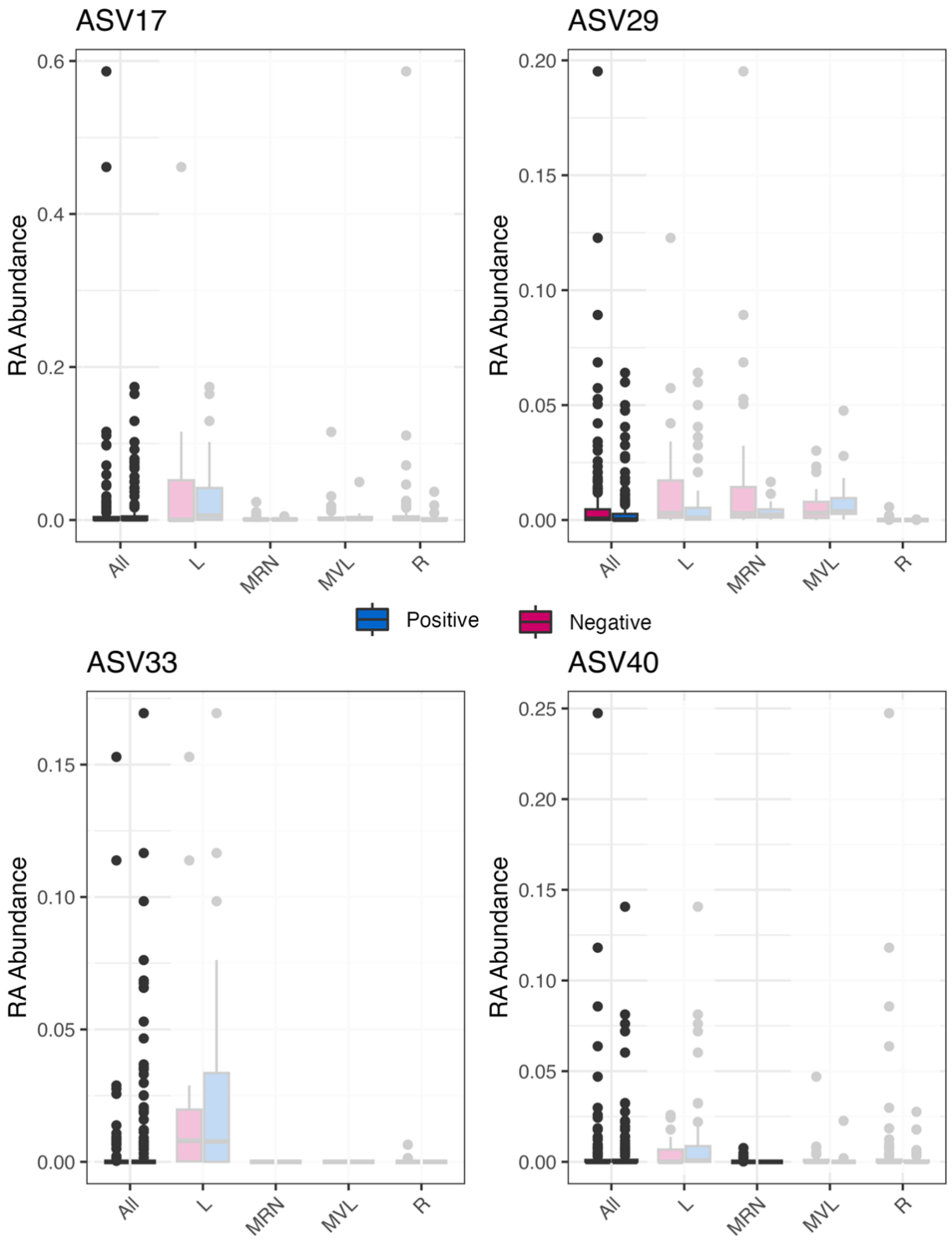


Figure S1.10. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for ASV17, ASV29, ASV33, ASV40 using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

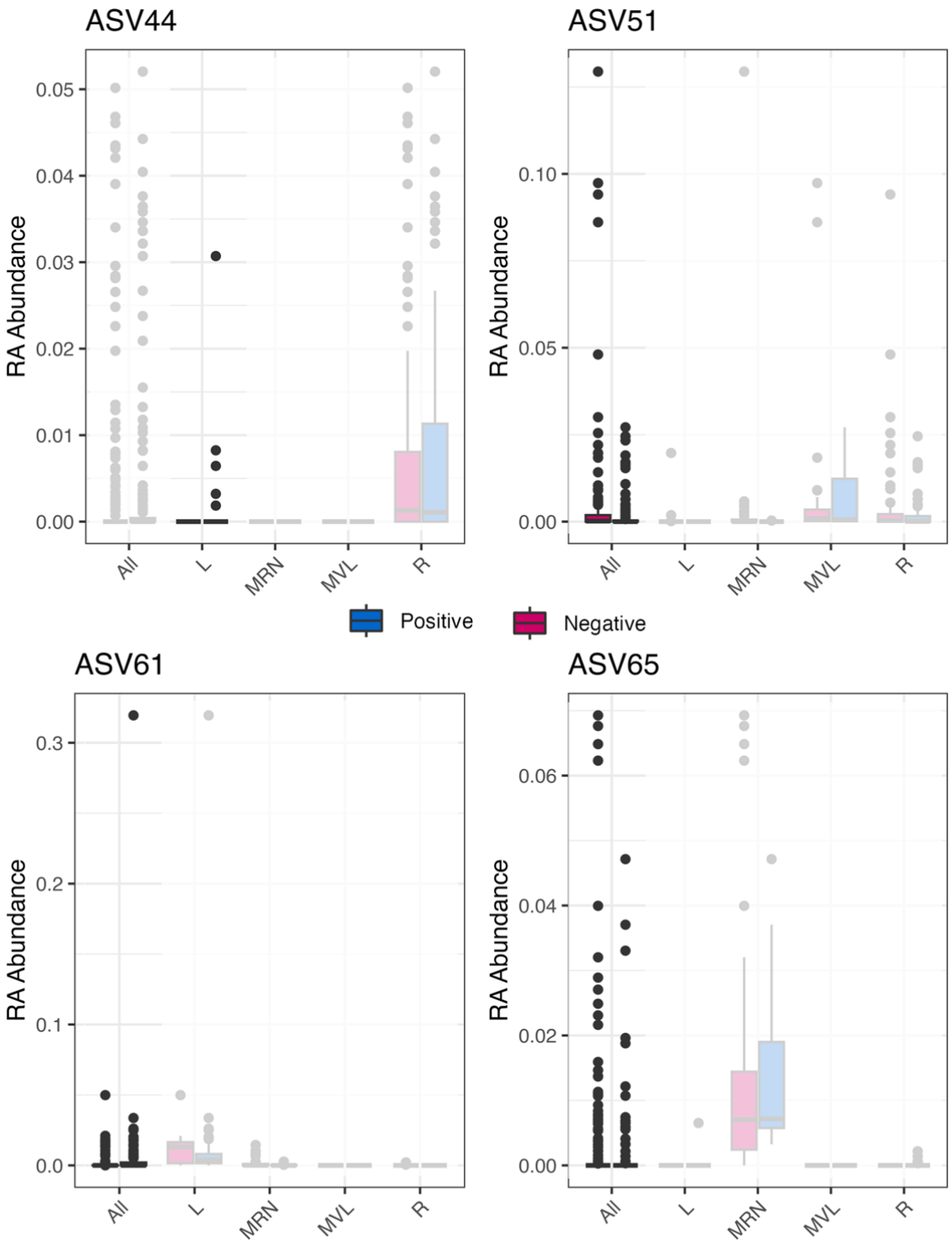


Figure S1.11. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for ASV44, ASV51, ASV61, ASV65 using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

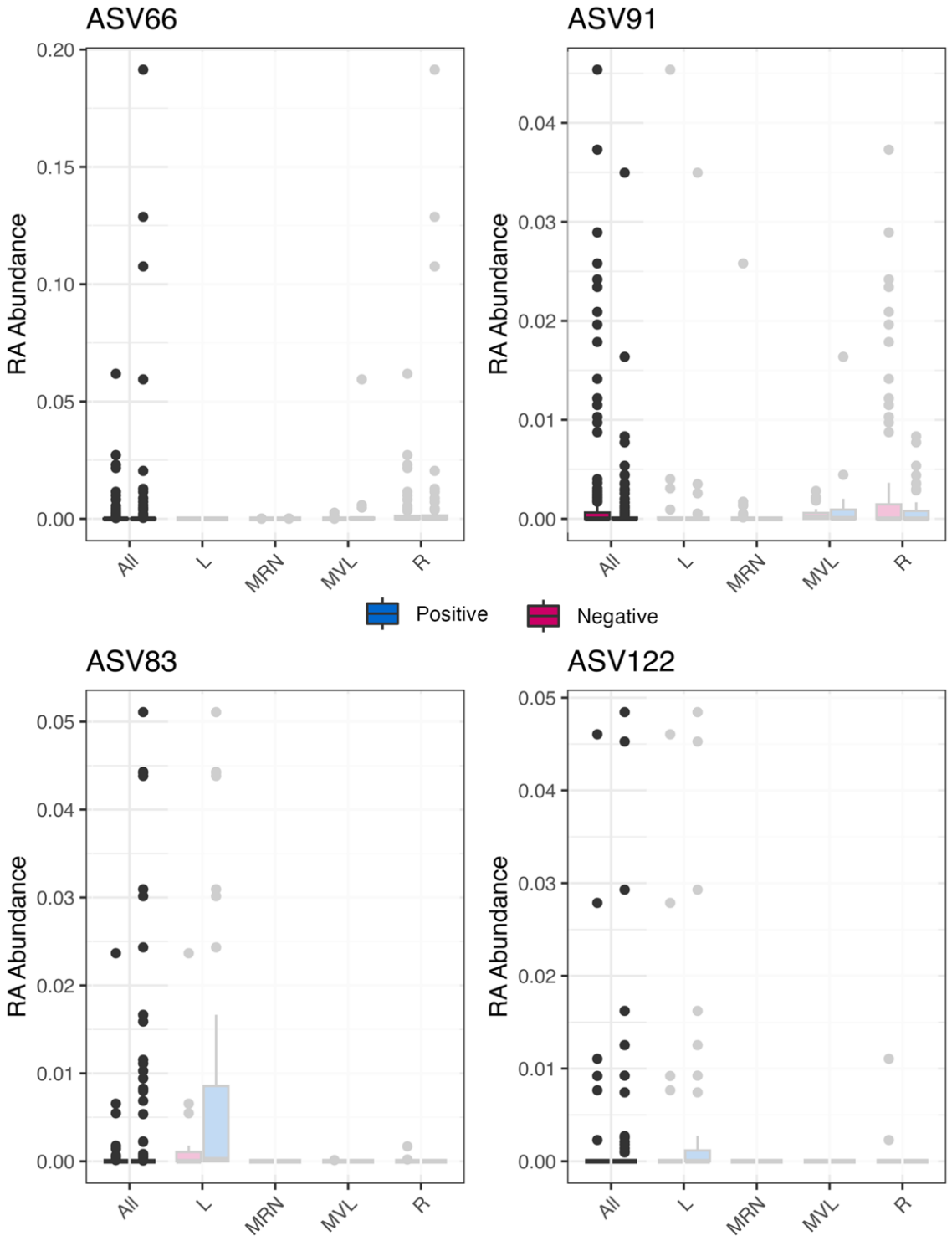


Figure S1.12. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for ASV66, ASV91, ASV83, ASV122 using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

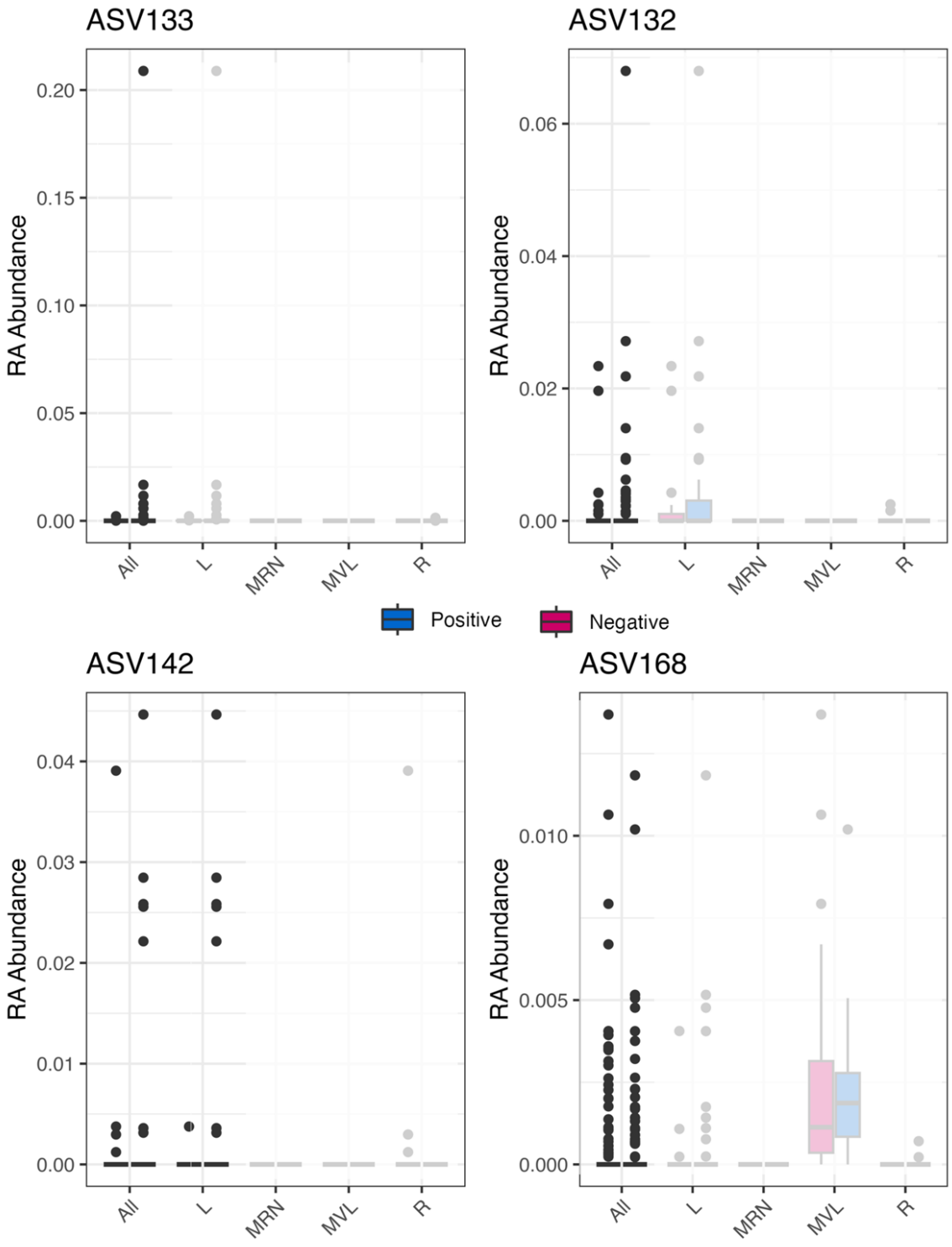


Figure S1.13. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for ASV133, ASV132, ASV142, ASV168 using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

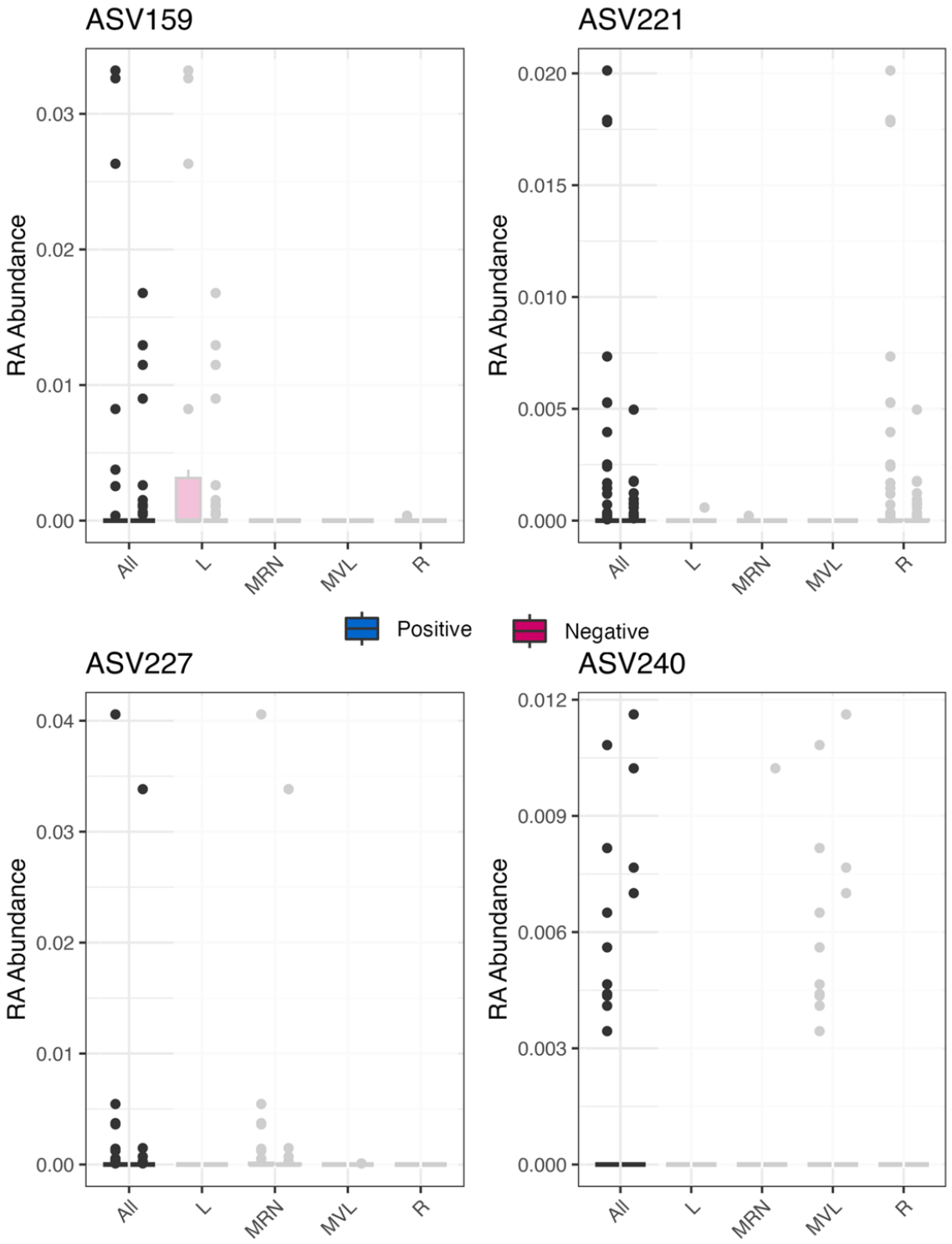


Figure S1.14. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for ASV159, ASV221, ASV227, ASV240 using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

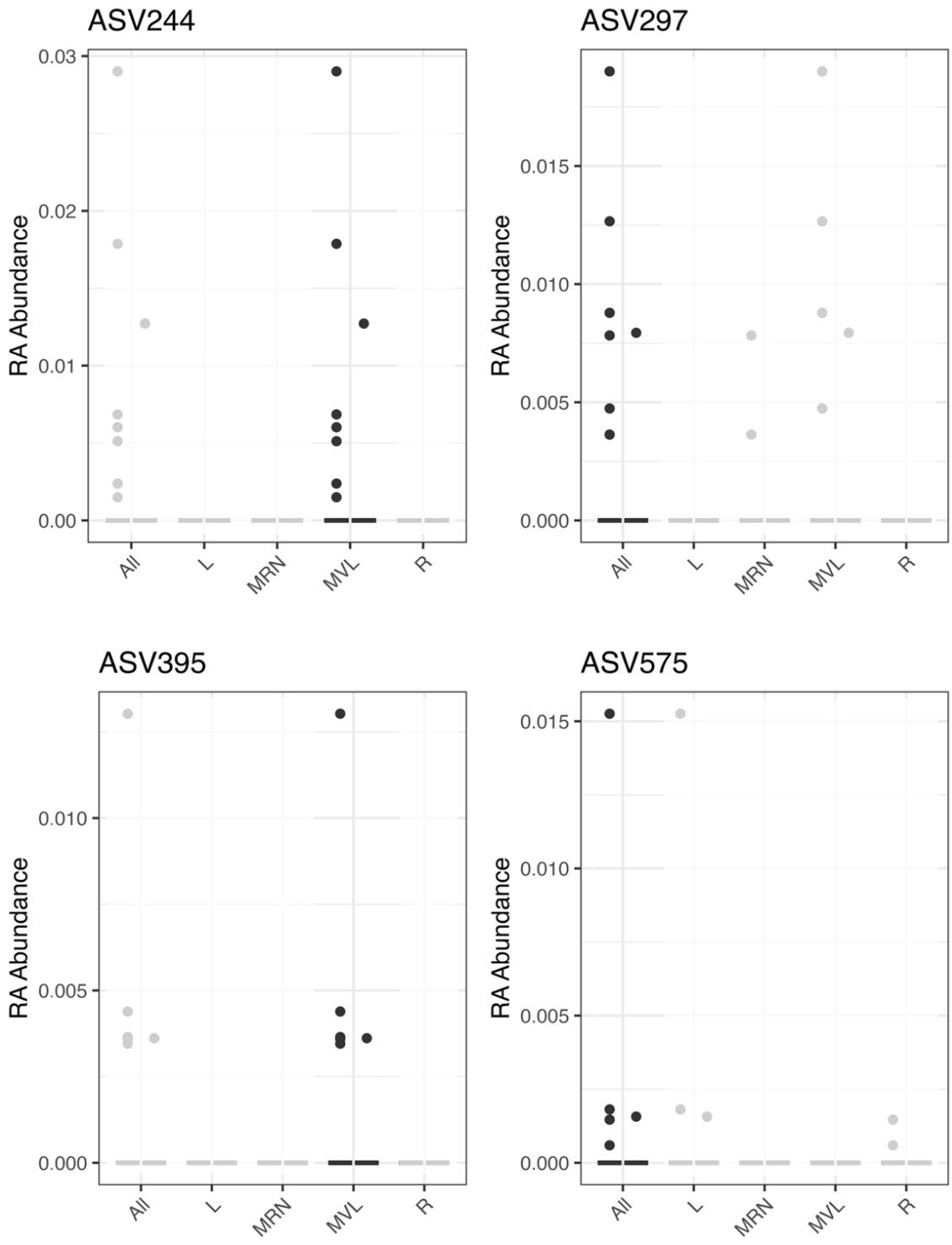


Figure S1.15. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for ASV244, ASV297, ASV395, ASV575 using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

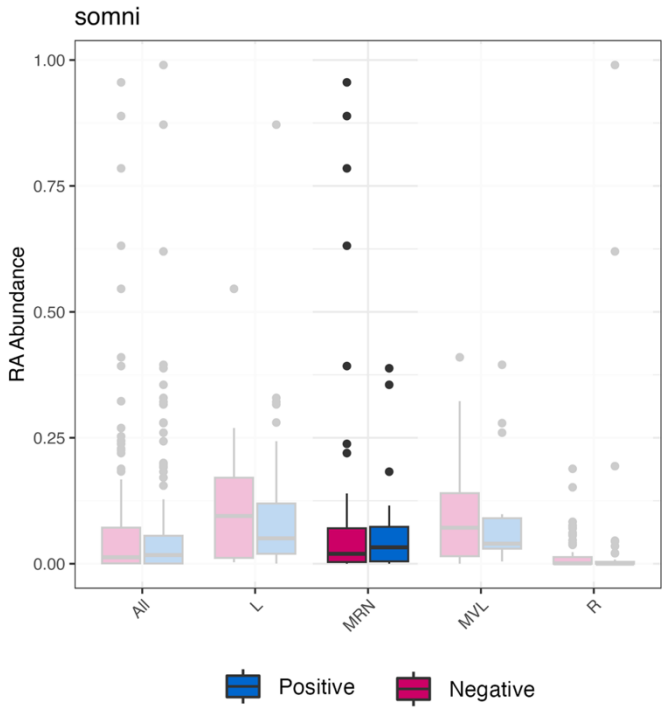


Figure S1.16. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for the species *Histophilus somni* using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

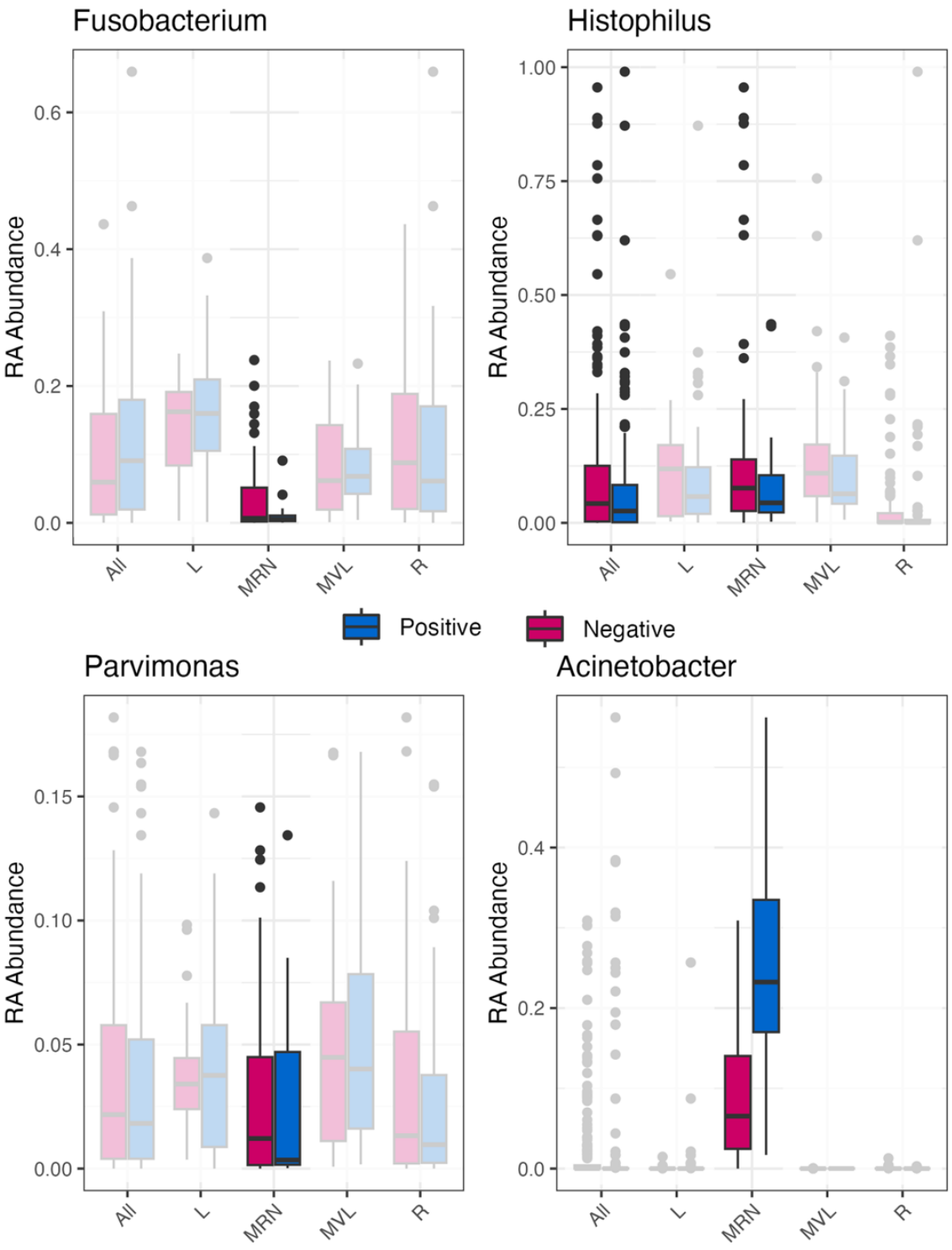


Figure S1.17. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for the genus *Fusobacterium, Histophilus, Parvimonas,* and *Acinetobacter* using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

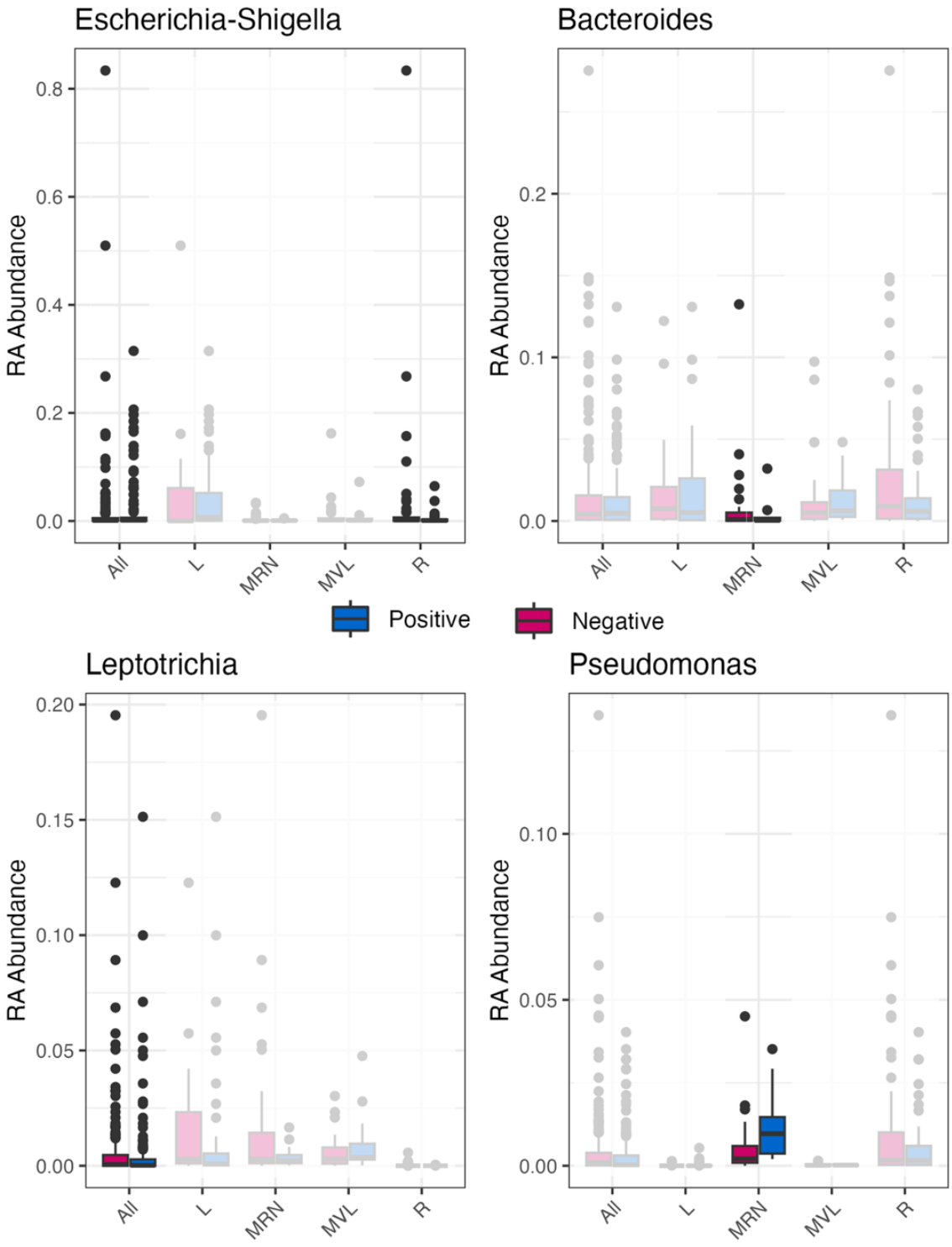


Figure S1.18. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for the genera *Escherichia-Shigella, Bacteroides, Leptotrichia,* and *Pseudomonas* using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

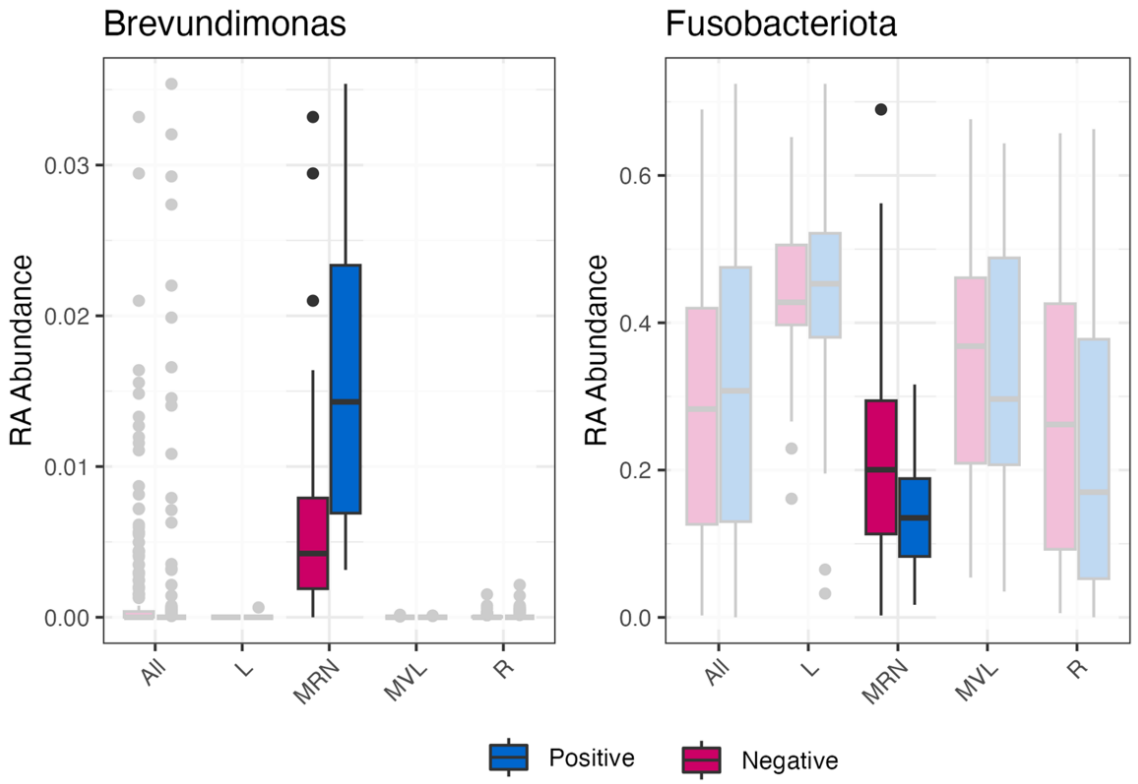


Figure S1.19. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for the genera *Brevundimonas* in the left and the phylum Fusobacteriota in the right using the 16S rRNA metabarcoding approach. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

Table S1.1. Pairwise comparisons for Herd groups using permutation MANOVAs on a distance matrix at ASV level

	L	MRN	MVL
MRN	**0.001**	-	-
MVL	**0.001**	**0.001**	-
R	**0.001**	**0.001**	**0.001**

Pairwise comparisons were performed based on Euclidean distance matrix at the ASV level. 999 permutations were implemented, and p-values were adjusted using the FDR method. Values in bold indicate significant results (adjusted P-value < 0.05).

Table S1.2. Pairwise comparisons for Age category groups using permutation MANOVAs on a distance matrix at ASV level

	< 2 years	2-3 years	3-4 years	4-5 years
2-3 years	0.2760	-	-	-
3-4 years	**0.0043**	0.0578	-	-
4-5 years	**0.0025**	**0.0050**	**0.0033**	-
> 5 years	**0.0025**	**0.0025**	**0.0025**	**0. 0033**

Pairwise comparisons were performed based on Euclidean distance matrix at the genus ASV. 999 permutations were implemented, and p-values were adjusted using the FDR method. Values in bold indicate significant results (adjusted P-value < 0.05).

Table S1.3. Pairwise comparisons for Parity category groups using permutation MANOVAs on a distance matrix at ASV level

	< 2	2-3	3-4	4-5
2-3	0.0643	-	-	-
3-4	0.1720	0.0643	-	-
4-5	**0.0025**	**0.0025**	0. 0737	-
> 5	**0.0025**	**0.0025**	0. 0967	0. 0643

Categories represent the number of parturitions. Pairwise comparisons were performed based on Euclidean distance matrix at the genus ASV. 999 permutations were implemented, and p-values were adjusted using the FDR method. Values in bold indicate significant results (adjusted P-value < 0.05).

Appendix for Study 2: characterisation of vaginal microbiota and microbial genes associated with pregnancy outcomes using metagenomics

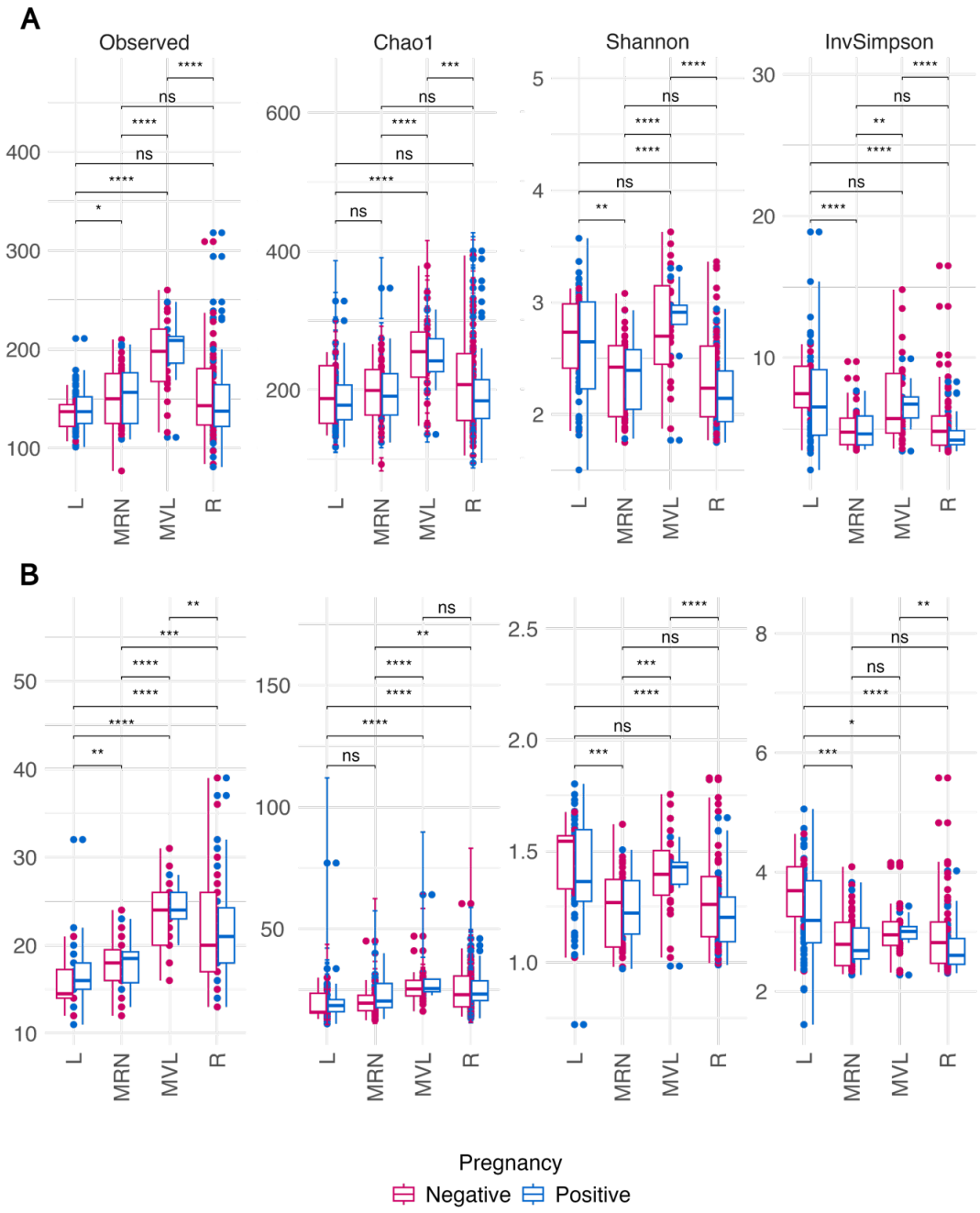


Figure S2.1. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for Pregnancy and herd at the Genus (**A**) and Phylum (**B**) level. Data were rarefied. Statistical significance is indicated by asterisks: * ($p < 0.05$), *** ($p < 0.001$). ns: not significant.

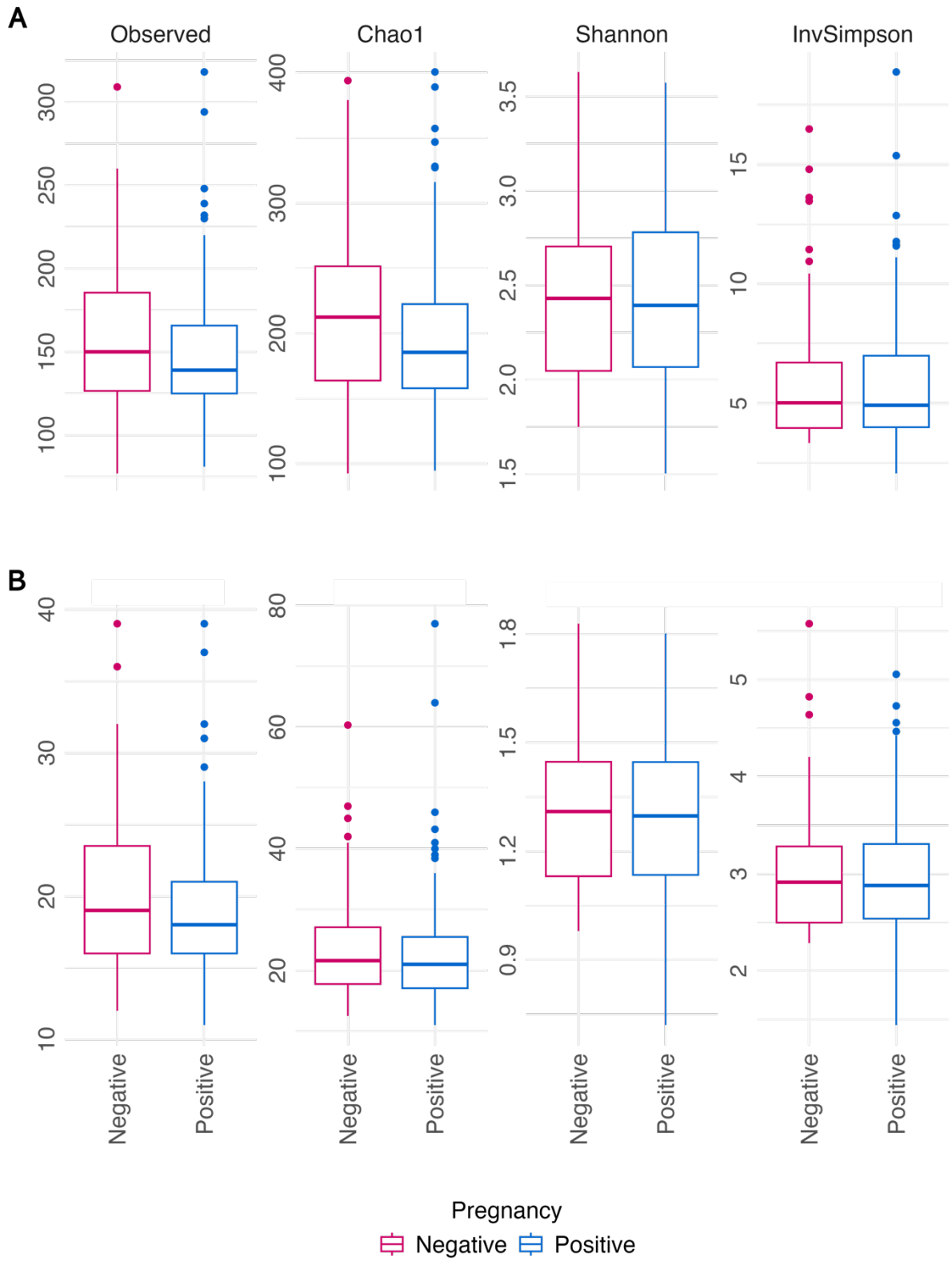


Figure S2.2. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for Pregnancy at the Genus (**A**) and Phylum (**B**) level. There were no significant differences between groups.

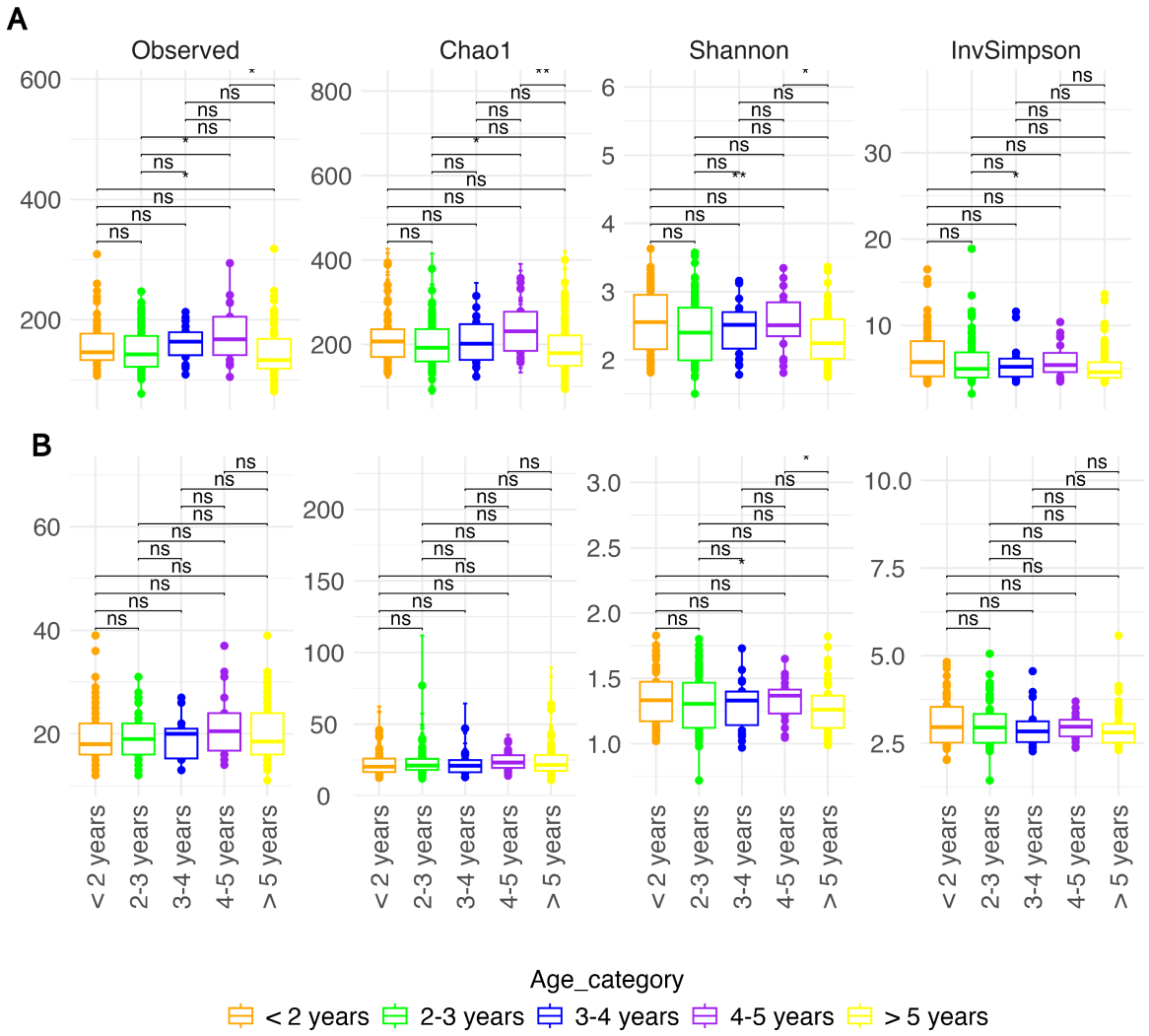


Figure S2.3. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for age category at the genus (**A**) and phylum (**B**) level. Data were rarefied. Statistical significance is indicated by asterisks: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). ns: not significant.

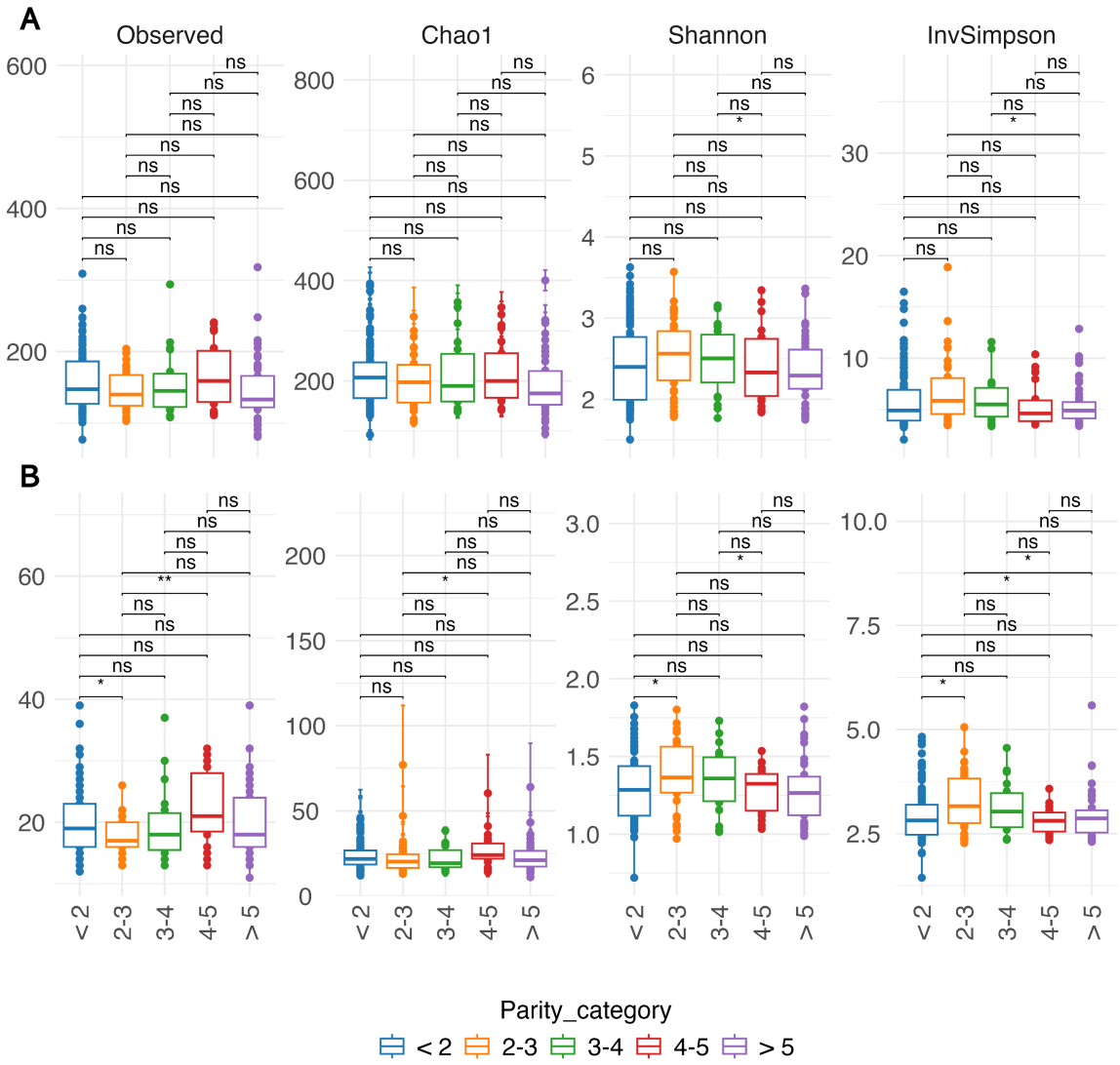


Figure S2.4. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for pariry category at the genus (**A**) and phylum (**B**) level. Data were rarefied. Statistical significance is indicated by asterisks: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). ns: not significant.

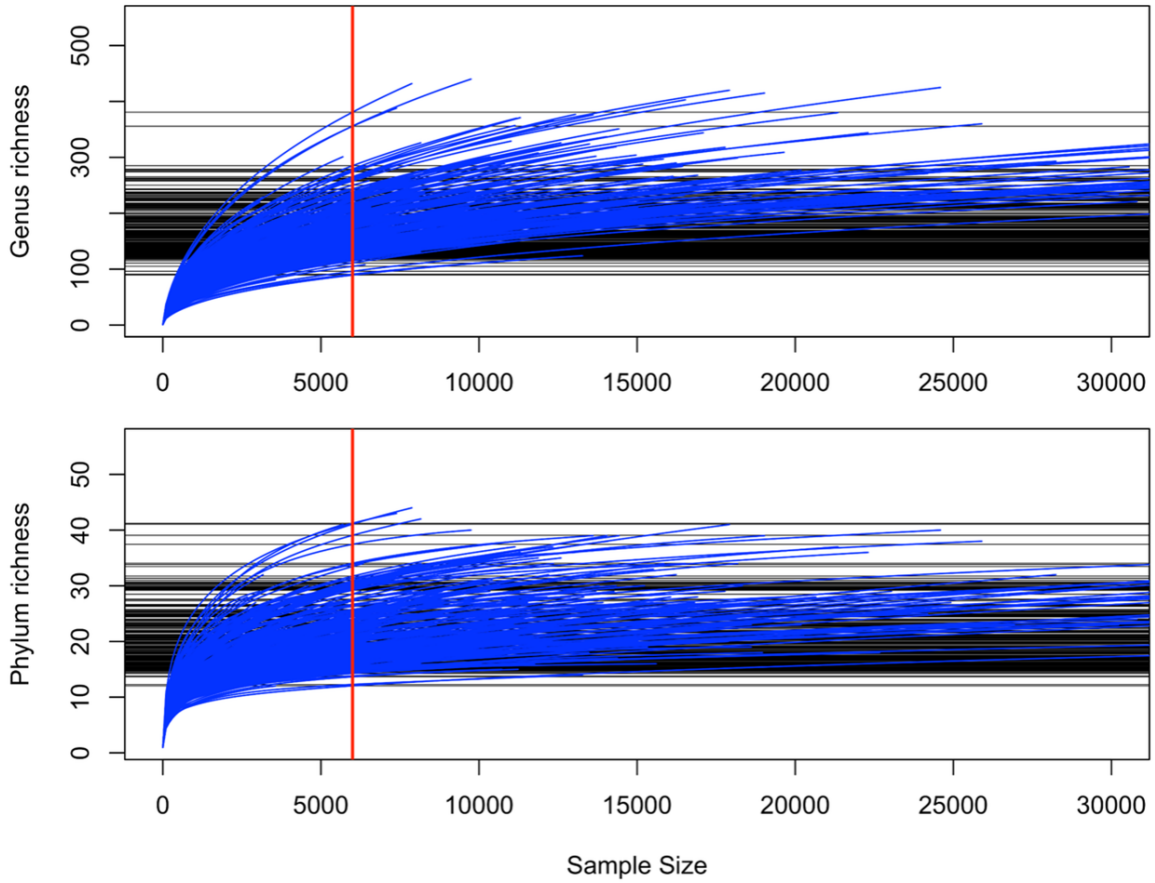


Figure S2.5. Rarefaction curves for all analysed samples for alpha-diversity analysis. Sequencing depth in the x-axis and observed taxa richness in the y-axis for phylum and genus level. Blue lines represent individual samples. Red vertical dashed line indicates the threshold value (6000). Black horizontal lines indicate the total sequencing depth for each sample.

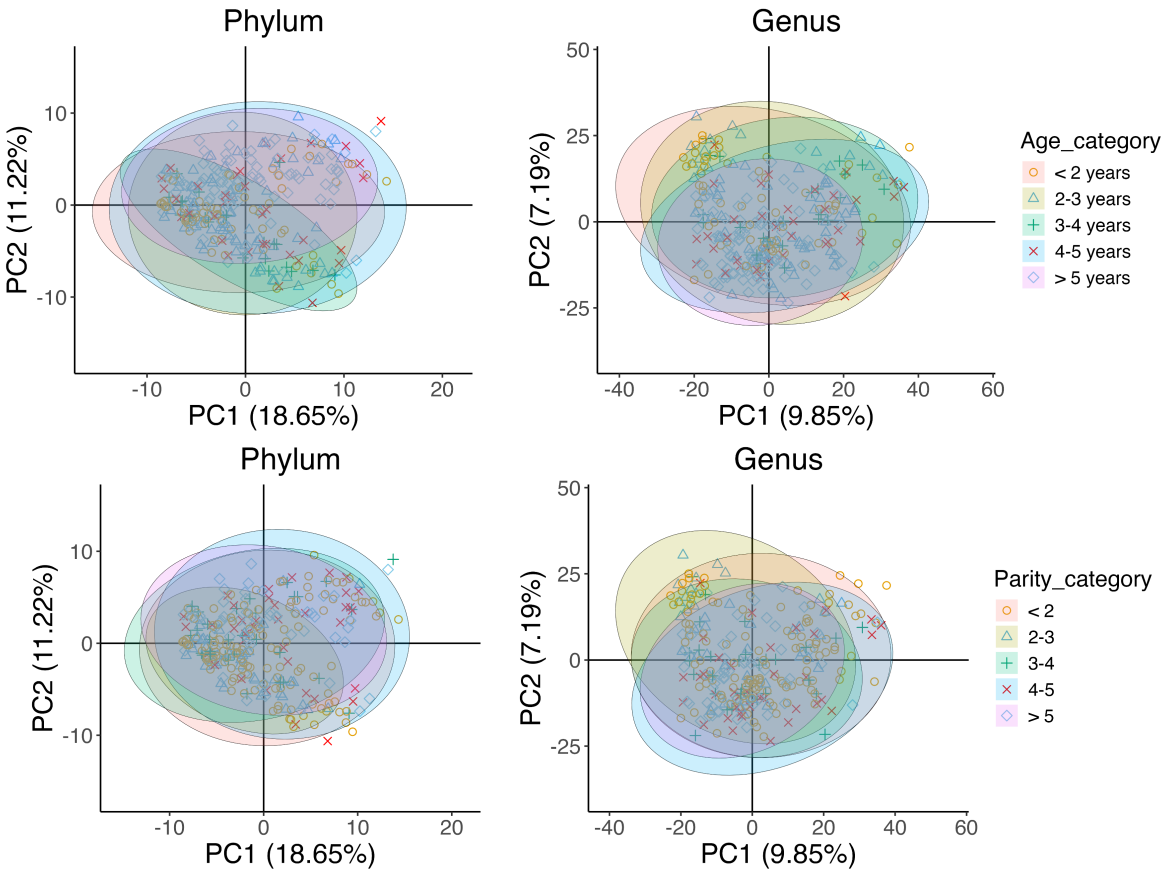


Figure S2.6. Principal component analysis of microbiota composition by Age and Parity categories. Phylum (a, c) and genus (b, d) levels were analysed. Ellipses were calculated using covariance to represent data variability within each age/parity category.

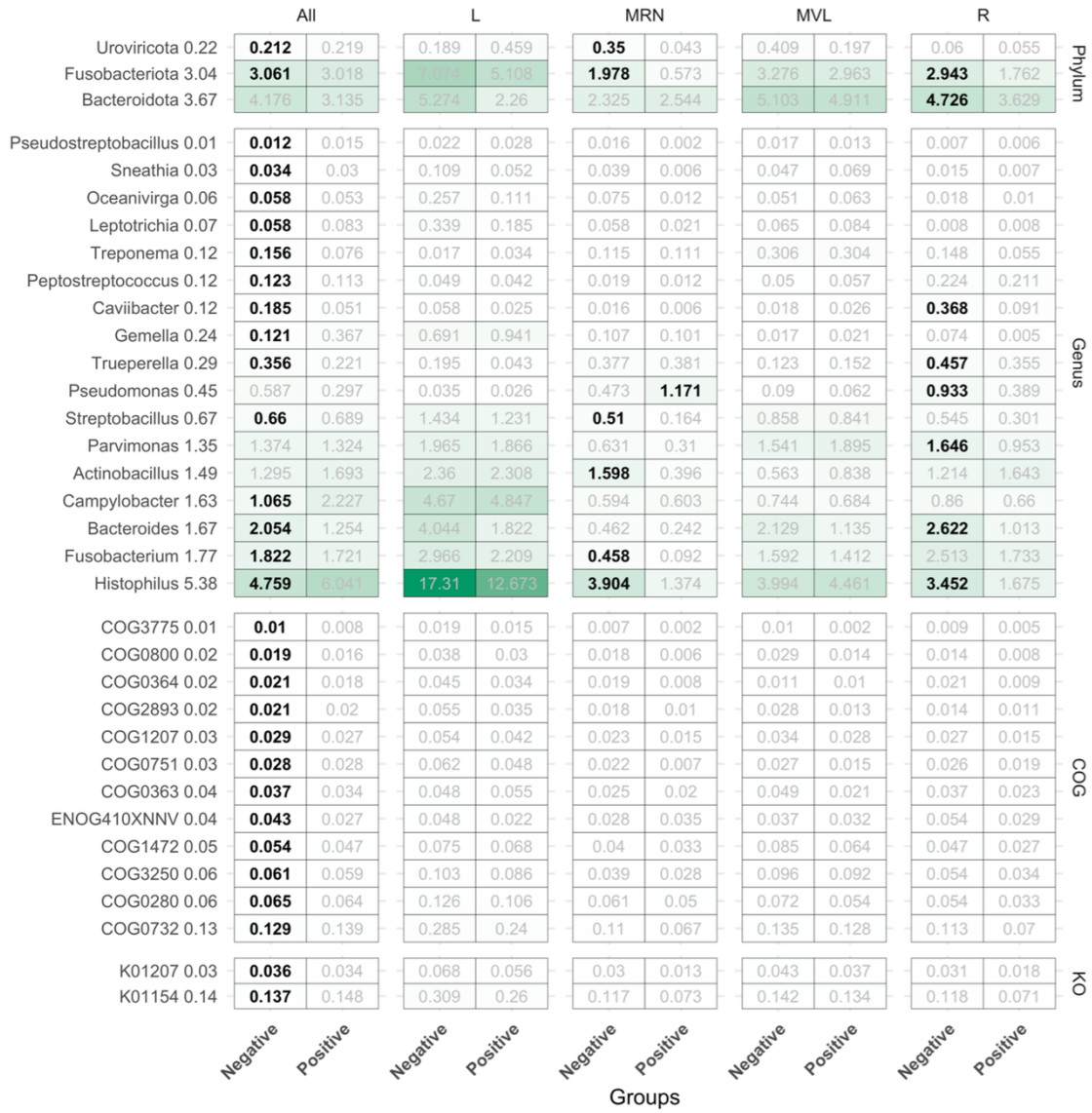


Figure S2.7. Heatmap representing the RA of taxa against group means (on a 100-point scale) from the differential abundance analysis for pregnancy, for the global model (ALL) and within each herd. Taxa levels include Genus, Phylum, COG, and KO. In bold, significantly differential abundant taxa between pregnant (positive) and non-pregnant (negative) ewes.



Figure S2.8. Prevalence of pregnancy by herd groups. Heatmap representation of prevalence of taxa normalised by group on a 100-point scale. Taxa levels include Genus, Phylum, COG and KO with significantly abundant pregnancy groups in differential analysis denoted in bold.

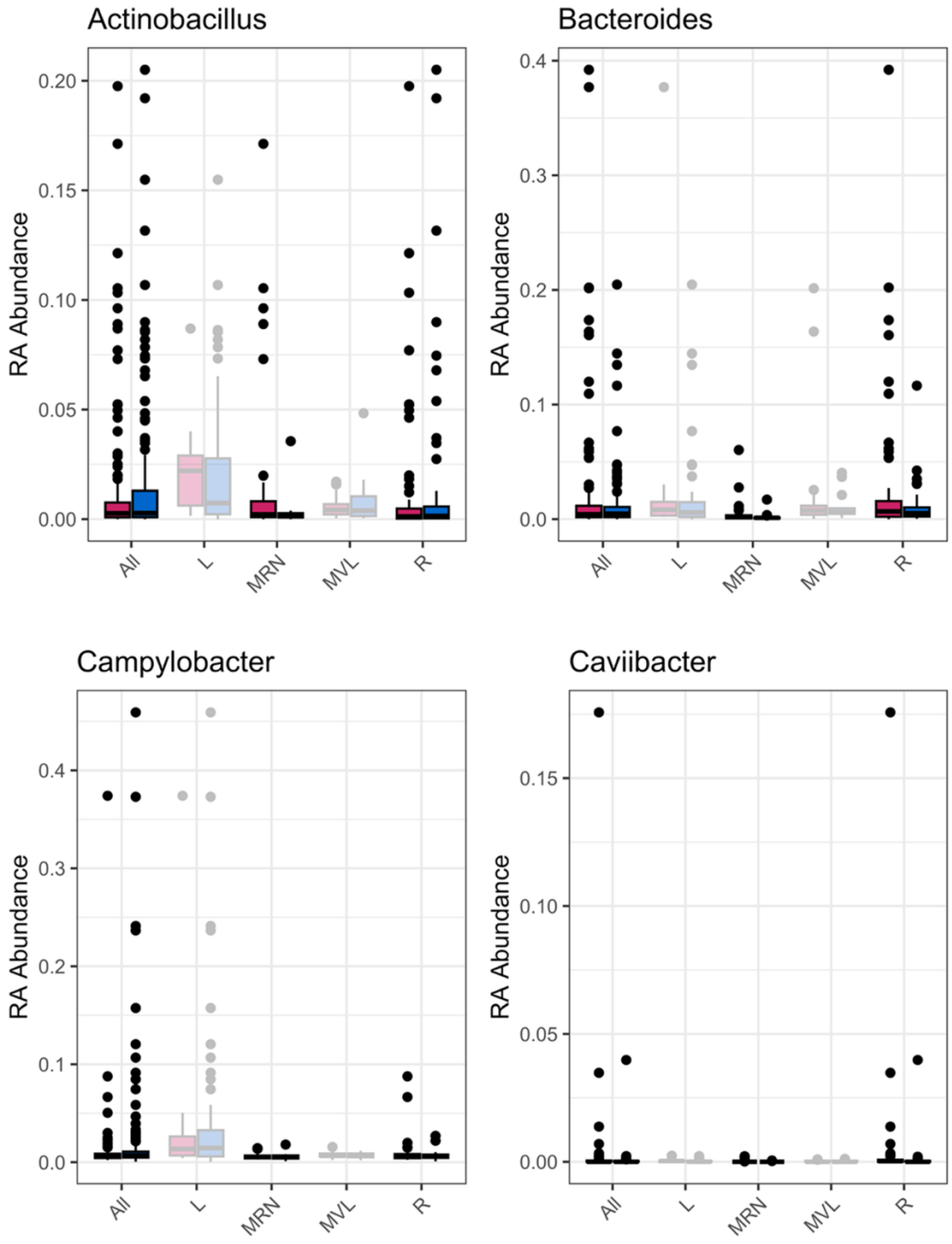


Figure S2.9. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for *Actinobacillus*, *Bacteroides*, *Campylobacter*, and *Caviibacter* genera. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

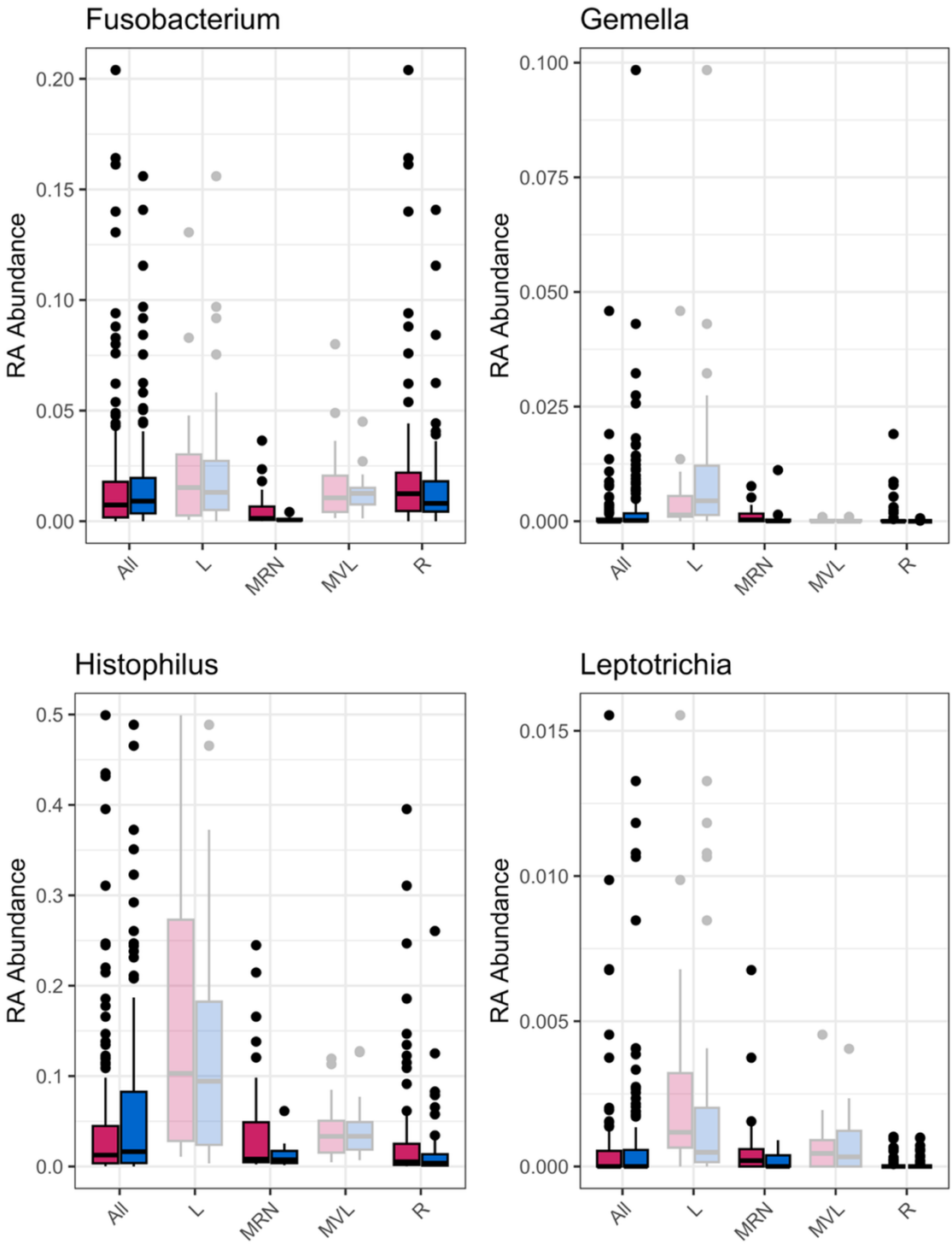


Figure S2.10. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for *Fusobacterium*, *Gemella*, *Histophilus*, and *Leptotrichia* genera. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

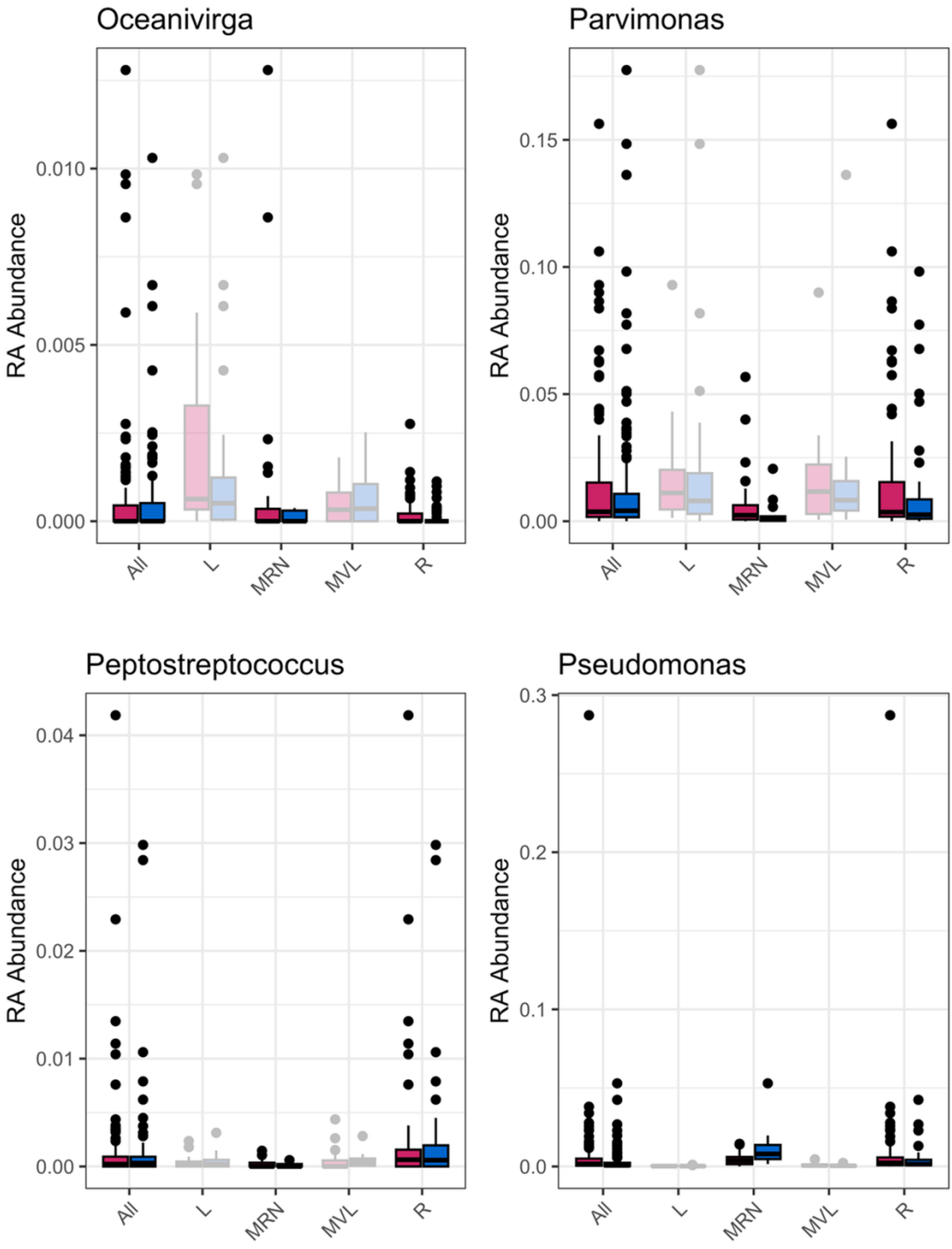


Figure S2.11. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for *Oceanivirga*, *Parvimonas*, *Peptostreptococcus*, and *Pseudomonas* genera. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

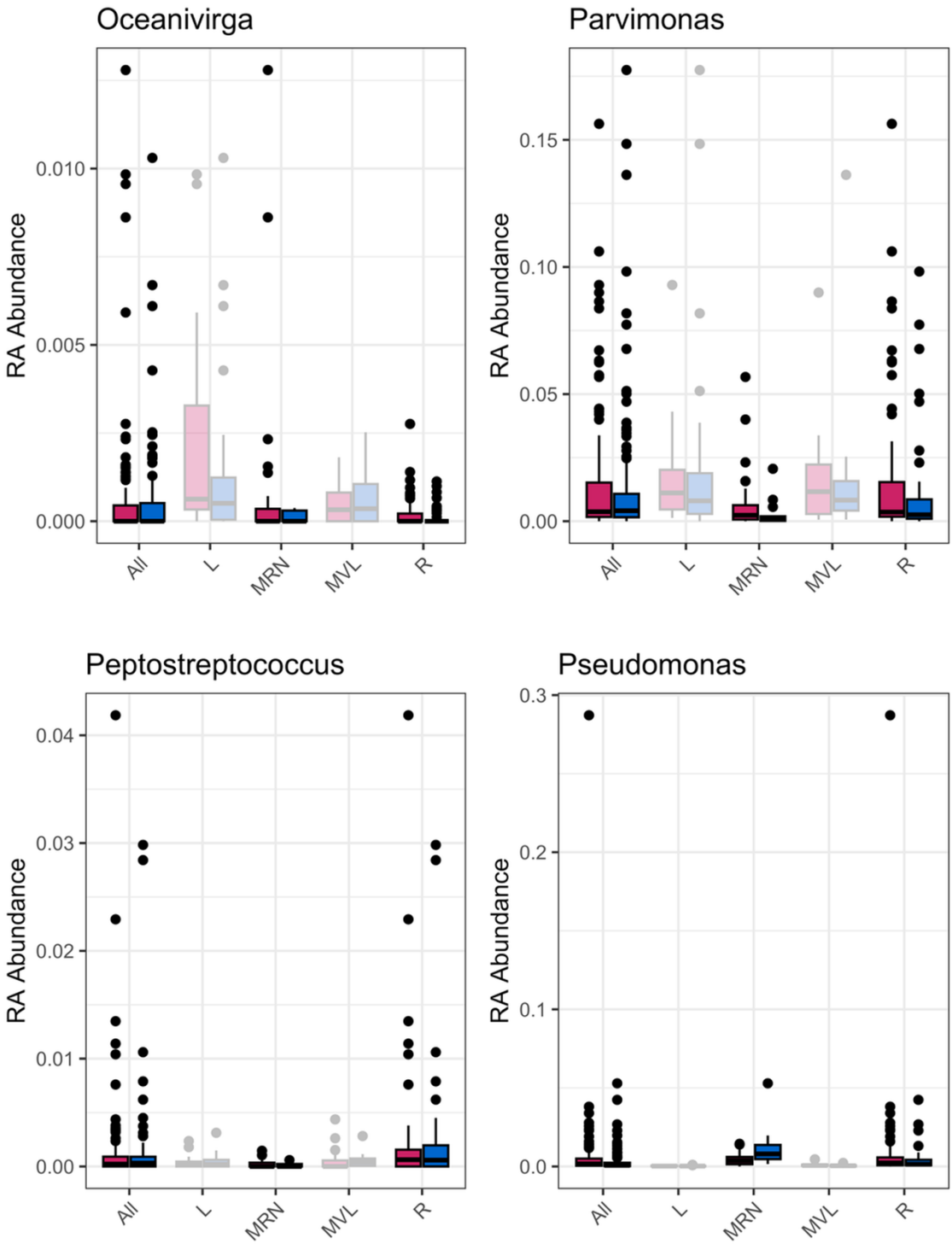


Figure S2.12. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for *Pseudostreptobacillus*, *Sneathia*, *Streptobacillus*, and *Treponema* genera. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

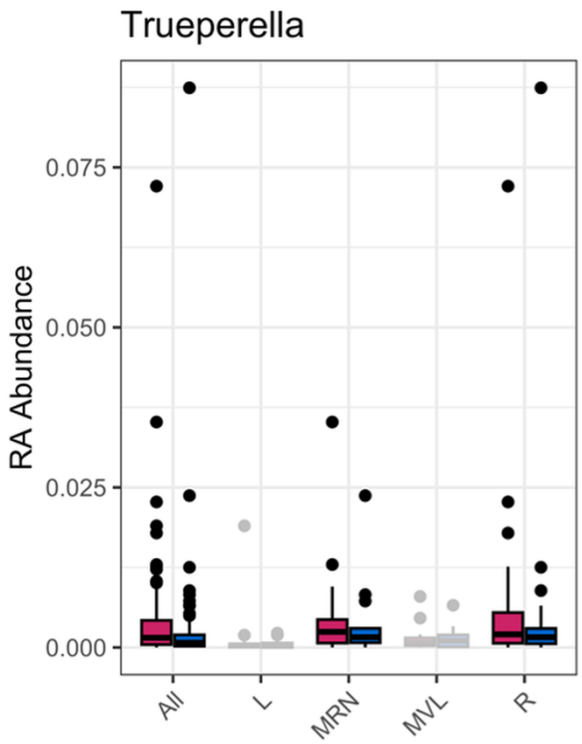


Figure S2.13. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for *Trueperella* Genus. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

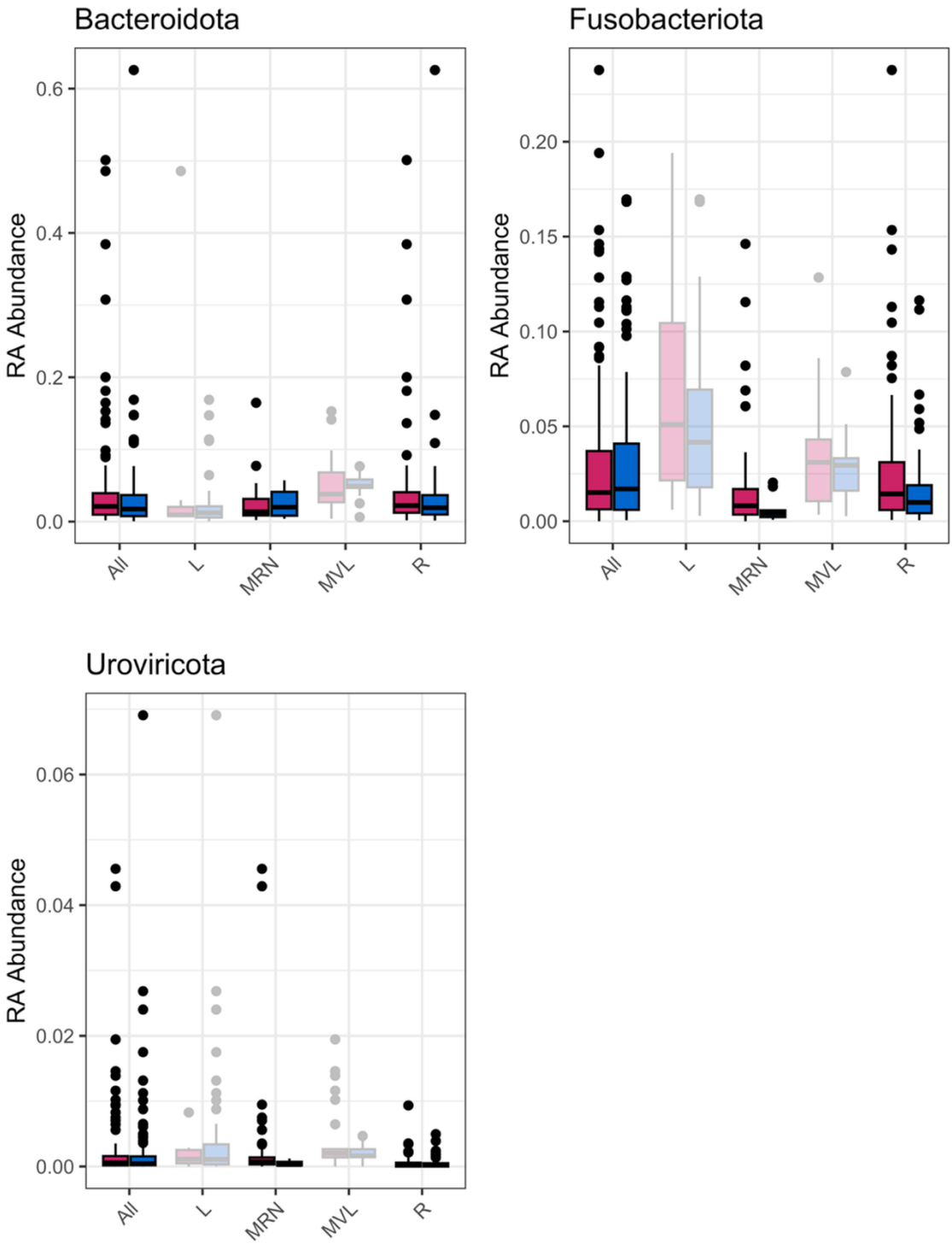


Figure S2.14. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for Bacteroidota, Fusobacteriota, and Uroviricota phyla. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

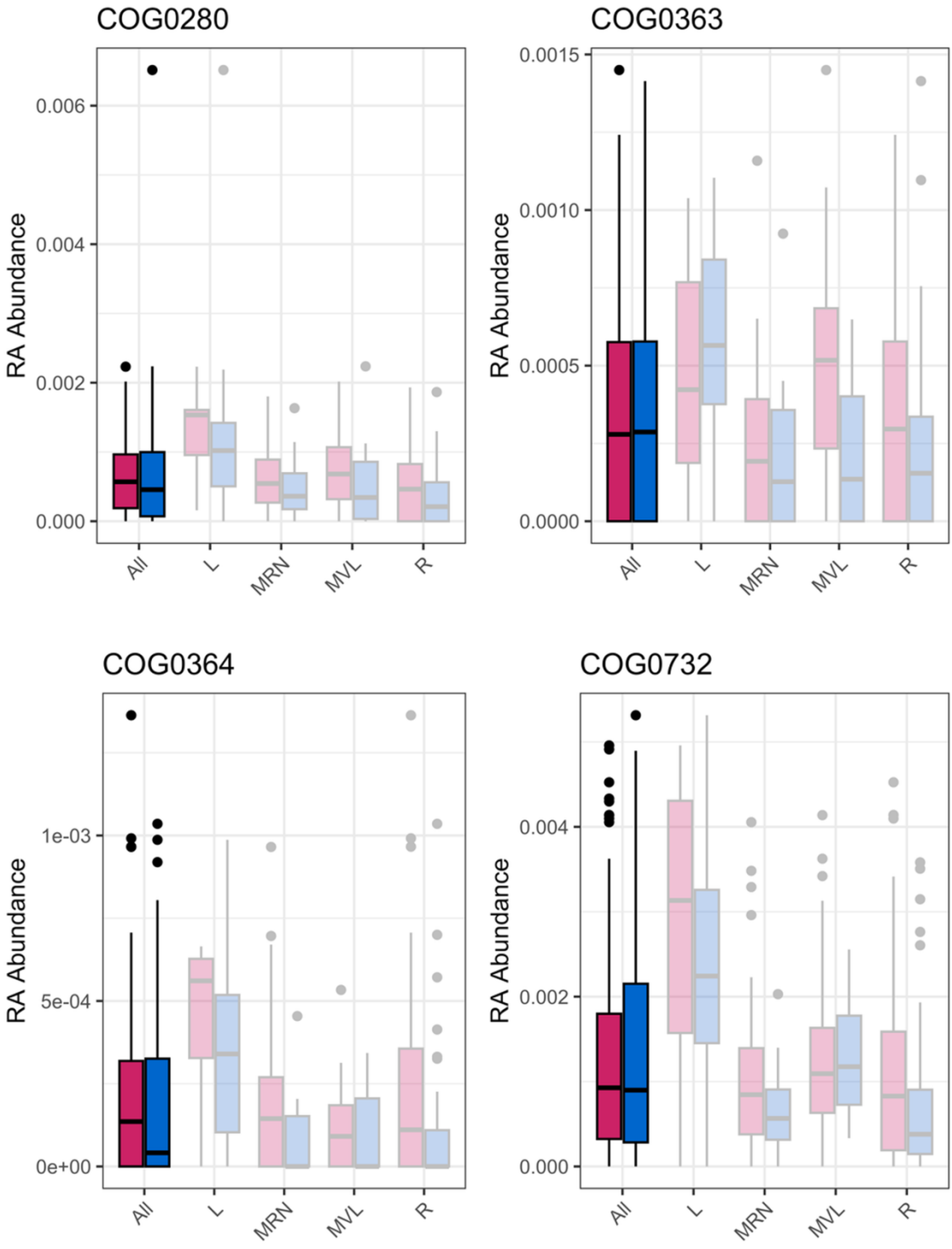


Figure S2.15. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for COG0280, COG0363, COG0364, and COG0732. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

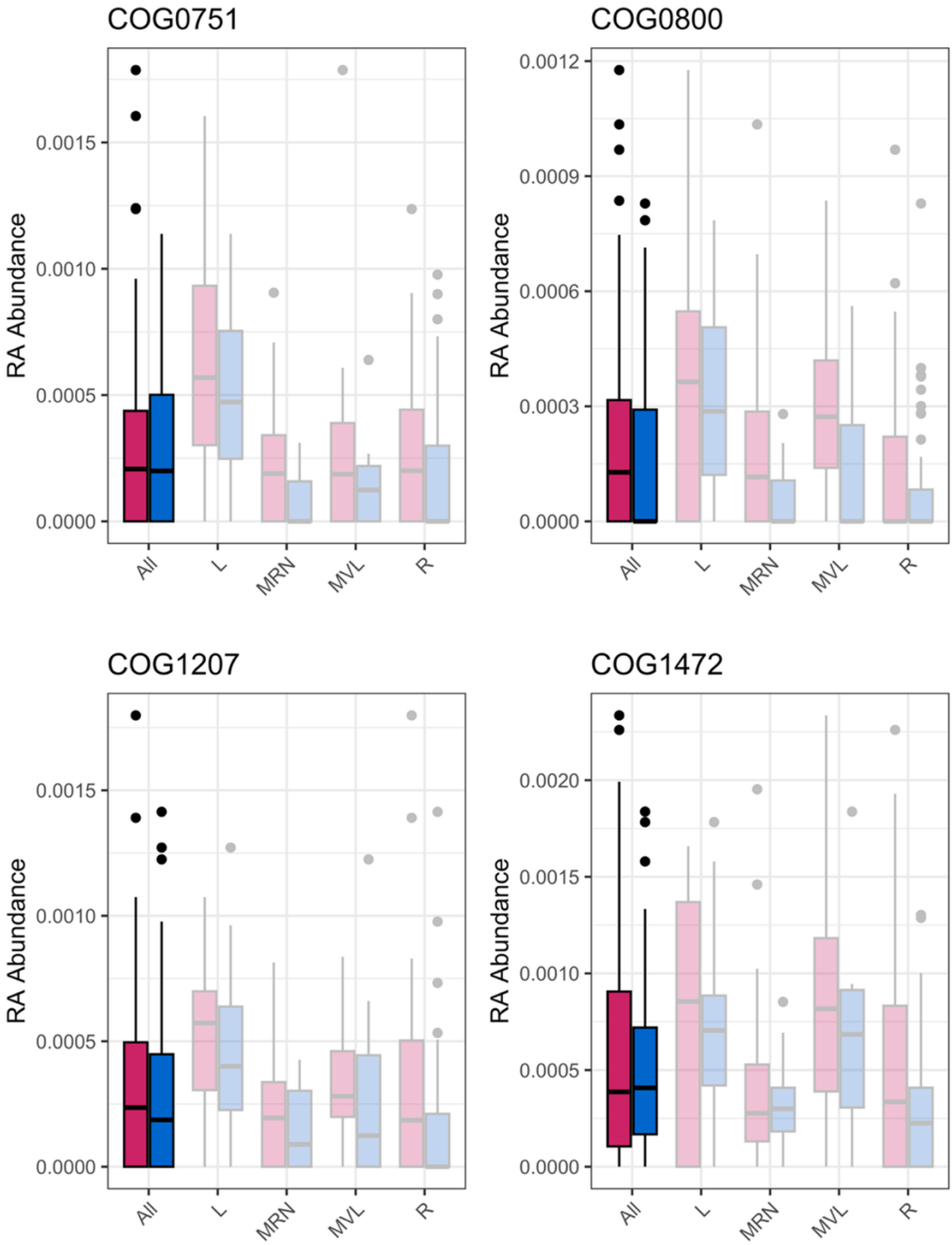


Figure S2.16. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for COG0751, COG0800, COG1207, and COG1472. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

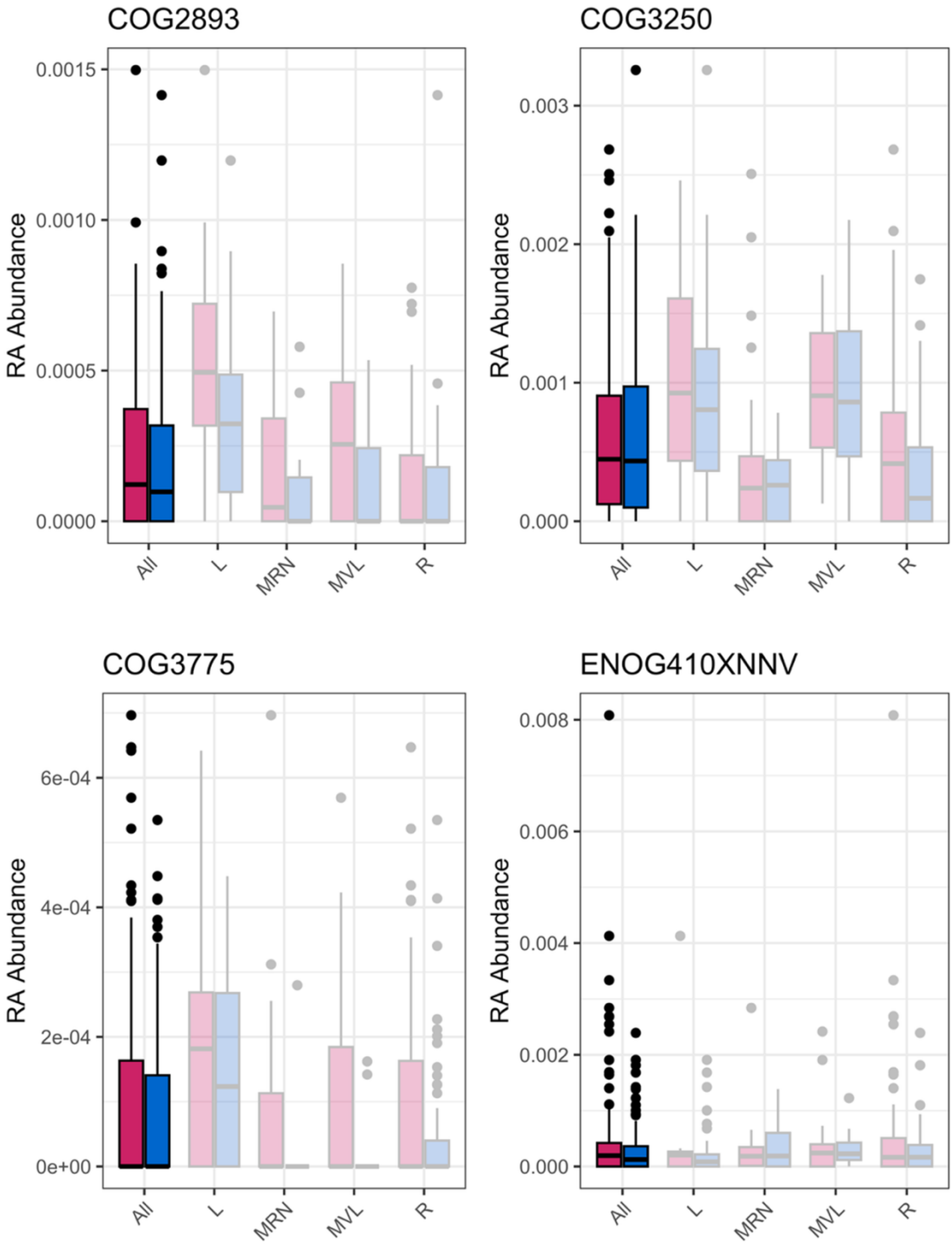


Figure S2.17. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for COG2893, COG3250, COG3775, and ENOG410XNNV. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

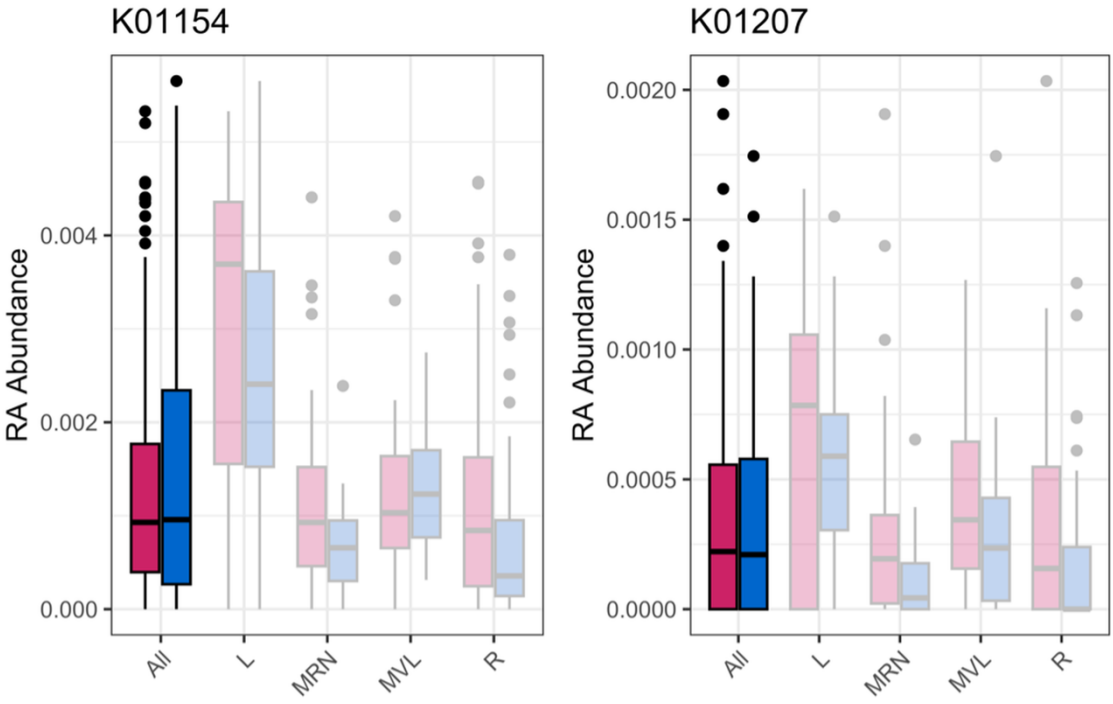


Figure S2.18. Boxplots of taxa with significant differential abundance between pregnant and non-pregnant ewes for KO1154 and KO01207. Pregnancy-positive samples in blue and pregnancy-negative samples in red. The x-axis categorizes samples by herd level. Saturated colours represent groups with statistically significant differences; desaturated colours are used for groups without significant differences.

Table S2.1. PERMANOVA results for the effects of Pregnancy, age category, and parity category on microbiota composition within each herd group at genus level

Variable		SumOfSqs	R2	F	Pr(>F)
Pregnancy		4,394.1	0.010	3.057	**0.001**
	L	643.90	0.010	0.647	0.997
	R	1,164.61	0.007	0.896	0.736
	MRN	1,260.31	0.020	1.074	0.247
	MVL	1,536.83	0.026	1.043	0.285
Age category		1,8897.07	0.031	2.352	**0.001**
	L	4,665.00	0.071	1.160	0.084
	MRN	2,381.40	0.080	1.109	0.170
	MVL	1,863.00	0.099	0.985	0.510
	R	10,027.00	0.038	1.282	**0.035**
Parity category		16,006.77	0.026	1.983	**0.001**
	L	4,779.00	0.072	1.191	0.062
	MRN	1,781.20	0.060	1.104	0.205
	MVL	1,800.50	0.095	0.949	0.666
	R	9,622.00	0.037	1.228	**0.011**

999 permutations were implemented. **SumOfSqs**: Sum of squares, a measure of variance. **R2**: Coefficient of determination, indicating the proportion of variance explained by the factor. **F**: F-statistic, a ratio used to determine the significance of the factor. **Pr(>F)**: P-value, indicating the statistical significance of the factor. Values in bold indicate significant results ($P < 0.05$).

Table S2.2. Pairwise comparisons for Herd groups using permutation MANOVAs on a distance matrix

	L	MRN	MVL
MRN	**0.001**	-	-
MVL	**0.001**	**0.001**	-
R	**0.001**	**0.001**	**0.001**

Pairwise comparisons were performed based on Euclidean distance matrix at the genus level. 999 permutations were implemented, and p-values were adjusted using the FDR method. Values in bold indicate significant results (adjusted P-value < 0.05).

Table S2.3. Pairwise comparisons for Age category groups using permutation MANOVAs on a distance matrix

	< 2 years	2-3 years	3-4 years	4-5 years
2-3 years	0.2810	-	-	-
3-4 years	**0.0357**	0.2456	-	-
4-5 years	**0.0250**	0.0900	**0.0250**	-
> 5 years	**0.0025**	**0.0025**	**0.0025**	**0.0025**

Pairwise comparisons were performed based on Euclidean distance matrix at the genus level. 999 permutations were implemented, and p-values were adjusted using the FDR method. Values in bold indicate significant results (adjusted P-value < 0.05).

Table S2.4. Pairwise comparisons for Parity category groups using permutation MANOVAs on a distance matrix

	< 2	**2-3**	**3-4**	**4-5**
2-3	**0.0060**	-	-	-
3-4	0.1363	**0.0117**	-	-
4-5	**0.0025**	**0.0025**	0.2980	-
> 5	**0.0025**	**0.0025**	0.2980	0.06710

Categories represent the number of parturitions. Pairwise comparisons were performed based on Euclidean distance matrix at the genus level. 999 permutations were implemented, and p-values were adjusted using the FDR method. Values in bold indicate significant results (adjusted P-value < 0.05).

Appendix for Study 3: impact of estrus synchronisation devices on vaginal microbiota and artificial insemination success under four treatments

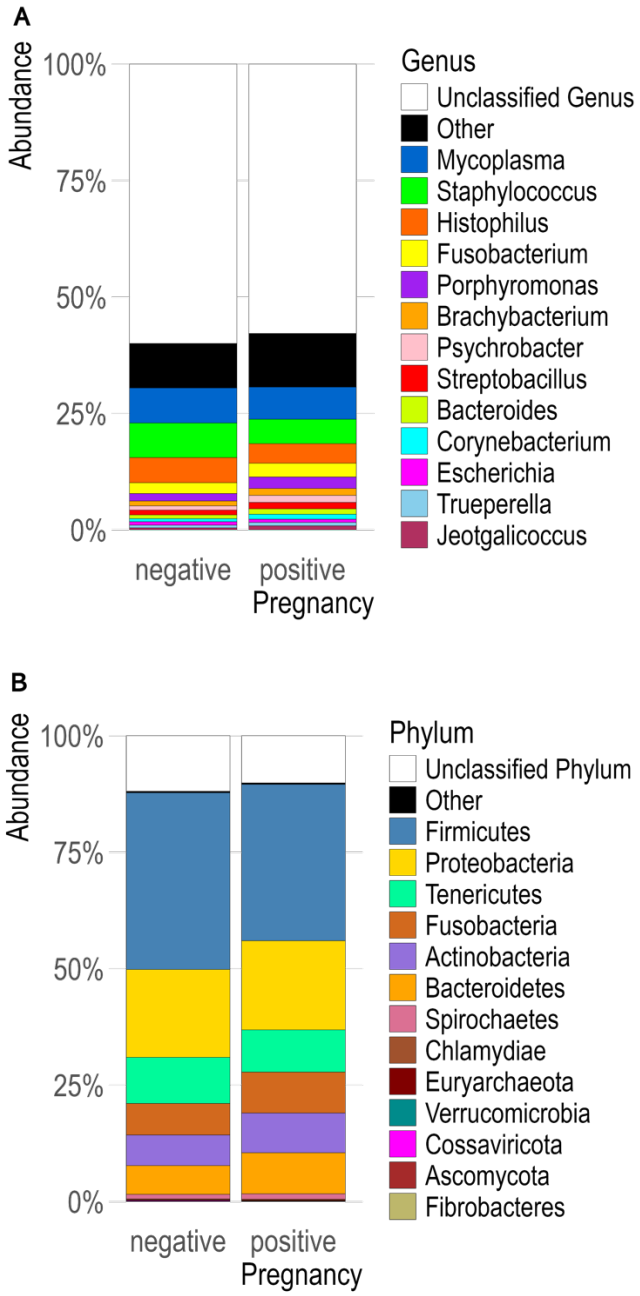


Figure S3.1. Composition of the microbiota for pregnancy groups from S2 sampling point group (after PRID removal) at genus (**A**) and phylum (**B**) levels.

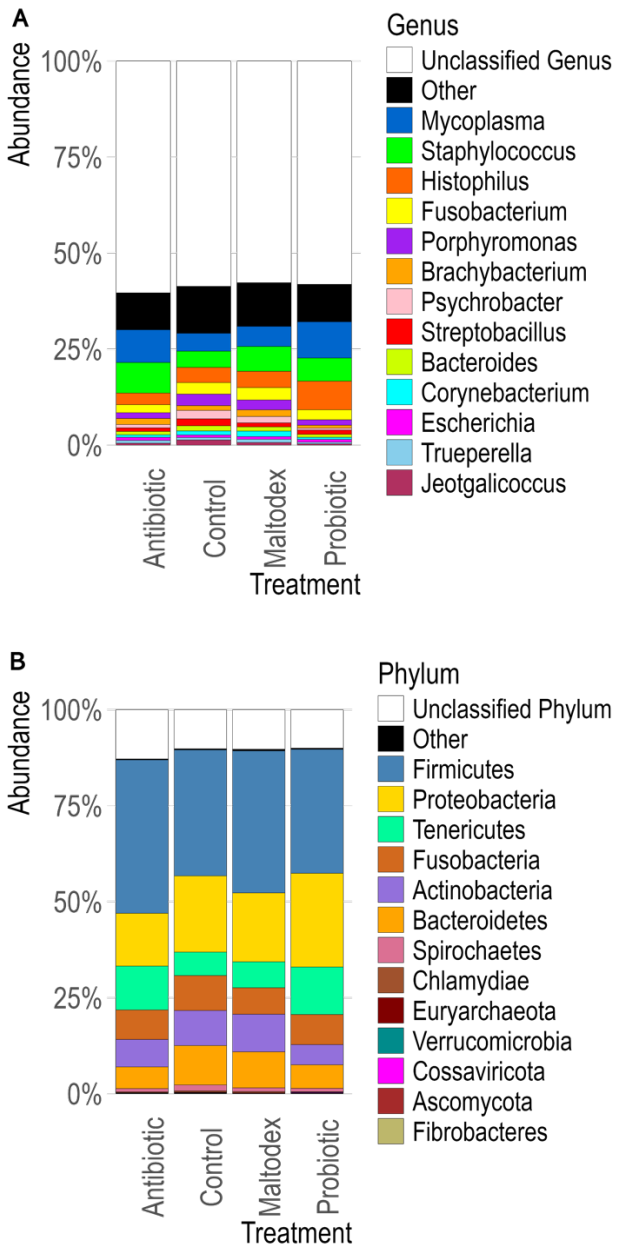


Figure S3.2. Composition of the microbiota for treatments groups from S2 sampling point group (after PRID removal) at genus (**A**) and phylum (**B**) levels.

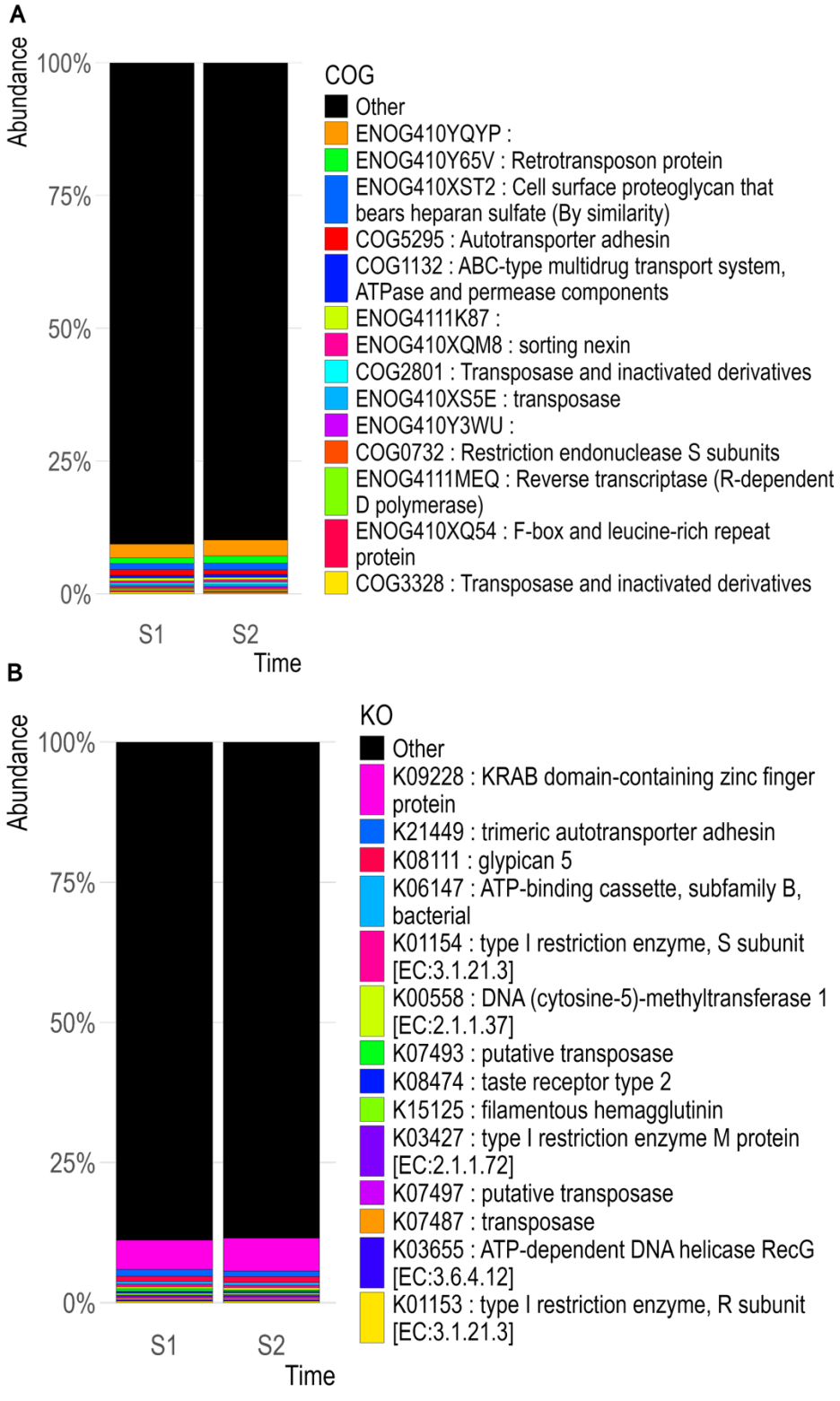


Figure S3.3. Composition of the microbiota for sampling point groups (S1 and S2) at COG (**A**) and KO (**B**) levels. The legend displays the gene name alongside its functional description. When no description is provided, it indicates that the function of the gene is currently unknown. KO: KEGG Orthology.

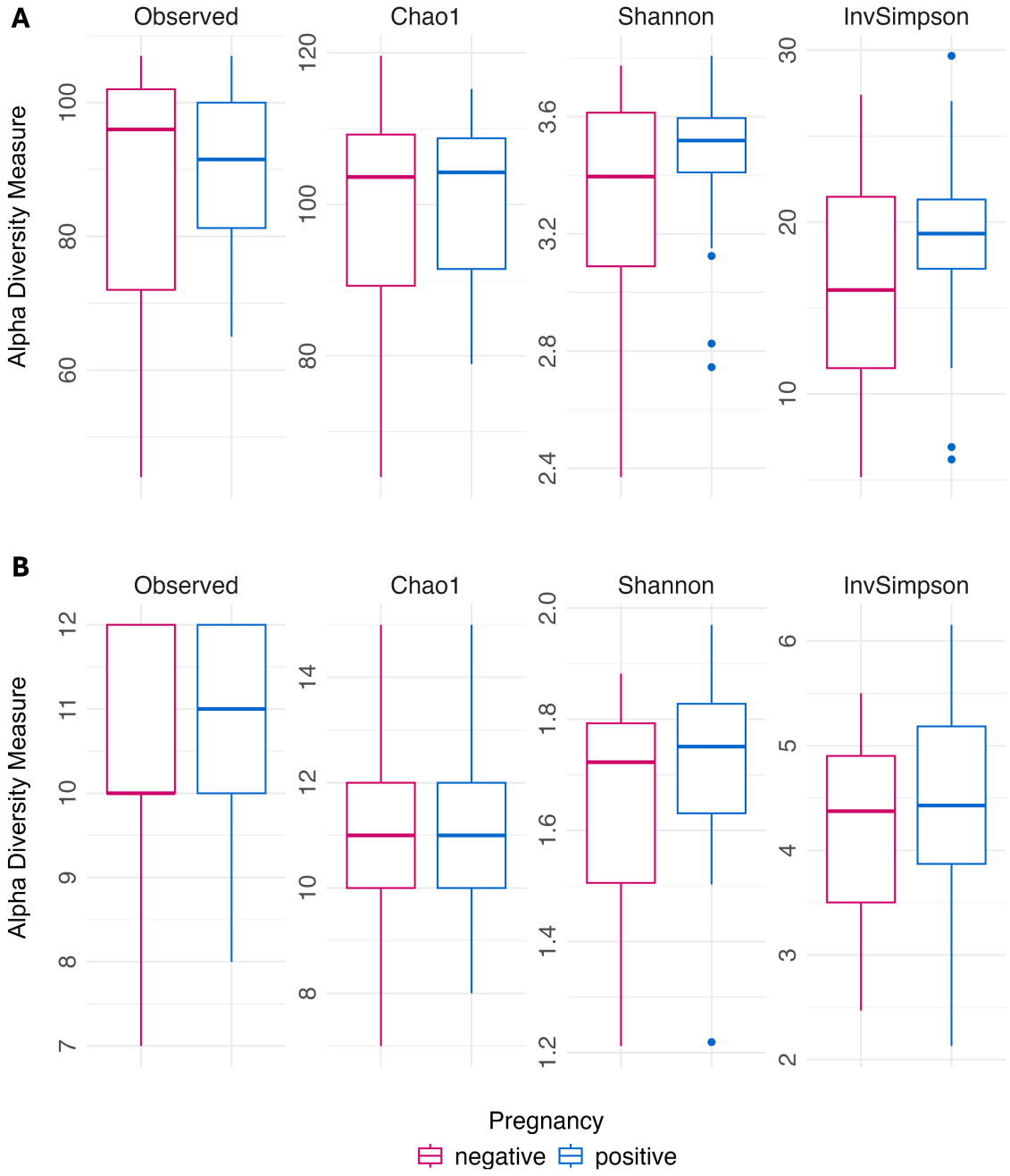


Figure S3.4. Alpha-diversity measures (Observed, Chao1, Shannon, and InvSimpson index) for Pregnancy at the Genus (**A**) and Phylum (**B**) level. There were no significant differences between groups.

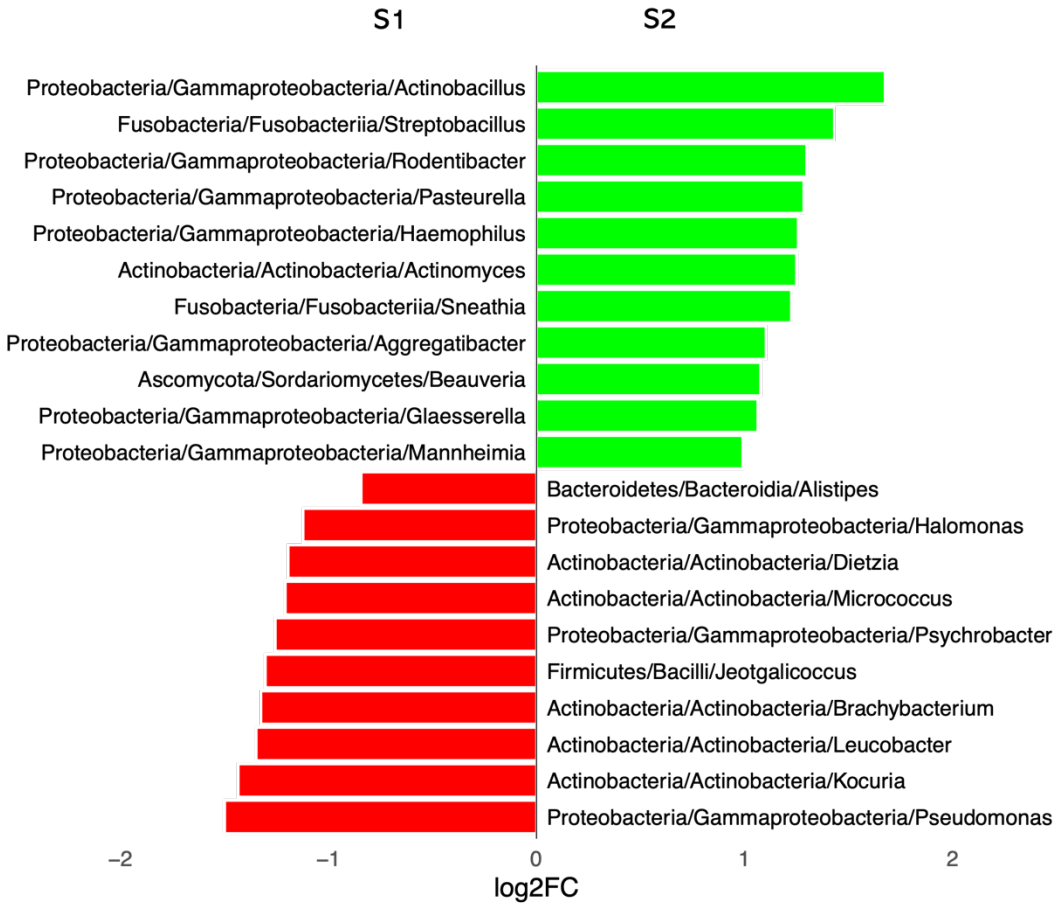


Figure S3.5. Stacked bar charts illustrating significant results (FDR 5%) from the differential abundance analysis at the genus level, comparing sampling point groups S1 and S2 within the antibiotic treatment group. Red represents a higher abundance for S1 compared to S2, while green represents a higher abundance for S2 compared to S1. Data were filtered to retain taxa with > 0.05% relative abundance.

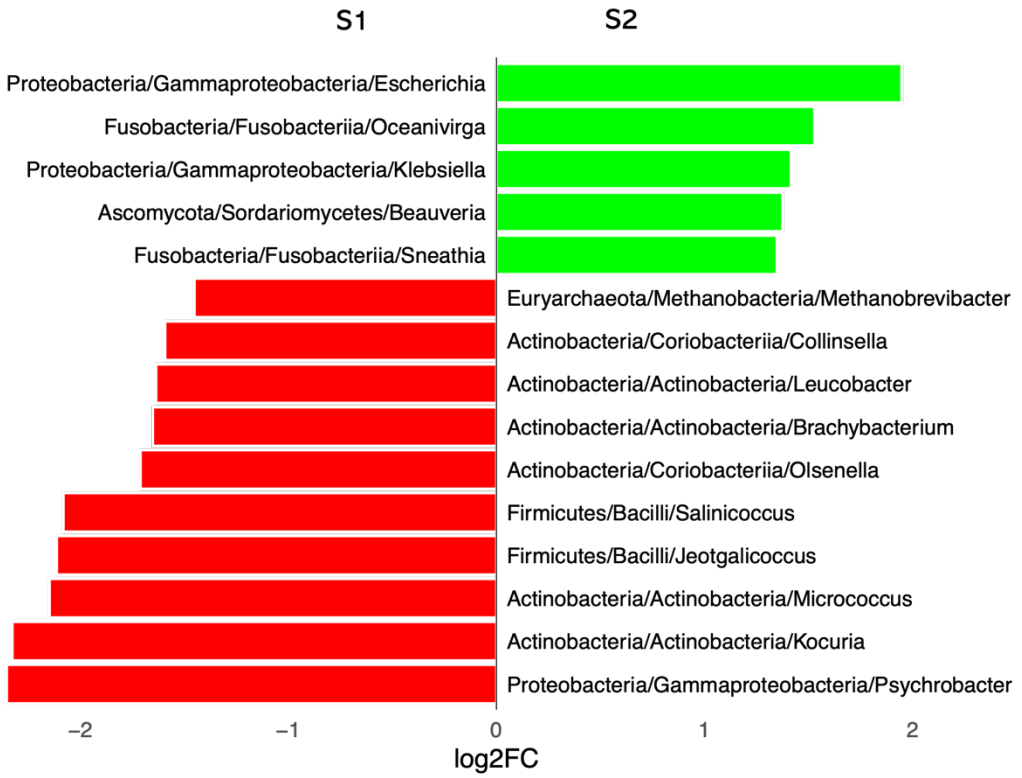


Figure S3.6. Stacked bar charts illustrating significant results (FDR 5%) from the differential abundance analysis at the genus level, comparing sampling point groups S1 and S2 within the control treatment group. Red represents a higher abundance for S1 compared to S2, while green represents a higher abundance for S2 compared to S1. Data were filtered to retain taxa with > 0.05% relative abundance.

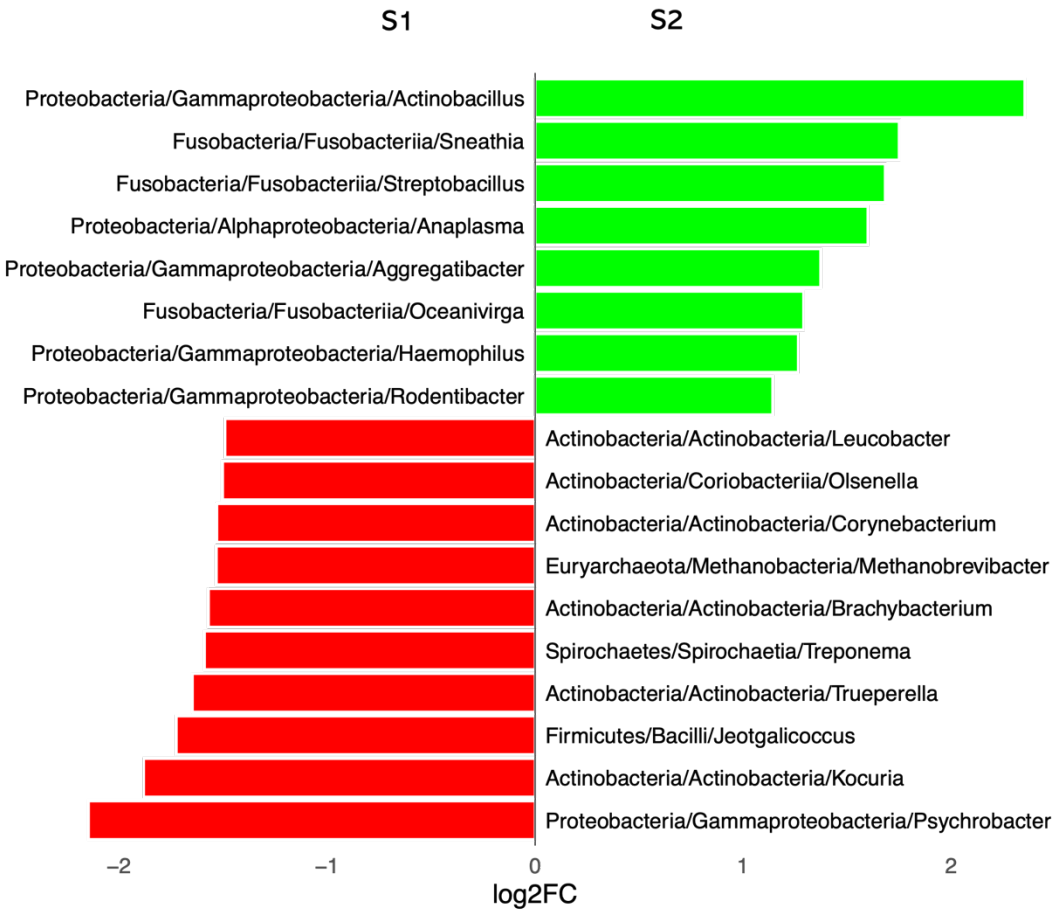


Figure S3.7. Stacked bar charts illustrating significant results (FDR 5%) from the differential abundance analysis at the genus level, comparing sampling point groups S1 and S2 within the maltodex treatment group. Red represents a higher abundance for S1 compared to S2, while green represents a higher abundance for S2 compared to S1. Data were filtered to retain taxa with > 0.05% relative abundance.

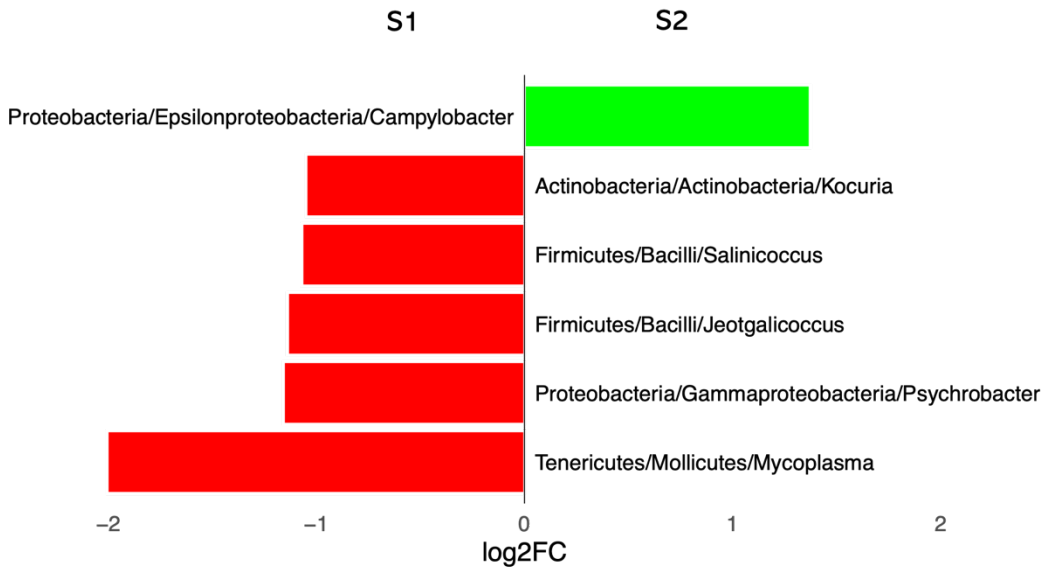


Figure S3.8. Stacked bar charts illustrating significant results (FDR 5%) from the differential abundance analysis at the genus level, comparing sampling point groups S1 and S2 within the probiotic treatment group. Red represents a higher abundance for S1 compared to S2, while green represents a higher abundance for S2 compared to S1. Data were filtered to retain taxa with > 0.05% relative abundance.

Table S3.2. Average alpha-diversity indexes (Observed, Chao1, Shannon and Simpson Inverse) estimated for the microbial composition for pregnancy (P and N) and for sampling time groups S1 (before PRID samples) and S2 (after PRID samples), specifying treatments within each group at the genus level

	Observed	SD	Chao1	SD	Shannon	SD	InvSimpson	SD
P	56.12	12.34	70.49	14.67	2.84	0.40	10.48	4.27
N	51.62	15.70	66.88	17.45	2.59	0.55	7.92	3.72
S1	38.45	14.12	50.15	18.81	2.06	0.52	5.19	2.55
S2	54.11	13.96	68.88	15.90	2.73	0.48	9.34	4.19
Antibiotic	48.00	12.92	61.63	15.31	2.61	0.62	8.30	4.22
Control	61.36	11.36	75.58	13.69	3.00	0.23	11.99	3.82
Maltodex	52.40	15.06	69.46	17.00	2.86	0.49	10.93	3.76
Probiotic	55.38	14.49	70.01	16.18	2.53	0.38	6.90	3.25

P: positive for pregnancy, N: negative for pregnancy

Table S3.3. Results from the PERMANOVA between pregnancy groups within treatments (antibiotic, control, maltodex, probiotic) considering the S1 group

	MeanSqs	**F Model**	**RZ**	**Pr(>F)**
Pregnancy	254.224	0.930	0.021	0.511
Antibiotic	401.328	1.358	0.131	0.127
Maltodex	353.172	1.351	0.144	0.180
Probiotic	271.377	1.121	0.101	0.333
Control	401.328	1.358	0.131	0.127

Appendix for Study 4: host genomic regions associated with vaginal bacterial communities

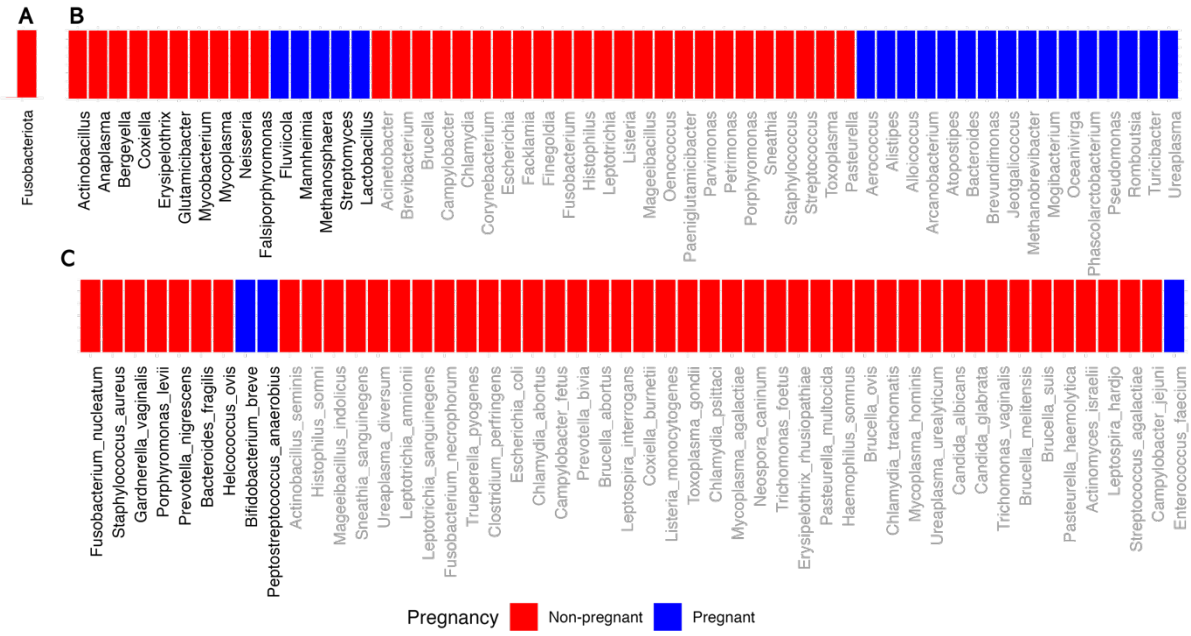


Figure S4.1. Bibliographic search of taxa at phylum (**A**), genus (**B**), and species (**C**) levels and their relationship with fertility. Presence and relevance of various taxa with fertility based on literature at phylum, family, and genus levels. Red colour indicates taxa associated with positive fertility outcomes, while blue colour represents taxa associated with negative fertility outcomes. Vibrant colours represent taxa significantly associated with host genes, while pale colours indicate taxa associated with fertility in the literature but without significant host gene associations.

Table S4.2. References for taxa associated with host-genome and that have been linked to fertility in previous research, including findings from both this thesis and the scientific literature

Taxa	Higher RA	Tax level	Author
Fusobacteriota	NP	Ph	**Study 1, Study 2**, Reinoso-Peláez et al. (2025a), Reinoso-Peláez et al. (2025b)
Actinobacillus	NP	G	Serrano et al. (2020)
Anaplasma	NP	G	Ruiz-Fons et al. (2014)
Bergeyella	NP	G	Koester et al. (2021)
Coxiella	NP	G	Poole et al. (2023), Ruiz-Fons et al. (2014)
Erysipelothrix	NP	G	Serrano et al. (2020)
Falsiporphyromonas	NP	G	Wagener et al. (2014)
Mycobacterium	NP	G	Ruiz-Fons et al. (2014)
Mycoplasma	NP	G	Ruiz-Fons et al. (2014)
Neisseria	NP	G	**Study 3**, Reinoso-Peláez et al. (2023)
Fluviicola	P	G	Serrano et al. (2020)
Glutamicibacter	P	G	Koester et al. (2021)
Lactobacillus	P	G	Quereda et al. (2020)
Mannheimia	P	G	Koester et al. (2021)
Methanosphaera	P	G	Koester et al. (2021)
Streptomyces	P	G	Serrano et al. (2020)
Bacteroides fragilis	NP	Sp	Evaldson et al. (1983)
Fusobacterium nucleatum	NP	Sp	Vander Haar et al. (2018)
Gardnerella vaginalis	NP	Sp	Catlin (1992)
Helcococcus ovis	NP	Sp	Locatelli et al. (2013)
Mycoplasma agalactiae	NP	Sp	Ruiz-Fons et al. (2014)
Mycoplasma hominis	NP	Sp	Hosny et al. (2017)
Porphyromonas levii	NP	Sp	Elad et al. (2004)
Prevotella nigrescens	NP	Sp	Huang et al. (2020)
Staphylococcus aureus	NP	Sp	Martinez-Ros et al. (2018)
Bifidobacterium breve	P	Sp	Bottacini et al. (2014)
Peptostreptococcus anaerobius	P	Sp	Pybus and Onderdonk (1998), Legaria et al. (2021)

Higher RA indicates the group in which the relative abundance was significantly greater. **NP**: more abundant in non-pregnant ewes; **P**: more abundant in pregnant ewes. **Tax level** refers to the taxonomic rank: **Ph**: phylum, **G**: genus, **Sp**: species.

Peer-reviewed publications and scientific contributions originating from the doctoral thesis

The doctoral research has resulted in a series of scientific contributions aimed at enhancing our understanding of the vaginal microbiota and its relationship with fertility and host genetics in sheep. These include two peer-reviewed journal articles already published, two manuscripts currently under review, and one book chapter prepared in collaboration with international researchers.

1. Reinoso-Peláez, Edgar Leonardo*, Serrano, M., Saura, M., Sanchez, J. P., Ramayo-Caldas, Y. & Ramon, M. (2025). The influence of sheep genome on reproductive microbiota. Genetics Selection Evolution (Under Review).
2. Reinoso-Peláez, Edgar Leonardo*, Saura, M., González, C., Ramón, M., Calvo, J. H. & Serrano, M. (2025). The influence of vaginal microbiota on ewe fertility: a metagenomic and functional genomic approach. Microbiome. https://doi.org/10.1186/s40168-025-02165-z
3. Reinoso-Peláez, Edgar Leonardo*, Puente-Sánchez, F., Serrano, M., Calvo, J. H., Ramon, M. & Saura, M. (2025). Characterization of bacterial communities of ewe's vaginal tract and its potential impact on reproductive efficiency. Animal Microbiome. https://doi.org/10.1186/s42523-025-00383-2
4. Reinoso-Peláez, Edgar Leonardo*, Saura, M., González-Recio, Ó., González, C., Fernández, A., Peiro-Pastor, R., López-García, A., Saborío-Montero, A., Calvo, J. H., Ramón, M. & Serrano, M. (2023). Impact of oestrus synchronization devices on ewes vaginal microbiota and artificial insemination outcome. Frontiers in Microbiology, 14, 1063807. https://doi.org/10.3389/fmicb.2023.1063807
5. Reinoso-Peláez, Edgar L., Gianola, D. & González-Recio, O. (2022). Genome-Enabled Prediction Methods Based on Machine Learning. Methods in Molecular Biology, 2467, 189–218. https://doi.org/10.1007/978-1-0716-2205-6_7

Communications at congresses related to this thesis

This section summarises the doctoral student's participation in scientific dissemination activities carried out during the course of the PhD. These contributions include oral communications, poster presentations, and accepted abstracts at national and international congresses in the fields of animal reproduction, microbiome research, and veterinary science.

International Conferences

1. Reinoso-Peláez, E. L., Saura, M., González, C., Ramón, M., Calvo, J., Beltrán de Heredia, I., Serrano, M.
   Title: Metagenomic and Genome-Wide Association Analysis in Ewe Fertility: A Comprehensive Study
   Conference: 75th EAAP Annual Meeting
   Location: Florence, Italy
   Date: September 1–5, 2024
   Organiser: European Federation of Animal Science (EAAP)
2. Reinoso Peláez, E. L., Ramón, M., Saura, M., González Recio, O., Arias, R., Pérez-Guzmán, M., Beltrán de Heredia, I., Calvo, J., González, C., Puente Sánchez, F., Serrano Noreña, M.
   Title: Host genomic regions associated with ewes' vaginal microbiota
   Conference: 39th International Society for Animal Genetics Conference (ISAG 2023)
   Location: Cape Town, South Africa
   Date: July 3rd, 2023
   Organiser: International Society for Animal Genetics
3. Reinoso Peláez, E. L., Puente Sánchez, F., González, C., Calvo, J., Serrano Noreña, M., Saura, M.
   Title: Impact of the vaginal microbiota on the pregnancy rate by artificial insemination in three Spanish sheep breeds
   Conference: 39th International Society for Animal Genetics Conference (ISAG 2023)
   Location: Cape Town, South Africa
   Date: July 3rd, 2023
   Organiser: International Society for Animal Genetics
4. Reinoso Peláez, E. L., Saura, M., González, C., Puente Sánchez, F., Serrano Noreña, M.
   Title: Optimising metagenomic sequencing: A comparative study of ONT Adaptive Sampling strategies to improve microbial DNA recovery

Conference: 39th International Society for Animal Genetics Conference (ISAG 2023)
Location: Cape Town, South Africa
Date: July 3rd, 2023
Organiser: International Society for Animal Genetics

5. Reinoso Peláez, E. L., González Recio, O., González, C., Saura, M., López García, A., Saborío Montero, A., Calvo, J., Ramón, M., Serrano Noreña, M.
Title: Effect of alternative synchronisation treatments on the vaginal microbiota of ewes and their impact on pregnancy rate
Conference: 12th World Congress on Genetics Applied to Livestock Production (WCGALP 2022)
Location: Rotterdam, Netherlands
Date: July 3rd, 2022
Organiser: WCGALP

National Conferences

1. Reinoso-Peláez, E. L., Serrano, M., Fernández, A., Villanueva, B., Saura, M.
Title: Impacto de la consanguinidad del huésped sobre la diversidad de la microbiota vaginal en ovinos y su relación con la tasa de preñez
Conference: XXII Jornadas sobre Producción Animal (AIDA 2025)
Location: Zaragoza, Spain
Date: June 2025
Organiser: Asociación Interprofesional para el Desarrollo Agrario
2. Reinoso-Peláez, E. L., Sánchez, J. P., Ramayo-Caldas, Y., Serrano, M., Saura, M., Ramón, M.
Title: El Genoma Ovino como Modulador de su Microbiota Reproductiva
Conference: XXI Reunión Nacional de Mejora Genética Animal
Location: Cartagena, Spain
Date: June 6–7, 2024
Organiser: Universidad Politécnica de Cartagena
3. Reinoso Peláez, E. L., Puente Sánchez, F., González, C., Serrano Noreña, M., Saura, M.
Title: Impacto de la microbiota vaginal sobre la tasa de preñez por inseminación artificial en tres razas ovinas españolas
Conference: XX Jornadas sobre Producción Animal (AIDA 2023)
Location: Zaragoza, Spain
Date: June 12th, 2023
Organiser: Asociación Interprofesional para el Desarrollo Agrario

4. Reinoso Peláez, E. L.
   Title: Estudio de la fertilidad ovina y su relación con el metagenoma: Desafíos y oportunidades
   Conference: I Jornadas INIciando CienciA 2023
   Location: Madrid, Spain
   Date: April 18th, 2023
   Organiser: Consejo de Doctorandos, INIA-CSIC
5. Reinoso Peláez, E. L., Saura, M., González, C., González Recio, O., Fernández, A., Ramón, M., Calvo, J., Beltrán de Heredia, I., Serrano Noreña, M.
   Title: Diversidad y abundancia de la microbiota vaginal asociada a diferentes razas de ovino
   Conference: XX Reunión Nacional de Mejora Genética Animal
   Location: Madrid, Spain
   Date: June 3rd, 2022
   Organiser: INIA-CSIC
6. Reinoso Peláez, E. L., Gianola, D., González Recio, O.
   Title: Aprendizaje Automático Aplicado a la Predicción de Caracteres Complejos
   Conference: XIX Jornadas sobre Producción Animal (AIDA 2021)
   Location: Online, Spain
   Date: June 2nd, 2021
   Organiser: Asociación Interprofesional para el Desarrollo Agrario
7. Reinoso Peláez, E. L., González Recio, O., González, C., Saura, M., López García, A., Saborío Montero, A., Serrano Noreña, M.
   Title: Efecto de tratamientos alternativos en las esponjas de sincronización sobre la microbiota vaginal de ovejas de raza Assaf y su impacto en la tasa de preñez por inseminación artificial
   Conference: XIX Jornadas sobre Producción Animal (AIDA 2021)
   Location: Online, Spain
   Date: June 2nd, 2021
   Organiser: Asociación Interprofesional para el Desarrollo Agrario

Research stays during the doctoral process

As part of the doctoral training, two international and national research stays were conducted to strengthen skills in bioinformatics, metagenomics, and microbial ecology:

1. SLU – Swedish University of Agricultural Sciences
   Entity Type: University
   City/Country: Uppsala, Stockholm, Sweden
   Duration: 01/09/2022 – 30/11/2022 (3 months)
   Objective: Doctoral Research Stay
   Activities and Achievements:
   Conducted research on metagenomic analysis using advanced bioinformatics tools. Tasks included sequence data preprocessing, optimization of filtering and taxonomic assignment workflows, and robust statistical interpretation of microbial community data. The stay greatly enhanced skills in applied microbial genomics and ecological data analysis.

2. Centro Nacional de Biotecnología (CNB-CSIC)
   Entity Type: State Research Agency
   City/Country: Madrid, Community of Madrid, Spain
   Duration: 01/07/2021 – 30/08/2021 (2 months)
   Objective: Doctoral Research Stay
   Activities and Achievements:
   Focused on microbial community analysis using bioinformatics platforms such as Mothur and SqueezeMeta. Gained hands-on experience in Linux environments, deploying pipelines for integrated metagenomics processing, and interpreting microbial ecosystem data. This stay laid the foundation for computational analyses applied throughout the doctoral thesis.