UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNRAVELLING THE ENVIRONMENTAL VARIANCE OF LITTER SIZE THROUGH THE GENOME AND GUT MICROBIOME

PhD. Thesis by

Cristina Casto Rebollo

Supervisor: **PhD. Noelia Ibáñez Escriche**

Institute for Animal Science and Technology

Universitat Politècnica de València

València, 13th January 2023

UNRAVELLING THE ENVIRONMENTAL VARIANCE OF LITTER SIZE THROUGH THE GENOME AND GUT MICROBIOME

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

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

By

Cristina Casto Rebollo

Thesis Supervisor

PhD. Noelia Ibáñez Escriche

Contact

Cristina Casto Rebollo

cricasre@posgrado.upv.es

Referees

Amélia Camarinha Silva

Nuria Mach

Isabel Cervantes Navarro

María Muñoz Muñoz

Thesis committee

Amélia Camarinha Silva

Lars Rönnegård

Jordi Estellé Fabrellas

Cover and illustrations design

Carolina Vilaplana

@ceuveuve_

"In biology, nothing is clear, everything is too complicated, everything is a mess, and just when you think you understand something, you peel off a layer and find deeper complications beneath. Nature is anything but simple."
Richard Preston-The Hot Zone

"Anyone who has never made a mistake
has never tried anything new"
Albert Einstein

ACKNOWLEDGEMENT

During the development of this thesis, I have had the pleasure of meeting many people who in one way or another have contributed to its successful completion. And I am not only talking about the scientific contributions, but also the personal ones that have allowed me to never give up and to enjoy the process. My first words of thanks go to my brother Fran, Alicia and my parents. This thesis is for you, for your unconditional support for all the decisions I have made and for the affection, albeit from afar, you have given me. It is also a small reminder that, although life sometimes plays tricks on us, we must not give up. The following words of thanks go to Noelia Ibáñez, my supervisor. Everything I can write here will be too little for everything you have given me. I know it has not been easy to deal with myself but thank you for being always there. I have always admired you as you are a great scientist and an exceptional person. During these four years I have learned a lot from you. You are a person who brings a lot to the table and whom I respect enormously. Thank you for the trust you have placed in me and for backing all the decisions I have made, even if they were sometimes wrong. You have been key to the success of this thesis and a great friend when I have needed it. Thank you for everything. Without a doubt, I would do the thesis with you as many times as necessary. I would also like to thank Agustín Blasco. I really appreciate all the time you have invested in trying to make me a more educated person. I value all those concerts very much, operas, discussions about art and poetry quotations we have had over the years. I believe that culture must also be the basis of a person, and with you I have learned a lot over the years. Thank you very much. I would also like to thank Pilar Hernández, for her willingness to help me whenever I needed it and for all those informal talks to let off steam that we have had. They were necessary. To Toñi Santacreu, for the joy she brings to the group and her unconditional help. I would also like to thank María José Argente and Mariluz García. Thank you for all your patience and time invested in resolving all my doubts as quickly as possible. The group of "Animal breeding" from the Institute for Animal Science and Technology of Valencia, in which I have worked during these years, has been a big family where all of you have contributed to improve me as a person.

On the other hand, I would like to thank my travelling partner, Agostina Zubiri. Together we started this roller coaster that is the thesis and that we did not know what it was going to bring us: deadlines at the last moment, DNA extractions, dark corridors, barbecues, fernet, beers, laughter, crying, etc. Although the pandemic distanced us a little, with you everything has been easier. You have always been there when I needed you. Thank you. I would also like to thank Samuel Sosa, Enrico Mancín and David López for all our

discussions about genetics, among other things. To Marina Martínez for her energy and cheerfulness, especially when I was in Edinburgh as she helped me feel at home. As well as María Martínez and Ana Villaplana. Thank you for helping me and being there when I needed it. And to the whole Highlander Lab group for making me feel like one of them. I would also like to thank Ximo, Andrea and the aquaculture group, Víctor, Marina, Germán and Miriam for all the moments and laughs we have shared. And to the latest additions, Pedro and Illyas. There will be good and bad moments during the thesis but enjoy and learn from all of them.

Outside the social circle of the thesis, I would like to thank all those people I have met and with whom I have shared unforgettable experiences. Henry, Karla, Paola, Sergio, Jesús, Facu and Aru. As well as all the people with whom I share my passion for rugby and hiking. Thank you for helping me to disconnect from the thesis and recharge my energy. To Carol Vilaplana by all the illustrations and her amazing cover design. And above all thanks to Noelia Pérez. Our relationship began at the most difficult stage of my PhD and you never stopped trusting me and supporting me. Thank you for being by my side and putting up with all my internal disputes during this period.

There are many people I have met on this journey, and they are no less important than those I have named here. Thank you all for crossing my path and helping me to become the person I am today.

AGRADECIMIENTOS

Durante el desarrollo de esta tesis he tenido el placer de conocer a muchas personas que de una forma u otra han contribuido a que ésta llegara con éxito a su final. Y no hablo sólo de las aportaciones científicas, sino también aquellas personales que han permitido que nunca desistiera y disfrutara del proceso. Mis primeras palabras de agradecimiento son para mi hermano Fran, Alicia y mis padres. Esta tesis es para ustedes, por ese apoyo incondicional a todas las decisiones que he ido tomando y por el cariño, aunque desde la lejanía, me habéis aportado. Además, es un pequeño recordatorio de que, aunque la vida a veces nos juega malas pasadas, no por ello hay que desistir. Las siguientes palabras de agradecimiento son para el pilar de esta tesis, Noelia Ibáñez, mi tutora. Todo lo que aquí pueda escribir será poco para todo lo que me has aportado. Sé que no ha sido fácil lidiar con mi persona, pero gracias por estar ahí en todo momento. Siempre te he admirado ya que eres una gran científica y una persona excepcional. Durante estos cuatro años he aprendido mucho de ti. Eres una persona que aporta muchísimo y a la cual respeto enormemente. Gracias por la confianza que has depositado en mí y por apostar por todas las decisiones que he ido tomando, aunque a veces fueran erróneas. Has sido clave para el éxito de esta tesis y una gran amiga cuando lo he necesitado. Gracias por todo, sin duda, realizaría la tesis contigo las veces que hicieran falta. También me gustaría agradecer a Agustín Blasco. Aprecio muchísimo todo el tiempo que has invertido en intentar hacer de mí una persona más culta. Valoro muchísimo todos esos conciertos, operas, discusiones sobre arte y citas de poesía que hemos tenido durante estos años. Creo que la cultura tiene que ser también la base de una persona, y contigo he aprendido muchísimo durante estos años. Muchas gracias. También agradecer a Pilar Hernández, por su disposición a ayudarme siempre que lo he necesitado y por todas esas charlas informales para desahogarnos que hemos tenido. Eran necesarias. A Toñi Santacreu, por la alegría que aporta al grupo y su ayuda incondicional. También agradecer a María José Argente y a Mariluz García. Gracias por vuestra paciencia y todo ese tiempo invertido en resolver todas mis dudas con la mayor brevedad posible. El grupo de mejora genética del Instituto de Ciencia Animal de Valencia, en el que he trabajado durante estos años, ha sido una gran familia donde todos habéis contribuido a mejorarme como persona.

Por otro lado, me gustaría agradecerle a mi compañera de viaje, Agostina Zubiri. Juntas empezamos esta montaña rusa que es la tesis y que no sabíamos que nos iba a deparar: deadlines en el último momento, extracciones de ADN, pasillos oscuros, asados, fernet, cervezas, risas, llantos, etc. Aunque la pandemia nos distanció un poco, contigo todo ha sido más sencillo. Siempre has estado cuando lo he necesitado. Gracias. También

agradecer a Samuel Sosa, Enrico Mancín y David López por todas nuestras discusiones sobre genética, entre otras cosas. A Marina Martínez por su energía y alegría, sobre todo cuando estaba en Edimburgo ya que me ayudó a sentirme como en casa. Al igual que María Martínez y Ana Vilaplana. Gracias por ayudarme y estar ahí cuando lo he necesitado. Y a todo el grupo de HighlanderLab por hacerme sentir una más entre ellos. También agradecer a Ximo, Andrea y al grupo de acuicultura, Víctor, Marina, Germán y Miriam por todos los momentos y las risas que hemos compartido. Y a las últimas incorporaciones, Pedro e Illyas. Habrá momentos buenos y malos durante la tesis, pero disfrutad y aprended de todos ellos.

Fuera del círculo social de la tesis, me gustaría agradecer a todas esas personas que he ido conociendo y con las he compartido vivencias inolvidables. Henry, Karla, Paola, Sergio, Jesús, Facu y Aru. Así como a todas las personas con las que comparto mi pasión por el rugby, y el senderismo. Gracias por ayudarme a desconectar de la tesis y cargarme de energías. A Carol Vilaplana por todas las ilustraciones y el espectacular diseño de la portada. Y sobre todo gracias a Noelia Pérez. Nuestra relación comenzó en la etapa más difícil del doctorado y no por ello dejaste de confiar en mí y apoyarme. Gracias por estar a mi lado y soportar todas mis disputas internas durante este periodo.

Son muchas las personas que he conocido en este viaje, y no por ello son menos importantes que las que aquí he nombrado. Gracias a todas por cruzarse en mi camino y ayudarme a convertirme en la persona que soy actualmente.

INDEX

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The residual variance of traits, referred to as V_E, is under genetic control. This means that some of the dissimilarities among individuals observed in the V_E result from a genetic variation in response to microenvironmental factors. Response to selection for the V_E was observed in divergent selection experiments. Therefore, including V_E traits as a breeding goal could homogenize production on farms by its reduction. Furthermore, the reduction of the V_E may improve health traits and animal resilience due to their negative correlation. Thus, V_E is a promising trait to reach a more sustainable livestock system by improving welfare and animal resilience. Disclosing the biological mechanisms of the V_E can help to gain some insight into the biological basics of animal resilience. In this thesis, genomic, metagenomic, and metabolomic analyses were performed on rabbit lines divergently selected for high and low V_E of litter size (LS). These animals showed differences in their resilience potential. Thus, these divergent populations are an excellent biological material for studying animal resilience through the V_E.

Genome-wide association studies (GWAS) were performed using single marker regression, and Bayesian multiple marker regression approaches. Four genomic regions were associated with the V_E in the *Oryctolagus cuniculus* chromosome (OCU) 3, OCU7, OCU10, and OCU14, explaining 8.6% of the total genetic variance for the V_E. In addition, the signature of selection (SS) study identified 134 genomic regions which could be under selection for V_E. Overlapping between both studies was placed in the OCU3, where functional mutations for the DOCK2, INSYN2B and FOXI1 genes were also found. Candidate genes from GWAS and SS were those with functional mutations identified using whole genome sequencing (WGS) analysis with pools of DNA. Highlighted candidate genes showed biological functions related to the development of sensory structures, the immune response, the stress response, and the nervous system. All of them are relevant functions to modulate animal resilience. On the other hand, metagenomic and metabolomic studies showed that the selection for V_E modified the gut microbiome and metabolome composition. Beneficial microbial species such as *Alistipes prutedinis*, *Alistipes shahii*, *Odoribacter splanchnicus* and *Limosilactobacillus fermentum* were more abundant in the resilient population. In contrast, harmful microbial species such as *Acetatifactor muris* and *Eggerthella sp* were more abundant in the non-resilient animals. Genes related to biofilm formation, aromatic amino acid metabolism (Phenylalanine, tryptophan, and tyrosine), and glutamate metabolism were also differentially expressed between the rabbit populations. Furthermore, 15 gut metabolites were identified as potential biomarkers to properly discriminate and predict between the resilient and non-resilient rabbit populations. Five of them, the equol, 3-(4-

hydroxyphenyl)lactate, 5-aminovalerate, N6-acetyllisine, and serine were microbial-derived metabolites.

This is the first study unravelling important biological mechanisms under the animal resilience generated by V_E of LS selection. Genome and gut microbiome and metabolome composition were modified throughout the selection process, affecting the immune and stress response. Overlapping results were found between the metagenomic and metabolome studies, suggesting the aromatic amino acid metabolism and the L-glutamate synthesis influence the phenotype of the V_E. Moreover, some of the candidate genes associated with the VE of LS such as DOCK2, ACE and HDAC9 could influence the microbiome composition in the rabbit populations. On the other hand, in this thesis, we developed a flexible tool for simulating the coevolution of the genome and microbiome across a selection process for the first time. The key of this tool was the implementation of the microbiome inheritance. It is constructed in R and based on AlphaSimR so the user can modify the code and implement different scenarios. We performed a first approximation, showing as the selection based on the phenotype is successful and generated phenotypic, genetic, and microbiome responses. The results highlighted the importance of microbial heritability, symbiosis, and microbial species fitness as fundamental actors to obtain a proper selection response of traits.

This thesis is the first step to develop future strategies and further research to improve animal resilience. A selection combining genomic and metagenomic information may improve the selection response. Moreover, gut-derived metabolites with evidence of crosstalk can be used as biomarkers to identify resilient animals by plasma, avoiding the extraction of faecal samples to determine the microbiome composition. If these studies suceed, these strategies could improve animal resilience with the aim of search a more sustainable livestock system. Lastly, the simulation tool developed could help unravel the microbiome's implications in animal breeding programs.

RESUMEN

La varianza residual de los caracteres, denominada V_E, está bajo control genético. Esto significa que algunas de las disimilitudes entre individuos observadas en la V_E son el resultado de una variación genética en respuesta a factores microambientales. La respuesta a la selección para la V_E se observó en experimentos de selección divergente. Por lo tanto, la inclusión de los rasgos de V_E como objetivo de selección podría homogeneizar la producción en las explotaciones ganaderas mediante su reducción. Además, esta reducción de la V_E podría mejorar los caracteres de salud y la resiliencia de los animales debido a su correlación negativa. Así pues, la V_E es un rasgo prometedor para alcanzar un sistema ganadero más sostenible al mejorar el bienestar y la resiliencia animal. Desvelar los mecanismos biológicos de la V_E podría ayudar a conocer los fundamentos biológicos de la resiliencia animal. En esta tesis, se realizaron análisis genómicos, metagenómicos y metabolómicos en líneas de conejo seleccionadas de forma divergente para alta y baja V_E del tamaño de la camada (TC). Estos animales mostraron diferencias en su potencial de resiliencia. Por ello, estas poblaciones divergentes son un excelente material biológico para estudiar la resiliencia animal a través de la V_E.

Se realizaron estudios de asociación del genoma (GWAS) utilizando la regresión de un solo marcador y la regresión bayesiana de múltiples marcadores. Cuatro regiones genómicas se asociaron con la V_E en el cromosoma 3 de *Oryctolagus cuniculus* (OCU), OCU7, OCU10 y OCU14, explicando el 8,6% de la varianza genética total para la V_E. Además, el estudio de huellas de selección (SS) identificó 134 regiones genómicas que podrían estar bajo selección para la V_E. El solapamiento entre ambos estudios se identificó en el OCU3, donde también se encontraron mutaciones funcionales para los genes *DOCK2*, *INSYN2B* y *FOXI1*. Los genes candidatos de GWAS y SS fueron aquellos con mutaciones funcionales identificadas mediante el análisis de secuenciación del genoma completo (WGS) con pools de ADN. Los genes candidatos destacados mostraron funciones biológicas relacionadas con el desarrollo de estructuras sensoriales, la respuesta inmunitaria, la respuesta al estrés y el sistema nervioso. Todas ellas son funciones relevantes para modular la resiliencia de los animales. Por otra parte, los estudios metagenómicos y metabolómicos mostraron que la selección para la V_E modificó el microbioma intestinal y la composición de su metaboloma. Las especies microbianas beneficiosas como *Alistipes prutedinis*, *Alistipes shahii*, *Odoribacter splanchnicus* y *Limosilactobacillus fermentum* eran más abundantes en la población resiliente. En cambio, las especies microbianas nocivas, como *Acetatifactor muris* y *Eggerthella sp*, fueron más abundantes en los animales no resistentes. Los genes

relacionados con la formación de biofilms, el metabolismo de aminoácidos aromáticos (fenilalanina, triptófano y tirosina) y el metabolismo del glutamato también se expresaron de forma diferencial entre las poblaciones de conejos. Además, se identificaron 15 metabolitos intestinales como potenciales biomarcadores para discriminar y predecir adecuadamente entre las poblaciones de conejos resistentes y no resistentes. Cinco de ellos, el equol, el 3-(4-hidroxifenil)lactato, el 5-aminovalerato, la N6-acetilisina y la serina son metabolitos de origen microbiano.

Este es el primer estudio que desvela importantes mecanismos biológicos de la resiliencia animal generada por la selección de la V_E de TC. El genoma y el microbioma intestinal y la composición del metaboloma se modificaron a lo largo del proceso de selección, afectando a la respuesta inmunitaria y al estrés. Se encontraron resultados coincidentes entre los estudios metagenómicos y del metaboloma, lo que sugiere que el metabolismo de aminoácidos aromáticos y la síntesis de L-glutamato influyen en el fenotipo de la V_E. Además, algunos de los genes candidatos asociados a la V_E de TC, como *DOCK2*, *ACE* y *HDAC9*, podrían influir en la composición del microbioma en las poblaciones de conejos. Por otro lado, en esta tesis desarrollamos por primera vez una herramienta flexible para simular la coevolución del genoma y el microbioma a través de un proceso de selección. La clave de esta herramienta fue la implementación de la herencia del microbioma. Está construida en R y basada en AlphaSimR para que el usuario pueda modificar el código e implementar diferentes escenarios. Realizamos una primera aproximación, mostrando como la selección basada en el fenotipo es exitosa y generamos respuestas fenotípicas, genéticas y del microbioma. Los resultados destacaron la importancia de la heredabilidad microbiana, la simbiosis y la aptitud de las especies microbianas como actores fundamentales para obtener una respuesta de selección adecuada.

Esta tesis es el primer paso para desarrollar futuras estrategias y nuevas investigaciones para mejorar la resiliencia de los animales. Una selección que combine información genómica y metagenómica puede mejorar la respuesta de selección. Además, los metabolitos derivados del intestino con evidencia de crosstalk pueden utilizarse como biomarcadores para identificar animales resilientes por plasma, evitando la extracción de muestras fecales para determinar la composición del microbioma. Si estos estudios tienen éxito, estas estrategias podrían mejorar la resiliencia de los animales con el objetivo de buscar un sistema ganadero más sostenible. Por último, la herramienta de simulación desarrollada podría ayudar a desentrañar las implicaciones del microbioma en los programas de cría de animales

RESUM

La variància residual dels caràcters, denominada V_E, està baix control genètic. Això significa que algunes de les dissimilituds entre individus observades en la V_E són el resultat d'una variació genètica en resposta a factors microambientals. La resposta a la selecció per a la V_E es va observar en experiments de selecció divergent. Per tant, la inclusió dels trets de V_E com a objectiu de selecció podria homogeneïtzar la producció en les explotacions ramaderes mitjançant la seua reducció. A més, aquesta reducció de la V_E podria millorar els caràcters de salut i la resiliència dels animals a causa de la seua correlació negativa. Així doncs, la V_E és un tret prometedor per a aconseguir un sistema ramader més sostenible en millorar el benestar i la resiliència animal. Revelar els mecanismes biològics de la V_E podria ajudar a conéixer els fonaments biològics de la resiliència animal. En aquesta tesi, es van realitzar anàlisis genòmiques, metagenòmiques i metabolòmiques en línies de conill seleccionades de manera divergent per a alta i baixa V_E de la grandària de la ventrada (GV). Aquests animals van mostrar diferències en el seu potencial de resiliència. Per això, aquestes poblacions divergents són un excel·lent material biològic per a estudiar la resiliència animal a través de la V_E.

Es van realitzar estudis d'associació del genoma (GWAS) utilitzant la regressió d'un solo marcador i la regressió bayesiana de múltiples marcadors. Quatre regions genòmiques es van associar amb la V_E en el cromosoma 3 de *Oryctolagus cuniculus* (OCU), OCU7, OCU10 i OCU14, explicant el 8,6% de la variància genètica total per a la V_E. A més, l'estudi de petjades de selecció (SS) va identificar 134 regions genòmiques que podrien estar sota selecció per a la V_E. El solapament entre tots dos estudis es va identificar en l'OCU3, on també es van trobar mutacions funcionals per als gens *DOCK2*, *INSYN2B* i *FOXI1*. Els gens candidats de GWAS i SS van ser aquells amb mutacions funcionals identificades mitjançant l'anàlisi de seqüenciació del genoma complet (WGS) amb pools d'ADN. Els gens candidats destacats van mostrar funcions biològiques relacionades amb el desenvolupament d'estructures sensorials, la resposta immunitària, la resposta a l'estrés i el sistema nerviós. Totes elles són funcions rellevants per a modular la resiliència dels animals. D'altra banda, els estudis metagenòmiques i *metabolòmiques van mostrar que la selecció per a la V_E va modificar el microbioma intestinal i la composició de la seua metaboloma. Les espècies microbianes beneficioses com *Alistipes prutedinis*, *Alistipes shahii*, *Odoribacter splanchnicus* i *Limosilactobacillus fermentum* eren més abundants en la població resilient. En canvi, les espècies microbianes nocives, com *Acetatifactor muris* i *Eggerthella sp*, van ser més abundants en els animals no resistents. Els gens relacionats amb la formació de biofilms, el metabolisme d'aminoàcids aromàtics

(fenilalanina, triptòfan i tirosina) i el metabolisme del glutamat també es van expressar de manera diferencial entre les poblacions de conills. A més, es van identificar 15 metabòlits intestinals com a potencials biomarcadores per a discriminar i predir adequadament entre les poblacions de conills resistents i no resistents. Cinc d'ells, el equol, el 3-(4-hidroxifenil)lactat, el 5-aminovalerato, la N6-acetilisina i la serina són metabòlits d'origen microbià.

Aquest és el primer estudi que revela importants mecanismes biològics de la resiliència animal generada per la selecció de la V_E de GC. El genoma i el microbioma intestinal i la composició del metaboloma es van modificar al llarg del procés de selecció, afectant la resposta immunitària i a l'estrés. Es van trobar resultats coincidents entre els estudis metagenòmiques i del metaboloma, la qual cosa suggereix que el metabolisme d'aminoàcids aromàtics i la síntesi de L-glutamat influeixen en el fenotip de la V_E. A més, alguns dels gens candidats associats a la V_E de GC, com *DOCK2*, *ACE* i *HDAC9*, podrien influir en la composició del microbioma en les poblacions de conills. D'altra banda, en aquesta tesi desenvolupem per primera vegada una eina flexible per a simular la coevolució del genoma i el microbioma a través d'un procés de selecció. La clau d'aquesta eina va ser la implementació de l'herència del microbioma. Està construïda en R i basada en AlphaSimR perquè l'usuari puga modificar el codi i implementar diferents escenaris. Realitzem una primera aproximació, mostrant com la selecció basada en el fenotip és reeixida i generem respostes fenotípiques, genètiques i del microbioma. Els resultats van destacar la importància de la heredabilidad microbiana, la simbiosi i l'aptitud de les espècies microbianes com a actors fonamentals per a obtindre una resposta de selecció adequada.

Aquesta tesi és el primer pas per a desenvolupar futures estratègies i noves investigacions per a millorar la resiliència dels animals. Una selecció que combine informació genòmica i metagenòmique pot millorar la resposta de selecció. A més, els metabòlits derivats de l'intestí amb evidència de crosstalk poden utilitzar-se com biomarcadores per a identificar animals resilients per plasma, evitant l'extracció de mostres fecals per a determinar la composició del microbioma. Si aquests estudis tenen èxit, aquestes estratègies podrien millorar la resiliència dels animals amb l'objectiu de buscar un sistema ramader més sostenible. Finalment, l'eina de simulació desenvolupada podria ajudar a desentranyar les implicacions del microbioma en els programes de cria d'animals.

CHAPTER 1: SCOPE OF THIS STUDY

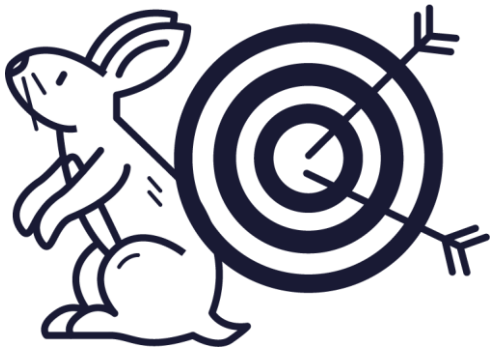
This thesis used rabbits from a divergent selection experiment for high and low V_E of litter size (LS) by the University of Miguel Hernández in Elche. These populations showed an outstanding genetic response to the selection for V_E of LS as well as differences in their resilience potential, being an exceptional biological material to study animal resilience through the V_E. Thus, this thesis aimed to investigate the molecular basics of these rabbit populations to gain insight into the biological mechanisms of animal resilience. This thesis focused on studying the genome and gut microbiome because both are important to shape phenotypes.

SPECIFIC OBJECTIVES

1. To identify genes and functional mutations associated with the V_E and animal resilience using genome-wide association approaches, signatures of selection methodologies, and whole genome sequencing analysis.

2. To determine the gut microbiome and metabolome composition underlying the differences in the resilience potential of the rabbits divergently selected for V_E of litter size using metagenomic and metabolomic analysis.

3. To develop a simulation tool for studying the effects of the genome and the microbiome coevolution in animal breeding programs.



CHAPTER 2: INTRODUCTION

The most frequent occurrence of extreme climate events such as heat waves is proof of climate change [1]. The exposure of the animal to the new climate conditions requires an adaptation of these to avoid worse livestock production. Additionally, less use of antibiotics, as well as an increase in the animal welfare, must be required in the short future in the livestock industry. Hence, in this reality, a more sustainable livestock system is needed. The study of animal resilience may solve this one because the improvement of animal resilience would allow animals to cope better with the environmental fluctuations.

1. ANIMAL RESILIENCE

Until now, resilience was considered a challenging trait because is highly difficult to measure it. There are many definitions for resilience depending on the nature of the environmental perturbation [2]. In this dissertation, we referred to resilience as the ability of animals to maintain or quickly recover their performance after an environmental perturbation [3]. Furthermore, there are different ways to quantify the resilience. It can be measured by identifying the non-additive variance due to genotype by environment interactions (GxE) [4-6], as deviations of the observed individuals' performance from an expected pattern [7-11], or using fluctuation traits [3, 12-15]. GxE needs to categorize or quantify the environmental conditions to correlate the phenotypic value from each level and extract the genotype by environment interaction [4, 16-18]. It is based on the intercept and slope of a reactive norm to calculate the estimated breeding values for resilience [4]. However, GxE is not a measure of overall resilience as strongly depends on the challenging environment or the nature of the disturbance categorized, i.e. heat stress [6, 19-20] or pathogen infection [5,21]. Deviations and fluctuations of traits seemed to fit better with our definition of resilience. Deviations from an expected pattern are measured with smoothing curves. Nevertheless, thresholds must be settled to determine the degree of resilience depending on the duration and the magnitude of the deviation [7-8]. The novel measure indicators are based on fluctuations traits such as the skewness of deviations, the autocorrelation of deviations, the variance of deviations, the residual variance [3, 22-23], the root mean square error, or quantile regression [24]. In all cases to measure animal resilience, repeated records such as daily feed intake [24-25] daily milk yield [14, 23], litter size [26,], and body weight [10-11] are needed. The use of one or other resilience indicators depends on the breeding goal. However, recent studies have proposed the environmental variance as a promising indicator of resilience considering their correlation, specifically with the residual

environmental variance [3, 14-15]. The environmental variance is easy to measure and showed low-to-moderate heritability [13, 27]. Moreover, it showed a negative correlation with health-related traits such as mortality [28], longevity, and fertility [23]. Likewise, animals with a low environmental variance showed less susceptibility to environmental perturbations for heat waves [14], pathogen infections [15], and highly stressful conditions as the parity [29]. The immune system seemed to be an important part of resilience since it allows to perceive the environmental factors and mitigate their effects [2].

Unravelling the biological mechanism underlying the environmental variance could be a strategy to decipher animal resilience. The selection for environmental variance may allow more resilient animals to improve the sustainability of the livestock system. For that, this dissertation aimed to disentangle the genetic and microbiome basics of the environmental variance with the perspective of extrapolating the results to animal resilience. There are a wide number of evidence highlighting the relevance of the genome and microbiome shaping the phenotypes [30]. Hence, we studied the genome and the microbiome of animals divergently selected by high and low environmental variance of litter size to obtain a global vision of the most important mechanisms that could modulate the animal resilience.

2. ENVIRONMENTAL VARIANCE OF TRAITS

The environmental variance or V_E is a measure of environmental sensitivity, how the individuals respond to the environment [31]. V_E is not homoscedastic and differs from individuals even if they share an environment. This is because each genome can respond differently to environmental factors [32-33]. That means that a part of the dissimilarities among individuals observed on the phenotype is a result of a genetic variation for environmental sensitivity which can affect the mean or the variance of traits (Fig. 1). The estimation of the environmental sensitivity depends on the ability to quantify or categorized the environmental factors. When the environmental factors can be measurable (for instance diet or temperature) are defined as macro-environmental factors [17, 33-34]. In this case, the environmental sensitivity can be estimated by computing a GxE interaction due to the phenotypic mean differences between environmental factors depend on the genotype (Fig 1B). When factors are unknown, they are known as micro-environmental factors. These factors are all unknown environmental factors in a short period or throughout the lifetime of individuals [17, 33-35]. The micro-environmental factors are studied using the residual variance of the traits that are the differences in the within-individual variance under the same environmental conditions,

also named V_E [32-34]. Thus, the term V_E is used to refer to the environmental sensitivity for micro-environmental factors measured as the residual variance of the traits (Fig. 1C). The genetic variation in the V_E is of interest in evolutionary ecology and animal and plant breeders to study the environmental sensitivity and use it as a selection criterion [30, 33].

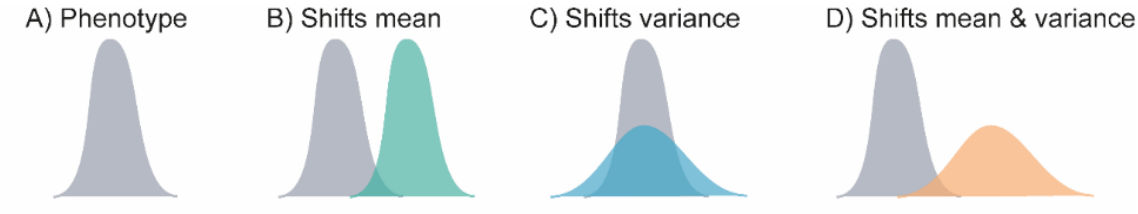


Figure 1. Differences in the phenotype value due to environmental sensitivity. A) Phenotype do not affect by environmental factors. B) Environmental factors shift the mean of the phenotype. C) Environmental factors shift the variance of the phenotype. D) Environmental factors affecting both the mean and the variance of the phenotype.

The V_E has been treated as a trait itself and different assumptions were suggested to model it, based on the approach made by SanCristobal-Gaudy et al. (1998) [36]. This approach assumed a hierarchical mixed model where an additive genetic effect is fitted on both the phenotypic mean and the exponential log of the V_E (for more details see Hill, 2010 [37]). To solve the mixed models for the V_E different methodologies such as the double hierarchical model (DGLM) solve by maximum likelihood [38], two-step restricted maximum likelihood (REML) [39] and Bayesian methodologies [40] has been proposed. Bayesian methodologies allow to directly estimate the parameters for the V_E. However, they require the implementation of highly complex MCM sampling algorithms (Metropolis Hasting with Langevin proposals) [41-42]. The heritability estimations for the V_E under these methods statistically supported the genetic control of the V_E. The values for these estimates can find in the review by Hills and Mulder (2010) [37] and most recently by Iung et al. (2019) [27], showing heritabilities for V_E ranging from 0 to 0.208 with a mean of 0.027. The large fluctuation in the heritabilities values is due to the extreme sensitivity of the estimate to the model applied (Yang et al., 2011) and the effects included on it [40, 43]).

The first empirical study to prove the segregation of genetic variance for the V_E was made by Mackay and Lyman (2005) [44] using the Drosophila Melanogaster. After that, divergent selection experiments for V_E in mice [28] and rabbits [26, 39] showed a substantial genetic response. In these experiments, a negative correlated response to the selection was observed for traits such as longevity [45], litter size [26, 46], implanted embryo, and foetal survival [46], and mortality and animal resilience [15]. A positive correlation response for productive traits such as birth weight was found for the divergent experiment in mice [28] instead no effect was observed in rabbit for this trait

[39]. Statistical models for V_E also found a negative correlation for health and reproductive trait such mortality, longevity and fertility [12-13, 23-24], as well as for animal resilience [13]. Likewise, a positive correlation was found for milk daily yield [12-13] and birth weight [47-48]. However, the correlation obtained using statistical model difficult the estimation of reliable genetic correlations because the model is highly complex and hyperparametrized, and the estimates are strongly dependent on the scale used for the trait. Correlation between the mean and the variance of productive traits was observed to range from -1 to 1 [27]. Then, a properly knowledge about the biological mechanisms influencing the V_E is necessary to unravel the animal resilience and its correlation with key trait for the livestock industry.

2.1. RABBITS FOR STUDYING THE V_E

Selection for deciphering animal resilience in livestock systems is necessary. The rabbit and the mouse are good animal models for studying it. Both species are easy-management, prolific and have short gestation periods; 21 days for the mouse [49] and 31 for the rabbit. However, the rabbit is of importance in the field of animal production. Animal breeding programs have been developed to improve the meat quality [50-52] and reproductive traits [53-55]. Furthermore, two divergent selection experiments for high and low V_E were successful performed in rabbit [26, 39]. In this thesis we used animals from the divergent selection experiment for V_E of litter size (LS) developed by Blasco et al. (2017) [26]. This divergent selection was performed considering the within-dam variance of LS pre-corrected for year-season and parity-lactation:

$$V_E = \frac{1}{n+1}\sum_{1}^{n}(x_i - \bar{x}) \quad (1)$$

where n is the total number of parities of the dam, x_i is the pre-corrected litter size, and $\bar{x}$ is the average of litter size of the dam. Although the correlation of the V_E using pre-corrected and uncorrected LS showed a genetic correlation of 0.99, meaning that no differences exist when pre-correction is made for LS. This result was subtle different from the found by Poppe et al. (2020) [23] where a moderate correlation from 0.39 to 0.45 was observed when the V_E is computed using the raw data. Anyway, this computation of the V_E avoids artifacts due to the hyperparametrization of the models to calculate the V_E.

The divergent selection experiment considered 125 breeding females and 25 breeding males per divergent line in each generation (see details in Blasco et al., 2017 [26]). The candidate females belonged to the best females with almost 3 or 4 parities. The males

were selected from the offspring of the best female for each sire (best mating). Animals from the divergent lines were coetaneous and under the same environmental conditions. After 10 generations, the rabbit populations showed an outstanding genetic response of 4.5% in the mean of the base populations. Another experiment for V_E in rabbits showed a genetic response of 16% in the mean of the trait after four generations [39]. In mice, a selection for V_E was also successful [28]. Hence, selecting animals for V_E is feasible. Furthermore, the rabbit lines from Blasco et al. (2017) [26] showed a negative genetic correlation with reproductive traits such as implanted embryo (-0.49), embryonic survival (-0.43), foetal survival (-.18) and prenatal survival (-0.45) [46]. A negative correlated response for litter size [46] and resilience was also found [15, 29]. The selection for V_E of LS modified the mean and the variance of the LS (Fig 1). That means that the rabbit line selected for low V_E of LS was more productive for LS and more resilient. Resilience was measured through the identification of immunological biomarkers under stressful conditions to know how the inflammatory response is triggered in the populations. The stressful conditions considered was a challenge of a vaccination with myxomatosis [15] and the moment of the first parity of the dam [29]. Both challenges showed that animals from the line with low V_E of LS were less susceptible to the stress generated since showed low levels of inflammatory biomarkers such as the CRP. Then, these rabbit populations are an exceptional biological material to unravel the mechanisms under the V_E and the animal resilience.

3. OMICS STUDIES

The omics studies emerged with the aim to unveil the biological mechanisms shaping the phenotypes. Next-generation sequencing (NGS) technologies drove the “Omics era”, dawning the field of genomics since it can sequence millions of DNA fragments in a cheaper way. The different omics disciplines emerged with the advancement of modern technology and bioinformatics, allowing to study the different components of a living organism on a large-scale. The different omics disciplines are:

- Genomic: the study of individual’s DNA
- Transcriptomic: the study of the individual’s DNA expression based on the number of copies of transcripts of RNA.
- Epigenomic: the study of the individual`s DNA expression comprising epigenetic marks.
- Proteomic: the study of proteins from the individual
- Metabolomic: the study of the metabolites produced from the individual’s metabolism.

All these disciplines were extended to the study of the microbiome since was demonstrated that living microorganisms in individuals can also contribute to shaping their phenotypes [30, 56]. Hence, an extension of the omics vocabulary was made with the incorporation of the terms: Meta-genomic, Metatranscriptomics, Metaproteomics, and Metabonomic (similar to metabolomic) [57]. Only 1% of the total microbial species can cultivate so these new omics allow to explore a field previously almost unknown.

4. GENOMIC STUDIES

The aim of genome studies is to try to explain the genetic basis of heritable traits. In this thesis, we focus on genome-wide association, signatures of selection, and whole genome sequencing studies.

4.1. GENOME-WIDE ASSOCIATION STUDIES

The genome-wide association studies (GWAS) try to identify variations in the individual's genome associated with phenotypes. Most common GWAS [58-59] are based on the identification of single-nucleotides polymorphisms (SNP). These SNPs are variations in a single position of the DNA which are in linkage disequilibrium (LD) or segregating together with the causal variants, those with an effect on the phenotype [60]. Extensively GWAS has been applied in livestock to identify quantitative trait loci (QTL) associated with key traits with economic repercussions in this industry. For instance, milk yield and somatic cell score in cattle [61-63], carcass weight, intramuscular fat, and fat composition in pigs [64-67], intramuscular fat in rabbits [68-69], and feed intake in chickens [70-71], among others. Most of these quantitative traits are influenced by a huge number of variants each one with a weak effect on the trait [72-74]. Thus, the identification of most of them is still challenging. GWAS only captures a small fraction of the genetic variance, usually due to the identification of those with major additive effects on the phenotype [75].

Limitations in the detection power of GWAS are mainly conditioned by multiple factors [76]. Most important are (i) the SNP chip, (ii) the population studied, (iii) the sample size, (iii) the quantitative trait, (iv) the effect size of the variants, and (v) the methodology applied. (i) SNP chips used to be developed for species using different genetic lines. In livestock, most used SNP chips have between 50K and 200K SNP. That means that for a genome of around 3 billion pair bases (length of the *Oryctolagus cuniculus* genome), the average SNP density ranged from 17 to 67 SNPs per megabase. In such a manner, they do not generally incorporate all SNPs needed to catch all the causal variants contributing genetically to a phenotype. Moreover, SNP density decreases substantially after the

quality control [77] and is greatly dependent on the (ii) population under study and how many SNPs are not polymorphic. Moreover, an inappropriate (iii) sample size hides the true allele frequencies of the population and difficult the identification of its family structure. This hinders the estimate of the variants' effects and increases the ratio of false positives (FP) and negatives (FN) [58-59, 61]. Additionally, the genetic basics of the quantitative trait analysed such as its heritability, the number of variants and their distribution through the genome, and (iv) their effect size determine the statistical power of the GWAS. There is no a (v) methodology consensus to estimate the genetic effect of the variants [78]. Thresholds used to determine if a variant is associated with the trait vary among studies. It must consider all the factors to properly control the ratio of FP and FN [61]. However, it is complicated due to the hyper parametrization of the models. In general, models use two different approximations based on the estimation of the genetic effect marker by marker or accounting for groups of SNPs. Both methodologies, single-marker regression and multiple-marker regression allow the implementation of GWAS [78]. However, we must take both their limitations and benefits into account.

4.1.1. SINGLE-MARKER REGRESSION

SMR consider that the SNPs are independent among them, and the causal variants or QTLs have a strong effect on the phenotype. Models are fitted marker by marker, including the SNP effect as a fixed effect. In general, all SMR models follow the structure:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{z_j}\alpha_j + \boldsymbol{e} \qquad (2)$$

where y is the vector of phenotypic values; X is the matrix of incidence of the systematic effect b; z is the vector with the allele dosage of the SNP j code as 0,1 and 2, according to the number of copies of the reference allele; α is the additive genetic effect of the SNP j; and e is the residual vector normally distributed with mean zero and standard deviation $I\sigma_e^2$. The population stratification can be corrected by including in the model the first five principal components as systematic effects. However, the best way to take the population stratification into account is to include in the model a polygenic effect (g) based on the genomic relationship matrix (GRM) such as the variance-covariance matrix of the distribution. In livestock, the GRM uses to be constructed using the VanRaden (2008) [79] or Yang et al. (2010) [80] estimators. Using these estimators, the SNPs are fitted double in the model, decreasing the detection power of the GWAS [81]. Thus, Yang et al. (2014) [81-82] proposed to leave the chromosome of the tested SNP out to avoid that. Maximum likelihood is used to adjust the models and test the level of "significance" (p-value) using a F-Fisher test or a chi-square test. The threshold to consider when a p-value

rejects the null hypothesis and determines that a SNP is associated with the trait used to be 0.05.

Caution must be taken with the p-value to avoid misinterpretation. The p-value is not a probability and does not show the degree or magnitude of the evidence to reject the null hypothesis. Moreover, it is based on the sample so their values can vary among experiments. Thus, the p-value must be considered only as a qualitative indicator of association or no association. Furthermore, a threshold of 0.05 must be adjusted to reduce the number of false positives due to the multiple testing. False discovery rate (FDR) [83] and Bonferroni correction [84] are the most common to correct for multiple testing, although in GWAS most implemented used to be the Bonferroni correction. Bonferroni correction adjusts the threshold of the p-value dividing it by the number of SNP that are independent. The easy way is to consider that all SNPs are independent, but this reduces drastically the detection power of the GWAS, increasing the number of false negatives. A more realistic approximation is to calculate the number of independent LD blocks to approximate the number of independent SNPs. On the other hand, an empirical p-value can calculate (based on a permutation test) to obtain the distribution of the SNPs under the null hypothesis to establish a proper threshold for the study. In livestock systems, the threshold should be fixed to 1e-4 because artificial selection has been applied to most of the species and their effective sample size are low, so they should have large blocks of LD. However, there is no consensus about the threshold since it highly depends on the genetic architecture of the sample population, so each GWAS should calculate its optimal threshold to control the number of FP and FN in their study.

SMR models do not consider the correlation among the SNPs, generating an overestimation of the additive genetic effects of the SNPs dependent on sample size [85]. In this way, multiple-marker regression might estimate better the SNP effects taking the LD of surrounding SNPs into account.

4.1.2. BAYESIAN MULTIPLE-MARKER REGRESSION

BMMR models were developed by Meuwissen et al. (2001) [86] with genomic prediction purposes and then were adapted for performing GWAS. Multi-maker regression (MMR) models take the correlation among SNPs into account to estimate the genetic effect of the SNPs. Bayesian inference is used to perform the MMR, fitting the effects of the SNPs as random effects. Bayesian inference is based on probabilities since the degree of uncertainty is expressed with a probability distribution of the beliefs about the unknown parameters in the model (θ), the prior distribution. The posterior distribution of the parameters given the data ($P(\theta|y)$ is constructed by joining the data distribution (a

likelihood function) and this prior distribution. This posterior distribution is used to make inferences since it reflects the knowledge about the value of the parameters given the observed data [87-88]. There are different methodologies to perform a BMMR model. Dissimilarities among them are in the information about the size of the SNP effects included in the prior distribution [89-90]. These methodologies are Bayes A, B, C, Cπ or Bayes LASSO, among others [90-93]. In all methods, Markov chain Monte Carlo (MCMC) algorithms (i.e. Gibbs sampling, Metropolis Hastings) are used to sample from the posterior distribution.

What methodology is best relies on the genetic architecture of the trait since the prior distribution highly affects the estimate of the SNPs effects [90]. The Bayes B was the most used by animal breeders because was easy to implement by the GenSel software developed by Garrick and Fernando (2013) [94]. Moreover, its computation efficiency was improved by Cheng et al. (2015) [95] and included in the software JWAS which also optimized other Bayesian Alphabet methods [96]. Bayes B considers that a percentage of the SNPs (π) do not influence the phenotype, and the rest have effect and variance with a probability of 1-π [78]. It is a good methodology for a low sample size to detect SNPs with large effects size on the phenotype [97]. However, the π value conditions the results so different priors must be used to obtain reliable results [93]. In this line, Bayes R might be a better approximation due to allow the use of a mixture of prior distributions [98].

BMMR models have a better estimation of the SNP effect with the use of SNP windows [99]. Additionally, the genetic variance explained by each window can be obtained. The SNP windows can be sliding or discrete. Sliding windows allow the overlapping of SNP among adjacent windows while discrete windows are independent windows of a fixed size. However, there is not reference size for the windows (by SNP number or base pair length), affecting the estimation of the parameters in the model. The detection of SNPs associated with the trait is performed by the computation of the Bayes Factor (BF). BF is the ratio of the marginal posterior probability of each SNP given the null hypothesis and the alternative hypothesis, each one multiplied by its prior [88, 100]. Kass and Raftery (2015) [100] suggested two thresholds to indicate the degree of the evidence to reject the hypothesis of no association: a BF higher than 10 as strong evidence of association and a BF higher than 100 as decisive evidence of association. However, as for the SMR models, the threshold is suggestive because is highly dependent on the assumptions considered to perform the model. Other measures such as the posterior probability of association (PPA) and window posterior probability association (WPPA) [100-101] can be used to identify an association (see the review Wolc et al. 2022 [93] for more details).

GWAS is a useful tool to discover novel biological mechanisms of traits due to a genetic effect on them. However, it is still complicated to extrapolate and validate the results in other populations [75, 102]. The degree of the LD between the SNP markers and the causal variant, the strength of their genetic effects, as well as the heritability of the trait, affect the ability of GWAS to identify and quantify the true QTLs. Therefore, GWAS must be considered an exploratory analysis. The magnitude of the genetic effect and the variance explained by the SNPs are highly influenced by the methodology applied. GWAS must be complemented with other genomic studies and with post-genomic analysis to properly interpret and disclose the biological mechanism under key traits. In this way, the study of signatures of selection may help to deep insight into the biological mechanisms of the genetic variance observed in the traits.

4.2. SIGNATURE OF SELECTION

Signatures of selection (SS) are patterns of positive selection on the DNA where there are advantageous alleles that were or are under selection pressure. SS is generated due to a hitchhiking effect, which is an extension of the selection pressure to neutral alleles surrounding the advantageous allele [103], forming haplotypes. Haplotypes are DNA segments where the patterns of neutral variation have been disrupted and both neutral and advantageous alleles are segregating jointly because are in LD. The increase in the prevalence of a haplotype is due to a selective sweep where the advantageous alleles become more common after selection and therefore their surrounding neutral alleles [104]. Selective sweep can be hard if a new mutation arises and quickly reaches fixation, partial if has not reached yet fixation, and soft if a neutral allele that is segregating in the population becomes favourable at a certain time and increase its frequency [105].

Hard selective sweeps act on rare mutations with a large effect on the phenotype. This generates specific signatures of selection easily trackable such as high levels of LD, long haplotypes, a reduction in the local variability, and a quick increase in allele frequencies that can reach fixation [105]. However, after generations of selection recombination makes LD decay and breaks haplotypes [106]. This event depends on the genomic regions because recombination rates vary widely across the genome [107]. Hard selective sweep is the classic and expected, but the most likely is the soft selective sweep due to the polygenic nature of complex traits. Artificial selection can act over many loci, increasing their allele frequencies at the same time so fixation is not reached. Furthermore, neutral or standing variants on the genome can become advantageous because of the selection target [108-109]. In this line, the signatures of selection in soft

selective sweeps are less pronounced and thus more difficult to identify. Thus, the detection of this one is highly dependent on the sample size.

GWAS is based on the estimation of the effect variants on the phenotype so fails to identify alleles under weak selection or fixed alleles [75]. Thus, mapping signatures of selection allows searching most of the patterns of positive selection regardless of the phenotype [110]. SS methodologies try to detect all pointedly patterns on the DNA produced by selective sweep which explain the evolution of the populations and the improvement of key traits in the livestock systems [111-112]. They focus on detecting SS due to the (i) reduction in local variability, (ii) deviation in the site frequency spectrum, (iii) increment in the LD or extended haplotype, and (iv) differentiation in a single site position. Furthermore, they can be classified into intra- (i, ii, and iii) and inter-population (iv) statistics [111-112]. A recent review of the most common program to detect these signatures of selection could be found in Saravana et al. (2020) [113].

4.2.1 INTRA-POPULATION STATISTICS

i. Reduction in local variability

A common pattern of SS is the increment of homozygous segments in the genome due to a local reduction of the variability in the regions under selection. The most common methodology is the ROH or Runs of Homozygosity. The ROH methodology tries to identify continuous segments where on average the individual showed a higher number of homozygous sites regardless of the entire genome [114]. This methodology also allows to identify of past and recent selection events, depending on the extension of the ROH, as well as determining in a better way the inbreeding coefficient in the population [115-116]. The heterozygosity depression across the genome through the pooled heterozygosity (H_P) can also use as a measure of the reduction in the local variability [117].

ii. Deviation in the site frequency spectrum

In this category, statistics compute the deviations of the allele frequencies from the distribution under the neutral model or no selection [118-119]. A selective sweep distorts the allele frequencies in the region harbouring the advantageous allele going out of the neutral distribution. The most common statistics are Tajima's D (TD) [120], Fay and Wu's H [121], and composite-likelihood ratio (CLR) [122-123]. TD cannot distinguish between rare and high frequency new mutations because no ancestral genome is used which is supplied by Fay and Wu's H with the inference of the ancestral alleles from an

external population. CLR implements a coalescence simulation to obtain the neutral distribution of the allele frequencies under the hypothesis of no selection [123]. The null distribution can be based on the data itself [104] or on the allele frequencies observed in the overall breed compared with the distribution of a subset of breed selected for a key trait [124].

iii. Variation in the linkage disequilibrium

The following statistic measures the extension in the LD patterns after a hitchhiking effect. Advantageous alleles take time to reach high frequencies. During this time LD trends decay and haplotypes break [106, 125]. In this line, Sabeti et al. (2002) [125] developed the long-range haplotype (LRH) method to compare the extension of homozygous haplotypes in a region with their frequency in the population. For that, a core haplotype using the LD decay is computed to then calculate the extended homozygous haplotype (EHH), that is the probability that the core haplotype in two random chromosomes will be identical by descent. The recombination rate is used to normalize its values [125] due LD differences across the genome [126]. The integrated haplotype score (iHS) is an expansion of EHH including the recombination distance and determining the ancestral and derived alleles to compare the EHH in the population [127]. However, it requires phasing haplotypes. In events where the inference of the haplotype is not possible or is not so accurate, we can detect high allele frequencies by the identification of excessive LD with the LD decay (LDD) statistics [128].

4.2.2 INTER-POPULATION STATISTICS

iv. Differentiation in a single site position

This pattern of positive SS is based on the differences in allele frequencies among populations. These differences can be measured by computing the fixation index or inbreeding coefficients F_{ST}, F_{IT} and F_{IS} [129]. Most used is the F_{ST} which harbouring the F_{IT} and F_{IS} in its definition and determines the correlation between random gametes of a subpopulation relative to the entire population. All used approximations are based on the allele frequencies and the degree of heterozygosity in the populations [130-132]. The Nei (1973) [130] and Weir & Cockerham (1984) [131] approximations are the most used. Although the former does not consider the sample size, and latter tends to overestimate the F_{ST} values [133]. Moreover, F_{ST} statistics are for populations that diverged from the same ancestor population. FLK statistic deal with and take the population tree structure into account with the inclusion of the coancestry matrix in the model [134] and can also compare haplotypes frequencies (hapFLK [135]). Bayesian approaches are also

implemented to compute the F_{ST} [136-138]. On the other hand, an extension of EHH (XP-EHH and Rsb) and CLR (XP-CLR) statistics were made to also allow comparisons among populations. They compare the EHH or CLR at the same position among populations [106, 139-140]. Ancient selection before the divergence can be tested using the 3P-CLR approach [141]. Finally, genomic diversity among populations can be also detected by the study of the variation in the LD between two populations [142-143]

SS identified by each methodology could differ among them because they used different assumptions to detect the SS [111,113]. For example, iHS identified segregating haplotypes whereas XP-EHH reveals haplotypes close to fixation [106] as well as Fay and Wu's H statistic can identify high allele frequencies and Tajima's D low-to-medium alleles frequencies [144]. The heritability of traits and their polygenic nature could hide some SS due to a soft selective sweep [105]. Low correlations among statistics were reported by Ma et al. (2015) [145], González-Rodríguez et al. (2016) [146], and Sosa-Madrid et al. (2020) [147]. Thus, the use of multiple methodologies was suggested to increase the potential to detect SS [111, 148]. Combining the information of the different statistics and taking their correlation into account could be implemented using the DCMS method [145, 149]. Overall, the statistics for the identification of positive SS are affected by an SNP ascertainment bias and SNP chip limitations [111, 150-151]. Furthermore, statistics not based on variations in the linkage disequilibrium tend to overestimate their values because the surrounding alleles are large LD with the causal variants. The improvement of the estimation of these parameters could be achieved by considering sliding windows for their calculation [111], as well as knowing the genomic structure of the ancestor population [105]. The use of whole-genome sequencing data could help to complete the gaps to achieve the entire genetic architecture of the populations and improve the detection of SS and causal variant by GWAS.

4.3. WHOLE-GENOME SEQUENCING

Whole-genome sequencing (WGS) is a promising technology that overcomes most of the limitations of SNP chips mentioned before. This technology allows to obtain millions of DNA fragments in parallel, sequencing the whole individual genome. WGS scrutinizes the entire genome and releases the identification of most of the variants, increasing the probability to discover causal variants and capture the overall genetic variation that shapes complex traits. WGS has become cheaper and cheaper over the years, being available for a wide range of experiments. Nowadays, second- (SGS) and third generations sequencing (TGS) technologies are competing in the market [152]. Differences between both technologies lie in the longitude of the sequenced reads, being

reads from SGS technologies shorter [152-154]. The short reads of SGS technologies allow to successfully identify single nucleotides variants (SNVs) and short indels but fails in the identification of most structural variations as copy number variations due to repeated sequences [155-156]. Moreover, SGS are highly influence by PCR bias such as the GC content across the genome, being this one inefficient to amplify regions with low or high GC content [157]. TGS technologies try to deal with all the SGS limitations [156, 158]. Their methodologies do not depend on PCR amplification, so they are not affected by its bias. Moreover, they might resolve the problem of identifying most of the structural variants due to their long reads [159-160] and of improving genome assembly [154]. The most relevant TGS technologies until now are the SMRT of PacBio [161] and Nanopore sequencing of Oxford [162] Nanopore Technologies while for SGS are the Illumina platforms which dominates the market [152].

In livestock, there is a strong LD among variants, so the identification of the causal variants using SNP chips is challenging. The massive information generated by WGS technologies can help to identify directly all the causal variants under the domestication process of livestock species [163-166], as well as those underlying adaptative and productive traits [167-170]. However, unless WGS technologies are nowadays cheaper, it is still costly compared with genotyping with SNP chips. Furthermore, proper sequencing coverage is needed to accurately call the variants across the genome and calculate their allele frequencies to estimate their effect and identify the causal variant [171]. In general, WGS studies use a quite low sample size per population or breed which limits the statistical power to extract reliable estimates and results. The polygenic nature of most traits makes it challenging. Variants with a small effect on the phenotype require a large sample size to detect them. Moreover, WGS generated massive useless information due to the large LD of the livestock populations, increasing the computational cost of this analysis. Resequencing of target regions with WGS technologies has been proposed as an alternative to screen feasibly thousands of individual genomes [172]. In this way, we can explore the non-coding and coding variants to identify the causal variant in candidate GWAS regions [173]. This also allows to reduce the WGS data, although we lost information from the regions that were not identified by the GWAS as QTL. Another approximation to reduce the cost of WGS is to perform whole exome sequencing (WES) since coding regions seem to harbour most of the variants with high effect on phenotypes [174]. WES enables to sequence all coding genomic regions which represent only 2% of the WGS data. However, the location of the causal variant depends on the genetic architecture of the population and the trait. Hence, with WES we lost information of the noncoding regions which can also contain important variants to shape phenotypes [175]. On the other hand, WGS was proposed to do in DNA pools of representative animals for

identifying most of the variants underlying traits with an acceptable sample size [176]. This approach was successfully implemented in pigs and sheep allowing to sequence the whole genome of a representative population of hundred animals [177]. However, although the use of DNA pools gives reliable estimates of the allele frequencies, this methodology confounds low frequent variants with sequencing errors avoiding to accurately identify rare variants.

Despite of the promising benefits of the WGS, no clear advantages of WGS over SNP chips have yet been identified to perform GWAS. Furthermore, the use of WGS instead of SNP chips did not increase substantially the prediction performance of the models, showing an average improvement of 2.5 to 4.2 percentage points depending on the trait and the population analysed [178]. Computational cost limits the sample population which avoids to properly accurate WGS data. A low sample size requires a selection of the animals to sequences avoiding well-defined genetic architecture of the trait in terms of LD, calling variants, and allele frequencies. Moreover, WGS pipelines must be optimized to process the data. Genotyping with SNP chips allows to successfully identify the common genomic regions associated with key traits [179], patterns of positive selection [113] as well as copy number variations regions [180]. Although it fails in identify rare variants and large structural mutations as deletions and insertions. Considering the limitation of the WGS, this one might be used as a complementary methodology to GWAS using SNP chips because those are more economically feasible and can be performed at large scale. Nowadays, WGS data might help to discover the causal variants in candidate genomic regions identified using SNP chips until WGS becomes feasible for large sample sizes.

4.4. PREVIOUS GENOMIC ANALYSIS IN V_E AND RESILIENCE

Candidate genes for animal resilience were identified using different resilience indicators. In small ruminants, genes of Toll Like Receptors (TLR), Heat Shock Protein (HSP) family, gonadotrophin hormone, as well as *BMP2*, *BMP7,* and *LEP* genes were associated with resilience to heat stress (see review Joy et al., 2020 [181]). Houda et al. (2022) [182] showed that prostaglandin receptor genes (*PTGDR2*), genes from membrane-spanning 4A family (MS4A), *GRB10*, *IKZF1*, *MYD88,* and *RNASEL* genes were some of the genes associated with disease resilience. The QTLs identified in that study explained 26.95% of the genetic variance for the resilience indicator used (deviations from the expected body weight) [182]. Likewise, *RSAD2* (virus inhibitory protein), *IL17RA, TRIM39* genes, and the MHC region were also associated with disease

resilience in pigs, among others [183]. Genes for a correct neural response and heat tolerance were found for resilience to heat stress in cattle [20] and for olfactory receptor, stress response, and immune response in sheep [184].

V_E could be a global measure of resilience. Several genomic analyses sought to identify quantitative trait loci associated with the V_E (vQTL) using different livestock species; pigs [185-186], cattle [187-189] and chicken [190]. Overall, vQTLs explained a low genetic variance of V_E ranging from 0.56 to 8%, except for Wolc et al. (2012) [190] and Iung et al. (2018) [189] which explained 16 and 20% of the genetic variance, respectively. Candidate genes found were involved in pathways related to the stress and immune response, and the development of sensory and neural structures. The genes from the heat-shock proteins (HSP) or related to them were also identified by Sell-Kubiak et al. (2015) [185] in pigs, Iung et al. (2018) [189] in cattle and Morgante et al. (2015) [191] in Drosophila Melanogaster. Independently the resilience biomarker used seemed that the HSP protein is critical to modulate the V_E. HSP are related to the response to heat stress by boosting the immune response and countering cell damage [192-194]so could also be key for modulating animal resilience. Candidate genes identified by all resilience biomarkers supported the relevance of the immune system to control animal resilience [2, 195].

5. MICROBIOME STUDIES

In a traditional framework, the phenotypes variability observed were assumed to be a consequence of the individual genome and the environment [32]. However, microorganisms (archaea, bacteria, eukaryotes, and viruses) which live in each individual can also shift their phenotypes [30, 196]. According to the definition proposed by Marchesi and Ravel [57], these microorganisms, their genomes, and their habitat or surrounding environment (e.g., tissues) assemble the individual microbiome. The consortium of the individual genome and the microbial genomes is called hologenome and can drive the evolution and adaption of animals and plants [197-199]. Because of this coevolution, animals and plant breeder became to incorporate in their models the microbiome as another source of phenotypic variability which can affect both the mean and the variance of traits (see review of Henry et al. 2021 [30]).

The largest community of functional microbial species in animals is from the gut with a density around 10^{13} cells/mL of content [200-202]. Its implications to health and disease have been widely studied (see reviews [56, 203-205]. However, gut microbiome inheritance is complex which complicate its modelling and reliability across traits and species. In mammals, its inheritance starts with the colonization of microbial species

from the dam. Gut microbiome from the dam can reach the placenta [206-207] and the amniotic fluid [207-208] through dam's blood stream. Thus, colonization of bacteria could begin in utero when the foetus absorbs amniotic fluid. This was supported by the high similarity between the microbiome composition of the meconium (earliest stool) and the amniotic fluid [209]. After that, birth route (caesarean section or vaginal-delivery) [210-212], suckling and maternal environment as well as other environmental conditions can influence the acquisition of microbial species by the offspring [213]. In rabbit, coprophagous behaviour in pups may also be important to establish a properly bacterial community in the gut [214-215].

Incursions of microbial species are produced until a mature immune system is completely developed. Education of host immunity is essential to establish the normal microbiota and develop an immune system which protects individual against pathogens [216]. Suckling is important for that because the dam can pass bacterial components and antibodies through the milk, which is an advantage for colonization of the gut by maternal microbial species [217]. Daft et al. (2015) [218] showed how nursing mothers influence the gut microbiome of fostered mice more than their biological dams. However, this study removed the solid feed which seems to be decisive for the composition of the stable microbiome. Bian et al. (2016) [219] showed in a cross-fostering model using piglets that the postweaning and the intake of solid feed are key establishing the gut microbiome composition and diversity of the piglets, decreasing the effect of the suckling. The nursing mother determined only a few microbial species in the adult pigs [219]. Likewise, Mulder et al. (2009) [220] suggested that housing environment is also a major factor for determining the stable microbiome, showing that animals under a high-hygiene conditions had higher microbial diversity than those animals housed outdoor. High microbial diversity is related to a healthy and resilient gut because allow individual to cope better with gut disturbances [221-222]. However, the acquisition of microbial species with healthy properties may be most important than the diversity because directly control the growth of harmful bacteria [223]. Although multiple factors can affect the ability of bacteria to respond to stressful conditions (see review Shade et al., 2012 [224]).

Variability in the microbiome composition among individuals can be also a consequence of the host genetics due to microbial heritability [225-227]. Host genome can affect around 5-10% of the microbiome variability [225, 227] which means that microbiome composition fluctuations seem to be dominated by environmental factors over the host genetics [228]. Grieneisen et al. (2021) [226] showed that most of the microbial species (97%) were heritable with low to moderate microbial heritability. Moreover, the

microbial heritability of the family *Christensenellaceae* seems to be the most consistent across experiments [226-227, 229]. However, correlation among microbial species is not considered in these models, leading to imprecise heritability estimates. Moreover, there is an overestimation of the number of heritable taxa due to the huge number of models fitted [230]. Ever, the microbiome is influenced by a multitude of factors so deciphering it is still a challenge (for a complete review of how the microbiome is influenced across the lifetime see Martino et al., 2022 [231]).

Phenotypes are shifted by the microbiome composition [197]. Differences in gut microbiome composition have been widely found in key traits for livestock such as feed intake [232-233], growth performance, and carcass quality in pigs [234], feed intake in chickens [235], milk protein yield [236] and methane emissions in cattle [237], and intramuscular fat in rabbits [238]. Faecal transplantation in germ free animals supported that because receptor animals showed similar phenotypes to the donor [239-242]. Hence, microbiome composition dissimilarities allow to explain a percentage of the total phenotypic variance. Diffort et al (2016) [243] was the first to name this phenomenon as microbiability (m^2), that is the proportion of the phenotypic variance explained by the microbiome. In humans, m^2 was 0.36 for lactose consumption and HDL cholesterol, between 0.22-0.36 to body mass index and 0.25 for glycaemic status [228].

Table 1. Heritability (h^2) and microbiability (m^2) in different traits for some livestock species.

Specie	Trait	h^2	m^2	Reference
Cattle				
	Fat %	0.19	0.08	Buitenhuis et al., 2019 [244]
	Protein %	0.18	0.08	Buitenhuis et al., 2019 [244]
	Methane emissions	0.21	0.13	Diffort et al., 2018 [243]
		0.14	0.16	Ramayo-Caldas et al., 2020 [245]
		0.15-0.22(e)	0.15-0.3(e)	Saborio-Montero et al., 2021 [246]
Pig				
	Feed Intake	0.11	0.16	Camarinha-Silva et al., 2017 [247]
	RFI(a)	0.30	0.11	Aliakbari et al., 2022 [248]
		0.24-0.32(f)	0.18-0.33(f)	Derú et al., 2022 [233]
	FCR(b)	0.19	0.21	Camarinha-Silva et al., 2017 [247]
		0.31	0.20	Aliakbari et al., 2022 [248]
		0.23-0.27(f)	0.14	Derú et al., 2022 [233]
	ADG(c)	0.47	0.03	Aliakbari et al., 2022 [248]
		0.22-0.29(f)	0.17-0.20(f)	Derú et al., 2022 [233]
	IMF(d)	0.54	0.03-0.06	Khanal et al., 2021 [249]
Rabbit				
	ADG(c)	0.05-0.14(e)	0.49-0.79(e)	Velasco-Galilea et al., 2021 [250]

(a)Residual feed intake; (b)Feed conversion rate; (c)Average of daily gain; (d)Intramuscular fat; (e)The estimates depend on the microbiota relationship matrix used; (f) The estimates depend on the composition of fiber in the diet.

In livestock, several studies reported values of microbiability for key traits such residual feed intake, methane emissions, and intramuscular fat (Table 1). The value for

microbiability is highly dependent on the trait and the species used. Moreover, the model approximation could generate subtle differences in their values [246, 250]. Most of the value for m^2 showed that microbiome contributes in a low to medium degree to traits with quite similar values for h^2 and m^2 [243, 245-248]. Exceptions were found for Average daily gain (ADG) and intramuscular fat (IMF) in pigs where genetic contribution dominates over the microbiome composition, unlike in rabbits for ADG (Table 1). High values from 0.49 to 0.79 were found for the m^2 of ADG in rabbits in contrast to that found by Aliakbari et al. (2022) [248] in pigs which were almost null. These studies indicate that m^2 is specific for traits and species and may not have the same value as the h^2.

5.1. IDENTIFYING THE MICROBIAL SPECIES DIVERSITY

Two different NGS methods are used to identify and quantify the microbiome: amplicon-based approach [251-252] and shotgun metagenomic sequencing [253-254]. The former is based on the sequencing of the 16s rRNA. The latter is based on the WGS approach since the whole metagenome of the entire microbial community is sequenced using untargeted primer sequences.

Amplicon-based approach sequences directly the 16s rRNA sequence, the stable and the hypervariable regions. The hypervariable region harbours the differences between microbial species allowing their assignment. To determine the microbial diversity in the sample, the reads are clustered in operational taxonomic units (OUT [255]) with a similarity in their sequences higher than 97% [256]. For that, different strategies such as close-related clustering, open-reference clustering and de *novo* clustering can be used [257]. Most used is de *novo* clustering because outperforms the other strategies [258-259] and was implemented in the most used software such as QIIME [260] and Mothur [261]. De *novo* clustering does not use an external, predefined database for clustering, so OTUs identification is not restricted to the number of reference OTU in the database, avoiding loss of information [257, 262]. However, it requires a higher computational cost and the OTUs cannot be reproduced in other studies because depend on the dataset used to define them. The threshold of 97% of similarity for clustering reads in the same OTU is adequate for classifying at species level but avoids detecting intra-species variations [263]. For that, amplicon single variant (ASV) was proposed as an optimal approach to detect most of the microbial variability in the samples [264]. This approach avoids using arbitrary thresholds [263] as cluster contains similar sequences only differing in a single nucleotide, capturing all the biological variation of the dataset. Furthermore, ASV can be compared among datasets [265]. QIIME software incorporates the identification of ASVs through the DADA2 pipeline [264]. After the clustering of the read sequences, the taxa

annotation is obtained by comparing the ASVs and OTUs with public databases as Silva [266]. Amplicon-based sequencing variants are useful to unravel the bacteria and archaea or fungi (using internal transcribed spacer; ITS [267]) diversity in the samples, but they cannot capture the differences in other microorganisms such as eukaryotes and viruses. Moreover, pipelines are not developed to determine the genes and function of the sequences in the sample. In this line, shotgun metagenomic is a more powerful methodology to extract all the information about the microbiome composition.

Metagenomic allows to extract information of both the taxonomic composition and the genetic resource (functional information) of all microorganisms in a specific environment. For that, the reads obtained after the shotgun sequencing process must be assembled in large genomic fragments named contigs to then assign their functions (genes) and taxonomies (see review Lapidus and Korobeynikov, 2021 [268]). The contig assembly is a highly computational cost step in metagenomic studies. It depends on the algorithm used (Overlap-layout-consensus or De Brujin graph [269]), the coverage depth, and the complexity of the sample (number of different species) [270]. Low-abundant species, bacterial strain mixtures and repetitive sequences should induce errors or chimeras (sequence incorrectly joined) in the assembly. Moreover, a lot of reads are lost in this process because they do not assemble in any contig [271]. Coassembly approach was proposed to improve the percentage of mapping reads because it considers all the reads of all the samples for the construction of contigs [271]. After contig assembly, gene prediction is performed on each contig [272] which will allow for annotating their taxonomies and functions through the software Diamond [273]. Public databases such as GeneBank nr database, NCBI taxonomy, or GTDB taxonomy [274] can be used for taxonomic annotation, and eggNOG database [275-276], KEGG ID database [277] or InterPro [278] for functional annotations. Taxonomic annotation requires another step to assign a taxonomy rank to each contig. The last common ancestor approach implemented in MEGAN [279] is one of the most used. In this case, the last common ancestor is used to define the taxonomy rank of the contig when it matches with multiple genomes. This approach is also used for the taxonomic assignment using amplicon-based sequencing [280] and seems to work properly for both short and long reads [281]. The tools QIIME [280], SqueezeMeta [271], and Kraken [282] implemented this methodology to assign the taxonomy.

Lastly, the abundance of each taxonomy and functional assignment is determined by counting the total number of reads annotated in each OUT, ASV or contig database. Taxonomy databases can be formed by phylum to species rank, depending on the degree of information available in each database.

5.1.1. NORMALIZATION

The absolute abundance of microbial species is simply impossible to obtain. NGS technologies allow to observe of a random sampling of the absolute microbiome composition. Hence, the observed counts are a part of the real absolute values for each microbiome feature (microbial species, genes, OTUs or ASVs), restricted to a constant sum imposed by the NGS instrument. In this line, microbiome data is compositional, and the use of proper normalization or data transformation is necessary to avoid misleading in the identification of the relevant microbiome features [283]. The most common normalization is the Total-Sum scaling (TSS) and rarefaction [284]. The former converts the data to "relative abundance" considering the total number of reads sequences and the latter makes a subsample of equal size to the observed data [285]. None of them correct or consider the compositional nature of the data since TSS is still a constant sum [283] and the rarefaction is a subcomposition of the data, still resulting in spurious correlations [283, 286]. Inference based on these kinds of normalized data can generate false positive [287] and misleading information [283]. The only way is to extract the information from the ratios between variables [283, 288].

Log-ratio transformations are based on Aitchison methodology (1986) [288] and it was proposed to resolve the problem of the compositional nature of microbiome data. This included a new challenge, because zero values are not supported by logarithms and in microbiome data has a large number of zeros [289]. In a previous step, filtering uninformative variables (low abundance variable or with a high number of zero) are necessary to avoid false positives [290]. After that, the remaining zeros can be imputed using different strategies. Pseudo-count imputation is easier since all zeros are replaced by a small value that is arbitrary. However, this imputation assumes that all zeros are non-biological or sampling zeros due to the sample collection or an inefficient amplification [291]. Moreover, the analyses are sensitive to the arbitrary pseudo-count used [291-292]. Other strategies try to model the zeros using zero-inflated Gaussian models [293], zero-inflated beta regression [294], zero-inflated negative binomial generalized linear model [295], negative binomial regression [296] or Bayesian-multiplicative inference which involves a Dirichlet distribution [297]. However, no consensus on the way to impute the zeros in microbiome data is reached.

For compositional data, different log-ratio transformations can be used: (i) additive log-ratio transformation (ALR), centered log-ratio transformation (CLR) or isometric log-ratio transformation (ILR). ALR considers a reference variable to construct all the ratios. However, depending on the reference variable, the transformed dataset could be no

isometry. CLR transformation is isometry because used the geometric mean of all taxa to perform the ratios. However, the interpretation of the result is complicated because all variables are included in the denominator. ILR constructs the ratio based on amalgamations among the variables, for instance, $ILR = \frac{a+b+c}{z+y}$, where a, b, c, z, y are different variables from the datasets. This is a good approximation when you know what ILR are interested. In another way, the calculation of all the ILR from all the variables has a high computational cost. Moreover, the biological interpretation of the ratio between the amalgamation of variables could be complicated. In a biological view, ALR is easier to interpret because the same reference variable is used to calculate all the ratios. Greenacre et al. (2021) [298] proposed that the reference variable must be the one with the lowest variance which after a Procrustes analysis allows maintaining the isometry in the data.

5.1.2. STATISTICAL ANALYSIS

(i) Differential analyses (DA), (ii) alpha- and beta-diversity indexes and (iii) linear mixed models are used to study the microbiome. (i) DA analysis and (ii) alpha- and beta-diversity allow identifying dissimilarities among groups in the microbiome composition. (i) DA analysis is more sensitive allowing to detect features which can act as biomarkers for specific group or phenotypes. (ii) Alpha- and beta-diversity is a more conservative method which indicate if there are or not global differences in the microbial species composition among groups. (iii) Linear mixed models are used to study the contribution of the microbiome to key phenotypes.

i. Differential abundance analysis

Differential abundance (DA) analyses test if there are dissimilarities in the microbiome due to the identification of relevant features among groups. Each methodology applied its own normalization or transformation of the data [299-300] to then models the data under different assumptions for hypothesis testing (see reviews Lutz et al. 2022 [301]). Most of the tests are univariate so a false discovery rate (FDR) correction [83] is performed in most of the methodologies to correct the multiple testing. Multivariate methods such as partial least square discriminant analysis (PLS-DA) [302] and the zero-inflated generalized Dirichlet-multinomial model [303] can avoid univariate models. However, no advantages of multivariate over univariate methods have been observed [304]. DA analysis can be implemented using different methodologies available in R: DESeq2 [305], edgeR [306], limma [307], ALDEx2 [287], metagenomeSeq [293],

Corncob [308] and ANCOM [299, 309], PLS-DA (mixOmics) [310] among others (see reviews from Calgaro et al. 2020 [311] and Nearing et al. 2022 [312]).

The identification of microbiome features is highly dependent on the methodology used. For instance, the methodologies used to identify differential abundant ASVs gave different results across 38 different datasets [312]. ALDEx2 and ANCOM-II was the most conservative methods because identified the lowest number of relevant ASVs. Moreover, most of the variables identified by them are also identified by all the other methodologies suggesting that these ASVs are likely true positives. Both ALDEX2 and ANCOM-II are methodologies that include log-ratio transformations; CLR in ALDEx2 [287] and ALR in ANCOM-II [309]. Another promising methodology for differential abundance is the partial least square discriminant analysis (PLS-DA). PLS-DA is a multivariate method that can reduce the dimensionality of the data to remove noise. It tries to maximize the covariance between a categorical variable (dependent variable) and its features (independent variables) [302]. This allow to properly calculate the contribution of each feature (metabolites, taxa or genes) on the prediction and classification of the populations, leading to the high collinearity of the datasets. However, this methodology seemed overfit when the number of features is larger than the number of samples, reaching an over-optimistic model [313]. Thus, validation test such as permutation test and cross-validation must be used to check the results. Nevertheless, there is not a consensus methodology to analyse this kind of data to search significant differences in the features abundance among samples. Nearing et al. (2022) [312] suggest used multiple methodologies and compare the result among them.

ii. Alpha- and beta-diversity

Alpha-diversity indicates the microbial species richness and abundance within a sample whereas beta-diversity determines the degree of (dis)similarities in microbial species composition between samples [314]. Indexes for estimating alpha-diversity are ACE index, Chao1 index, Shannon index, and (inverse) Simpson index. ACE [315] and Chao1 [316] index measure the number of observed microbial species in the sample, corrected by a factor to avoid underestimation of the real richness due to the presence of rare microbial species [317]. Simpson index [318] measures the presence of dominant species in the sample. Both richness and dominance are combined in the Shannon index [319]. Non-parametric Wilcoxon test can use to obtain the relevance of the difference in alpha-diversity among samples. To measure the beta-diversity Bray-Curtis distance, UniFrac distance and Jaccard distance can be calculate. Bray Curtis distance [320] determines the dissimilarity in microbial species composition between groups or treatments. It

considers both the abundance and the number of microbial species in each group. UniFrac [321] distance uses the phylogenetic tree to compare the samples. Jaccard distance measures the difference in presence or absence of microbial species between the groups. Jaccard distance does not consider the abundance of the microbial species. Beta-diversity values can be plotted using ordination plots such PCoA or non-multidimensional scaling (NMDS) and the relevance for the differences in their values among groups can be tested using a permutational multivariate analysis of the variance (PERMANOVA), prior verification of homoscedasticity of the variances.

Each diversity measure gives different information about the microbiome. For that, a combination of them must be used to evaluate the microbiome diversity in the sample. The *vegan* packages in R [322] implements most of the alpha- and beta-diversity measures in an easy way.

iii. Linear mixed models

Linear mixed models allow to estimate the proportion of the phenotypic variance that is explained by the microbiome i.e., the microbiability or m^2. There is not a consensus model to estimate it because it depends to a large extent on the trait under study. Moreover, there are fluctuations in the m^2 estimates due to the use of different approaches to compute the microbial relationship matrix (see Table 1). The first approach used to compute the microbial relationship matrix (M) was developed by Ross et al. (2012) [323], being $M = \frac{XX'}{n}$; where X is the log relative abundance of the OTU autoscaled. Different research used this approach to compute the m^2 of key traits in pigs [248-249], cows [243, 245-246], and rabbits [250]. Camarinha-Silva et al. (2017) [247] proposed a quite similar approach to Ross et al. (2012) [323] where $M = \frac{1}{n \cdot XX'}$ for off-diagonal values and $M = \frac{1}{n}\sum_{k=1}^{n} X_{ik}^2$ for diagonal values. Moreover, distance matrix based on beta-diversity estimates such as Bray Curtis dissimilarity [245, 250] and Weighted Unifrac distance [250] were used to compute M, as well as other matrix distance such as Jensen-Shannon distance [324], and Mahalanobis distance, among others [246]. There is not consensus on the approach to compute the microbial relationship matrix. Saborio-Montero et al. (2021) [246] showed that the correlation among the different M deeply vary among approaches, explaining the differences in the estimation of m^2. Furthermore, the algorithm used for estimating it (restricted maximum likelihood or Bayesian alphabet) also influence the m^2 estimates. Hence, simulation studies could be useful to define the proper methodology to estimate the m^2 [325].

5.2. GUT METABOLOME

One of the principal functions of the microbial species from the gut is to convert the non-digestible nutrients into absorbable components for the host. Digestion of microbial species produces a wide gut-derived metabolite with detrimental or beneficial effects on the host [202, 326]. The production and abundance of these metabolites depend on the diet and the microbiome composition, which is affected by a huge number of factors [327]. Likewise, these metabolites also affect the microbial species community due to the regulation of functions such as the quorum sensing. Quorum sensing is important to establish gut microbiome behaviour and affect the pathogenesis of some microbial species [328-329]. Thus, both microbial species communities and their derived metabolites determine the microbiome composition which establishes a crosstalk with the host to shape phenotypes.

The identification of gut-metabolites can be targeted or untargeted depending on if we know the compounds that we want to identify or by contrast we want an overview of the whole metabolites. Different technologies and sample preparations are needed to identify most of the metabolites due to their chemical nature. Gas or liquid chromatography coupled to mass spectrometry (MS) are the most used techniques to separate and identify small molecules of less than 1,500 Da [330-331]. Gas chromatography showed a high performance but is only for volatile organic compounds such as short chain fatty acids (SCFAs). Liquid chromatography techniques are easier than the gas chromatographic because it does not require sample volatility. Moreover, it allows identifying a wide number of compounds in the sample [332-333]. Other techniques as capillary electrospray coupled to MS and fourier transform infrared spectroscopy as well as nuclear magnetic resonance (NMR) spectroscopy can be used in metabolome studies (see Vernocchi et al. 2016 [331] for more details).

De Vos et al. (2022) [56] made a complete review of the most important metabolites identified in the gut as well as their implications on the host. Short chain fatty acids (SCFAs), bile acid derivates and metabolites from the aromatic amino acid metabolism (AAAs) are the metabolites more studied [56, 334-335]. They can interplay with the brain, liver, and immune system modulating host's health. Beneficial effects have been extensively reported for gut SCFAs production due to their anti-inflammatory properties [336-339]. The effect of the metabolites from the AAAs is still unclear as depending on the compound and its target they could induce disease or not in the host [340]. For instance, indole-3-acetic acid (I3A) has anti-inflammatory effect in the liver [341] whereas the end-product 3-(4-hydroxyphenyl)lactate is a biomarker of hepatic inflammation and steatosis [342]. Moreover, gut metabolites allow to establish a dialog

between the gut microbiota and the intestinal epithelial cells (IECs) to maintain the integrity of the intestinal epithelial barrier [343]. The integrity of IECs is crucial to modulate gut immunity and preventing infections [344]. Gut metabolites can act directly or indirectly on the organ and tissues and the effect could be influenced by other molecules and factors in the host. However, the biological mechanisms underlying the crosstalk between the gut and the host are still unclear, so the effect of the gut metabolites must be studied carefully. Moreover, their origin (microbiota, host or diet) must be properly assigned to know their real implications [345-346].

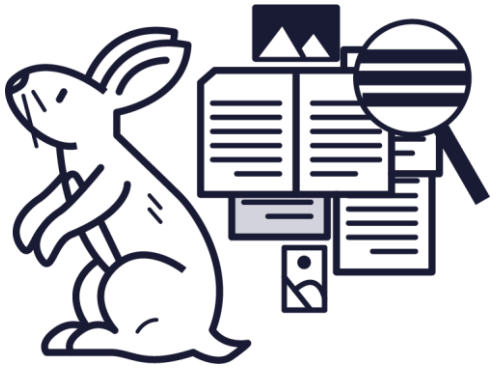
In livestock, the study of the gut metabolome is not as common as the metagenome analysis. Recent studies in cattle and pigs tried to identify the gut metabolites differentially abundant for feed efficiency [347-349]. The results in the pigs' studies were consistent since they found metabolites or derivates of the bile acids and vitamin D in the animals with high feed efficiency [348-349]. For cattle, the results were different to the pig studies because animals with high feed efficiency showed a great abundance of essential amino acids such as L-Serine, L-Tyrosine, L-Glutamate and L-Lysine, among others [347]. However, there are species with highly different gastrointestinal tract. On the other hand, the rumen metabolome was studied for underlaying its relationship with the milk yield protein in cattle [350].

5.3. MICROBIOME COMPOSITION FOR V_E AND RESILIENCE.

Alberdi et al. (2016) [351] suggested that gut microbiota is a key factor to detect environmental fluctuations. Genetic selection is a slow process so the modulation of the microbiome composition could allow faster acclimatisation of species to environmental disturbances [351]. A challenging experiment in rats suggested that the genus *Lactobacillus* could develop a protective mechanism to cope with heat stress conditions [352]. In *Daphnia magna,* their inoculation with different microbiomes (after an exposition with toxic cyanobacteria) showed an effect on individual fitness and survival [353]. Moreover, in coral reefs, the microbiome was identified as a potential mechanism to cope with the environmental factor due to the horizontal gene transfer among bacteria, mutations in the microbial genomes, modification of the abundance of the microbial species and/or the acquisition of beneficial microbial species from the environment [354].

Gut microbiome is still an emerging field in the livestock industry, so no studies have been performed specifically for environmental variance or resilience yet. However, the implications of the gut microbiome in the homeostasis of the immune system and stress

susceptibility are widely studied [355-358], supporting its effect on the health status of the individuals [56, 205]. A comparison between germ-free mice and those with transferred wild gut microbiota showed how wild gut microbiota help to cope with the infection by the influenza virus. The mice with transferred microbiota showed high survival compared with the germ-free mice [359]. Furthermore, the ability of the gut microbiome to quickly restore its initial taxonomy and functionality (to be resilient) also affects health and disease [360-361] Thus, the gut microbiome composition could partially influence the animal resilience through the modulation of the immune system [195]. Thus, promising results may obtain by the study of how the gut microbiome affects animal resilience.



6. REFERENCES

1. Hansen J, Sato M, Ruedy R. Perception of climate change. Proc Natl Acad Sci U S A. 2012;109(37):E2415-E2423. doi:10.1073/pnas.1205276109

2. Colditz IG, Hine BC. Resilience in farm animals: biology, management, breeding and implications for animal welfare. Anim Prod Sci. 2016;56:1961–83. doi:10.1071/AN15297

3. Berghof TVL, Poppe M, Mulder HA. Opportunities to improve resilience in animal breeding programs. Front Genet. 2019;9:692. doi: 10.3389/fgene.2018.00692

4. Mulder HA. Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. Front Genet. 2016;7:178. Published 2016. doi:10.3389/fgene.2016.00178

5. Dekkers JCM. Multiple trait breeding programs with genotype-by-environment interactions based on reaction norms, with application to genetic improvement of disease resilience. Genet Sel Evol 2021;53:93. https://doi.org/10.1186/s12711-021-00687-2

6. Sánchez-Molano E, Kapsona VV, Ilska JJ, et al. Genetic analysis of novel phenotypes for farm animal resilience to weather variability. BMC Genet. 2019;20(1):84. doi:10.1186/s12863-019-0787-z

7. de Haas Y, Veerkamp RF, Barkema HW, et al. Associations between pathogen-specific cases of clinical mastitis and somatic cell count patterns. J Dairy Sci. 2004;87(1):95-105. doi:10.3168/jds.S0022-0302(04)73146-X.

8. Codrea MC, Højsgaard S, Friggens NC. Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: An example using milk yield profiles in dairy cows. J Anim Sci. 2011;89(10):3089-3098. doi:10.2527/jas.2010-3753.

9. Nguyen-Ba H, van Milgen J, Taghipoor M. A procedure to quantify the feed intake response of growing pigs to perturbations. Animal. 2020;14(2):253-260. doi:10.1017/S1751731119001976

10. Berghof TVL, Bovenhuis H, Mulder HA. Body Weight Deviations as Indicator for Resilience in Layer Chickens. Front Genet. 2019b Dec 13;10:1216. doi: 10.3389/fgene.2019.01216.

11. Laghouaouta H, Pena RN, Ros-Freixedes R, et al. A Methodology to Quantify Resilience in Growing Pigs. Animals (Basel). 2021;11(10):2970. doi:10.3390/ani11102970

12. Elgersma GG, de Jong G, van der Linde R, Mulder HA. Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. J Dairy Sci. 2018;101(2):1240-1250. doi:10.3168/jds.2017-13270

13. Poppe M, Bonekamp G, van Pelt ML, Mulder HA. Genetic analysis of resilience indicators based on milk yield records in different lactations and at different lactation stages. J Dairy Sci. 2021;104(2):1967-1981. doi:10.3168/jds.2020-19245

14. Poppe M, Mulder HA, van Pelt ML, et al. Development of resilience indicator traits based on daily step count data for dairy cattle breeding. Genet Sel Evol 54, 21 (2022). doi:10.1186/s12711-022-00713-x.

15. Argente MJ, García ML, Zbyňovská K, et al. Correlated response to selection for litter size environmental variability in rabbits' resilience. Animal. 2019; 13:2348-55. doi:10.1017/S1751731119000302

16. Kolmodin R, Strandberg E, Madsen P, et al. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. Acta Agric Scand A. 2002;52:11–24. doi:10.1080/09064700252806380

17. Mulder HA, Rönnegård L, Fikse WF, et al. Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models. Genet Sel Evol. 2013;45(1):23. doi:10.1186/1297-9686-45-23

18. Bergsma R, Hermesch S. Exploring breeding opportunities for reduced thermal sensitivity of feed intake in the lactating sow. J Anim Sci. 2012;90:85-98. doi:10.2527/jas.2011-4021.

19. Ravagnolo O, Misztal I. Genetic component of heat stress in dairy cattle, parameter estimation. J Dairy Sci. 2000;83(9):2126-2130. doi:10.3168/jds.S0022-0302(00)75095-8

20. Cheruiyot EK, Haile-Mariam M, Cocks BG, et al. New loci and neuronal pathways for resilience to heat stress in cattle. Sci Rep. 2021;11(1):16619. doi:10.1038/s41598-021-95816-8.

21. Cheng J, Lim K, Putz AM, et al. Genetic analysis of disease resilience of wean-to-finish pigs under a natural disease challenge model using reaction norms. Genet Sel Evol. 2022;54(1):11. doi:10.1186/s12711-022-00702-0

22. Scheffer M, Carpenter SR, Dakos V, Van Nes EH. Generic indicators of ecological resilience: inferring the chance of a critical transition. Annu. Rev. Ecol. Evol. Syst. 2015;46:145–167. doi:10.1146/annurev-ecolsys-112414-054242

23. Poppe M, Veerkamp RF, van Pelt ML, Mulder HA. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. J Dairy Sci. 2020;103(2):1667-1684. doi:10.3168/jds.2019-17290

24. Putz AM, Harding JCS, Dyck MK, et al. Novel Resilience Phenotypes Using Feed Intake Data From a Natural Disease Challenge Model in Wean-to-Finish Pigs. Front Genet. 2019;9:660. doi:10.3389/fgene.2018.00660

25. Cheng J, Putz AM, Harding JCS, et al. Genetic analysis of disease resilience in wean-to-finish pigs from a natural disease challenge model. J Anim Sci. 2020;98(8):skaa244. doi:10.1093/jas/skaa244

26. Blasco A, Martínez-Álvaro M, García ML, et al. Selection for environmental variance of litter size in rabbit. Genet Sel Evol. 2017;49:48. doi:10.1186/s12711-017-0323-4

27. Iung LHS, Carvalheiro R, Neves HHR, Mulder HA. Genetics and genomics of uniformity and resilience in livestock and aquaculture species: A review. J Anim Breed Genet. 2020;137(3):263-280. doi:10.1111/jbg.12454

28. Formoso-Rafferty N, Cervantes I, Ibáñez-Escriche N, Gutiérrez JP. Correlated genetic trends for production and welfare traits in a mouse population divergently selected for birth weight environmental variability. Animal. 2016;10(11):1770-1777. doi:10.1017/S1751731116000860.

29. Beloumi D, Blasco A, Muelas R, et al. Inflammatory Correlated Response in Two Populations of Rabbit Selected Divergently for Litter Size Environmental Variability. Animals. 2020;10:1540. doi:10.3390/ani10091540

30. Henry LP, Bruijning M, Forsberg SKG, Ayroles JF. The microbiome extends host evolutionary potential. Nat Commun. 2021;12(1):5141. doi:10.1038/s41467-021-25315-x.

31. Waddington CH. Canalization of development and the inheritance of acquired characters. Nature, 1942;150(3811):563-565. doi:10.1038/150563a0

32. Falconer D, Mackay TFC. (1996) Introduction to Quantitative Genetics, 4th ed. Edinburgh Gate, Harlow, U.K.Addison Wesley Longman Limited.

33. Lynch M, and Walsh B. (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland.

34. Jinks JL, Pooni HS: The genetic basis of environmental sensitivity. Proceedings of the Second International Conference on Quantitative Genetics: 31 May - 5 June 1987; Raleigh. Edited by: Weir BS, Eisen EJ, Goodman MM, Namkoong G. 1988, North Carolina State University, 505-522.

35. Ros M, Sorensen D, Waagepetersen R, et al. Evidence for Genetic Control of Adult Weight Plasticity in the Snail Helix aspersa. Genetics. 2004;168:2089–2097. doi:10.1534/genetics.104.032672

36. SanCristobal-Gaudy M, Elsen JM, Bodin L, Chevalet C. Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. Genet. Sel. Evol. 1998;30:423–451. doi:10.1051/gse:19980502

37. Hill WG, Mulder HA. Genetic analysis of environmental variation. Genet Res (Camb). 2010;92(5-6):381-395. doi:10.1017/S0016672310000546

38. Rönnegård L, Felleki M, Fikse F, et al. Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. Genet Sel Evol. 2010;42(1):8. doi:10.1186/1297-9686-42-8

39. Garreau H, Bolet G, Larzul C, et al. Results of four generations of a canalising selection for rabbit birth weight. Livest Sci, 2008;119(1–3):55–62. doi:10.1016/j.livsci.2008.02.009

40. Sorensen D, Waagepetersen R. Normal linear models with genetically structured residual variance heterogeneity: a case study. Genet Res. 2003;82(3):207-222. doi:10.1017/s0016672303006426

41. Waagepetersen R, Ibáñez-Escriche N, Sorenson D. A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics. Genet. Sel. Evol. 2008;40:161–176. doi:10.1051/gse:2007042

42. Ibáñez-Escriche N, Garcia M, Sorensen D. GSEVM v.2: MCMC software to analyze genetically structured environmental variance models. J. Anim. Breed. Genet. 2010;127:249–251. doi:10.1111/j.1439-0388.2009.00846.x

43. Ibáñez-Escriche N, Sorensen D, Waagepetersen R, Blasco A. Selection for environmental variation: a statistical analysis and power calculations to detect response. Genetics. 2008;180(4):2209-2226. doi:10.1534/genetics.108.091678

44. Mackay TF, Lyman RF. Drosophila bristles and the nature of quantitative genetic variation. Philos Trans R Soc Lond B Biol Sci. 2005;360(1459):1513-1527. doi:10.1098/rstb.2005.1672

45. Formoso-Rafferty N, Gutiérrez JP, García-Álvarez A, et al. Impact of selection for birth weight variability on reproductive longevity: A mice model. J Anim Breed Genet. 2022;139(4):370-379. doi:10.1111/jbg.12676

46. Argente MJ, Calle EW, García ML, Blasco A. Correlated response in litter size components in rabbits selected for litter size variability. J Anim Breed Genet. 2017;134(6):505-511. doi:10.1111/jbg.12283

47. Sell-Kubiak E, Bijma P, Knol EF, Mulder HA. Comparison of methods to study uniformity of traits: Application to birth weight in pigs. J Anim Sci, 2015;93(3):900–911. doi:10.2527/jas.2014-8313

48. Sell-Kubiak E, Wang S, Knol EF, Mulder HA. Genetic analysis of within-litter variation in piglets' birth weight using genomic or pedigree relationship matrices. J Anim Sci. 2015;93(4):1471-1480. doi:10.2527/jas.2014-8674

49. Peters LL, Robledo RF, Bult CJ, et al. The mouse as a model for human biology: a resource guide for complex trait analysis. Nat Rev Genet. 2007;8(1):58-69. doi:10.1038/nrg2025

50. Hernández P, Aliaga S, Pla M, Blasco A. The effect of selection for growth rate and slaughter age on carcass composition and meat quality traits in rabbits. J Anim Sci. 2004;82(11):3138-3143. doi:10.2527/2004.82113138x

51. Martínez-Álvaro M, Hernández P, Blasco A. Divergent selection on intramuscular fat in rabbits: Responses to selection and genetic parameters. J Anim Sci. 2016;94(12):4993-5003. doi:10.2527/jas.2016-0590

52. Zubiri-Gaitán A, Blasco A, Ccalta R, et al. Intramuscular Fat Selection in Rabbits Modifies the Fatty Acid Composition of Muscle and Liver Tissues. Animals (Basel). 2022;12(7):893. doi:10.3390/ani12070893

53. Ziadi C, Mocé ML, Laborda P, et al. Genetic selection for ovulation rate and litter size in rabbits: Estimation of genetic parameters and direct and correlated responses. J Anim Sci. 2013;91(7):3113-3120. doi:10.2527/jas.2012-6043

54. Piles M, Garreau H, Rafel O, et al. Survival analysis in two lines of rabbits selected for reproductive traits. J Anim Sci. 2006;84(7):1658-1665. doi:10.2527/jas.2005-678

55. Argente MJ, Santacreu MA, Climent A, Bolet G, Blasco A. Divergent selection for uterine capacity in rabbits. J Anim Sci. 1997;75(9):2350-2354. doi:10.2527/1997.7592350x

56. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. Gut. 2022;71(5):1020-1032. doi:10.1136/gutjnl-2021-326789

57. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome 2015;3:31. doi:10.1186/s40168-015-0094-5

58. Uffelmann E, Huang QQ, Munung NS, et al. Genome-wide association studies. Nat Rev Methods Primers 2021;1:59. doi:10.1038/s43586-021-00056-9

59. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol. 2012;8(12):e1002822. doi: 10.1371/journal.pcbi.1002822.

60. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005;6(2):95-108. doi:10.1038/nrg1521

61. Strillacci MG, Frigo E, Schiavini F, et al. Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. BMC Genet. 2014;15:106. doi:10.1186/s12863-014-0106-7.

62. Atashi H, Salavati M, De Koster J, et al. Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. J Anim Breed Genet. 2020;137(3):292-304. doi:10.1111/jbg.12442

63. Jiang J, Ma L, Prakapenka D, et al. A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. Front Genet. 2019;10:412. doi:10.3389/fgene.2019.00412

64. Ros-Freixedes R, Gol S, Pena RN, et al. Genome-Wide Association Study Singles Out SCD and LEPR as the Two Main Loci Influencing Intramuscular Fat Content and Fatty Acid Composition in Duroc Pigs. PLoS One. 2016;11(3):e0152496. doi:10.1371/journal.pone.0152496

65. Eusebi PG, González-Prendes R, Quintanilla R, et al. A genome-wide association analysis for carcass traits in a commercial Duroc pig population. Anim Genet. 2017;48(4):466-469. doi:10.1111/age.12545.

66. Pena RN, Noguera JL, García-Santana MJ et al. Five genomic regions have a major impact on fat composition in Iberian pigs. Sci Rep 2019;9:2031. doi:10.1038/s41598-019-38622-7.

67. Crespo-Piazuelo D, Criado-Mesas L, Revilla M, et al. Identification of strong candidate genes for backfat and intramuscular fatty acid composition in three crosses based on the Iberian pig. Sci Rep. 2020;10(1):13962. doi:10.1038/s41598-020-70894-2

68. Sosa-Madrid BS, Hernández P, Blasco A, et al. Genomic regions influencing intramuscular fat in divergently selected rabbit lines. Anim Genet. 2020;51(1):58-69. doi:10.1111/age.12873

69. Laghouaouta H, Sosa-Madrid BS, Zubiri-Gaitán A, et al. Novel Genomic Regions Associated with Intramuscular Fatty Acid Composition in Rabbits. Animals (Basel). 2020;10(11):2090. doi:10.3390/ani10112090.

70. Yuan J, Wang K, Yi G, et al. Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. Genet Sel Evol. 2015;47:82. doi:10.1186/s12711-015-0161-1

71. Marchesi JAP, Ono RK, Cantão ME, et al. Exploring the genetic architecture of feed efficiency traits in chickens. Sci Rep. 2021;11(1):4622. doi:10.1038/s41598-021-84125-9

72. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1918;52:339-433.

73. Holland D, Frei O, Desikan R, et al. Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. PLoS Genet. 2020;16(5):e1008612. doi:10.1371/journal.pgen.1008612

74. McCartney DL, Min JL, Richmond RC, et al. Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. Genome Biol. 2021;22(1):194. doi:10.1186/s13059-021-02398-9

75. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356-369. doi:10.1038/nrg2344

76. Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467-484. doi:10.1038/s41576-019-0127-1

77. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608. doi:10.1002/mpr.1608

78. Gondro C, Werf JVD, Hayes, B. (2013) Genome-Wide Association Studies and Genomic Prediction, in: Methods in Molecular Biology. Media Springer Science+ Business LLC. doi:10.1007/978-1-62703-447-0

79. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91(11):4414-4423. doi:10.3168/jds.2007-0980

80. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565-569. doi:10.1038/ng.608

81. Yang J, Zaitlen NA, Goddard ME, et al. Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. 2014;46:100–106. doi:10.1038/ng.2876

82. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 2011;88;76–82. doi:10.1016/j.ajhg.2010.11.011

83. Benjamini Y, and Hochberg Y. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing, J. Royal Stat. Soc. Ser. B. 1995;57(1):289-300. doi: 10.1111/j.2517-6161.1995.tb02031.x

84. Ryan TA. Significance tests for multiple comparison of proportions, variances, and other statistics. Psychol Bull. 1960;57:318-328. doi:10.1037/h0044320

85. Xu S. Theoretical basis of the Beavis effect. Genetics. 2003;165(4):2259-2268. doi:10.1093/genetics/165.4.2259

86. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819-1829. doi:10.1093/genetics/157.4.1819

87. van de Schoot R, Kaplan D, Denissen J, et al. A gentle introduction to bayesian analysis: applications to developmental research. Child Dev. 2014;85(3):842-860. doi:10.1111/cdev.12169.

88. Blasco, A. Bayesian Data Analysis for Animal Scientists: The Basics. 2017. Springer, Cham. doi: 10.1007/978-3-319-54274-4

89. van de Schoot R, Depaoli S, King R et al. Bayesian statistics and modelling. Nat Rev Methods Primers. 2021;1:1. doi:10.1038/s43586-020-00001-2

90. Gianola D. Priors in whole-genome regression: the bayesian alphabet returns. Genetics. 2013;194(3):573-596. doi:10.1534/genetics.113.151753

91. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186. doi:10.1186/1471-2105-12-186

92. Fernando RL, Garrick D. Bayesian methods applied to GWAS. Methods Mol Biol. 2013;1019:237-274. doi:10.1007/978-1-62703-447-0_10

93. Wolc A, Dekkers JCM. Application of Bayesian genomic prediction methods to genome-wide association analyses. Genet Sel Evol. 2022;54(1):31. doi:10.1186/s12711-022-00724-8

94. Garrick DJ, Fernando RL. Implementing a QTL detection study (GWAS) using genomic prediction methodology. Methods Mol Biol. 2013;1019:275–98. doi:10.1007/978-1-62703-447-0_11

95. Cheng H, Qu L, Garrick DJ, Fernando RL. A fast and efficient Gibbs sampler for BayesB in whole-genome analyses. Genet Sel Evol. 2015;47:80. doi:10.1186/s12711-015-0157-x

96. Cheng H, Fernando R, Garrick D. JWAS: Julia implementation of whole-genome analyses software. In Proceedings of the 11th World Congress on Genetics Applied to Livestock Production: 11–16 February 2018; Auckland. 2018.

97. López de Maturana E, Ibáñez-Escriche N, González-Recio Ó, et al. Next generation modeling in GWAS: comparing different genetic architectures. Hum Genet. 2014;133(10):1235-1253. doi:10.1007/s00439-014-1461-1

98. Moser G, Lee SH, Hayes BJ, et al. Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. PLoS Genet. 2015;11(4):e1004969. doi:10.1371/journal.pgen.1004969

99. Sahana G, Guldbrandtsen B, Janss L, Lund MS. Comparison of association mapping methods in a complex pedigreed population. Genet Epidemiol. 2010;34(5):455-462. doi:10.1002/gepi.20499

100. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90:773–95. doi:10.1080/01621459.1995.10476572

101. Fernando R, Toosi A, Wolc A, et al. Application of whole-genome prediction methods for genome-wide association studies: a Bayesian approach. J Agric Biol Environ Stat. 2017;22:172–93. doi:10.1007/s13253-017-0277-6

102. Wray NR, Yang J, Hayes BJ, et al. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet. 2013;14(7):507-515. doi:10.1038/nrg3457

103. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974;23(1):23-35. doi:10.1017/S0016672308009579

104. Nielsen R, Williamson S, Kim Y, et al. Genomic scans for selective sweeps using SNP data. Genome Res. 2005;15(11):1566-1575. doi:10.1101/gr.4252305

105. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010;20(4):R208-R215. doi:10.1016/j.cub.2009.11.055

106. Sabeti PC, Varilly P, Fry B, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449(7164):913-918. doi:10.1038/nature06250

107. Simianer H, Szyda J, Ramon G, Lien S. Evidence for individual and between-family variability of the recombination rate in cattle. Mamm Genome. 1997;8(11):830-835. doi:10.1007/s003359900587

108. Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 2005;169(4):2335-2352. doi:10.1534/genetics.104.036947

109. Pennings PS, Hermisson J. Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol. 2006;23(5):1076-1084. doi:10.1093/molbev/msj117

110. Storz JF. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol. 2005;14(3):671-688. doi:10.1111/j.1365-294X.2005.02437.x

111. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. Livest Sci. 2014;166:133-143. doi:10.1016/j.livsci.2014.05.003

112. de Simoni Gouveia JJ, da Silva MV, Paiva SR, de Oliveira SM. Identification of selection signatures in livestock species. Genet Mol Biol. 2014 Jun;37(2):330-42. doi: 10.1590/s1415-47572014000300004.

113. Saravanan KA, Panigrahi M, Kumar H, et al. Selection signatures in livestock genome: A review of concepts, approaches and applications. Livest Sci. 2020;241:104257. doi:10.1016/j.livsci.2020.104257

114. McQuillan R, Leutenegger AL, Abdel-Rahman R, et al. Runs of homozygosity in European populations [published correction appears in Am J Hum Genet. 2008 Nov;83(5):658]. Am J Hum Genet. 2008;83(3):359-372. doi:10.1016/j.ajhg.2008.08.007

115. Saura M, Fernández A, Varona L, et al. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genet Sel Evol. 2015;47(1):1. doi:10.1186/s12711-014-0081-5

116. Ceballos FC, Joshi PK, Clark DW, et al. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018;19(4):220-234. doi:10.1038/nrg.2017.109

117. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet. 2006;15(5):789-795. doi:10.1093/hmg/ddi493

118. Kimura M. The neutral theory of molecular evolution and the world view of the neutralists. Genome. 1989;31(1):24-31. doi:10.1139/g89-009

119. Achaz G. Frequency spectrum neutrality tests: one for all and all for one. Genetics. 2009;183(1):249-258. doi:10.1534/genetics.109.104042

120. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585-595. doi:10.1093/genetics/123.3.585

121. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics. 2000;155(3):1405-1413. doi:10.1093/genetics/155.3.1405

122. Lindsay B. Composite Likelihood Methods. Contemp Math. 1988;80:221-239. doi:10.1090/conm/080/999014

123. Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics. 2002;160(2):765-777. doi:10.1093/genetics/160.2.765

124. Stella A, Ajmone-Marsan P, Lazzari B, Boettcher P. Identification of selection signatures in cattle breeds selected for dairy production. Genetics. 2010;185(4):1451-61. doi: 10.1534/genetics.110.116111.

125. Sabeti PC, Reich DE, Higgins JM, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002;419(6909):832-837. doi:10.1038/nature01140

126. Simianer H, Szyda J, Ramon G, Lien S. Evidence for individual and between-family variability of the recombination rate in cattle. Mamm Genome. 1997;8(11):830-835. doi:10.1007/s003359900587

127. Voight BF, Kudaravalli S, Wen XQ, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006;4:e72. doi: 10.1371/journal.pbio.0040072

128. Wang ET, Kodama G, Baldi P, Moyzis RK. Global landscape of recent inferred Darwinian selection for Homo sapiens. Proc Natl Acad Sci U S A. 2006;103(1):135-140. doi:10.1073/pnas.0509691102

129. Wright S. The genetical structure of populations. Ann Eugen. 1951;15(4):323-354. doi:10.1111/j.1469-1809.1949.tb02451.x

130. Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A. 1973;70(12):3321-3323. doi:10.1073/pnas.70.12.3321

131. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(6):1358-1370. doi:10.1111/j.1558-5646.1984.tb05657.x

132. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132(2):583-589. doi:10.1093/genetics/132.2.583

133. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. Genome Res. 2013;23(9):1514-21. doi: 10.1101/gr.154831.113

134. Bonhomme M, Chevalet C, Servin B, et al. Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics. 2010;186(1):241-262. doi:10.1534/genetics.104.117275

135. Fariello MI, Boitard S, Naya H, et al. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics. 2013;193(3):929-941. doi:10.1534/genetics.112.147231

136. Gianola D, Simianer H, Qanbari S. A two-step method for detecting selection signatures using genetic markers. Genet Res (Camb). 2010;92(2):141-155. doi:10.1017/S0016672310000121

137. Beaumont MA, Balding DJ. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. 2004;13(4):969-980. doi:10.1111/j.1365-294x.2004.02125.x

138. Holsinger KE. Analysis of genetic diversity in geographically structured populations: A Bayesian perspective. Hereditas 1999;130:245–255. doi:10.1111/j.1601-5223.1999.00245.x

139. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010;20(3):393-402. doi:10.1101/gr.100545.109

140. Tang R, Feng T, Sha Q, Zhang S. A variable-sized sliding-window approach for genetic association studies via principal component analysis. Ann Hum Genet. 2009;73(6):631-7. doi: 10.1111/j.1469-1809.2009.00543.x

141. Racimo F. Testing for Ancient Selection Using Cross-population Allele Frequency Differentiation. Genetics. 2016;202(2):733-50. doi: 10.1534/genetics.115.178095

142. Ong RT, Teo YY. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics. 2010;26(9):1269-1270. doi:10.1093/bioinformatics/btq125

143. Teo YY, Fry AE, Bhattacharya K, et al. Genome-wide comparisons of variation in linkage disequilibrium. Genome Res. 2009 Oct;19(10):1849-60. doi: 10.1101/gr.092189.109

144. Cadzow M, Boocock J, Nguyen HT, Wilcox P, Merriman TR, Black MA. A bioinformatics workflow for detecting signatures of selection in genomic data. Front Genet. 2014;5:293. doi:10.3389/fgene.2014.00293

145. Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different selection signature statistics and a new strategy for combining them. Heredity (Edinb). 2015;115(5):426-436. doi:10.1038/hdy.2015.42

146. González-Rodríguez A, Munilla S, Mouresan EF, et al. On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. Genet Sel Evol. 2016;48(1):81. doi:10.1186/s12711-016-0258-1

147. Sosa-Madrid BS, Varona L, Blasco A, et al. The effect of divergent selection for intramuscular fat on the domestic rabbit genome. Animal. 2020;14(11):2225-2235. doi:10.1017/S1751731120001263

148. Staubach F, Lorenc A, Messer PW, et al. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (Mus musculus). PLoS Genet. 2012;8(8):e1002891. doi:10.1371/journal.pgen.1002891

149. Grossman SR, Andersen KG, Shlyakhter I, et al. Identifying recent adaptations in large-scale genomic data. Cell. 2013;152(4):703-713. doi:10.1016/j.cell.2013.01.035

150. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. Bioessays. 2013;35(9):780-786. doi:10.1002/bies.201300014

151. Albrechtsen A, Nielsen FC, Nielsen R. Ascertainment biases in SNP chips affect measures of population divergence. Mol Biol Evol. 2010 Nov;27(11):2534-47. doi: 10.1093/molbev/msq148

152. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol. 2021;82(11):801-811. doi:10.1016/j.humimm.2021.02.012

153. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418-426. doi:10.1016/j.tig.2014.07.001

154. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nat Rev Genet. 2020;21(10):597-614. doi: 10.1038/s41576-020-0236-x

155. Salzberg SL, Yorke JA. Beware of mis-assembled genomes. Bioinformatics. 2005;21(24):4320-4321. doi:10.1093/bioinformatics/bti769

156. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-351. doi:10.1038/nrg.2016.49

157. van Dijk EL, Jaszczyszyn Y, Naquin D. Thermes C. The Third Revolution in Sequencing Technology. Trends Genet. 2018;34(9):666-681. doi:10.1016/j.tig.2018.05.008

158. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing [published correction appears in Hum Mol Genet. 2011 Feb 15;20(4):853]. Hum Mol Genet. 2010;19(R2):R227-R240. doi:10.1093/hmg/ddq416

159. Seo JS, Rhie A, Kim J, et al. De novo assembly and phasing of a Korean human genome. Nature. 2016;538(7624):243-247. doi:10.1038/nature20098

160. Chaisson MJP, Sanders AD, Zhao X, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat Commun. 2019;10(1):1784. Published 2019 Apr 16. doi:10.1038/s41467-018-08148-z

161. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13(5):278-89. doi: 10.1016/j.gpb.2015.08.002

162. Wang Y, Zhao Y, Bollas A, et al. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348-1365. doi:10.1038/s41587-021-01108-x

163. Groenen MA, Archibald AL, Uenishi H, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491(7424):393-398. doi:10.1038/nature11622

164. Weldenegodguad M, Popov R, Pokharel K, et al. Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions. Front Genet. 2019;9:728. Published 2019 Jan 11. doi:10.3389/fgene.2018.00728

165. Xia X, Zhang S, Zhang H, et al. Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. BMC Genomics. 2021;22(1):43. doi:10.1186/s12864-020-07340-0

166. Hou Y, Qi F, Bai X, et al. Genome-wide analysis reveals molecular convergence underlying domestication in 7 bird and mammals. BMC Genomics. 2020;21(1):204. doi:10.1186/s12864-020-6613-1

167. Ramírez-Ayala LC, Rocha D, Ramos-Onsins SE, et al. Whole-genome sequencing reveals insights into the adaptation of French Charolais cattle to Cuban tropical conditions. Genet Sel Evol. 2021;53(1):3. doi:10.1186/s12711-020-00597-9

168. Jiang L, Kon T, Chen C, et al. Whole-genome sequencing of endangered Zhoushan cattle suggests its origin and the association of MC1R with black coat colour. Sci Rep. 2021;11(1):17359. doi:10.1038/s41598-021-96896-2

169. Wang K, Liu X, Qi T, et al. Whole-genome sequencing to identify candidate genes for litter size and to uncover the variant function in goats (Capra hircus). Genomics. 2021;113(1 Pt 1):142-150. doi:10.1016/j.ygeno.2020.11.024

170. Wang X, Ran X, Niu X, et al. Whole-genome sequence analysis reveals selection signatures for important economic traits in Xiang pigs. Sci Rep. 2022;12(1):11823. doi:10.1038/s41598-022-14686-w

171. Sims D, Sudbery I, Ilott NE, Heger A. Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15(2):121-132. doi:10.1038/nrg3642

172. Stratton M. Genome resequencing and genetic variation. Nat Biotechnol. 2008;26(1):65-66. doi:10.1038/nbt0108-65

173. Beaudoin M, Goyette P, Boucher G, et al. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. PLoS Genet. 2013;9(9):e1003723. doi:10.1371/journal.pgen.1003723

174. Majewski J, Schwartzentruber J, Lalonde E, et al. What can exome sequencing do for you?. J Med Genet. 2011;48(9):580-589. doi:10.1136/jmedgenet-2011-100223

175. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet. 2015;24(R1):R102-10. doi: 10.1093/hmg/ddv259

176. Bansal V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics. 2010;26(12):i318-i324. doi:10.1093/bioinformatics/btq214

177. Bovo S, Ribani A, Muñoz M et al. Whole-genome sequencing of European autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. Genet Sel Evol. 2020;52:33. doi:10.1186/s12711-020-00553-7

178. Ros-Freixedes R, Johnsson M, Whalen A, et al. Genomic prediction with whole-genome sequence data in intensely selected pig lines. Genet Sel Evol. 2022;54(1):65. doi:10.1186/s12711-022-00756-0

179. Sharma A, Lee JS, Dang CG, Sudrajad P, Kim HC, Yeon SH, Kang HS, Lee SH. Stories and Challenges of Genome Wide Association Studies in Livestock - A Review. Asian-Australas J Anim Sci. 2015;28(10):1371-9. doi: 10.5713/ajas.14.0715

180. Bhanuprakash V, Chhotaray S, Pruthviraj DR, et al. Copy number variation in livestock: A mini review. Vet World. 2018;11(4):535-541. doi: 10.14202/vetworld.2018.535-541

181. Joy A, Dunshea FR, Leury BJ, et al. Resilience of Small Ruminants to Climate Change and Increased Environmental Temperature: A Review. Animals (Basel). 2020;10(5):867. doi:10.3390/ani10050867

182. Laghouaouta H, Fraile L, Suárez-Mesa R, et al. A genome-wide screen for resilient responses in growing pigs. Genet Sel Evol. 2022;54(1):50. doi:10.1186/s12711-022-00739-1

183.Cheng J, Fernando R, Cheng H, et al. Genome-wide association study of disease resilience traits from a natural polymicrobial disease challenge model in pigs identifies the importance of the major histocompatibility complex region. G3 (Bethesda). 2022;12(3):jkab441. doi:10.1093/g3journal/jkab441

184. Tsartsianidou V, Sánchez-Molano E, Kapsona VV et al. A comprehensive genome-wide scan detects genomic regions related to local adaptation and climate resilience in Mediterranean domestic sheep. Genet Sel Evol. 2021:53, 90. doi:10.1186/s12711-021-00682-7

185. Sell-Kubiak E, Duijvesteijn N, Lopes MS, et al. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. BMC Genomics. 2015;16:1049. doi:10.1186/s12864-015-2273-y

186. Wang Y, Ding X, Tan Z, et al. Genome-Wide Association Study of Piglet Uniformity and Farrowing Interval. Front Genet. 2017;8:194. doi: 10.3389/fgene.2017.00194.

187. Wijga S, Bastiaansen JW, Wall E, et al. Genomic associations with somatic cell score in first-lactation Holstein cows. J Dairy Sci. 2012;95(2):899-908. doi:10.3168/jds.2011-4717

188. Mulder HA, Crump RE, Calus MPL, Veerkamp RF. Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. J Dairy Sci. 2013;96(11):7306-7317. doi:10.3168/jds.2013-6818

189. Iung LHS, Mulder HA, Neves HHR et al. Genomic regions underlying uniformity of yearling weight in Nellore cattle evaluated under different response variables. BMC Genomics 2018;19:619. doi:10.1186/s12864-018-5003-4

190. Wolc A, Arango J, Settar P, et al. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. Anim Genet. 2012;43(1):87-96. doi:10.1111/j.1365-2052.2012.02381.x

191. Morgante F, Sørensen P, Sorensen DA, et al. Genetic Architecture of Micro-Environmental Plasticity in Drosophila melanogaster. Sci Rep. 2015;5:9785. doi:10.1038/srep09785

192. Gibson OR, Tuttle JA, Watt PW, et al. Hsp72 and Hsp90α mRNA transcription is characterised by large, sustained changes in core temperature during heat acclimation. Cell Stress Chaperones. 2016;21(6):1021-1035. doi:10.1007/s12192-016-0726-0

193. Nava R, Zuhl MN. Heat acclimation-induced intracellular HSP70 in humans: a meta-analysis. Cell Stress Chaperones. 2020;25(1):35-45. doi:10.1007/s12192-019-01059-y

194. Sareh H, Tulapurkar ME, Shah NG, et al. Response of mice to continuous 5-day passive hyperthermia resembles human heat acclimation. Cell Stress Chaperones. 2011;16(3):297-307. doi:10.1007/s12192-010-0240-8

195. Dantzer R, Cohen S, Russo SJ, Dinan TG. Resilience and immunity. Brain Behav Immun. 2018;74:28-42. doi:10.1016/j.bbi.2018.08.010

196. Sommer F, Bäckhed F. The gut microbiota--masters of host development and physiology. Nat Rev Microbiol. 2013;11(4):227-238. doi:10.1038/nrmicro2974

197. Rosenberg E, Zilber-Rosenberg I. The hologenome concept of evolution after 10 years. Microbiome. 2018;6(1):78. doi:10.1186/s40168-018-0457-9

198. Shapira M. Gut Microbiotas and Host Evolution: Scaling Up Symbiosis. Trends Ecol Evol. 2016;31(7):539-549. doi:10.1016/j.tree.2016.03.006

199. Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev. 2008;32(5):723-735. doi:10.1111/j.1574-6976.2008.00123.x

200. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016;14(8):e1002533. doi: 10.1371/journal.pbio.1002533.

201. Jacobs DM, Gaudier E, van Duynhoven J, Vaughan EE. Non-digestible food ingredients, colonic microbiota and the impact on gut health and immunity: a role for metabolomics. Curr Drug Metab. 2009;10(1):41-54. doi:10.2174/138920009787048383

202. Tremaroli V, Bäckhed F. Functional interactions between the gut microbiota and host metabolism. Nature. 2012;489(7415):242-249. doi:10.1038/nature11552

203. Relman DA. The human microbiome: ecosystem resilience and health. Nutr Rev. 2012;70(Suppl 1):S2-S9. doi:10.1111/j.1753-4887.2012.00489.x

204. Kho ZY, Lal SK. The Human Gut Microbiome - A Potential Controller of Wellness and Disease. Front Microbiol. 2018;9:1835. doi:10.3389/fmicb.2018.01835

205. Fan, Y., Pedersen, O. Gut microbiota in human metabolic health and disease. Nat Rev Microbiol 19, 55–71 (2021). doi:10.1038/s41579-020-0433-9

206. Aagaard K, Ma J, Antony KM, et al. The placenta harbors a unique microbiome. Sci Transl Med. 2014;6(237):237ra65. doi:10.1126/scitranslmed.3008599

207. Collado MC, Rautava S, Aakko J, et al. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. Sci Rep. 2016;6:23129. doi:10.1038/srep23129

208. Stinson LF, Boyce MC, Payne MS, Keelan JA. The not-so-sterile womb: evidence that the human fetus is exposed to bacteria prior to birth. Front Microbiol. 2019;10:1124. doi:10.3389/fmicb.2019.01124

209. He Q, Kwok LY, Xi X, et al. The meconium microbiota shares more features with the amniotic fluid microbiota than the maternal fecal and vaginal microbiota. Gut Microbes. 2020;12(1):1794266. doi:10.1080/19490976.2020.1794266

210. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A. 2010;107(26):11971-11975. doi:10.1073/pnas.1002601107

211. Coelho GDP, Ayres LFA, Barreto DS, et al. Acquisition of microbiota according to the type of birth: an integrative review. Rev Lat Am Enfermagem. 2021;29:e3446. doi:10.1590/1518.8345.4466.3446

212. Penders J, Thijs C, Vink C, et al. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics. 2006;118(2):511-521. doi:10.1542/peds.2005-2824

213. Senn V, Bassler D, Choudhury R, et al. Microbial Colonization From the Fetus to Early Childhood-A Comprehensive Review [published correction appears in Front Cell Infect Microbiol. 2021;11:715671]. Front Cell Infect Microbiol. 2020;10:573735. doi:10.3389/fcimb.2020.573735

214. Combes S, Gidenne T, Cauquil L, et al. Coprophagous behavior of rabbit pups affects implantation of cecal microbiota and health status. J Anim Sci. 2014;92(2):652-665. doi:10.2527/jas.2013-6394

215. Bogatyrev SR, Rolando JC, Ismagilov RF. Self-reinoculation with fecal flora changes microbiota density and composition leading to an altered bile-acid profile in the mouse small intestine. Microbiome. 2020;8(1):19. doi:10.1186/s40168-020-0785-4

216. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30(6):492-506. doi:10.1038/s41422-020-0332-7

217. Donnet-Hughes A, Perez PF, Doré J, et al. Potential role of the intestinal microbiota of the mother in neonatal immune education. Proc Nutr Soc. 2010;69(3):407-15. doi:10.1017/S0029665110001898

218. Daft JG, Ptacek T, Kumar R, et al. Cross-fostering immediately after birth induces a permanent microbiota shift that is shaped by the nursing mother. Microbiome. 2015;3:17. doi:10.1186/s40168-015-0080-y

219. Bian G, Ma S, Zhu Z, et al. Age, introduction of solid feed and weaning are more important determinants of gut bacterial succession in piglets than breed and nursing mother as revealed by a reciprocal cross-fostering model. Environ Microbiol. 2016;18(5):1566-77. doi:10.1111/1462-2920.13272

220. Mulder IE, Schmidt B, Stokes CR, et al. Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. BMC Biol. 2009;7:79. doi:10.1186/1741-7007-7-79

221. Mosca A, Leclerc M, Hugot JP. Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? Front Microbiol. 2016;7:455. doi:10.3389/fmicb.2016.00455

222. Griffiths B, Ritz K, Bardgett R D, et al. Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity–ecosystem function relationship. Oikos. 2003;90:279–294. doi:10.1034/j.1600-0706.2000.900208.x

223. Liévin-Le Moal V, Servin AL. Anti-infective activities of lactobacillus strains in the human intestinal microbiota: from probiotics to gastrointestinal anti-infectious biotherapeutic agents. Clin Microbiol Rev. 2014;27(2):167-99. doi:10.1128/CMR.00080-13

224. Shade A, Peter H, Allison SD, et al. Fundamentals of microbial community resistance and resilience. Front Microbiol. 2012;3:417. doi:10.3389/fmicb.2012.00417

225. Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. Nat Rev Genet. 2017;18(11):690-699. doi:10.1038/nrg.2017.63

226. Grieneisen L, Dasari M, Gould TJ, et al. Gut microbiome heritability is nearly universal but environmentally contingent. Science. 2021;373(6551):181-186. doi:10.1126/science.aba5483

227. Goodrich JK, Waters JL, Poole AC, et al. Human genetics shape the gut microbiome. Cell. 2014;159(4):789-99. doi:10.1016/j.cell.2014.09.053

228. Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555(7695):210-215. doi:10.1038/nature25973

229. Waters JL, Ley RE. The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. BMC Biol. 2019;17(1):83. doi:10.1186/s12915-019-0699-4

230. Bruijning M, Ayroles JF, Henry LP, et al. Relative abundance data can misrepresent heritability of the microbiome. (Pre-print) bioRxiv 2022. doi:10.1101/2022.04.26.489345

231. Martino C, Dilmore AH, Burcham ZM, Metcalf JL, Jeste D, Knight R. Microbiota succession throughout life from the cradle to the grave. Nat Rev Microbiol. 2022. doi:10.1038/s41579-022-00768-z

232. Bergamaschi M, Tiezzi F, Howard J et al. Gut microbiome composition differences among breeds impact feed efficiency in swine. Microbiome. 2020; 8:110. doi:10.1186/s40168-020-00888-9

233. Déru V, Tiezzi F, Carillier-Jacquin C, et al. Gut microbiota and host genetics contribute to the phenotypic variation of digestive and feed efficiency traits in growing pigs fed a conventional and a high fiber diet. Genet Sel Evol. 2022;54(1):55. doi:10.1186/s12711-022-00742-6

234. Maltecca C, Dunn R, He Y, et al. Microbial composition differs between production systems and is associated with growth performance and carcass quality in pigs. Anim Microbiome. 2021;3(1):57. doi:10.1186/s42523-021-00118-z

235. Liu J, Stewart SN, Robinson K, et al. Linkage between the intestinal microbiota and residual feed intake in broiler chickens. J Anim Sci Biotechnol. 2021;12(1):22. doi:10.1186/s40104-020-00542-2

236. Xue MY, Sun HZ, Wu XH, et al. Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. Microbiome. 2020;8(1):64. doi:10.1186/s40168-020-00819-8

237. Martínez-Álvaro M, Auffret MD, Stewart RD, et al. Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. Front Microbiol. 2020;11:659. doi:10.3389/fmicb.2020.00659

238. Martínez-Álvaro M, Zubiri-Gaitán A, Hernández P, et al. Comprehensive functional core microbiome comparison in genetically obese and lean hosts under the same environment. Commun Biol. 2021;4(1):1246. doi:10.1038/s42003-021-02784-w

239. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-31. doi:10.1038/nature05414

240. Yan H, Diao H, Xiao Y, et al. Gut microbiota can transfer fiber characteristics and lipid metabolic profiles of skeletal muscle from pigs to germ-free mice. Sci Rep. 2016;6:31786. doi:10.1038/srep31786

241. Vrieze A, Van Nood E, Holleman F, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology. 2012;143(4):913-6.e7. doi:10.1053/j.gastro.2012.06.031. Erratum in: Gastroenterology. 2013;144(1):250

242. Niederwerder MC. Fecal microbiota transplantation as a tool to treat and reduce susceptibility to disease in animals. Vet Immunol Immunopathol. 2018;206:65-72. doi:10.1016/j.vetimm.2018.11.002

243. Difford GF, Plichta DR, Løvendahl P, et al. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLoS Genet. 2018;14(10):e1007580. doi:10.1371/journal.pgen.1007580

244. Buitenhuis B, Lassen J, Noel SJ, et al. Impact of the rumen microbiome on milk fatty acid composition of Holstein cattle. Genet Sel Evol 2019;51;23. doi:10.1186/s12711-019-0464-8

245. Ramayo-Caldas Y, Zingaretti L, Popova M, et al. Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. J Anim Breed Genet. 2020;137(1):49-59. doi:10.1111/jbg.12427

246. Saborío-Montero A, Gutiérrez-Rivas M, López-García A, et al. Holobiont effect accounts for more methane emission variance than the additive and microbiome effects on dairy cattle. Livest Sci. 2021;250:104538. doi:10.1016/j.livsci.2021.104538.

247. Camarinha-Silva A, Maushammer M, Wellmann R, et al. Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. Genetics. 2017;206(3):1637-1644. doi:10.1534/genetics.117.200782

248. Aliakbari A, Zemb O, Cauquil L, et al. Microbiability and microbiome-wide association analyses of feed efficiency and performance traits in pigs. Genet Sel Evol 2022;54:29. doi:10.1186/s12711-022-00717-7

249. Khanal P, Maltecca C, Schwab C, et al. Microbiability of meat quality and carcass composition traits in swine. J Anim Breed Genet. 2021;138(2):223-236. doi:10.1111/jbg.12504

250. Velasco-Galilea M, Piles M, Ramayo-Caldas Y, et al. The value of gut microbiota to predict feed efficiency and growth of rabbits under different feeding regimes. Sci Rep. 2021;11:19495. doi:10.1038/s41598-021-99028-y

251. Handelsman J. Metagenetics: spending our inheritance on the future. Microb Biotechnol. 2009;2(2):138-9. doi:10.1111/j.1751-7915.2009.00090_8.x

252. Schloss PD, Westcott SL. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol. 2011;77(10):3219-26. doi:10.1128/AEM.02810-10

253. Breitwieser FP, Lu J, Salzberg SL. A review of methods and databases for metagenomic classification and assembly. Brief Bioinform. 2019;20(4):1125-1136. doi:10.1093/bib/bbx120

254. Xia LC, Cram JA, Chen T, et al. Accurate genome relative abundance estimation based on shotgun metagenomic reads. PLoS One. 2011;6(12):e27992. doi:10.1371/journal.pone.0027992

255. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804-10. doi:10.1038/nature06244

256. Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 1994;44:846-849. doi:10.1099/00207713-44-4-846

257. Wei ZG, Zhang XD, Cao M, et al. Comparison of Methods for Picking the Operational Taxonomic Units From Amplicon Sequences. Front Microbiol. 2021 Mar 24;12:644012. doi:10.3389/fmicb.2021.644012

258. Schloss PD. Application of a Database-Independent Approach To Assess the Quality of Operational Taxonomic Unit Picking Methods. mSystems. 2016;1(2):e00027-16. doi:10.1128/mSystems.00027-16

259. Jackson MA, Bell JT, Spector TD, Steves CJ. A heritability-based comparison of methods used to cluster 16S rRNA gene sequences into operational taxonomic units. PeerJ. 2016;4:e2341. doi:10.7717/peerj.2341

260. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-857. doi:10.1038/s41587-019-0209-9. Erratum in: Nat Biotechnol. 2019;37(9):1091.

261. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41. doi:10.1128/AEM.01541-09

262. Rideout JR, He Y, Navas-Molina JA, et al. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ. 2014;2:e545. doi:10.7717/peerj.545

263. Edgar RC. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics. 2018;34(14):2371-2375. doi:10.1093/bioinformatics/bty113

264. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3. doi:10.1038/nmeth.3869

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

266. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013 Jan;41(Database issue):D590-6. doi:10.1093/nar/gks1219

267. Schoch CL, Seifert KA, Huhndorf S, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A. 2012;109(16):6241-6246. doi:10.1073/pnas.1117018109

268. Lapidus AL, Korobeynikov AI. Metagenomic Data Assembly - The Way of Decoding Unknown Microorganisms. Front Microbiol. 2021 Mar 23;12:613791. doi: 10.3389/fmicb.2021.613791

269. Li Z, Chen Y, Mu D, et al. Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. Brief Funct Genomics. 2012;11(1):25-37. doi:10.1093/bfgp/elr035

270. Ayling M, Clark MD, Leggett RM. New approaches for metagenome assembly with short reads. Brief Bioinform. 2020;21(2):584-594. doi:10.1093/bib/bbz020

271. Tamames J, Puente-Sánchez F. SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front Microbiol. 2019 Jan 24;9:3349. doi:10.3389/fmicb.2018.03349

272. Hyatt D, Chen GL, Locascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119. doi:10.1186/1471-2105-11-119

273. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60. doi:10.1038/nmeth.3176

274. Parks DH, Chuvochina M, Chaumeil PA, et al. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol. 2020 Sep;38(9):1079-1086. doi:10.1038/s41587-020-0501-8

275. Powell S, Szklarczyk D, Trachana K, et al. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. 2012;40(Database issue):D284-9. doi:10.1093/nar/gkr1060

276. Huerta-Cepas J, Forslund K, Coelho LP, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol. 2017 Aug 1;34(8):2115-2122. doi:10.1093/molbev/msx148

277. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27-30. doi:10.1093/nar/28.1.27

278. Mitchell A, Chang HY, Daugherty L, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 2015;43(Database issue):D213-21. doi:10.1093/nar/gku1243

279. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007 Mar;17(3):377-86. doi:10.1101/gr.5969107

280. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6. doi:10.1038/nmeth.f.303

281. Bağcı C, Patz S, Huson DH. DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences. Curr Protoc. 2021;1(3):e59. doi:10.1002/cpz1.59

282. Wood, D.E., Salzberg, S.L. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46. doi:10.1186/gb-2014-15-3-r46

283. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017;8:2224. Published 2017 Nov 15. doi:10.3389/fmicb.2017.02224

284. Sanders HL. Marine benthic diversity: a comparative study. Am. Nat. 1968;102:243–282. doi:10.1086/282541

285. Hurlbert SH. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. Ecology. 1971;52(4):577–586. doi:10.2307/1934145

286. K. Pearson. Mathematical contributions to the theory of evolution. – on a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc. R. Soc. Lond.. 1897;60:489–498. doi: 10.1098/rspl.1896.0076

287. Fernandes AD, Reid JN, Macklaim JM, et al. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2:15. doi:10.1186/2049-2618-2-15

288. Aitchison J. The Statistical Analysis of Compositional Data, Chapman & Hall, London – New York (1986), XII.

289. Kaul A, Mandal S, Davidov O, Peddada SD. Analysis of Microbiome Data in the Presence of Excess Zeros. Front Microbiol. 2017;8:2114. doi:10.3389/fmicb.2017.02114

290. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci U S A. 2010 May 25;107(21):9546-51. doi: 10.1073/pnas.0914005107

291. Silverman JD, Roche K, Mukherjee S, David LA. Naught all zeros in sequence count data are the same. Comput Struct Biotechnol J. 2020;18:2789-2798. doi:10.1016/j.csbj.2020.09.014

292. Costea PI, Zeller G, Sunagawa S, Bork P. A fair comparison. Nat Methods. 2014;11(4):359. doi: 10.1038/nmeth.2897

293. Paulson JN, Olson ND, Braccia DJ, et al. metagenomeSeq: Statistical analysis for sparse high-throughput sequncing. Bioconductor package. 2013. http://www.cbcb.umd.edu/software/metagenomeSeq.

294. Chen EZ, Li H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. Bioinformatics. 2016;32(17):2611-7. doi:10.1093/bioinformatics/btw308

295. Xinyan Z, Himel M, Nengjun Y. Zero-inflated negative binomial regression for differential abundance testing in microbiome studies. J Bioinforma Genomics. 2016; 2(2). doi:10.18454/jbg.2016.2.2.1

296. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

297. Martín-Fernández JA, Hron K, Templ M, et al. Bayesian-multiplicative treatment of count zeros in compositional data sets. Statistical Modelling. 2015;15(2):134-158. doi:10.1177/1471082X14535524

298. Greenacre M, Martínez-Álvaro M, Blasco A. Compositional Data Analysis of Microbiome and Any-Omics Datasets: A Validation of the Additive Logratio Transformation. Front Microbiol. 2021 Oct 11;12:727398. doi:10.3389/fmicb.2021.727398

299. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. NPJ Biofilms Microbiomes. 2020;6(1):60. doi:10.1038/s41522-020-00160-w

300. Pereira MB, Wallroth M, Jonsson V, Kristiansson E. Comparison of normalization methods for the analysis of metagenomic gene abundance data. BMC Genomics. 2018;19(1):274. doi:10.1186/s12864-018-4637-6

301. Lutz KC, Jiang S, Neugent ML, et al. A Survey of Statistical Methods for Microbiome Data Analysis. Front. Appl. Math. Stat. 2022;8:884810. doi:10.3389/fams.2022.884810

302. Barker M, Rayens W. Partial least squares for discrimination. J Chemometrics. 2003;17(3):166–73. doi:10.1002/cem.785

303. Tang ZZ, Chen G. Zero-inflated generalized Dirichlet multinomial regression model for microbiome compositional data analysis. Biostatistics. 2019;20:698–13. doi:10.1093/biostatistics/kxy025

304. Lê Cao KA, Costello ME, Lakis VA, et al. MixMC: a multivariate statistical framework to gain insight into microbial communities. PLoS ONE. 2016;11:e0160169. doi:10.1371/journal.pone.0160169

305. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

306. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. doi: 10.1093/bioinformatics/btp616

307. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47. doi:10.1093/nar/gkv007

308. Martin BD, Witten D, Willis AD. Modeling microbial abundances and dysbiosis with beta-binomial regression. Ann Appl Stat. 2020;14(1):94-115. doi:10.1214/19-aoas1283

309. Mandal S, Van Treuren W, White RA, et al. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis. 2015;26:27663. doi:10.3402/mehd.v26.27663

310. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 2017;13(11):e1005752. doi: 10.1371/journal.pcbi.1005752

311. Calgaro M, Romualdi C, Waldron L, et al. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data. Genome Biol. 2020;21(1):191. doi:10.1186/s13059-020-02104-1

312. Nearing JT, Douglas GM, Hayes MG, et al. Microbiome differential abundance methods produce different results across 38 datasets. Nat Commun. 2022;13(1):342. doi:10.1038/s41467-022-28034-z. Erratum in: Nat Commun. 2022;13(1):777

313. Brereton RG, Lloyd GR. Partial least squares discriminant analysis: Taking the magic away. J. Chemom. 2014;28:213-225. doi:10.1002/cem.2609

314. Magurran AE. Measuring biological diversity. Curr Biol. 2021;31(19):R1174-R1177. doi:10.1016/j.cub.2021.07.049

315. Chao A, Lee S-M. Estimating the number of classes via sample coverage. J Am Stat Assoc. 1992;87:210–217. doi:10.1080/01621459.1992.10475194

316. Chao A. Non-parametric estimation of the number of classes in a population. Scand J Stat. 1984;11:265–270. doi:10.2307/4615964

317. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJ. Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol. 2001;67(10):4399-406. doi:10.1128/AEM.67.10.4399-4406.2001

318. Simpson E. Measurement of Diversity. Nature 1949;163:688. doi:10.1038/163688a0

319. Shannon CE. A Mathematical Theory of Communication. Bell Syst. Tech. J. 1948;27: 379-423. doi:10.1002/j.1538-7305.1948.tb01338.x

320. Bray JR, Curtis JT. An ordination of upland forest communities of southern Wisconsin. Ecol. Monogr. 1957;27:325-349. doi: 10.2307/1942268

321. Lozupone C, Lladser ME, Knights D, et al. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011;5(2):169-72. doi:10.1038/ismej.2010.133

322. Oksanen J, Blanchet FG, Friendly M, et al. vegan: Community Ecology Package. R package version 2.5-7. 2020. https://CRAN.R-project.org/package=vegan

323. Ross EM, Moate PJ, Bath CR, et al. High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. BMC Genet. 2012;13:53. doi:10.1186/1471-2156-13-53

324. Maltecca C, Lu D, Schillebeeckx C, et al. Predicting Growth and Carcass Traits in Swine Using Microbiome Data and Machine Learning Algorithms. Sci Rep. 2019;9(1):6574. doi:10.1038/s41598-019-43031-x

325. Pérez-Enciso M, Zingaretti LM, Ramayo-Caldas Y, de Los Campos G. Opportunities and limits of combining microbiome and genome data for complex trait prediction. Genet Sel Evol. 2021;53(1):65. doi: 10.1186/s12711-021-00658-7. Erratum in: Genet Sel Evol. 2021;20;53(1):97.

326. Nicholson JK, Holmes E, Kinross J, et al. Host-gut microbiota metabolic interactions. Science. 2012;336(6086):1262-1267. doi:10.1126/science.1223813

327. Alberdi A, Andersen SB, Limborg MT, et al. Disentangling host-microbiota complexity through hologenomics. Nat Rev Genet. 2022;23(5):281-297. doi:10.1038/s41576-021-00421-0

328. Rutherford ST, Bassler BL. Bacterial quorum sensing: its role in virulence and possibilities for its control. Cold Spring Harb Perspect Med. 2012;2(11):a012427. doi:10.1101/cshperspect.a012427

329. Thompson JA, Oliveira RA, Xavier KB. Chemical conversations in the gut microbiota. Gut Microbes. 2016;7(2):163-70. doi:10.1080/19490976.2016.1145374

330. Vernocchi P, Vannini L, Gottardi D, Del Chierico F, Serrazanetti DI, Ndagijimana M, Guerzoni ME. Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome. Front Cell Infect Microbiol. 2012;2:156. doi:10.3389/fcimb.2012.00156

331. Vernocchi P, Del Chierico F, Putignani L. Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health. Front Microbiol. 2016;7:1144. doi:10.3389/fmicb.2016.01144

332. Dunn WB, Bailey NJ, Johnson HE. Measuring the metabolome: current analytical technologies. Analyst. 2005;130(5):606-25. doi:10.1039/b418288j

333. Macedo AN, Faccio AT, Fukuji TS, et al. Analytical Platforms for Mass Spectrometry-Based Metabolomics of Polar and Ionizable Metabolites. Adv Exp Med Biol. 2021;1336:215-242. doi: 10.1007/978-3-030-77252-9_11

334. Sanna, S, van Zuydam NR, Mahajan A, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet 2019;51:600–605. doi:10.1038/s41588-019-0350-x

335. Lavelle A, Sokol H. Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2020;17(4):223-237. doi:10.1038/s41575-019-0258-z

336. Parada Venegas D, De la Fuente MK, Landskron G, et al. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases [published correction appears in Front Immunol. 2019;10:1486]. Front Immunol. 2019;10:277. doi:10.3389/fimmu.2019.00277

337. Visekruna A, Luu M. The Role of Short-Chain Fatty Acids and Bile Acids in Intestinal and Liver Function, Inflammation, and Carcinogenesis. Front Cell Dev Biol. 2021;9:703218. doi: 10.3389/fcell.2021.703218

338. Silva YP, Bernardi A, Frozza RL. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. Front Endocrinol (Lausanne). 2020;11:25. doi: 10.3389/fendo.2020.00025

339. Kim CH. Control of lymphocyte functions by gut microbiota-derived short-chain fatty acids. Cell Mol Immunol. 2021;18:1161–1171. doi:10.1038/s41423-020-00625-0

340. Liu Y, Hou Y, Wang G, et al. Gut Microbial Metabolites of Aromatic Amino Acids as Signals in Host-Microbe Interplay. Trends Endocrinol Metab. 2020;31(11):818-834. doi:10.1016/j.tem.2020.02.012

341. Krishnan S, Ding Y, Saeidi N, et al. Gut Microbiota-Derived Tryptophan Metabolites Modulate Inflammatory Response in Hepatocytes and Macrophages. Cell Rep. 2019;28(12):3285. doi:10.1016/j.celrep.2019.08.080

342. Caussy C, Hsu C, Lo MT, et al. Genetics of NAFLD in Twins Consortium. Link between gut-microbiome derived metabolite and shared gene-effects with hepatic steatosis and fibrosis in NAFLD. Hepatology. 2018;68(3):918-932. doi:10.1002/hep.29892

343. Soderholm AT, Pedicord VA. Intestinal epithelial cells: at the interface of the microbiota and mucosal immunity. Immunology. 2019;158(4):267-280. doi:10.1111/imm.13117

344. Zhou A, Yuan Y, Yang M, et al. Crosstalk Between the Gut Microbiota and Epithelial Cells Under Physiological and Infectious Conditions. Front Cell Infect Microbiol. 2022;12:832672. doi:10.3389/fcimb.2022.832672

345. Krautkramer KA, Fan J, Bäckhed F. Gut microbial metabolites as multi-kingdom intermediates. Nat Rev Microbiol. 2021;19(2):77-94. doi: 10.1038/s41579-020-0438-4

346. Postler TS, Ghosh S. Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. Cell Metabolism 2017;26:110–30. doi:10.1016/j.cmet.2017.05.008

347. Liu Y, Liu C, Wu H, et al. Small Intestine Microbiome and Metabolome of High and Low Residual Feed Intake Angus Heifers. Front Microbiol. 2022;13:862151. doi:10.3389/fmicb.2022.862151

348. Wu J, Ye Y, Quan, J. et al. Using nontargeted LC-MS metabolomics to identify the Association of Biomarkers in pig feces with feed efficiency. Porc Health Manag 2021;7:39. doi:10.1186/s40813-021-00219-w

349. Ye Y, Wu J, Quan J, et al. Lipids and organic acids in three gut locations affect feed efficiency of commercial pigs as revealed by LC-MS-based metabolomics. Sci Rep. 2021;11(1):7746. doi:10.1038/s41598-021-87322-8

350. Xue MY, Sun HZ, Wu XH et al. Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. Microbiome 2020;8:64. doi:10.1186/s40168-020-00819-8

351. Alberdi A, Aizpurua O, Bohmann K, et al. Do Vertebrate Gut Metagenomes Confer Rapid Ecological Adaptation?. Trends Ecol Evol. 2016;31(9):689-699. doi:10.1016/j.tree.2016.06.008

352. Cao Y, Liu Y, Dong Q, et al. Alterations in the gut microbiome and metabolic profile in rats acclimated to high environmental temperature. Microb Biotechnol. 2022;15(1):276-288. doi:10.1111/1751-7915.13772

353. Houwenhuyse S, Stoks R, Mukherjee S et al. Locally adapted gut microbiomes mediate host stress tolerance. ISME J 2021;15:2401–2414. doi:10.1038/s41396-021-00940-y

354. Webster NS, Reusch TBH. Microbial contributions to the persistence of coral reefs. ISME J. 2017;11(10):2167-2174. doi:10.1038/ismej.2017.66

355. Weiss S. Microbiota as mediator and moderator of stress. BioEssays, 2020;43:1. doi:10.1002/bies.202000277.

356. Rea K, Dinan TG, Cryan JF. The microbiome: A key regulator of stress and neuroinflammation. Neurobiol Stress. 2016;4:23-33. doi: 10.1016/j.ynstr.2016.03.001.

357. Dos Santos Guilherme M, Valeri F, Winter J, et al. Resilience and the Gut Microbiome: Insights from Chronically Socially Stressed Wild-Type Mice. Microorganisms. 2022;10(6):1077. doi:10.3390/microorganisms10061077

358. Gareau MG, Sherman PM, Walker WA. Probiotics and the gut microbiota in intestinal health and disease. Nat Rev Gastroenterol Hepatol. 2010;7(9):503-14. doi: 10.1038/nrgastro.2010.117

359. Rosshart SP, Vassallo BG, Angeletti D, et al. Wild Mouse Gut Microbiota Promotes Host Fitness and Improves Disease Resistance. Cell. 2017;171(5):1015-1028.e13. doi:10.1016/j.cell.2017.09.016

360. Sommer F, Anderson JM, Bharti R, et al. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017;15(10):630-638. doi:10.1038/nrmicro.2017.58

361. Fassarella M, Blaak EE, Penders J, et al. Gut microbiome stability and resilience: elucidating the response to perturbations in order to modulate gut health. Gut. 2021 Mar;70(3):595-605. doi:10.1136/gutjnl-2020-321747

CHAPTER 3: Identification of functional mutations associated with environmental variance of litter size in rabbits

Authors

Cristina Casto-Rebollo[1], María José Argente[2], María Luz García[2], Romi Pena[3], Noelia Ibáñez-Escriche[1*]

Institutional affiliations

[1]Institute for Animal Science and Technology, Universitat Politècnica de València, València, Spain

[2]Departamento de Tecnología Agroalimentaria, Universidad Miguel Hernández de Elche, Orihuela, Spain

[3]Departament de Ciència Animal, Universitat de Lleida-AGROTECNIO Center, Lleida, Catalonia, Spain

The content of this chapter has been published in Genetic Selection Evolution.

doi: 10.1186/s12711-020-00542-w

1. ABSTRACT

1.1. BACKGROUND

Environmental variance (V_E) is partly under genetic control and has recently been proposed as a measure of resilience. Unravelling the genetic background of the V_E of complex traits could help to improve resilience of livestock and stabilize their production across farming systems. The objective of this study was to identify genes and functional mutations associated with variation in V_E of litter size (LS) in rabbits. To achieve this, we combined the results of a genome-wide association study (GWAS) and a whole-genome sequencing (WGS) analysis using data from two divergently selected rabbit lines for high and low V_E of LS. These lines differ in terms of biomarkers of immune response and mortality. Moreover, rabbits with a lower V_E of LS were found to be more resilient to infections than animals with a higher V_E of LS.

1.2. RESULTS

By using two GWAS approaches (single-marker regression and Bayesian multiple-marker regression), we identified four genomic regions associated with V_E of LS, on chromosomes 3, 7, 10, and 14. We detected 38 genes in the associated genomic regions and, using WGS, we identified 129 variants in the splicing, UTR, and coding (missense and frameshift effects) regions of 16 of these 38 genes. These genes were related to the immune system, the development of sensory structures, and stress responses. All of these variants (except one) segregated in one of the rabbit lines and were absent (n = 91) or fixed in the other one (n = 37). The fixed variants were in the *HDAC9*, *ITGB8*, *MIS18A*, *ENSOCUG00000021276* and *URB1* genes. We also identified a 1-bp deletion in the 3'UTR region of the *HUNK* gene that was fixed in the low V_E line and absent in the high V_E line.

1.3. CONCLUSIONS

This is the first study that combines GWAS and WGS analyses to study the genetic basis of V_E. The new candidate genes and functional mutations identified in this study suggest that the V_E of LS is under the control of functions related to the immune system, stress response, and the nervous system. These findings could also explain differences in resilience between rabbits with homogeneous and heterogeneous V_E of litter size.

2. BACKGROUND

Understanding the effect of the environment on the phenotype of farm animals is important to improve responses to genetic selection. The environment can affect both the mean of a trait and its variance (environmental variance or V_E). Many studies in various species have provided statistical evidence that V_E is partly under genetic control: pigs [1], mice [2], chickens [3], snails (*Helix aspersa*) [4] and cattle [5], among others. For instance, the V_E can differ between genotypes under the same environment [6]. Successful divergent selection experiments for V_E support these findings in both livestock and model animals [7-9].

Recently, V_E was proposed as a measure of resilience [10], which is the ability of an animal to maintain or quickly recover their performance in spite of environmental perturbations [11, 12]. Previous genome-wide association studies (GWAS) for V_E have identified relevant contributions from candidate genes that are related to important phases of the inflammatory response, such as *Hsp90* [13, 14], *p22-PHOX*, *GNG11,* and *GNGT1* [15], which are triggered by tissue damage and the entry of pathogens [16]. In humans, the *FTO* gene, which affects the variability of body mass index [17], was also found to be associated with sensitivity to infections [18]. All these results support the role of the immune system in the detection and response to environmental perturbations such as pathogen infections [12].

Unravelling the genetic background of the V_E of complex traits could help to improve resilience of livestock and stabilize their production across farming systems [19]. The aim of this study was to identify genes and functional mutations associated with variation in the V_E of litter size (LS) in rabbits. We performed a GWAS and a whole-genome sequencing (WGS) analysis using data from two rabbit lines that have been divergently selected for high and low V_E of LS [9]. These lines show a remarkable response (4.5% of the mean of base population), as well as differences in mortality, in biomarkers of immune response (plasma cortisol, leukocytes and acute-phase protein levels), and in concentrations of plasma cholesterol and triglycerides [20]. Moreover, the line with a low V_E of LS was found to cope better with environmental stressors such as infections than the line with a high V_E of LS, which suggests that the homogeneous line is more resilient.

3. METHODS

3.1. ANIMALS

The rabbits used in this study belong to a divergent selection experiment for high and

low V_E of LS over 12 generations at the University Miguel Hernandez of Elche, Spain. Each divergent line had approximately 125 female and 25 male parents per generation (for more details see Blasco et al. [9]). The total number of litters over these generations was equal to 13,788 for 3070 does: 6094 from the line with a low V_E of LS, 6682 from the line with a high V_E of LS, and 1012 from the base population. In total, 1658 records of litter size from generations 11 and 12, and genotypes for 384 does were used for the GWAS: 96 from the base population (404 parities), 149 from the line with a high V_E of LS (649 parities), and 139 from the line with a low V_E of LS (605 parities). The average litter size (total number born; TNB includes live born plus stillborn) for the base population and the lines with a low and high V_E of LS was 8.72 (±3.05), 7.71 (±2.38) and 6.51 (±3.06), respectively.

3.2. PHENOTYPE

In this study, we investigated genomic regions that were associated with the V_E of LS, which was the selection criterion in the divergent selection experiment [9]. The V_E of LS was calculated as the within-doe variance of TNB, after correction of TNB by year-season (47 levels) and parity-lactation status (3 levels) to avoid the effect of systematic effects on V_E. The mean estimate of residuals for a doe across parities was used to calculate the V_E of LS for a doe, using the minimum quadratic risk estimator:

$$V_E = \frac{1}{n+1}\sum_{1}^{n}(x_i - \bar{x}) \quad (1)$$

where x_i is the pre-corrected TNB at parity i of a doe and n is the total number of parities of the doe (ranging from 2 to 12). V_E was calculated by assuming that the additive genetic and permanent effects are approximately the same for each parity of a doe [21]. The average of the V_E of LS was 4.24 (±3.41), 2.27 (±1.97) and 3.84 (±3.69) for the base population and for the low and high lines, respectively.

3.3. GENOTYPES

Genomic DNA was isolated from blood sampled from does using standard procedures. Genotyping was performed with the 200K Affymetrix Axiom OrcunSNP array (ThermoFisher Scientific). Quality control of genotypes was performed using the platform Axiom Analysis Suite 3.1 of ThermoFisher Scientific and the PLINK v1.9 software [22]. Animals with a call rate lower than 97% and SNPs with a minor allele frequency lower than 0.05, with missing genotypes higher than 0.05, or with unknown positions on the rabbit reference genome (OryCun v2.0.96) were removed. After quality control, 367 animals (1589 parities) and 96,329 SNPs remained in the dataset. The

missing genotypes were imputed with the Beagle v.4.1 software [23]. Finally, we identified outliers and checked the population structure by applying a principal component analysis (PCA) based on the genotypes [see Additional file 1].

3.4. GWAS

Two approaches were used for GWAS: single-marker regression (SMR) and Bayesian multiple-marker regression (BMMR). SMR was performed using the linear mixed model method, which is available in the GCTA v1.91.4 beta software [24]. To correct for population stratification, GCTA considers the genomic relationship matrix built, but without SNPs on the chromosome of the tested SNP [25]. The SNPs that were associated with V_E of LS were identified at a conservative p-value threshold of 0.0001 [26]. Using the same method, we also tested the effect of ignoring differences in number of parities between does. In order to do that, we performed the SMR using a weighted linear mixed model method implemented in the R software. Instead of a genomic relationship matrix, the model included the first five principal components based on genotype to correct for population stratification. The V_E of LS was weighted according to Blasco et al. [9]:

$$w_i = \frac{(n_i + 1)^2}{2 * (n_i - 1)} \quad (2)$$

where n_i is the total number of parities for doe i.

BMMR was performed using a Bayes B model that is implemented in the GenSel software [27]. This model assumed that, in a given iteration of the Monte Carlo Markov chain, many SNPs have no effect and variance, with a prior probability of π = 0.999, and approximately 100 SNPs have an effect and a variance on the V_E of LS. The analysis was done using a chain length of 550,000, with a lag of 100 and a burn-in of 150,000. The means of the priors of the genotypic and environmental variances were equal to 5.1 and 4.3, respectively. A Bayes factor was calculated to determine statistical significance of the SNP associations, as:

$$BF = \frac{\hat{p}_i/(1 - \hat{p}_i)}{(1 - \pi)/\pi} \quad (3)$$

where π is the prior probability and $\hat{p}_i$ is the posterior probability of a SNP in locus i having an effect. A threshold for BF higher than 10 was used to identify SNPs that are associated with V_E of LS [28]. The contribution of each of the 2125 non-overlapping 1-Mb genomic windows to the genetic variance was computed as the posterior mean of the percentage of the genomic variance explained by all markers across the genome (total genomic variance).

3.4.1. ADDITIONAL EVIDENCE FOR ASSOCIATED SNPS

The SNPs identified to be associated with V_E using both the SMR and BMMR approach were further tested using a permutation test and a GWAS within each population to determine whether they were spurious associations due to drift. Only SNPs that passed these additional tests were considered as displaying a true association with V_E of LS.

The permutation test was performed using the PLINK v1. 9 software [22]. In total, 100,000 random permutations were performed to remove the true association between V_E of LS and the genotype. Each permutated dataset was analysed using a linear-mixed model and the p-value of each SNP was calculated. The resulting distribution of p-values was used to calculate an empirical p-value (EMP1) for each SNP in the original data based on the number of times that the p-value of that SNP was declared to have a significant association with V_E of LS under the null hypothesis of no association in the permutated data. The minimum EMP1 that could be registered was 1/N, where N is the number of permutation tests. Thus, only SNPs with an EMP1 close to 0.00001 were considered to be associated with V_E of LS.

Within-population GWAS was performed for each population using the SMR approach [24]. The same reference alleles were established in the two lines and in the base population to estimate allele substitution effects. Confidence intervals (CI) of the SNP effects within a population were estimated as the allele substitution effect estimate ±2SE. Overlapping CI for a SNP between lines was declared to signify no evidence of differences in allele effects on the phenotype across populations.

3.4.2. IDENTIFICATION OF ASSOCIATED GENOMIC REGIONS

The tested SNPs were used to perform a linkage disequilibrium (LD) study using the PLINK v1.9 software [22]. For this purpose SNPs within 0.5 Mb of a significant SNP were grouped, which were then expanded to genomic regions ±1 Mb for the LD study. Genomic regions associated with the trait were considered to be blocks of SNPs with r^2 higher than 0.7 among each other. We established this strong threshold following Vanliere et al. [29], who determined that two SNPs were dependent when r^2 was equal or higher than 0.43.

3.5. WHOLE-GENOME SEQUENCING

To identify which variants were present in one line but not in the other, due to the selection process, a pool of DNA from the breeding males in the 10^{th} generation was created for each line (27 animals per line) and sequenced with Illumina Technology at an

average depth of 27x. These males were all fathers of animals from the 11th generation, which were used in the GWAS.

Pre-processing of the WGS data was performed following Elston [30]. The BWA algorithm [31] was used to index the OryCun v2.0.96 reference genome from the raw data. Illumina adapters and low-quality read ends were removed using Trimmomatic v0.39 [32]. The BWA-MEM algorithm was used to align reads to OryCun v2.0.96. The sorted BAM files were obtained by SAMtools [33]. Duplicates were marked using Picard MarkDuplicates [34].

Variant calling was performed using the GATK Best Practices pipeline [35] by applying GATK to the BAM files using HaplotypeCaller and GenotypeGVCF to obtain the raw VCF files for the high and low V_E of LS lines. Variants were filtered using SelectVariant from GATK. Single nucleotide variants (SNVs) that were filtered out were labelled using VariantFiltration with the following filter expressions: QD < 2, FS > 60, MQ < 40, MQRankSum < -12.5, ReadPosRankSum < -8. INDELs were filtered out according to QD < 2, FS > 200, and ReadPosRankSum < -20. Finally, variants were annotated using the snpEFF software [36].

3.6. IDENTIFICATION OF CANDIDATE GENES AND FUNCTIONAL MUTATIONS

The Ensembl release 97 database [37] was used to investigate candidate genes in the genomic regions associated with V_E of LS, using OryCun 2.0.97 as the reference genome. SNVs and INDELs that were present in the genomic regions associated with V_E of LS were also detected. Variants that segregated differently between the two lines and that had a greater impact on gene function, were proposed as functional mutations for V_E of LS. We considered that a variant had a greater impact if it affected: (a) the amino acid sequence of the protein (missense or frameshift mutations), (b) the UTR regions of the mRNAs, or (c) the splicing pattern of the transcripts. Genes that contained such possible functional mutations were identified as candidate genes for V_E of LS were. The biological functions and the gene ontology of the candidate genes were reviewed in GeneCards [38].

4. RESULTS

4.1. GENOMIC REGIONS ASSOCIATED WITH V_E OF LS

GWAS identified SNPs associated with V_E of LS using two approaches, SMR and BMMR. SMR identified 12 SNPs with a p-value less than 0.0001 on *Oryctolagus cuniculus* chromosomes (OCU) 3, 7, 10, and 14 (Fig. 1a). The same results were obtained with the weighted SMR analysis of V_E to take differences in number of parities between does into

account (data not shown). With BMMR, we identified 60 SNPs on several chromosomes that had a Bayes factor (*BF*) higher than 10 (Fig. 1b), including all the SNPs that were identified by SMR (Table 1). These latter SNPs were in genomic windows on OCU3 (50-52 Mb), OCU7 (141-142 Mb), OCU10 (4-5.7 Mb), and OCU14 (163-164 Mb), which explained 4.0, 0.2, 3.2 and 0.5% of the total genomic variance, respectively. The three most significant SNPs on OCU9 were also considered because they reached a p-value close to the threshold of 0.0001 (0.00018) and a *BF* greater than 10 (Fig. 1). The genomic window that contained these SNPs on OCU9 (4-6 Mb) explained 0.9% of the total genomic variance. In summary, 15 SNPs were identified to be associated with V_E of LS by both methods.

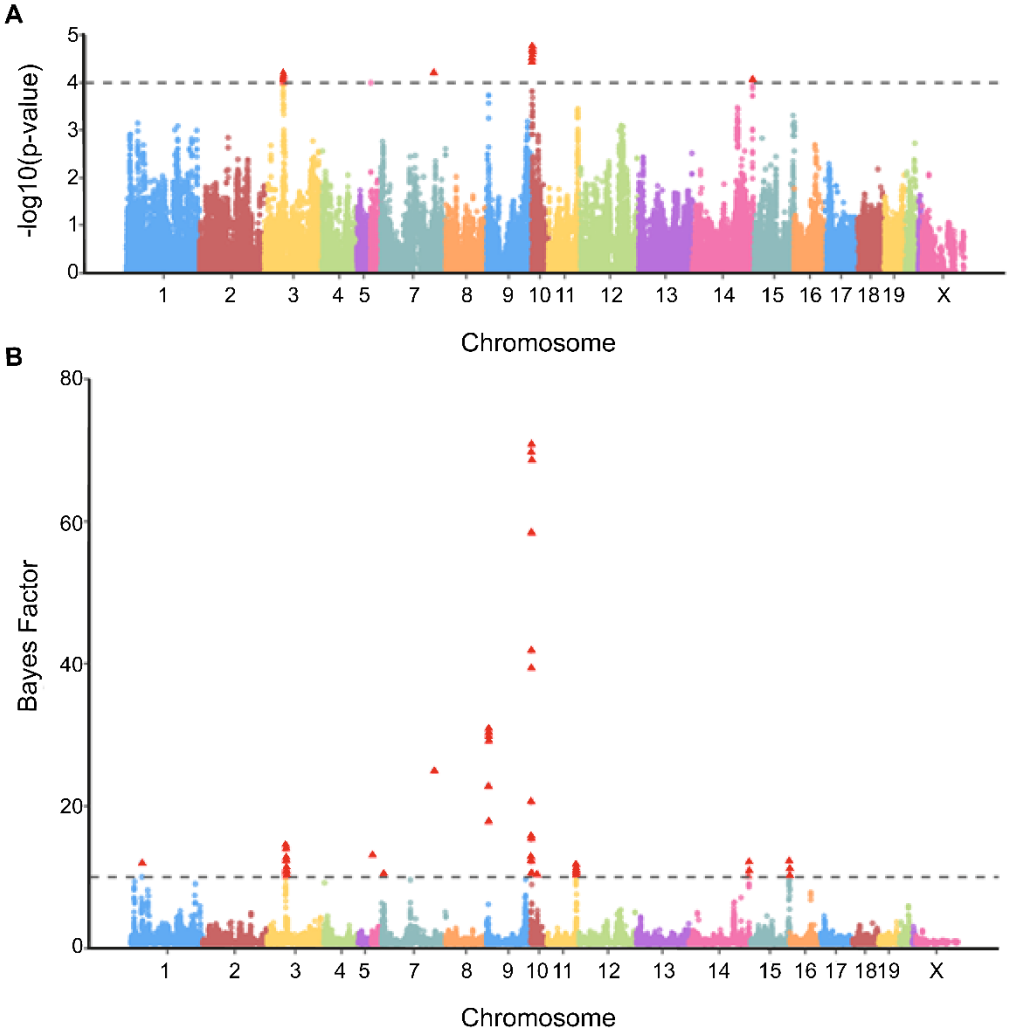


Figure 1. Manhattan plots for genome-wide association analyses for environmental variance of litter size. A) -$\log_{10}$(p-value) for association of SNPs using the single-marker regression approach**.** B) Bayes factor (BF) for association of SNPs using the Bayesian multiple-marker regression approach. The dashed lines represent significance thresholds a p-value of 0.0001 (A) and *BF* of 10 (B). The red triangles highlight the SNPs that pass the threshold.

These 15 SNPs were further evaluated by comparing their within-population allele substitution effects and by a permutation test. The allele substitution effect estimates for the 15 SNPs did not differ significantly between lines. However, five of these SNPs, located on OCU14 and 9, did not pass the permutation test because of their high empirical p-value (EMP1). The 10 SNPs that passed the additional tests were used to perform an LD analysis and determine the V_E-associated genomic regions (vQTL), as described in methods (Table 1), resulting in associated LD blocks of 1.2, 1.8 and 2.4 Mb on OCU14, 10 and 3, respectively. On OCU7, no associated LD block was detected [see Additional files 2, 3, 4 and 5]. Hence, vQTL were identified on OCU3 at 50.4-52.8 Mb, on OCU10 at 3.9-5.7 Mb, on OCU14 at 162-163.2, and on OCU7, close to 141,236,037 bp (Table 1).

Table 1. Genomic regions associated with environmental variance of litter size in rabbits

OCU[a]	Position (Mb)	Significant SNPs	p-value	*BF*[b]	Genes located in the region
3	50.4-52.8	Affx-151987366	7.02e-5	13.88	*SPDL1*, *DOCK2*[c], *INSYN2B*[c], *FOXI1*[c], *LCP2*, *KCNMB1*, *ENSOCUG00000020826*, *KCNIP1*, *GABRP*, *RANBP17*, *TLX3*, *FGF18*[c], *ENSOCUG00000022678*, *ENSOCUG00000011117*[c], *ENSOCUG00000018666*
		Affx-151799106	7.66e-5	14.39	
		Affx-151959457	8.79e-5	12.54	
7	141.2	Affx-151820818	6.26e-5	24.88	*AOX1*
10	3.9-5.7	Affx-151981327	1.76e-5	58.37	*HDAC9*[c], *FERD3L*[c], *TWISTNB*, *TMEM196*[c], *ENSOCUG00000019989*, *ENSOCUG00000018779*, *MACC1, ITGB8*[c]
		Affx-151890261	1.97e-5	58.48	
		Affx-151932936	2.29e-5	69.79	
		Affx-151906185	2.29e-5	70.82	
		Affx-151891719	2.29e-5	68.65	
14	162-163.2	Affx-151919621[d]	8.65e-5	10.01	*HUNK*[c], *MIS18A*[c], *URB1*[c], *ENSOCUG00000021276*[c], *EVA1C*[c], *CFAP298*, *SYNJ1*, *PAXBP1*[c], *C21orf62*[c], *ENSOCUG00000011671*, *OLIG1*, *PLCXD1*, *GTPBP6*, *ENSOCUG00000017611*
		Affx-151789209	1.14e-4	12.13	
		Affx-151983021[d]	1.34e-4	10.91	

[a]*Oryctolagus cuniculus* chromosome
[b]Bayes factor
[c]Candidate genes with relevant variants identified by whole-genome sequencing analysis
[d]SNPs that did not pass the additional tests

4.2. CANDIDATE GENES FOR V_E OF LITTER SIZE

In total, 38 genes were located in the genomic regions that were associated with V_E of LS (Table 1). We used WGS data of each line to identify 18,729 variants (SNVs + INDELs) in these regions (Table 2). From these, 129 were relevant (112 SNVs and 17 INDELs) based on their location in the transcription unit and/or splicing sites, which were located in 16 of the 38 genes identified in the GWAS [see Additional file 6]. These 16 genes (proposed as candidate genes) are involved in biological processes related to inflammatory response, development of sensory structures, and regulation of gene expression [see Additional file 7]).

Table 2. Classification and total number of variants (SNVs + INDELs) in each vQTL region

Region	OCU3	OCU7	OCU10	OCU14	Total
Upstream	141	21	34	328	524
5'UTR	2	0	5	0	7
Synonymous	9	2	10	34	55
Missense	1	0	6	28	35
Frameshift	0	0	0	4	4
Splicing	4	0	1	9	14
Intron	2019	267	1622	3145	7053
3'UTR	28	0	3	39	70
Downstream	193	34	57	486	770
Intergenic	3961	112	2237	3887	10,197
Total					18,729

All 129 relevant variants segregated in one of the two lines and were absent (91) or fixed (37) in the other line, except for one INDEL in the 3'UTR of the *HUNK* gene [see Additional file 6]. This latter was a 1-bp deletion that was fixed in the line with a low V_E of LS and absent in the line with a high V_E of LS. The other variants that were fixed for the alternative allele were identified in the line with a high V_E of LS in the *ITGB8*, *MIS18A*, *ENSOCUG00000021276*, and *URB1* genes, and in the line with a high V_E of LS for the *HDAC9* gene [see Additional file 6]. These variants could affect biological processes that are related to immune (*HDAC9*, *ITGB8,* and *HUNK*) and stress (*ENSOCUG00000021276*) responses, to regulation of gene expression (*HDAC9*, *MIS18A,* and *URB1*), and to phosphorylation of proteins (*HUNK*).

5. DISCUSSION

Our aim was to identify candidate genes and functional mutations associated with V_E of litter size in rabbits. In GWAS, estimates of the effect of genomic variants on the phenotype depends on the model used [39]. In our study, we identified associated genomic regions using SMR and BMMR analyses. The SMR analysis does not consider the dependencies between SNPs, so the effects were overestimated. In addition, the number of false negatives increases when a correction such as Bonferroni is applied and variants with small effects cannot be detected. In the BMMR analysis, the shrinkage parameter of the model (π = 0.999) increases the power to detect variants with small effects but also increases the number of false positives [39]. Thus, in our study, only genomic regions that were identified by both methods were considered as candidate regions for identifying relevant genes.

Several genomic regions were associated with V_E of LS (Table 1). The highlighted SNPs

in these regions were further evaluated using within-population GWAS and a permutation test. However, both these tests have some limitations. The within-population GWAS, accurate estimation of the allele substitution effect was limited by the small number of individuals per population (base = 91; low = 134; high = 142), which did not represent the allele and genotype frequencies in each population. For the permutation test, the highest EMP1 of the SNPs retained in the analysis was 0.00097 (OCU14). This means that in 97 of the 100,000 permutation tests performed, the SNP was associated with V_E of LS by chance under the null hypothesis. This could be due to the high level of relationship between animals in each population, which hinders elimination of true associations between genotype and phenotype.

The GWAS results were combined with WGS to identify candidate genes and functional mutations associated with V_E of LS. We screened for SNVs and INDELs between the two divergent rabbit lines in the vQTL that were detected by GWAS (Table 2). A variant was considered as a potential functional mutation when it caused a missense or frameshift mutation or affected the UTR regions in the mRNAs or the splicing pattern of the transcripts. Such variants are expected to have a critical effect on the function of a gene because of a change in mRNA stability or in the amino acid sequence of the protein it encodes. Sixteen of the 38 genes identified in the GWAS contained at least one of these variants [see Additional file 6]. Most of these variants segregated in one of the rabbit lines and were absent (91) or fixed (37) in the other line [see Additional file 6]. The use of DNA pools for WGS allowed us to have more coverage to identify different variants between the lines with high and low V_E of LS. However, the use of pools does not allow estimation of the frequency of a variant in a line, or the genotype of each animal used in the pool. For this reason, although we classified 129 variants as functional mutations, we focused on the variants that were fixed in one line and not in the other [see Additional file 6].

The 16 candidate genes identified in this study are involved in functions that are related to immune (*DOCK2*, *HDAC9*, *ITGB8,* and *HUNK*) and stress (*ENSOCUG00000021276*) responses, development of sensory structures (*FOXI1*, *FGF18*, and *EVA1C*), regulation of gene expression (*PAXBP1*, *FERD3L*, *HDAC9*, and *FOXI1*), and phosphorylation of proteins (*HUNK*), among others [see Additional file 7]. A recent study by Argente et al. [20] found differences in levels of plasma leukocytes and cortisol between the divergent rabbit lines used here but from generation 8 and showed that the line with a low V_E of LS was less sensitive to infection and stress than the high line. Our results confirm the importance of immune and stress responses for V_E of LS through the *DOCK2*, *ITGB8, HDAC9,* and *ENSOCUG00000021276* genes. For instance, *DOCK2* is involved in

extravasation of monocytes (entry into the affected tissue) by promoting polarization of the cell membrane and remodelling of the actin cytoskeleton needed for this function. In addition, *DOCK2* controls the monocyte inflammatory response via FcyR receptors [40], such as ITGB8, through TGF-β activation [41]. The *HDAC9* gene may play a role in hematopoiesis and self-tolerance through the control of T_{reg} cells [42]. The *ENSOCUG00000021276* gene, which is orthologous to the human *MRAP (melanocortin 2 receptor accessory protein*) gene, could modulate stress response though the production of glucocorticoids in the adrenal gland but experimental analyses are needed to verify this inferred function [37].

Previous GWAS for V_E in pigs and cows also identified genes that are involved in the immune response, more specifically in the inflammatory response [13-15]. Sell-Kubiak et al. [13] and Morgante et al. [14] identified genes of the HSP (heat shock protein) family to be associated with V_E, which regulate activation of leukocytes and protect cells against reactive-oxygen species (ROS) [43]. In mice, the candidate gene *HDAC9* regulates expression of a gene of the HSP family (*HSP70*) [42]. Wijga et al. [15] also found genes involved in the phagocytosis process to be associated with the standard deviation of milk somatic cell count in cattle, which is in line with functions related to the *DOCK2* gene [40]. Thus, there are several lines of evidence that support the importance of the immune system in the control of V_E.

For the other genes identified here (*FOXI1, EVAC1, FGF18,* and *HUNK*), we found no evidence in the literature that links them to a biological function associated with V_E. A recent study proposed to use V_E as a measure of animal resilience [10], which is supported by the results of Argente et al. [20], who suggested that the low V_E of LS line is more resilient to general stressors than the high V_E of LS line. According to Colditz and Hine [12], animals can better maintain performance (be more resilient) when they can properly discriminate environmental stimuli from the background. In this context, the nervous system, cell receptors, and the immune system act as sensors of environmental disturbances. Thus, the immune system is required to perceive and properly respond to environmental stimuli that occur on farms, as well as correct development of the sensory organs and the neuron system [12]. Along the same line, candidate genes such as *FOXI1*, *EVAC1,* and *FGF18* would be important to develop sensory structures and parts of the nervous system. The *FOXI1* gene encodes an important transcriptional factor, which is necessary for normal development of the inner ear, with mice that lack this gene developing deafness [44]. The *EVA1C* gene is involved in correct development of olfactory and optic sensory axons and other neural structures [45]. The growth factor FGF18 regulates development of the neural system, specifically the midbrain structure

[46]. Finally, HUNK is a serine/threonine kinase, which was recently shown to be associated with control of expression of E-cadherin [47], a molecule that can act as a receptor for pathogens [48]. We identified a 1-bp deletion in the 3'UTR region of the *HUNK* gene, which was fixed in the line with a low V_E of LS and absent in the line with a high V_E of LS. Mutations in the 3'UTR region can affect expression of the gene and/or translation rate of the mRNA. This suggests that different levels of expression of *HUNK* between the two lines could influence V_E of LS. The role of the identified candidate genes on modulation of V_E of LS and, therefore, on resilient responses, requires further study to complement current evidence on the relevance of the immune system on V_E [13-15, 20].

6. CONCLUSIONS

A combined GWAS and WGS analysis allowed us to identify 16 new candidate genes that carry 129 putative functional mutations that are associated with V_E of LS in rabbits. These findings provide support for the control of V_E of LS through regulation of the immune system and suggest that development of the nervous system and sensory structures may also be important to modulate animal resilience. This study advances our understanding of the genetic background of V_E. However, further studies are needed to validate the true effect of the putative functional mutations in these genes on V_E of LS, as well as the relationship of V_E with animal resilience.

7. DECLARATIONS

Ethics approval and consent to participate

All experimental procedures were approved by the Committee of Ethics and Animal Welfare of the Miguel Hernández University, according to Council Directives 98/58/EC and 2010/63/EU.

Consent for publication

Not applicable.

Availability of data and materials

Data are available upon request to the corresponding author.

Funding

This study was funded by the Spanish Ministry of Economy and Competitiveness (MINECO) with the Projects AGL2014-55921, C2-1-P and C2-2-P, and AGL2017-86083, C2-1-P and C2-2-P and the grant RYC-2016-19764.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CCR evaluated the experiments, analyzed the data and wrote the manuscript. MJA conceived and designed the study and contributed to the discussion of the results. MLG contributed to the study design and the discussion of the results. RP analyzed the lab data, contributed to the discussion of the results and edited the manuscript. NIE conceived the study, evaluated the experiments, contributed to the discussion of the results and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge the contributions of Roger Ros-Freixedes and Francisco Rosich of ASIC to the whole-genome sequencing pipeline and the useful comments of Agustin Blasco and Jack Dekkers. Cristina Casto-Rebollo acknowledges the FPU17/01196 scholarship of the Spanish Ministry of Science, Innovation and Universities.

8. REFERENCES

1. Ibáñez-Escriche N, Varona L, Sorensen D, Noguera JL. A study of heterogeneity of environmental variance for slaughter weight in pigs. Animal. 2008;2(1):19-26. doi:10.1017/S1751731107001000

2. Ibáñez-Escriche N, Moreno A, Nieto B, et al. Genetic parameters related to environmental variability of weight traits in a selection experiment for weight gain in mice; signs of correlated canalised response. Genet Sel Evol. 2008;40(3):279-93. doi:10.1186/1297-9686-40-3-279

3. Mulder HA, Hill WG, Vereijken A, Veerkamp RF. Estimation of genetic variation in residual variance in female and male broiler chickens. Animal. 2009;3(12):1673-80. doi:10.1017/S1751731109990668

4. Ros M, Sorensen D, Waagepetersen R, et al. Evidence for genetic control of adult weight plasticity in the snail Helix aspersa. Genetics. 2004;168(4):2089-97. doi:10.1534/genetics.104.032672

5. Rönnegård L, Felleki M, Fikse WF, et al. Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. J Dairy Sci. 2013;96(4):2627-2636. doi:10.3168/jds.2012-6198

6. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Harlow: Prentice Hall; 1996.

7. Garreau H, Bolet G, Larzul C,, et al. Results of four generations of a canalising selection for rabbit birth weight. Livest Sci. 2008;119:55–62. doi:10.1016/j.livsci.2008.02.009

8. Formoso-Rafferty N, Cervantes I, Ibáñez-Escriche N, Gutiérrez JP. Genetic control of the environmental variance for birth weight in seven generations of a divergent selection experiment in mice. J Anim Breed Genet. 2016;133(3):227-37. doi:10.1111/jbg.12174

9. Blasco A, Martínez-Álvaro M, García ML, et al. Selection for environmental variance of litter size in rabbits. Genet Sel Evol. 2017;49(1):48. doi:10.1186/s12711-017-0323-4

10. Berghof TVL, Poppe M, Mulder HA. Opportunities to Improve Resilience in Animal Breeding Programs. Front Genet. 2019;9:692. doi:10.3389/fgene.2018.00692

11. Mulder HA. Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. Front Genet. 2016;7:178. doi:10.3389/fgene.2016.00178

12. Colditz IG, Hine BC. Resilience in farm animals: biology, management, breeding and implications for animal welfare. Anim Prod Sci. 2016;56:1961–83. doi:10.1071/AN15297

13. Sell-Kubiak E, Duijvesteijn N, Lopes MS, et al. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. BMC Genomics. 2015;16:1049. doi:10.1186/s12864-015-2273-y

14. Morgante F, Sørensen P, Sorensen DA, et al. Genetic Architecture of Micro-Environmental Plasticity in Drosophila melanogaster. Sci Rep. 2015;5:9785. doi:10.1038/srep09785

15. Wijga S, Bastiaansen JW, Wall E, et al. Genomic associations with somatic cell score in first-lactation Holstein cows. J Dairy Sci. 2012;95(2):899-908. doi:10.3168/jds.2011-4717

16. Owen JA, Punt J, Stranford SA, Jones PP, Kuby J. Immunology. 7th ed. New York: Freeman; 2013.

17. Yang J, Loos RJ, Powell JE, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012;490(7419):267-72. doi:10.1038/nature11401

18. Feng Y, Wang F, Pan H, et al. Obesity-associated gene FTO rs9939609 polymorphism in relation to the risk of tuberculosis. BMC Infect Dis. 2014;14:592. doi:10.1186/s12879-014-0592-2

19. Hermesch S, Dominik S, editors. Breeding focus 2014 – Improving resilience. Armidale: University of New England; 2014.

20. Argente MJ, García ML, Zbyňovská K, et al. Correlated response to selection for litter size environmental variability in rabbits' resilience. Animal. 2019;13(10):2348-2355. doi:10.1017/S1751731119000302

21. Piles M, García ML, Rafel O, et al. Genetics of litter size in three maternal lines of rabbits: repeatability versus multiple-trait models. J Anim Sci. 2006;84(9):2309-15. doi:10.2527/jas.2005-622.

22. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. doi:10.1186/s13742-015-0047-8

23. Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples. Am J Hum Genet. 2016;98(1):116-26. doi:10.1016/j.ajhg.2015.11.020

24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011

25. Yang J, Weedon MN, Purcell S, et al; GIANT Consortium. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011;19(7):807-12. doi:10.1038/ejhg.2011.39

26. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11(3):241-7. doi:10.1038/ng1195-241

27. Garrick DJ, Fernando RL. Genome-wide association studies and genomic prediction. Methods Mol Biol. 2013;1019:275–98.

28. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90:773-95. doi:10.1080/01621459.1995.10476572

29. VanLiere JM, Rosenberg NA. Mathematical properties of the r2 measure of linkage disequilibrium. Theor Popul Biol. 2008;74(1):130-7. doi:10.1016/j.tpb.2008.05.006.

30. Elston RC. Preprocessing and quality control for whole-genome sequences from the Illumina HiSeq X platform. In: Wright MN, Gola D, Ziegler A, editors. Statistical human genetics. Methods in Molecular Biology, vol 1666. New York: Humana Press; 2017. p .629-47.

31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi:10.1093/bioinformatics/btp324

32. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi:10.1093/bioinformatics/btu170

33. Li H, Handsaker B, Wysoker A, et al; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi:10.1093/bioinformatics/btp352

34. Broad Institute. Picard tools. version 2.17.8; Broad Institute, GitHub repository. http://broadinstitute.github.io/picard/. Accessed: 2018/02/21

35. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. doi:10.1101/gr.107524.110

36. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92. doi:10.4161/fly.19695

37. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. Nucleic Acids Res. 2018;46(D1):D754-D761. doi:10.1093/nar/gkx1098

38. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54:1.30.1-1.30.33. doi:10.1002/cpbi.5

39. López de Maturana E, Ibáñez-Escriche N, González-Recio Ó, et al. Next generation modeling in GWAS: comparing different genetic architectures. Hum Genet. 2014;133(10):1235-53. doi:10.1007/s00439-014-1461-1

40. Chen Y, Meng F, Wang B, et al. Dock2 in the development of inflammation and cancer. Eur J Immunol. 2018;48(6):915-922. doi:10.1002/eji.201747157

41. Kelly A, Gunaltay S, McEntee CP, et al. Human monocytes and macrophages regulate immune tolerance via integrin αvβ8-mediated TGFβ activation. J Exp Med. 2018;215(11):2725-2736. doi:10.1084/jem.20171491

42. de Zoeten EF, Wang L, Sai H, et al. Inhibition of HDAC9 increases T regulatory cell function and prevents colitis in mice. Gastroenterology. 2010;138(2):583-94. doi:10.1053/j.gastro.2009.10.037

43. Zhang X, Mosser DM. Macrophage activation by endogenous danger signals. J Pathol. 2008;214(2):161-78. doi:10.1002/path.2284

44. Enerbäck S, Nilsson D, Edwards N, et al. Acidosis and Deafness in Patients with Recessive Mutations in FOXI1. J Am Soc Nephrol. 2018;29(3):1041-1048. doi:10.1681/ASN.2017080840

45. James G, Foster SR, Key B, Beverdam A. The expression pattern of EVA1C, a novel Slit receptor, is consistent with an axon guidance role in the mouse nervous system. PLoS One. 2013;8(9):e74115. doi:10.1371/journal.pone.0074115

46. Liu A, Li JY, Bromleigh C, et al. FGF17b and FGF18 have different midbrain regulatory properties from FGF8b or activated FGF receptors. Development. 2003;130(25):6175-85. doi:10.1242/dev.00845

47. Williams CB, Phelps-Polirer K, Dingle IP, et al. HUNK phosphorylates EGFR to regulate breast cancer metastasis. Oncogene. 2020;39(5):1112-1124. doi:10.1038/s41388-019-1046-5

48. Costa AM, Leite M, Seruca R, Figueiredo C. Adherens junctions as targets of microorganisms: a focus on Helicobacter pylori. FEBS Lett. 2013;587(3):259-65. doi:10.1016/j.febslet.2012.12.008

9. ADDITIONAL FILES

The online version contains supplementary material available at https://doi.org/10.1186/s1271 1-020-00542 -w.

Additional file 1

File format: .tiff

Title: Principal component analysis applied to the genotype data.

Description: Representation of the first (PC1) and second (PC2) component for the genotypes in the base population (red) and in the high (blue) and low (green) selection lines for environmental variance (V_E) of litter size (LS).

Additional file 2

File format: tiff

Title: Linkage disequilibrium of SNPs on OCU3 at 49-53 Mb.

Description: Representation of the linkage disequilibrium (LD) in the associated genomic region on OCU3. SNPs in this region were plotted according to their Bayes factor (BF). The colours of the SNPs indicate their LD with the SNP with the highest BF in this region (highlighted with a black triangle). Colours between red and green indicate an r^2 between 1 and 0.5. Colours between green and blue indicate an r^2 between 0.5 and 0. Genes in this region are plotted at the bottom of the graph according to their position on the genome.

Additional file 3

File format: tiff

Title: Linkage disequilibrium of SNPs on OCU7 at 140-142 Mb.

Description: Representation of the linkage disequilibrium (LD) in the associated genomic region on OCU7. SNPs in this region were plotted according to their Bayes factor (BF). The colours of the SNPs indicate their LD with the SNP with the highest BF in this region (highlighted with a black triangle). Colours between red and green indicate an r^2

between 1 and 0.5. Colours between green and blue indicate an r^2 between 0.5 and 0. Genes in this region are plotted at the bottom of the graph according to their position on the genome.

Additional file 4

File format: tiff

Title: Linkage disequilibrium of SNPs on OCU10 at 3-7 Mb.

Description: Representation of the linkage disequilibrium (LD) in the associated genomic region on OCU10. SNPs in this region were plotted according to their Bayes factor (BF). The colours of the SNPs indicate their LD with the SNP with the highest BF in this region (highlighted with a black triangle). Colours between red and green indicate an r^2 between 1 and 0.5. Colours between green and blue indicate an r^2 between 0.5 and 0. Genes in this region is plotted at the bottom of the graph according to their position on the genome.

Additional file 5

File format: tiff

Title: Linkage disequilibrium of SNPs on OCU14 at 161-164 Mb.

Description: Representation of the linkage disequilibrium (LD) in the associated genomic region in OCU14. SNPs in the region were plotted according to their Bayes Factor (BF). The colours of the SNPs indicate their LD with the SNP with the higher BF in this region (highlighted with a black triangle). Colours between red and green indicate r^2 between 1 and 0.5. Colours between green and blue indicate r^2 between 0.5 and 0. Genes in this region was plotted at the bottom of the graphic according their position in the genome.

Additional file 6

File format: xlsx

Title: Candidate genes and their biological function based on the GeneCards and Ensembl databases.

Additional file 7

File format: xlsx

Title: Total number of relevant variants identified in the associated genomic regions for the environmental variance of litter size.

Description: INDELs and SNVs identified by WGS analysis in a UTR or splicing region or with a missense effect. For each variant, the position is indicated according to the chromosome (OCU) and base pair (bp) location. REF show the allele in the reference genome (*Oryctolagus cuniculus* v2.0.96) and ALT the alternative variant identified in the rabbit lines. Low and High show the allelic distribution in each line where 0 indicates the reference allele and 1 the alternative allele.

CHAPTER 4: Selection for environmental variance of litter size in rabbits involves genes in pathways controlling animal resilience

Authors

Cristina Casto-Rebollo[1], María José Argente[2], María Luz García[2], Agustín Blasco[1], Noelia Ibáñez-Escriche[1]*

Institutional affiliations

[1]Institute for Animal Science and Technology, Universitat Politècnica de València, València, Spain

[2]Departamento de Tecnología Agroalimentaria, Universidad Miguel Hernández de Elche, Orihuela, Spain

The content of this chapter has been published in Genetic Selection Evolution.

doi: 10.1186/s12711-021-00653-y

1. ABSTRACT

1.1. BACKGROUND

Environmental variance (V_E) is partially under genetic control, which means that the V_E of individuals that share the same environment can differ because they have different genotypes. Previously, a divergent selection experiment for V_E of litter size (LS) during 13 generations in rabbit yielded a successful response and revealed differences in resilience between the divergent lines. The aim of the current study was to identify signatures of selection in these divergent lines to better understand the molecular mechanisms and pathways that control V_E of LS and animal resilience. Three methods (F_{ST}, ROH and varLD) were used to identify signatures of selection in a set of 473 genotypes from these rabbit lines (377) and a base population (96). A whole-genome sequencing (WGS) analysis was performed on 54 animals to detect genes with functional mutations.

1.2. RESULTS

By combining signatures of selection and WGS data, we detected 373 genes with functional mutations in their transcription units, among which 111 had functions related to the immune system, stress response, reproduction and embryo development, and/or carbohydrate and lipid metabolism. The genes *TTC23L*, *FBXL20*, *GHDC*, *ENSOCUG00000031631*, *SLC18A1*, *CD300LG*, *MC2R*, and *ENSOCUG00000006264* were particularly relevant, since each one carried a functional mutation that was fixed in one of the rabbit lines and absent in the other line. In the 3'UTR region of the *MC2R* and *ENSOCUG00000006264* genes, we detected a novel insertion/deletion (INDEL) variant.

1.3. CONCLUSIONS

Our findings provide further evidence in favour of V_E as a measure of animal resilience. Signatures of selection were identified for V_E of LS in genes that have a functional mutation in their transcription units and are mostly implicated in the immune response and stress response pathways. However, the real implications of these genes for V_E and animal resilience will need to be assessed through functional analyses.

2. BACKGROUND

The environmental variance (V_E) of a trait is the within-individual variation of the phenotypic values of that trait due to environmental factors [1, 2]. V_E is partially under genetic control, which means that individuals sharing the same environment can have different V_E because they have different genotypes [1]. Indeed, there have been successful divergent selection experiments for V_E in mice [3] and rabbits [4]. V_E was recently proposed as a measure of animal resilience [5], which has been defined as an animal's ability to cope with environmental disturbances and the rapid recovery of its productive performance [6, 7]. Differences in resilience have been reported in rabbit lines that were divergently selected for high and low V_E of litter size (LS), with the line with a low V_E of LS being more resilient [8]. According to Colditz and Hine [7], the immune system, nervous system and cell receptors are essential for modulating animal resilience and allowing detection of and response to environmental perturbations, such as pathogen infections.

Genome-wide associations studies (GWAS) for V_E in livestock have identified candidate genes that are involved in the immune response, which boosts the inflammatory response [9-11]. In rabbit lines that were divergently selected for V_E of LS, Casto-Rebollo et al. [11] identified functional mutations in candidate genes that are involved in the immune system, the nervous system, and the development of sensory structures. However, V_E is a complex trait with a low heritability [12] and low phenotype accuracy, which makes the identification of all the loci that affect V_E by GWAS only, difficult [13]. Analyses of signatures of selection, which do not require phenotype data, could help to identify more loci that affect V_E of LS. Several methods for the detection of signatures of selection have been proposed and are based on different assumptions according to the pattern of positive selection to be detected [14]: (1) reduction of the local genomic variability, (2) modification of the spectrums of allele frequency, or (3) variation of the linkage disequilibrium (LD). However, because of these different assumptions, the three methods are not strongly correlated [15-16] and, thus, have to be used in conjunction to identify the largest possible number of signatures of selection.

The aim of this study was to identify signatures of selection in rabbits that were divergently selected for high and low V_E of LS during 13 generations [4], to determine the molecular mechanisms and pathways that control the V_E of LS and animal resilience. We used three methods to identify signatures of selection in combination with whole-genome sequencing (WGS) analysis to highlight the genes with functional mutations. This study complements a previous GWAS for V_E of LS using the same rabbit lines [11].

3. METHODS

3.1. ANIMALS AND GENOTYPING DATA

The rabbits used in this study were from generations 11 and 13 and from the base population of a divergent selection experiment for high and low V_E of LS that was carried out at the University Miguel Hernández in Elche, Spain [4]. In total, 473 genotypes were used from 96 does from the base population, 282 from generation 11 (147 from the line with high V_E of LS and 135 from the line with low V_E of LS), and 95 from generation 13 (46 from the line with high V_E of LS and 49 from the line with low V_E of LS). Genomic DNA was isolated from blood and tissue samples using standard protocols. Genotyping was performed with the 200K Affymetrix Axiom Orcun Single Nucleotide Polymorphism (SNP) array (Thermo Fisher Scientific) and quality control was done with the Axiom Analysis Suite 3.1 platform from Thermo Fisher Scientific and PLINK v.1.9 software [17]. Quality control removed animals with a call rate lower than 97% and SNPs that had a minor allele frequency (MAF) lower than 0.05, a missing genotype higher than 0.05, or an unknown position in the rabbit reference genome (OryCun v2.0.103). The missing genotypes were imputed with the Beagle v4.1 software [18]. After quality control, 452 genotypes and 97,155 SNPs remained in the dataset. A principal component analysis (PCA) was performed to study population structure and to identify outliers using the R package SNPRelate available from Bioconductor [19].

3.2. IDENTIFICATION OF SIGNATURES OF SELECTION

Statistical analyses were performed to search for signatures of selection using the 274 genotypes from generation 11 (139 from the line with high V_E of LS and 135 from the line with low V_E of LS) and 90 genotypes from the base population. The 93 genotypes from generation 13 were kept for the validation analysis. Three methods were used to identify the patterns of signatures of selection (Fig. 1): (a) detection of runs of homozygosity (ROH), (b) quantification of the variation in LD patterns (VarLD), and (c) estimation of the fixation index (F_{ST}).

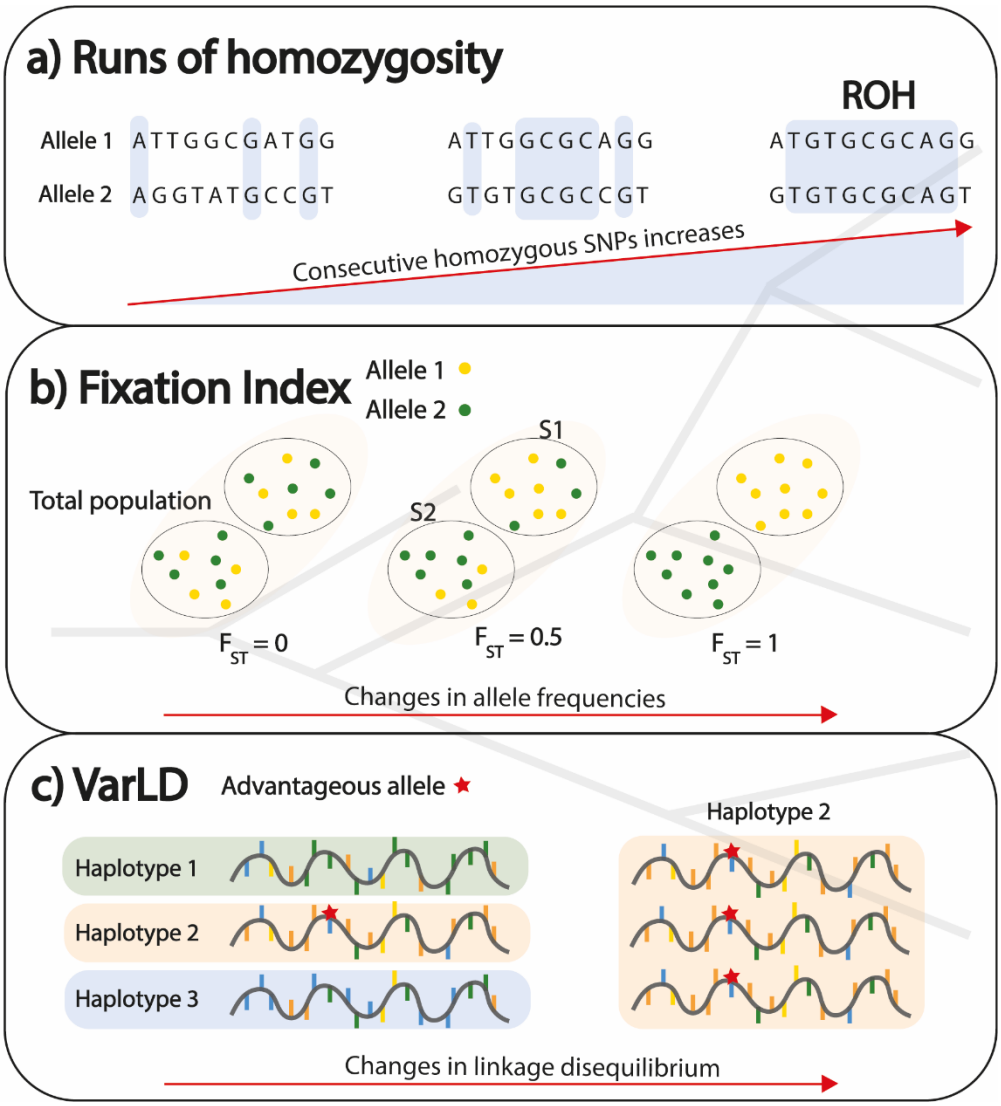


Figure 1. Methods of identifying patterns of signatures of selection. a) Runs of homozygosity (ROH). From left to right: the number of consecutive homozygous SNPs increases, generating a genomic region where the individual is homozygous at all sites, i.e. a ROH. b) Fixation index (FST). From left to right: allele frequencies of individuals in the population change until it differentiates into two different sub-populations (FST = 1). c) Variation in linkage disequilibrium (VarLD), which searches for differences in linkage disequilibrium (LD) patterns between populations. From left to right: an advantageous allele (red star) can modify the LD in a population because of the selective sweep containing the SNPs surrounding it, i.e. the haplotype of this advantageous allele (Haplotype 2).

3.2.1. DETECTION OF RUNS OF HOMOZYGOSITY

A ROH is a region of the genome that displays a local reduction of genetic variation, i.e. a genomic region for which the individual is homozygous at all sites, which indicates the presence of a locus that is affected by selection (Fig. 1a) [20]. Using an algorithm implemented in PLINK v1.9 [17], we identified the ROH in all the individuals from the base population and from the lines with high and low V_E of LS of generation 11. The parameters were set according to Ceballos et al. [21]. This algorithm searches for stretches of consecutive homozygous SNPs on each chromosome using sliding windows of 500 kb that contain around 50 SNPs. SNPs with missing calls and more than one

heterozygous SNP were not allowed in a window. The proportion of the overlapping windows that must be called homozygous to define any given SNP as in a homozygous segment was set to 0.05%. Two SNPs separated by more than 1 Mb belonged to two different homozygous segments. A homozygous segment was considered as a ROH if the number of consecutive SNPs exceeded 50 and the SNP density was higher than one SNP per 30 kb. A ROH must be a consensus genomic region in the selected animals to be a candidate signature of selection, i.e. it had to be identified in 50% of the animals in the line with low V_E of LS (65), and in 50% of the animals in the line with high V_E of LS (70).

3.2.2. ESTIMATION OF THE FIXATION INDEX

The fixation index (F_{ST}) was used to estimate the differences in allele frequencies between the lines with high and low V_E of LS (Fig. 1b). The F_{ST} was calculated using Weir and Cockerham's pairwise estimator method [22], implemented in the VCFtools v1.16 software [23]. The F_{ST} values were estimated in 500-kb sliding windows with a step size of 250 kb. Windows with less than ten SNPs were excluded from the analysis. F_{ST} values were weighted to take differences in sample sizes between populations into account (for further details see Weir and Cockerham [22]). Relevant F_{ST} windows were those with a weighted F_{ST} value equal or above the weighted F_{ST} value in the 99.9^{th} percentile of the distribution for all the genomes. MAF was calculated in the base population and the lines with high and low V_E of LS for the relevant F_{ST} windows. Those that showed divergent changes in MAF between the rabbit lines relative to the base population were considered to be putative signatures of selection. These windows were considered as resulting from an effect of genetic drift if the MAF between the lines with high and low V_E of LS at generation 11 displayed the same change relative to the base population (increase or decrease) or if one of the lines did not show any change (i.e. had a MAF equal to that in the base population).

3.2.3. QUANTIFICATION OF VARLD SCORES

We used the VarLD software [24] to evaluate the magnitude of the differences in LD patterns (Fig. 1c) between two populations. We analysed the pairwise comparison of the three populations: base population with the line with high V_E of LS (Base-High), base population with the line with low V_E of LS (Base-Low), and between the lines with high and low V_E of LS (High-Low). Sliding windows of 50 SNPs with a step size of one SNP were used to calculate the correlation matrix of each population per chromosome. The program computed the r2 metric for each pair of SNPs to determine the strength of LD in each window. The difference between the eigenvalues of the correlation matrices of both populations determined the VarLD score, which was standardized by the mean and

the standard deviation of all the scores along each chromosome. A genomic window was relevant when its standardized VarLD scores were equal or above the standardized VarLD score in the 99.9th percentile distribution for all the genomes. The relevant windows identified in both the Base-High and Base-Low comparisons were considered as putative signatures of selection and the relevant windows identified only in the High-Low comparison were considered as resulting from the effect of gene drift.

3.3. VALIDATION

The putative signatures of selection were validated by identifying those detected in the animals of the base population and of generation 13 (45 from the line with a high V_E of LS and 48 from the line with a low V_E of LS) applying the methods described above (Section on "*Identification of signatures of selection")*. Only those that were detected in both analyses (i.e. in generations 11 and 13) were considered as true signatures of selection.

3.5. IDENTIFICATION OF CANDIDATE GENES

Candidate genes were detected by searching for functional mutations in the genomic regions considered as true signatures of selection. Functional mutations were identified using whole-genome sequencing (WGS) data from two pools of DNA from breeding males in generation 10, i.e. all the fathers of animals from generation 11. Pools of DNA were prepared for each rabbit line (27 animals per line) and sequenced by Illumina Technology with an average depth of 27x.

WGS data were pre-processed following Elston et al. [25] with the following steps: (1) indexation to the reference genome (OryCun v2.0.103), (2) removal of adapters and low-quality read ends, (3) alignment to OryCun v2.0.103, and (4) identification of duplicates. Then, variant calling was performed using the GATK Best Practices pipeline [26] in three steps: (5) creation of raw VCF files for the high and low V_E of LS lines, (6) variant filtering, and (7) variant annotation (for further information see Casto-Rebollo et al. [11]).

A variant was considered as a functional mutation if it affected the transcription unit of a gene, i.e. (a) if it was located in the UTR regions, (b) if it was a missense or frameshift mutation, or (c) if it affected a splicing site. The gene ontologies (GO) of each candidate gene were extracted using the biomaRt package available from Bioconductor to R [27]. The Ensembl 103 release database [28] was used to access the *Oryctolagus cuniculus* v2.0.103 information.

4. RESULTS

Analysis of the population structure using principal component analysis showed a clear separation between the base population and the lines with high and low V_E of LS (Fig. 2). The individuals from generation 13 displayed the same family structure than that of their ancestors from generation 11.

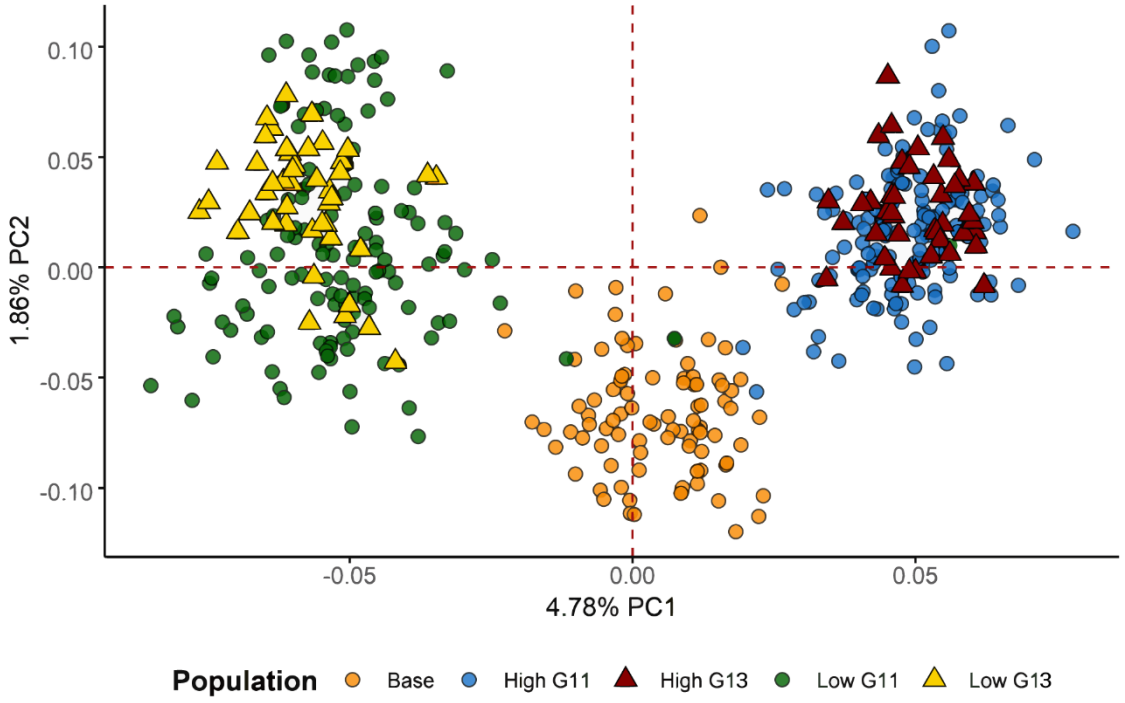


Figure 2. Principal component analysis of the genotyped data. Representation of the first (PC1) and second principal component (PC2) of the genotype data from the base population (orange) and the lines with high (right) and low (left) VE of LS in generations 11 (dot) and 13 (triangle).

4.1. SIGNATURES OF SELECTION FOR V_E OF LS

The ROH, F_{ST} and VarLD methods (Fig. 1) identified putative signatures of selection for V_E of LS, using the animals from generation 11 and the base population. Analysis of the contiguous homozygous segments identified 6230 consensus ROH, which were detected in at least two animals of the base population and of the lines with high and low V_E of LS. Of these 6230 consensus ROH, 720 were identified in at least 70 does of the line with high V_E of LS and 65 does of the line with low V_E of LS. These 720 consensus ROH were considered as putative signatures of selection because they were detected in at least 50% of the animals of each rabbit line. The F_{ST} analysis identified eight genomic regions with a weighted F_{ST} value equal or above 0.35 (99.9th percentile) on *Oryctolagus cuniculus* chromosome (OCU)2 (104.5-105 Mb), OCU9 (89-89.75 Mb), OCU12 (8.75-9.5 Mb), OCU14 (121-121.5 Mb), and OCUX (81-81.75 Mb) [see Additional file 1]. However, only

the regions on OCU2 (104.5-105 Mb), OCU12 (8.75-9.5 Mb) and OCUX (81-81.75 Mb) were considered as signatures of selection for V_E of LS because they showed consistent divergent changes in MAF between the lines with high and low V_E of LS relative to the base population [see Additional file 1]. The VarLD analysis identified three genomic regions that showed differences in LD patterns and overlapped between the lines with high and low V_E of LS on OCU13 (89.31-90.54 Mb), OCU14 (0.014-2.27 Mb), and OCU17 (28.78-29.92 Mb). The highest VarLD scores, 12.55 and 12.25, were obtained for OCU14 in the Base-High and Base-Low comparison, respectively. These three genomic regions were proposed as putative signatures of selection because their LD patterns differed relative to the base population.

The 726 putative signatures of selection identified in the ROH, F_{ST} and VarLD analyses were validated in 93 animals from generation 13. Among these 726 putative signatures of selection, 134 [see Additional file 2] were considered as true signatures of selection (Fig. 1), i.e. 129 ROH based on patterns of homozygous segments, two VarLD regions based on differences in LD patterns on OCU13 (89.31-90.54 Mb) and OCU14 (0.014-2.27 Mb), and three F_{ST} regions based on changes in allele frequencies on OCU2 (104.5-105 Mb), OCU12 (8.75-9.5 Mb) and OCUX (81-81.75 Mb). Finally, among these 134 true signatures of selection, the genomic regions did not overlap among the three methods.

4.2. CANDIDATE GENES FOR V_E OF LS

Nine hundred genes were identified in the genomic regions with positive selection patterns for V_E of LS. Candidate genes were identified by searching for functional mutations in the 900 genes using WGS data. In total, 212,845 variants (single-nucleotide variants (SNVs) and insertion/deletion variants (INDEL) were identified in the true signatures of selection (Table 1). Among these 212,845 variants, 1196 were relevant (207 INDEL and 989 SNVs) based on their location in the transcription unit of 373 of the 900 identified genes. These 373 genes (proposed as candidate genes) are involved in many biological processes [see Additional file 3]. Most of them (237) are found in basic common gene ontologies (GO) such as: protein binding (GO:0005515), cytoplasm (GO:0005737) and nucleus (GO:0005634), although, they also have a pleiotropic effect on other biological processes [see Additional file 3]. One hundred and eleven genes (29.76%) are involved in biological processes related to immune response (e.g., GO:0030335 and GO:0071356), stress response (GO:0042594), reproduction and embryo development (GO:0001701), and/or carbohydrate and lipid metabolism (e.g., GO:0005739 and GO:0055114).

We highlighted the *GATA3*, *FKBP10*, *KAT2A*, *CYP1B1*, *BRCA1*, *PGM3*, and *ACE* genes

that have a pleiotropic effect on the immune system, lipid and carbohydrate metabolism, and reproduction and embryo development.

Table 1. Effects of the variants (SNVs and INDEL) identified in the genomic regions with true signatures of selection for V_E of LS

Effect	Total
Upstream	10,925
Downstream	10,309
Intergenic	122,789
3'UTR	434
5'UTR	170
Intron	81,536
Splicing	156
Synonymous	657
Missense	368
Frameshift	38
Inframe	20
ncRNA exon	443
Stop gained	8
Start lost	2

A variant can affect more than one gene because they can share their DNA sequence, thus although the total number of variants identified was equal to 212,845, a total of 227,855 effects were found.

Among the 1196 functional mutations, we found 10 INDEL that were fixed (1/1) in one of the rabbit lines and absent (0/0) in the other and that affect the *TTC23L*, *FBXL20*, *GHDC*, *ENSOCUG00000031631*, *SLC18A1*, *CD300LG*, *MC2R*, and *ENSOCUG00000006264* genes (Table 2). These genes are involved in biological processes related to stress response (*MC2R*), energy, carbohydrate, and lipid metabolism (*ENSOCUG00000006264* and *MC2R*), nervous system (*MC2R, SLC18A1*, and *FBXL20*), immune response (*ENSOCUG00000000626)*, behaviour (*FBXL20*), cell maintenance (*TTC23L* and *GHDC*), and other processes (*ENSOCUG00000031631* and *CD300LG*). For each of the *MC2R* and *ENSOCUG00000006264* genes, one of the variants was a novel INDEL in the 3'UTR (Table 2) based on the two alleles in the reference rabbit genome OryCun v2.0.103. These novel alleles were fixed (2/2) in the line with high V_E of LS and absent in the line with low V_E of LS (Table 2).

Table 2. Functional mutations (INDEL or SNVs) fixed in one of the rabbit lines and absent in the other line

OCU[a]	bp[b]	Low[c]	High[d]	Region	Gene	Mutation[h]
9	48280960	1/1	2/2	3'UTR	*MC2R*	3-bp deletion*
11	56199847	0/0[e]	1/1[f]	Frameshift	*TTC23L*	2-bp deletion
12	138684799	1/1	0/0	Frameshift	*ENSOCUG00000031631*	2-bp deletion
12	138685173	1/1	0/0	Frameshift	*ENSOCUG00000031631*	1-bp insertion
12	138685179	1/1	0/0	Frameshift	*ENSOCUG00000031631*	2-bp deletion
14	33195918	1/1	2/2[g]	3'UTR	*ENSOCUG00000006264*	25-bp deletion*
15	4723998	1/1	0/0	3'UTR	*SLC18A1*	4-bp deletion
19	40686575	0/0	1/1	5'UTR	*FBXL20*	90-bp insertion
19	43006505	0/0	1/1	Frameshift	*GHDC*	2-bp deletion
19	44276359	1/1	0/0	3'UTR	*CD300LG*	8-bp deletion

[a]*Oryctolagus cuniculus* (OCU) chromosome
[b]Functional mutation location in base pairs
[c]Genotype of line with low V_E of LS
[d]Genotype of line with high V_E of LS
[e]0/0 indicates that the functional mutation is the homozygous for the reference allele
[f]1/1 indicates that the functional mutation is homozygous for the alternative allele
[g]2/2 indicates that the functional mutation is homozygous for a new allele not present in the reference genome
[h]All INDELs were marked according to the reference allele of OryCun v2.0.103
*With reference to the alternative allele of OryCun v2.0.103

5. DISCUSSION

Divergent lines provide good biological material for genomic studies since they are selected for a unique trait and share the same environment. Previous studies on intramuscular fat (IMF) in rabbits and pigs [16, 29], and on antibody response and feather pecking behaviour in chickens [30, 31], using divergently selected lines, successfully detected signatures of selection, and identified associated genomic regions. The principal component analysis based on genotype data in our study showed a clear separation between the two divergent rabbit lines (Fig. 2), which agreed with the remarkable phenotypic differentiation in V_E of LS (4.5% from the mean in the base population; Blasco et al. [4]). Thus, we searched for signatures of selection for V_E of LS, by combining the ROH, VarLD and F_{ST} methods (Fig. 1) to identify genes and pathways that were modified during 13 generations of selection.

In the analysis of signatures of selection, the identification of genomic regions under positive selection depends on the method applied [14]. Using the ROH, F_{ST} and VarLD methods, we identified 134 true signatures of selection for V_E of LS with no overlapping between methods. Indeed, as each method is based on different assumptions (Fig. 1), correlations between results are low [15, 16], which makes it difficult to detect overlaps between the identified signatures of selection. The methods used to detect signatures of selection should be considered independently of the sources of QTL detection. By

combining the three methods, ROH, F_{ST} and VarLD, we were able to identify most of the selection forces that affect the trait of interest, and we considered these 134 true signatures of selection as independent patterns of positive selection for V_E of LS. However, only one F_{ST} window on OCU3 (51-51.75 Mb) agreed with a variance-controlling locus (vQTL), which was previously identified in a GWAS [11] that used the same animals from the base and the generation 11 populations. The three genes (*SLIT3*, *FOXI1*, and *FGF18*) with functional mutations located in this vQTL could be the most relevant genes that play a role in the control of V_E of LS. They are involved in biological processes related to immune response, stress response, and/or development of sensory structures, which are relevant pathways to modulate resilience [7]. However, we identified this signature of selection in animals from generation 13 using a less conservative threshold of 99.5th percentile (weighted F_{ST} of 0.37). This F_{ST} window showed a weighted F_{ST} of 0.21 at generation 11 and 0.39 at generation 13. This stronger differentiation in allele frequencies in the population from generation 13 highlights the importance of this region for V_E of LS. However, the difference between the weighted F_{ST} at generations 11 and 13 could also be an effect of the reduced sample size (95) at generation 13, which may hide the true changes in allele frequencies between the generations.

Previous studies in divergent populations showed that some overlapping occurred between the signatures of selection obtained by F_{ST} analysis and a few QTL identified by GWAS [16, 29-31]. In contrast to our study, those studies used populations from a long-term divergent selection (during 40 generations) [30, 31], or from a selection for a highly heritable trait (intramuscular fat; IMF) [16, 29]. The fact that V_E of LS has a low heritability and accuracy [12] could hinder the identification of the genomic regions under selection in GWAS. Moreover, the small or moderate size of the effect of the variants could produce a sweep that is not large or strong enough to be detected as a signature of selection [32].

The identification of relevant loci for complex traits results in a large number of candidate genes due to their polygenic nature [33]. These genes are usually involved in multiple pathways or biological processes that may not be interrelated, such that searching for a relationship between these and the trait of interest is challenging. Along the same line, WGS analysis could be useful to identify the most relevant candidate genes that underlie the complex traits under study. In this study, we identified 900 genes that spanned genomic regions with patterns of signatures of selection for V_E of LS, and among these, 373 presented functional mutations that affect their transcription unit [see Additional files 4 and 5]. However, given these 373 genes that are implicated in a wide

range of functional categories [see Additional file 6], it remains difficult to identify the most relevant molecular mechanism involved in V_E of LS. Moreover, these genes may not have a clear relationship with V_E of LS since they may be acting indirectly, by modulating the core genes underlying V_E of LS [34]. For this reason, we could only make hypotheses based on the genes directly involved in previously identified biological pathways for V_E of LS.

Previous studies, developed by Argente et al. [8, 35] and Beloumi et al. [8] on the same rabbit lines as those used in this study, showed line-differences in immune response biomarkers (plasma cortisol, leukocytes, and acute-phase protein levels), in plasma concentrations of cholesterol and triglycerides, and in mortality [8, 35]. Among the 373 genes, 59 were related to immune response, six to stress response, and 49 to energy metabolism, carbohydrate metabolism or lipid metabolism [see Additional file 3], which could explain the differences reported by Argente et al. [8] and Beloumi et al. [36]. In addition, we found 38 genes involved in reproduction and embryo development [see Additional file 3] that could clarify the correlated response of V_E of LS with embryo implantation, embryo survival and litter size traits [36, 37]. In our study, we highlighted the genes, *GATA3*, *FKBP10*, *KAT2A*, *CYP1B1*, *BRCA1*, *PGM3*, and *ACE*, because they have a pleiotropic effect [see Additional file 3]. The ontologies of these genes are related to the immune system, lipid and carbohydrate metabolism, and reproduction and embryo development, supporting all the previously reported evidence [8, 35-37].

By searching for the most relevant functional mutations, we found seven promising genes that contained an INDEL with the alternative allele fixed in one rabbit line and absent in the other (Table 2), i.e., *TTC23L*, *FBXL20*, *GHDC*, *ENSOCUG00000031631*, *SLC18A1*, *CD300LG*, *MC2R*, and *ENSOCUG00000006264*, and which are the most relevant for V_E of LS. However, the functional mutation detected in the *MC2R* and *ENSOCUG00000006264* genes were even more interesting (Table 2). In our rabbit lines, both genes have lost the reference allele present in the rabbit reference genome OryCun v2.0.103 and display two different variants of the INDEL in their 3'UTR region. In both genes, one of the INDEL variant (the alternative variant for the reference genome) was fixed in the line with a high V_E of LS. The other (a new variant) was fixed in the line with a low V_E of LS (Table 2). The *MC2R* gene encodes the adrenocorticotropic hormone (ACTH) receptor, which controls ACTH and the level of cortisol [38]. Cortisol is an important molecule that regulates fat metabolism to mobilize glucose and mediates stress response and inflammatory response [39]. *ENSOCUG00000006264* is an orthologue of the *retinol binding protein 1* (*RBP1*) gene that is involved in the homeostasis of retinoid acid (RA) and the regulation of the vitamin A metabolism.

Retinoid acid could be involved in many immunological functions, such as the control of inflammatory and tolerogenic immune response [40].

When comparing the results with previously identified V_E loci (vQTL), we found evidence of the implication of the immune system in V_E, in line with the results reported by Argente et al. [8] and Beloumi et al [35]. In their GWAS, they identified genes that are related to the triggering of inflammatory response [9, 11] and belong to the HSP (heat shock protein) gene family [10, 11, 41], which can also modulate stress response, inflammatory response as well as the levels of glucose and fatty acids [42], and the fertilization and preimplantation of embryos [43]. Elevated levels of cortisol induce the synthesis of HSP to trigger stress response and cellular adaptation [44-46]. The INDEL variant on the *MC2R* gene could affect the stability of the transcribed mRNA, affecting the expression of the ACTH receptor that modulates the cortisol response. Thus, differential cortisol response could affect the stress and inflammatory response of animals, supporting the evidence of an effect on animal resilience [8, 35-37].

Animals can maintain their performance (be more resilient) when they can discriminate environmental stimuli from background. Berghof et al. [5] proposed V_E as a measure of animal resilience, while Argente et al. [8] showed that the line with a low V_E of LS was more resilient than the line with a high V_E of LS. According to Colditz and Hine [7], the immune system, cell receptors, and nervous and sensory structures are essential for coping with environmental disturbances. In this study, we identified 59 genes that are involved in the immune system, 23 in sensory perception, six in animal behaviour and 35 in the nervous system [see Additional file 3], which support the correlated response of V_E of LS in rabbit resilience [8]. Among these 59 genes, *SLC18A1*, *FBXL20,* and again *MC2R* were highlighted because they had also GO related to the modulation of the nervous system and/or behaviour [see Additional file 6]. However, the role of the highlighted candidate genes in controlling V_E of LS and animal resilience requires further studies to investigate their direct effect on the V_E of LS. Although the 373 identified genes with functional mutations were considered as candidate genes for V_E of LS, their direct or indirect implication in modulating the V_E of LS needs to be assessed.

6. CONCLUSIONS

We identified 373 candidate genes with functional mutations for V_E of LS in rabbits by combining independent methods of detection of signatures of selection and WGS data. These genes supported the biological pathways that were previously reported to be related to V_E of LS and involved in immune response, lipid and carbohydrate metabolism and stress response. These candidate genes could also explain the correlated response of

the V_E of LS in embryo implantation, embryo survival and litter size. Two novel INDEL variants were fixed in the line with high V_E of LS and absent in the line with low V_E of LS, one in the *MC2R* gene and one in the *ENSOCUG00000006264* gene. These promising functional mutations are located in genes that are involved in stress response and in the retinoid acid biosynthetic process, which could also control the immune response, respectively. This study expands on a previous GWAS for V_E of LS in rabbits and identified additional molecular mechanisms and pathways for V_E of LS and animal resilience. However, the real implications of these genes in V_E of LS still need to be assessed through functional analyses.

7. DECLARATIONS

Ethics approval and consent to participate

All the experimental procedures were approved by the Committee of Ethics and Animal Welfare of the Miguel Hernández University, according to Council Directives 98/58/EC and 2010/63/EU.

Consent for publication

Not applicable.

Availability of data and materials

Data are available upon request to the corresponding author.

Funding

This study was supported by Projects AGL2014-5592, C2-1-P and C2-2-P, and AGL2017-86083, C2-1-P and C2-2-P, funded by the Spanish *Ministerio de Ciencia e Innovación (MIC)-Agencia Estatal de Investigación* (AEI) and the European Regional Development Fund (FEDER).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CCR analyzed the data and wrote the manuscript. AB and MJA conceived and designed the study and contributed to the discussion of the results. MLG contributed to the study design and the discussion of the results. NIE conceived the study, contributed to the discussion of the results, and edited the manuscript. All the authors read and approved the final manuscript.

Acknowledgements

We are grateful to CEGEN-PRB3-ISCIII for their genotyping service, supported by Grant Nº PT17/0019 of the PE I+D+i 2013-2016, funded by ISCIII and ERDF. Cristina Casto-Rebollo acknowledges a FPU17/01196 scholarship from the Spanish Ministry of Science, Innovation and Universities.

8. REFERENCES

1. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Harlow: Prentice Hall; 1996.

2. Walsh B, Lynch M. Evolution and selection of quantitative traits. Oxford: Oxford University Press; 2018.

3. Formoso-Rafferty N, Cervantes I, Ibáñez-Escriche N, Gutiérrez JP. Genetic control of the environmental variance for birth weight in seven generations of a divergent selection experiment in mice. J Anim Breed Genet. 2016;133(3):227-37. doi:10.1111/jbg.12174

4. Blasco A, Martínez-Álvaro M, García ML, et al. Selection for environmental variance of litter size in rabbits. Genet Sel Evol. 2017;49(1):48. doi:10.1186/s12711-017-0323-4

5. Berghof TVL, Poppe M, Mulder HA. Opportunities to Improve Resilience in Animal Breeding Programs. Front Genet. 2019;9:692. doi:10.3389/fgene.2018.00692

6. Mulder HA. Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. Front Genet. 2016;7:178. doi:10.3389/fgene.2016.00178

7. Colditz IG, Hine BC. Resilience in farm animals: biology, management, breeding and implications for animal welfare. Anim Prod Sci. 2016;56:1961–83. doi:10.1071/AN15297

8. Argente MJ, García ML, Zbyňovská K, et al. Correlated response to selection for litter size environmental variability in rabbits' resilience. Animal. 2019;13(10):2348-2355. doi:10.1017/S1751731119000302

10. Wijga S, Bastiaansen JW, Wall E, et al. Genomic associations with somatic cell score in first-lactation Holstein cows. J Dairy Sci. 2012;95(2):899-908. doi:10.3168/jds.2011-4717

10. Sell-Kubiak E, Duijvesteijn N, Lopes MS, et al. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. BMC Genomics. 2015;16:1049. doi:10.1186/s12864-015-2273-y

11. Casto-Rebollo C, Argente MJ, García ML, et al. Identification of functional mutations associated with environmental variance of litter size in rabbits. Genet Sel Evol. 2020;52(1):22. doi:10.1186/s12711-020-00542-w

12. Hill WG, Mulder HA. Genetic analysis of environmental variation. Genet Res (Camb). 2010;92(5-6):381-95. doi:10.1017/S0016672310000546

13. Crouch DJM, Bodmer WF. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. Proc Natl Acad Sci U S A. 2020;117(32):18924-18933. doi:10.1073/pnas.2005634117

14. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. Livest Sci. 2014;166:133-143. doi:10.1016/j.livsci.2014.05.003

15. González-Rodríguez A, Munilla S, Mouresan EF, et al. On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. Genet Sel Evol. 2016;48(1):81. doi:10.1186/s12711-016-0258-1

16. Sosa-Madrid BS, Varona L, Blasco A, et al. The effect of divergent selection for intramuscular fat on the domestic rabbit genome. Animal. 2020;14(11):2225-2235. doi:10.1017/S1751731120001263

17. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. doi:10.1186/s13742-015-0047-8

18. Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples. Am J Hum Genet. 2016;98(1):116-26. doi:10.1016/j.ajhg.2015.11.020

19. Zheng X, Levine D, Shen J, et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28(24):3326-8. doi:10.1093/bioinformatics/bts606

20. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974;23(1):23-35. doi:10.1017/S0016672308009579

21. Ceballos FC, Joshi PK, Clark DW, et al. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018;19(4):220-234. doi:10.1038/nrg.2017.109

22. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(6):1358-1370. doi:10.1111/j.1558-5646.1984.tb05657.x

23. Danecek P, Auton A, Abecasis G, et al; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8. doi:10.1093/bioinformatics/btr330

24. Ong RT, Teo YY. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics. 2010;26(9):1269-1270. doi:10.1093/bioinformatics/btq125

25. Elston RC. Preprocessing and quality control for whole-genome sequences from the Illumina HiSeq X platform. In: Wright MN, Gola D, Ziegler A, editors. Statistical human genetics. Methods in Molecular Biology; vol 1666. New York: Humana Press; 2017. p .629-47.

26. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. doi:10.1101/gr.107524.110

27. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184-91. doi:10.1038/nprot.2009.97

28. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. Nucleic Acids Res. 2018;46(D1):D754-D761. doi:10.1093/nar/gkx1098

29. Kim ES, Ros-Freixedes R, Pena RN, et al, Rothschild MF. Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations. J Anim Sci. 2015;93(7):3292-302. doi:10.2527/jas.2015-8879

30. Lillie M, Sheng Z, Honaker CF, et al. Genome-wide standing variation facilitates long-term response to bidirectional selection for antibody response in chickens. BMC Genomics. 2017;18(1):99. doi:10.1186/s12864-016-3414-7

31. Johansson AM, Pettersson ME, Siegel PB, Carlborg O. Genome-wide effects of long-term divergent selection. PLoS Genet. 2010;6(11):e1001188. doi:10.1371/journal.pgen.1001188

32. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010;20(4):R208-15. doi:10.1016/j.cub.2009.11.055

33. Fisher RA. The correlation between relatives on the supposition of mendelian inheritance. Trans R Soc Edinb. 1918;53:399–433.

34. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177-1186. doi:10.1016/j.cell.2017.05.038

35. Beloumi D, Blasco A, Muelas R, et al. Inflammatory Correlated Response in Two Populations of Rabbit Selected Divergently for Litter Size Environmental Variability. Animals. 2020;10:1540. doi:10.3390/ani10091540

36. Argente MJ, Calle EW, García ML, Blasco A. Correlated response in litter size components in rabbits selected for litter size variability. J Anim Breed Genet. 2017;134(6):505-511. doi:10.1111/jbg.12283

37. Calle EW, García ML, Blasco A, Argente MJ. Correlated response in early embryonic development in rabbits selected for litter size variability. World Rabbit Sci. 2017;25:323-7.

38. Yang Y, Chen M, Ventro G, Harmon CM. Amino acid residue L112 in the ACTH receptor plays a key role in ACTH or α-MSH selectivity. Mol Cell Endocrinol. 2019;482:11-17. doi:10.1016/j.mce.2018.12.002

39. Hannibal KE, Bishop MD. Chronic stress, cortisol dysfunction, and pain: a psychoneuroendocrine rationale for stress management in pain rehabilitation. Phys Ther. 2014;94(12):1816-25. doi:10.2522/ptj.20130597

40. Larange A, Cheroutre H. Retinoic acid and retinoic acid receptors as pleiotropic modulators of the immune system. Annu Rev Immunol. 2016;34:369-94. doi:10.1146/annurev-immunol-041015-055427

41. Morgante F, Sørensen P, Sorensen DA, et al. Genetic architecture of micro-environmental plasticity in Drosophila melanogaster. Sci Rep. 2015;5:9785. doi:10.1038/srep09785

42. Collier RJ, Collier JL, Rhoads RP, Baumgard LH. Invited review: genes involved in the bovine heat stress response. J Dairy Sci. 2008;91(2):445-54. doi:10.3168/jds.2007-0540

43. Celi M, Vazzana M, Sanfratello MA, Parrinello N. Elevated cortisol modulates Hsp70 and Hsp90 gene expression and protein in sea bass head kidney and isolated leukocytes. Gen Comp Endocrinol. 2012;175(3):424-31. doi:10.1016/j.ygcen.2011.11.037

44. Neuer A, Spandorfer SD, Giraldo P, et al. Heat shock protein expression during gametogenesis and embryogenesis. Infect Dis Obstet Gynecol. 1999;7(1-2):10-6. doi:10.1155/S1064744999000034

45. Ravikumar S, Muthuraman P. Cortisol effect on heat shock proteins in the C2C12 and 3T3-L1 cells. In Vitro Cell Dev Biol Anim. 2014;50(7):581-6. doi:10.1007/s11626-014-9774-x

46. Pires BV, Stafuzza NB, Lima SBGPNP, et al. Differential expression of heat shock protein genes associated with heat stress in Nelore and Caracu beef cattle. Livest Sci. 2019;30:103839. doi: 10.1016/j.livsci.2019.103839

9. ADDITIONAL FILES

The online version contains supplementary material available at https://doi.org/10.1186/s12711-021-00653-y

Additional file 1

File format: xlsx

Title: Relevant F_{ST} windows with a weighted F_{ST} value higher than 0.35.

Description: Signatures of selection identified using F_{ST}. For each relevant F_{ST} window, the position is indicated according to the chromosome (OCU) and location in Mb. This table summarises the minor allele frequency (MAF) for each population and the weighted F_{ST} value for each F_{ST} window.

Additional file 2

File format: xlsx

Title: Localization of true-signatures of selection identified with each method of detection in generation 11 and validated in generation 13.

Additional File 3

File format: xlsx

Title: Classification of candidate genes in general biological processes according to their gene ontologies (GO).

Description: An in-house R script was used to group the gene ontology (GO) of the candidate genes in general biological pathways. For that, we created a dictionary for each biological pathway with their keyworks (for example, cytokine for immune system). Then, we matched these keywords with the GO description of each gene extracted from Ensembl 103 [see Additional file 6] using the package biomaRt. Each description containing the keyword was assigned to its biological pathway. Gene with GO terms for which it was difficult to assign a biological process were included in the category of "Other Processes", for example, MutSbeta complex. All keywords were extracted from

the literature. The percentage of each biological pathway was calculated as the ratio between the number of GO identified in the biological pathway (Xi) relative to the total number of GO (N); $\frac{X_i}{N} x100$. Using this in-house R script, we obtained a general vision of the pathways based on the GO of the candidate genes. This is an initiative named PATHionary. You can support it with your knowledge through the following link; https://forms.gle/yAt3S2JDUEzQxoLu9

Additional file 4

File format: xlsx

Title: Total number of INDEL identified in the true signatures of selection for the environmental variance of litter size.

Description: INDEL identified by WGS data analysis to be located in a UTR or splicing region or with a frameshift effect. For each variant, the position is indicated according to the chromosome (OCU) and base pair (bp) location. REF shows the allele in the reference genome (*Oryctolagus cuniculus* v2.0.103) and ALT the alternative variant identified in the rabbit lines. Low and High show the allelic distribution in each line where 0 indicates the reference allele, 1 the alternative allele, and 2 a new variant.

Additional file 5

File format: xlsx

Title: Total number of SNVs identified in the true signatures of selection regions for environmental variance of litter size.

Description: SNVs identified by WGS analysis to be located in a UTR or splicing region or with a frameshift effect. For each variant, the position is indicated according to the chromosome (OCU) and base pair (bp) location. REF shows the allele in the reference genome (*Oryctolagus cuniculus* v2.0.103) and ALT the alternative variant identified in the rabbit lines. Low and High show the allelic distribution in each line where 0 indicates the reference allele, 1 the alternative allele, and 2 a new variant.

Additional file 6

File format: xlsx

Title: Gene ontologies and descriptions of each candidate gene identified in the true signatures of selection, using biomaRt.

CHAPTER 5: Selection for environmental variance shifted the gut microbiome composition driving animal resilience

Authors

Cristina Casto-Rebollo[1], María José Argente[2], María Luz García[2], Ramona Natacha Pena[3], Agustín Blasco[1], Noelia Ibáñez-Escriche[1].

Institutional affiliations

[1]Institute for Animal Science and Technology, Universitat Politècnica de València, València, Spain

[2]Centro de Investigación e Innovación Agroalimentaria y Agroambiental (CIAGRO_UMH), Miguel Hernández University, Orihuela 03312, Spain

[3]Departament de Ciència Animal, Universitat de Lleida-AGROTECNIO Center, Lleida, Catalonia, Spain.

The content of this chapter is under review by Microbiome (Version 1).

Submitted 26th April 2022. The preprint is available at https://doi.org/10.21203/rs.3.rs-1597523/v1

1. ABSTRACT

1.1. BACKGROUND

Understanding how the host's microbiome shapes phenotypes and participates in the host response to selection is fundamental for evolutionists and animal and plant breeders. Currently, selection for resilience is considered a critical step in improving the sustainability of livestock systems. Environmental variance (V_E), the within-individual variance of a trait, has been successfully used as a proxy for animal resilience. Selection for reduced V_E could effectively shift gut microbiome composition, reshape the inflammatory response, triglyceride and cholesterol levels, and drive animal resilience. This study aimed to determine the gut microbiome composition underlying the V_E of litter size (LS), for which we performed a metagenomic analysis in two rabbit populations divergently selected for low (n=36) and high (n=34) V_E of LS. Partial least square-discriminant analysis, and alpha- and beta-diversity were computed to determine the differences in gut microbiome composition among the rabbit populations.

1.2. RESULTS

We identified 116 KEGG IDs, 164 COG IDs, and 32 species with differences in abundance between the two rabbit populations studied. These variables achieved a classification performance of the V_E rabbit populations of over than 80%. Compared to the high V_E population, the low V_E (resilient) population was characterised by an underrepresentation of *Megasphaera sp., A. muris, B. rodentium, R. bromii, B. togonis* and *Eggerthella sp.*, and greater abundances of *A. shahii, A. prutedinis, O. splanchnicus, L. fermentum* and *Sutterella,* among others. Differences in abundance were also found in pathways related to biofilm formation, quorum sensing, glutamate, and amino acid aromatic metabolism. All these results suggest differences in gut immunity modulation, closely related to resilience.

1.4. CONCLUSIONS

This is the first study to show that selection for V_E of LS can shift the gut microbiome composition. The results revealed differences in microbiome composition related to gut immunity modulation, which could contribute to the differences in resilience among rabbit populations. The selection-driven shifts in gut microbiome composition should make a substantial contribution to the remarkable genetic response observed in the V_E rabbit populations.

2. BACKGROUND

The dynamics and composition of gut microbiome have a substantial impact on the host's phenotypes. Previous studies on livestock have suggested that microbial variation contributes to production phenotypes, explaining between 13% and 33% of key traits [1-2]. It is thus fundamental for evolutionists and animal and plant breeders to understand how the host's microbiome shapes phenotypes and contributes to host response to selection [3], even though the complexity of microbiome inheritance and microbiome heritability (host genetics controlling microbiome) make this a challenging topic.

The livestock industry is demanding more sustainable production systems and resilience is one of the critical traits to be improved, this being the ability of individuals to maintain or quickly recover their performance after environmental disruptions [4]. In the last few years environmental variance (V_E) has successfully been used as a key measure of animal resilience [5-7]. V_E is defined as the within-individual variance of a trait. Animals showing a low V_E for a given trait seem to cope better with environmental disturbances that affect this trait and show lower mortality [6-7]. Quantitative genetics and genomic studies in different species underline the close association of V_E with the inflammatory response [8-11], while variations in gut microbiome composition can regulate the health status of individuals [12]. Conversely, the immune system, particularly the inflammatory signals, play an important role in the development of intestinal disorders and autoimmunity [13-15]. Selection for V_E might therefore effectively shift gut microbiome composition, affecting the inflammatory response and driving animal resilience [7].

This study aimed to determine the microbiome composition underlying the V_E of litter size (LS). For this we performed a metagenomic analysis considering the compositional nature of the data in two rabbit populations divergently selected for high and low V_E of LS [16]. The populations were selected from the same environmental conditions, this being an exceptional biological material to confirm the host-microbial evolution. They showed a notable genetic response to disruptive selection for V_E of LS, with a notable correlated response in resilience, mortality, and biomarkers of the immune response [7,11]. The genomic analysis also found relevant genes associated with the variation in the V_E of LS, supporting the link between the inflammatory response and the V_E [9-10].

3. METHODS

A divergent selection experiment for high and low V_E of LS was carried out on generation 13 rabbits at the Miguel Hernández University in Elche, Spain [16]. The rabbits were kept in the same room under the same environmental conditions. Cecum samples were

collected from 70 does (36 from the population with low V_E of LS and 34 from the population with high V_E of LS) slaughtered after their first parity. These samples were homogenized in 50 mL Falcon tubes and aliquoted in 2mL cryotubes for immediate snap-freezing in liquid nitrogen and storage at -80ºC until processed.

Bacterial DNA was isolated from 0.15 g of cecum samples using the DNeasy PowerSoil Kit (QIAGEN Inc, Hilden, Germany). DNA concentration and purity were estimated by measuring the 260/280 ratio with a Nanodrop ND-1000 and verifying by a Qubit™ 4 Fluorometer (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, USA). Whole bacterial genomes were sequenced at the FISABIO Sequencing and Bioinformatic Service (Valencia, Spain) by Illumina NextSeq 500 in 150 bp paired-end reads. Average coverage was set to 4,000,000 million paired-end reads per sample with a minimum of 2,000,000 paired-end reads. The shotgun library was made by the Nextera XT DNA Library Preparation Kit (Illumina Inc., San Diego, CA, USA).

Quality control of raw FASTQ files was done on FASTQ v0.11.06 software [17] and two raw FASTQ files were discarded from the analysis to low sequencing quality. Before analysing the whole metagenome data, the raw FASTQ files were pre-processed. The host genome (*Oryctolagus cuniculus* genome v.2.0.101) was removed by a pipeline that included the Bowtie2 v4.1.2 [18], SAMtools v1.2.1 [19], and BEDTools v2.29.0 software [20]. The full pipeline is available in Additional File 1. Illumina adapter removal and quality trimming of reads were performed on Trimmomatic v0.39 software [21] using "leading" and "trailing" settings of 8 bases with a minimum length of 96, a sliding window of 10, and a minimum quality score of Q15 [see Additional File 2]. The cleaned FASTQ files were analysed with the "default" settings of the "seqmerge" mode of SqueezeMeta v1.3.1 software [22] [see Additional File 3]. This software is a fully automatic metagenomic analysis pipeline that uses the latest publicly available version of the GeneBank nr, eggNOG, KEGG and PFAM database for taxonomic and functional assignment (for further details see Tamames & Puente-Sánchez, 2019 [22]). Each output dataset had the count abundance of j variables; the KEGG IDs (j=5008), COG IDs (j=14,577), or the taxonomic ranks per sample. The taxonomic rank dataset was split into four different groups: phylum (j=108), family (j=277), genus (j=647), and species (j=573).

All the statistical analyses were done in R [23]. A principal component analysis of each dataset was computed to remove outlier animals, according to the population structure. Of the 70 animals, 34 from the low V_E of LS (resilient) population and 28 from the high V_E (non-resilient) population remained in the datasets. Variables with almost 20% zeros

[24] within each population or in total (without a relevant difference of zeros among populations higher than 0.5) were removed, and one count was added to all datasets to deal with the remaining zeros. The data were transformed by the additive log-ratio (ALR) transformation to consider their compositional nature [25], using one fixed variable as denominator or reference variable (x_{ref}), with all the other variables as numerator (x_j):

$$ALR(\mathrm{j}|ref) = log\left(\frac{x_{\mathrm{j}}}{x_{ref}}\right) = \log(x_{\mathrm{j}}) - \log(x_{ref}) \quad (1)$$

where the number of total ALR is j-1, j being the total number of variables in the dataset. The dataset reference variable (KEGG IDs, COG IDs, and taxonomic ranks) was selected according to three requirements suggested by Greenacre et al. (2021) [25]: (a) the lowest variance of the log-count abundance (log (x_j)), (b) a high-count abundance (x_j), and (c) a Procrustes correlation higher than 0.9 to avoid lack of isometric in the transformed datasets. We used the lowest coefficient of variation of the log (x_j) to select the reference variables instead of the lowest variance. For KEGG and COG IDs datasets, we used the count abundance of the *RecA* gene (K03553 and COG0468, respectively) as the reference variable, as suggested in the SqueezeMeta software manual [22]. The *RecA* gene is present in most bacteria, archaea, and eukaryotes organisms and has a low copy number variation between taxa [26]. In our dataset the gene RecA for KEGG IDs and COG IDs overcame all the requirements to be a reference variable [25]. We selected the following reference variables for each taxonomic rank dataset; the phylum Firmicute, the family *Lachnospiraceae*, the genus *Butytivibrio*, and the species *Clostridium bacterium*. ALR transformed data was auto-scaled to mean 0 and standard deviation 1 before performing any statistical analysis.

Partial least square-discriminant analysis (PLS-DA) was used to identify the relevant genes and taxa to classify the rabbits among high and low V_E of LS. The PLS-DA models were computed on the mixOmics package in R [27]. A categorical vector Y of length n was used as input, indicating the rabbit population of each sample (resilient = 34 and non-resilient = 28), and an X matrix $n \times$ j dimensions, where n is the number of samples and j the number of ALR. A PLS-DA model with ten components was fitted for each ALR-transformed dataset (KEGG IDs; j=4,150, COG IDs; j=10,893, phylum; j=35, family; j=96, genus; j=212, species; j=196). An iterative process was done until each model reached the highest classification performance or a balanced error rate (BER) lower than 0.02. In each iteration, the optimal number of components for each model was selected considering the BER displayed for the Mahalanobis distance, computed by 4-fold cross-validation repeated 100 times. Feature/Variable selection was performed using the

variable important prediction (VIP) i.e., the influence of the variables on the model projection and classification for the number of components previously selected. The optimal number of variables to select were those with a VIP higher than 1 [28].

The prediction performance of the final models was validated by two tests: a confusion matrix and a permuted-confusion matrix. The former was constructed by 4-fold cross-validation repeated 10,000 times. The models' accuracy and precision were calculated considering the resilient population as the true positive value. We also computed a permuted-confusion matrix randomizing the categorical Y vector of the rabbit populations to check whether the classification performance of the final models was spurious. These were considered spurious when the percentage of true positives in the permuted-confusion matrix was far from 50% (random probability of two categories). A full record of the method used is included in Additional Files 4-6.

Bayesian statistics [29] was used to determine the relevance of the difference between the two rabbit populations in the microbial genes and taxonomy selected by PLS-DA [see Additional File 7]. The analysis was by four chains with a length of 50,000 iterations, a lag of 10, a burn-in of 1,000 iterations and flat priors. To check whether the model converged the R-hat statistic had to be below 1.05 [30]. The marginal posterior distribution of the differences among the resilient minus non-resilient population was computed to estimate its posterior mean and the probability of the difference being higher (if the difference is positive) or lower (if negative) than 0 (P_o). The posterior mean of the differences was indicated as units of standard deviations (SD) of each variable (unit of SD). Variables with an SD higher than 0.5 and a P_o higher than 0.9 were considered the most relevant for the classification and differentiation of the two rabbit populations.

The alpha- and beta-diversity were computed using the ALR at the species level to measure the differences in microbiome composition among the rabbit populations. The alpha-diversity was measured by Shannon's (H') and inverse Simpson indexes. The same indexes analysed the species diversity and evenness in the samples. Differences in the distribution of alpha-diversity among rabbit populations were considered when the p-value of a Mann-Whitney U test was lower than 0.05. Beta-diversity was measured by the Bray-Curtis dissimilarity matrix and a nonmetric multidimensional scaling (NMDS) was carried out to retrieve the loadings of the first two dimensions. Differences in microbial species composition were tested by the permutational multivariate analysis of variance (PERMANOVA; p-value <0.05) on the loadings of the two first MDS dimensions. A full record of the alpha- and beta-diversity calculation is included in Additional file 8.

4. RESULTS AND DISCUSSION

After the additive log-ratio (ALR) transformation, the partial least square discriminant analysis (PLS-DA) identified 361 relevant variables, including: 116 KEGG IDs, 164 COG IDs, 6 phyla, 15 families, 28 genera, and 32 species. Most models achieved a high classification performance of the rabbit populations in terms of resilience potential, given that rabbits with high V_E are considered less resilient than those with low V_E (Table 1) [see Additional Files 4-6]. The best models were those using the KEGG and COG IDs for functional assignment and the species level for taxonomic assignment (Table 1). The models using counts from functional assignment allowed higher discrimination than those from the taxonomic assignment (Table 1). The taxonomic ranks were inferred from the functional assignment [22], having a lower statistical power for discriminating between the two rabbit populations (Fig.1) due to the loss of information in the assignment.

Table 1. PLS-DA model specifications using counts from genes and taxa of the resilient and non-resilient rabbit populations

PLS-DA model	N[(a)]	Component[(b)]	Classification performance[(c)]		Accuracy*	Precision*
			Resilient[(d)]	Non-resilient[(e)]		
Phylum	6	2	66.68	66.32	0.67	0.66
Family	15	3	78.97	78.89	0.79	0.79
Genus	28	1	74.28	82.70	0.79	0.81
Species	32	2	84.87	88.87	0.87	0.89
KEGG	116	3	99.83	99.77	1	1
COG	164	3	99.88	99.99	1	1

PLS-DA models with taxa were those with phylum, family genus and species. PLS-DA models with genes were those with the KEGG and COG IDs
[(a)] Number of variables in the final model
[(b)] Number of components in the final model
[(c)] Final PLS-DA model classification performance (%) of each rabbit population. (True positive value from the total assignation to each rabbit population).
[(d)]Population with low V_E of LS
[(e)]Population with high V_E of LS
* Accuracy and precision of the final model considering the resilient population as true positive.

Likewise, the higher (Fig.1A) had less discrimination power than the lower taxonomic ranks (Fig. 1B). Clustering counts in the higher taxonomic ranks (phylum or family) could hide their variation between the populations due to grouping bacteria with dissimilarity in their functions. The results show that a few species (32) were relevant for the classification among the two rabbit populations, obtaining an accuracy of 0.87 and a precision of 0.89 (Table 1). These results were supported by the alpha- and beta-diversity scores, which did not differ between the two rabbit populations (Fig. 1D-F), indicating that in general both populations have a similar microbiome composition except for a few species identified by the PLS-DA.

The Bayesian statistical analysis [see Additional File 7] showed that 303 variables (including both genes and taxa) from the initial 361 identified by PLS-DA analysis (Table 1) had a posterior mean of the differences among the rabbit populations of at least 0.5 of the SD of the variable [see Additional File 9] in which the probability of differences being higher or lower than 0 (P0) was higher than 0.97. The Bayesian results showed that most of the variables included in the PLS-DA models (Table 1) are key variables for discriminating between rabbit populations, with relevant differences in mean abundance [see Additional File 9]. These differences must have arisen because of the divergent selection for VE of LS, since the animals were coetaneous and kept under the same environmental conditions (diet, management, temperature, etc).

Relevant results in the PLS-DA models using the taxonomic ranks are detailed below. The species *Alistipes shahii* (0.60 unit of SD), *Alistipes putredinis* (0.51), *Odoribacter splanchnicus* (0.58), and *Limosilactobacillus fermentum* (0.57) was more abundant in the resilient animals (Fig.2), as were the higher taxonomic ranks of these species (Fig. 2): genera *Odoribacter* (0.83), *Alistipes* (0.75), *Lactobacillus* (0.56) and *Rikenella* (0.51); families *Odoribacteraceae* (0.84) and *Rikenellaceae* (0.74); and the phylum *Bacteroidetes* (0.59). Health-beneficial properties have been reported from these taxa, in part due to their effects on the inflammatory and immune adaptive response [31-33]. These effects on the immune system have been suggested to be mediated by short-chain fatty acids (SCFs) and Th17 cells. SCFs have anti-inflammatory properties [12, 34] and differentiation of Th17 cells is essential for the host to develop a correct tolerance to foreign and non-pathogenic commensal species, playing an important role in gut immunity [35].

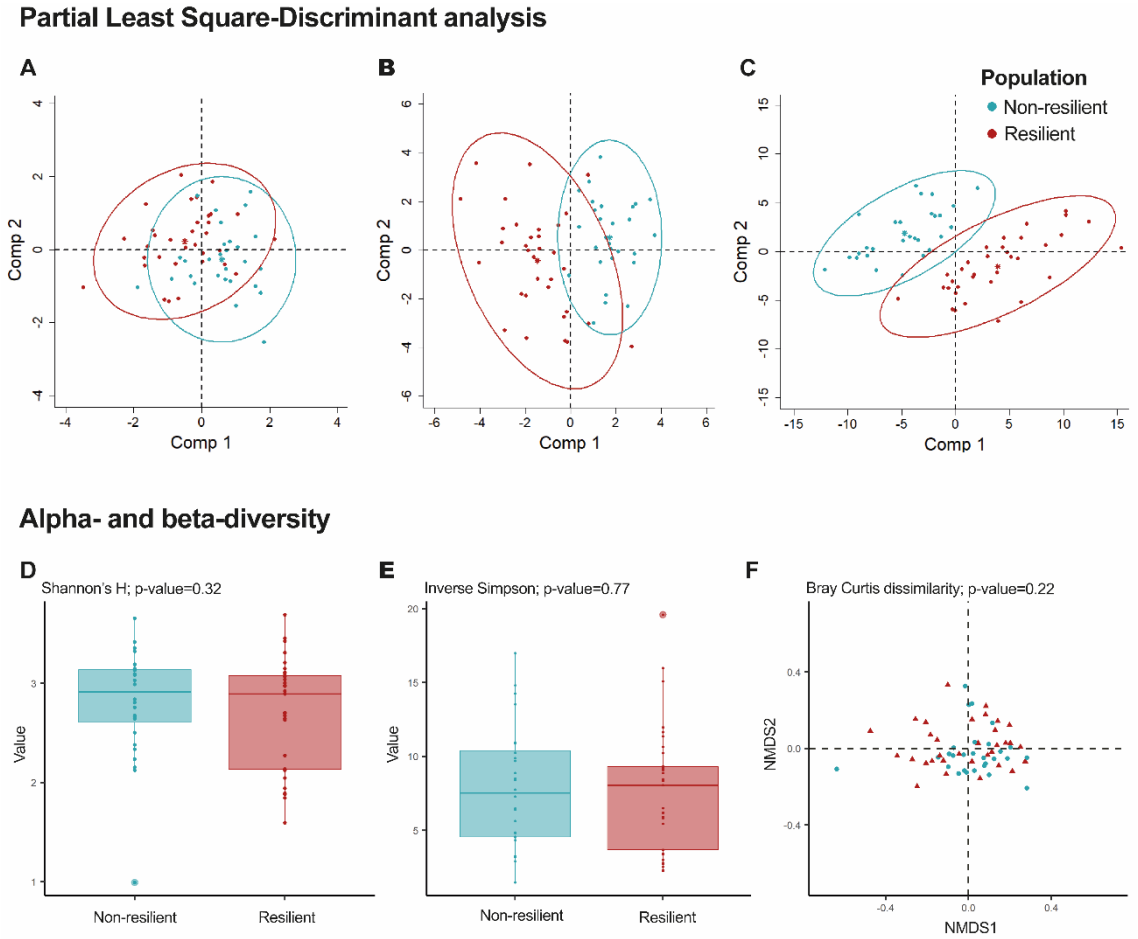


Figure 1. Gut microbiome composition dissimilarity. Representation of the first (Comp 1) and second components (Comp 2) of the final partial least square-discriminant analysis (PLS-DA) models (Table 1) and alpha- and beta-diversity scores from the resilient (red) and non-resilient (blue) rabbit populations. PLS-DA plotting was performed using three different datasets: (A) phyla abundances, (B) species abundances, and (C) KEGG IDs abundances. The alpha- and beta-diversity scores were calculated with the additive log-ratio of each species abundance according to a reference species (*Clostridium bacterium*). Alpha-diversity was computed using (D) Shannon's H index and (E) Inverse Simpson index. Beta-diversity was computed by calculating (F) the Bray Curtis dissimilarity matrix. Differences among populations were established with a p-value lower than 0.05.

Harmful microbial species such as *Acetatifactor muris* (-0.72 of SD unit) and *Eggerthella sp.* (-0.63) were more abundant in the non-resilient rabbits (Fig.2), which was consistent with their associations with autoimmunity and inflammatory diseases [36-37]. Species like *Megasphaera sp.* (-0.75), *Bacteroidetes rodentium* (-0.70), *Ruminococcus bromii* (-0.67), and *Bacteroides togonis* (-0.63) also showed higher abundance in the non-resilient population (Fig.2). Although we did not find any evidence in the literature of a possible negative effect of these species on the host-health status, the effect of microbial species on individual health still remains unclear. There are discrepancies in the literature on the gut microbiome composition related to health and disease. Species such as *Alistipes putredinis* have been identified as both beneficial and harmful species [31, 36], as have the genus *Sutterella* (0.83) and the family *Sutterellaceae* (0.60), with a higher abundance in resilient rabbits (Fig.2). The *Sutterella*

genus was identified in healthy and inflammatory bowel disease patients [38], particularly in those with Ulcerative Colitis [39]. The studies could only suggest the participation of the *Sutterella* genus in the modulation of the inflammatory response [39] through the alteration of IgA levels, an important immunoglobulin to neutralize pathogens and prevent infections [40]. All relevant species for the classification between the resilient and non-resilient animals can find in the Additional File 9.

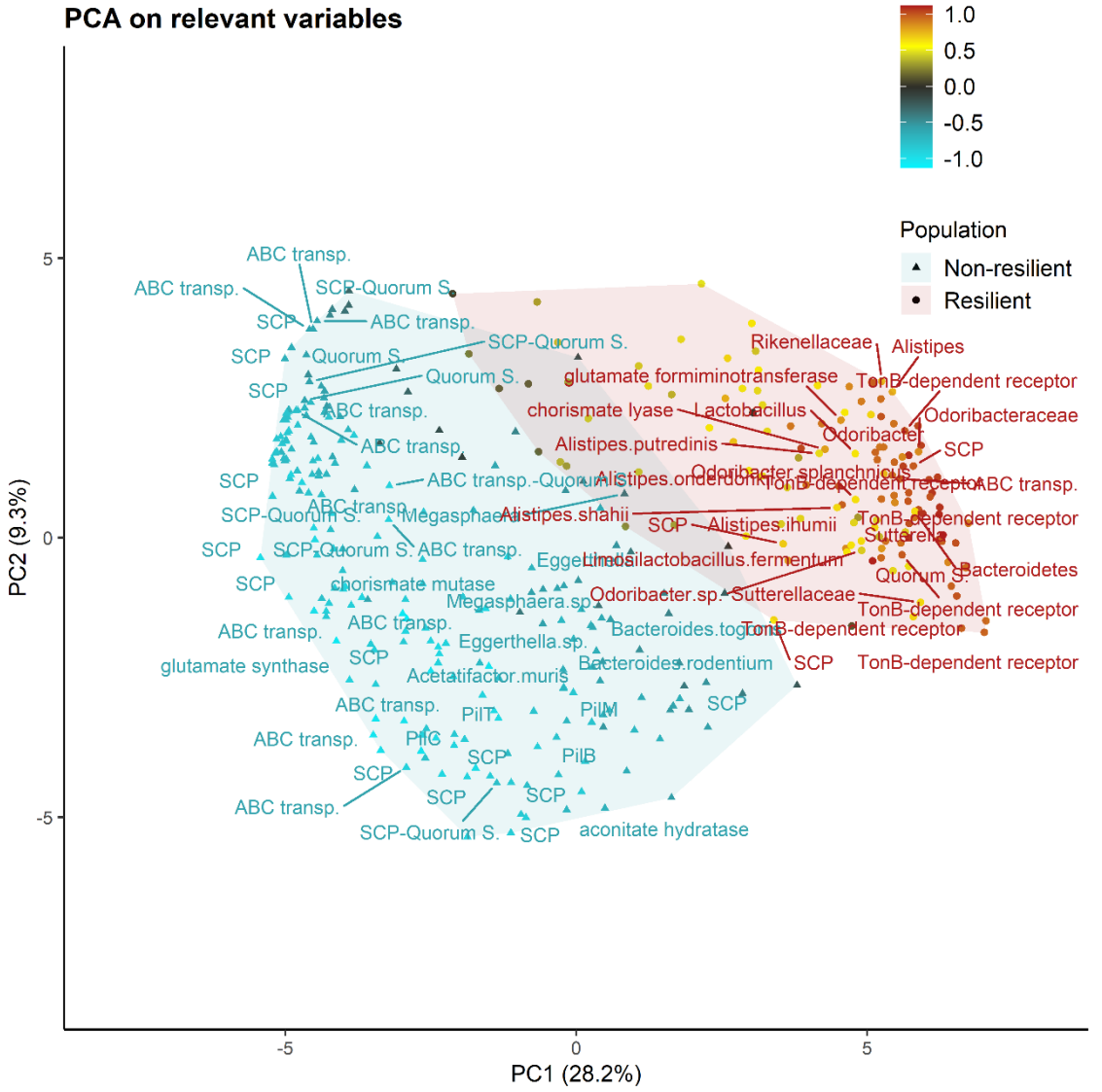


Figure 2. Principal component analysis of gut microbiome composition. Representation of first (PC1) and second principal component (PC2) of the additive log-ratio transformation of the relevant variables for distinguishing between resilient and non-resilient populations. SD indicates the unit of standard deviations of each variable from the posterior mean of the differences between the marginal posterior distributions of the rabbit populations. The SD colour gradient highlights the degree of difference, blue and red being the greatest differences among the rabbit populations. The blue-shaded area indicates that higher microbiome abundance in the non-resilient population. The red-shaded area indicates higher microbiome abundance in the resilient population. SCP: Proteins involved in signalling and cellular processing; ABC transp: Proteins of the family ABC transporter; Quorum S.: Proteins involved in quorum sensing. PilB, PilC, PilM and PilT: proteins that conform the pilus.

We also identified differences in relevant pathways that might contribute to the differences in V_E and resilience between the rabbit populations. We identified 42 KEGG

IDs in signalling and cellular processing [see Additional File 4-5], which were generally more abundant in the non-resilient population (Fig.2). We also highlighted those KEGG and COG IDs related to the ABC transporters (50), Quorum sensing (11) and pilus protein conformation (4), three components essential to form biofilms [41-42], which have been associated with both an ill and a healthy gut, so again it was necessary to identify the tipping point between a beneficial or harmful effect [43]. The genes aconitate hydratase (K01681; -0.73), glutamate synthase (K00284; -1.1), and glutamate formiminotransferase (K13990; 0.80) also showed differences between the rabbit populations (Fig.2), differences that could suggest different ways of synthesizing L-glutamate, according to the substrate used. Non-resilient animals could have reduced levels of glutamine, which might shift their microbiome balance and inflammatory response towards a worse health status [44]. Differences in the genes belonging to the chorismite metabolism were also found for chorismate mutase (K14170; -0.94) and chorismate lyase (K18240; 0.78). There are few studies in the literature on the impact of these enzymes' end-products (prephenate and 4-hydroxybenzoate, respectively), even though these genes are important for the metabolism of the aromatic amino acids phenylalanine, tyrosine, and tryptophan, which are linked to mucosal integrity and immune homeostasis in the gut [45]. They could thus be important for the modulation of V_E and animal resilience. All relevant genes for the classification between the resilient and non-resilient animals can find in the Additional File 9.

Since microbial inheritance is complex, more research is required to understand the implication of the differences in the microbiome composition found in this study. The microbiome variability between these two rabbit populations could be an effect or a cause of the remarkable genetic response for V_E of LS [16]. Microbial species with a contribution to the phenotype can be selected throughout generations, while selection could also modify the microbiome composition of species with microbial heritability; i.e., influenced by the genome of the host and not necessary with a contribution to the selected trait [3, 46]. A number of studies show how the host genome shapes the microbial abundance of around 10-97% of total microbial species, microbial heritability ranging between 0.008 and 0.64 [47-48]. In these rabbit populations several genomic regions were associated with the differences in V_E [9-10], so that the underlying genes might also affect the gut microbiome composition. For instance, the *DOCK2* gene identified as associated with the rabbit population on the rabbit chromosome 3 [9-10] has been suggested to modify gut microbiome composition in a study on knockout mice [49]. Further studies are needed to determine the implications of the host genome on shaping the V_E of LS and animal resilience.

Microbiome composition is multifactorial and different species could have different roles in health, according to the host genotype, diet, microbial interactions, and environmental factors, among others [12, 50]. Standardized factors affecting the gut microbiome composition are necessary to obtain reproducible results. This study has an advantage over other studies, as diet and environmental conditions were the same for both rabbit populations for 13 generations and the rabbits were coetaneous. Controlling these factors allowed us to decipher the commensal consortia or microbiome composition possibly associated with the V_E selection studied. Our results suggest that modulating gut immunity functions mediate some of the differences in resilience among rabbit populations [7, 11].

5. CONCLUSIONS

This is the first study to show that selection for V_E of LS can shift the gut microbiome in animals under the same environmental conditions. We identified 116 KEGG IDs, 164 COGs IDs and 32 species with differences in abundance between two rabbit populations with outstanding differences of V_E for LS after 13 generations of selection as a result of the V_E selection performed. The resilient rabbit population (with low V_E of LS) had lower abundance of *Megasphaera sp., A. muris, B. rodentium, R. bromii, B. togonis* and *Eggerthella sp.*, and greater abundance of *A. shahii, A. prutedinis, O. splanchnicus, L. fermentum* and *Sutterella,* among others. Differences in abundance were also found in pathways related to biofilm formation, quorum sensing, glutamate, and amino acid aromatic metabolism. The results suggest that differences in gut immunity modulation could drive the differences in resilience among rabbit populations. We also suggest that *DOCK2* could be one of the host's genes that influence gut microbiome composition. Due to the limited information in this field, further studies should be carried out to validate these results.

6. DECLARATIONS

Ethics approval and consent to participate

All the experimental procedures were approved by the Committee of Ethics and Animal Welfare of the Miguel Hernández University, according to Council Directives 98/58/EC and 2010/63/EU.

Consent for publication

Not applicable.

Availability of data and materials

Data are available upon request to the corresponding author.

Funding

This study was supported by Projects AGL2014-5592, C2-1-P and C2-2-P, AGL2017-86083, C2-1-P and C2-2-P, and PID2020-115558GB-C21, funded by the Spanish *Ministerio de Ciencia e Innovación (MIC)-Agencia Estatal de Investigación* (AEI) and the European Regional Development Funds (FEDER)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CCR analyzed the data and wrote the manuscript. AB and MJA conceived and designed the study and contributed to the discussion of the results. MLG contributed to the study design and the discussion of the results. RP analysed the lab data, contributed to the discussion of the results, and edited the manuscript. NIE conceived the study, contributed to the discussion of the results, and edited the manuscript. All the authors have read and approved the final manuscript.

Acknowledgements

We are grateful to Fernando Puente-Sánchez and Javier Tamames for their support with the SqueezeMeta software. We also acknowledge the contribution of Francisco Rosich of ASIC to the optimization of the metagenomic pipeline. Cristina Casto-Rebollo acknowledges a FPU17/01196 scholarship from the Spanish Ministry of Science, Innovation and Universities and PAID-12-21 funding for open access charge: Universitat Politècnica de València.

7. REFERENCES

1. Camarinha-Silva A, Maushammer M, Wellmann R, et al. Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. Genetics. 2017;206(3):1637-44. doi:10.1534/genetics.117.200782.

2. Difford GF, Plichta DR, Løvendahl P, et al. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLoS Genet. 2018;14(10):e1007580. doi:10.1371/journal.pgen.1007580.

3. Henry LP, Bruijning M, Forsberg SKG, Ayroles JF. The microbiome extends host evolutionary potential. Nat Commun. 2021 Aug 26;12(1):5141. doi:10.1038/s41467-021-25315-x.

4. Colditz IG, Hine BC. Resilience in farm animals: biology, management, breeding and implications for animal welfare. Anim Prod Sci. 2016;56:1961–83. doi:doi.org/10.1071/AN15297

5. Berghof TVL, Poppe M, Mulder HA. Opportunities to improve resilience in animal breeding programs. Front Genet. 2019;9:692. doi:10.3389/fgene.2018.00692

6. Poppe M, Mulder HA, Veerkamp RF. Validation of resilience indicators by estimating genetic correlations among daughter groups and with yield responses to a heat wave and disturbances at herd level. J Dairy Sci. 2021 Jul;104(7):8094-8106. doi:10.3168/jds.2020-19817.

7. Argente MJ, García ML, Zbyňovská K, et al. Correlated response to selection for litter size environmental variability in rabbits' resilience. Animal. 2019; 13:2348-55. doi:10.1017/S1751731119000302

8. Iung LHDS, Carvalheiro R, Neves HHDR, Mulder HA. Genetics and genomics of uniformity and resilience in livestock and aquaculture species: A review. J Anim Breed Genet. 2020;137:263–80. doi:10.1111/jbg.12454

9. Casto-Rebollo C, Argente MJ, García ML, et al. Identification of functional mutations associated with

environmental variance of litter size in rabbits. Genet Sel Evol. 2020;52(1):22. doi:10.1186/s12711-020-00542-w

10. Casto-Rebollo C, Argente MJ, García ML, et al. Selection for environmental variance of litter size in rabbits involves genes in pathways controlling animal resilience. Genet Sel Evol. 2021;53(1):59. doi:10.1186/s12711-021-00653-y

11. Beloumi D, Blasco A, Muelas R, et al. Inflammatory Correlated Response in Two Populations of Rabbit Selected Divergently for Litter Size Environmental Variability. Animals. 2020;10:1540. doi:10.3390/ani10091540

12. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. Gut. 2022;71(5):1020-32. doi:10.1136/gutjnl-2021-326789

13. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30: 492–506.doi:10.1038/s41422-020-0332-7

14. Manor O, Dai CL, Kornilov SA, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun. 2020 Oct 15;11(1):5206. doi:10.1038/s41467-020-18871-1

15. Rinninella E, Raoul P, Cintoni M, et al. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms. 2019;7(1):14. doi:10.3390/microorganisms7010014

16. Blasco A, Martínez-Álvaro M, García ML, et al. Selection for environmental variance of litter size in rabbit. Genet Sel Evol. 2017;49:48. doi:10.1186/s12711-017-0323-4

17. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

18. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012; 9:357-359. doi: 10.1038/nmeth.1923

19. Li H, Handsaker B, Wysoker A, et al; 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics. 2009;25:2078-9. doi:10.1093/bioinformatics/btp352

20. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033

21. Bolger AM, Lohse M, Usadel B.. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. 2014;30(15):2114-20. doi:10.1093/bioinformatics/btu170.

22. Tamames J and Puente-Sánchez F. SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front. Microbiol. 2019;9:3349. doi:10.3389/fmicb.2018.03349.

23. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

24. Bijlsma S, Bobeldijk I, Verheij ER, et al. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. Anal Chem. 2006;78(2):567-74. doi:10.1021/ac051495j.

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

26. Wu D, Jospin G, Eisen JA. Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. PLoS One. 2013;8(10):e77033. doi:10.1371/journal.pone.0077033.

27. Rohart F, Gautier B, Singh A, Le Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Computational Biology 2017;13(11):e1005752. doi:10.1371/journal.pcbi.1005752

28. Galindo-Prieto B, Eriksson L, Trygg J. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). Journal of Chemometrics. 2014;28:623– 632. doi:10.1002/cem.2627

29. Blasco A. Bayesian Data Analysis for Animal Scientists: The Basics. 2017. Springer, Cham. doi:10.1007/978-3-319-54274-4

30. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statist. Sci. 1992;7(4):457-72. doi:10.1214/ss/1177011136

31. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. Front. Immunol. 2020;11:906. doi:10.3389/fimmu.2020.00906

32. Rodríguez-Sojo MJ, Ruiz-Malagón AJ, Rodríguez-Cabezas ME, et al. Limosilactobacillus fermentum CECT5716: Mechanisms and Therapeutic Insights. Nutrients. 2021;13:1016. doi:10.3390/ nu13031016.

33. Xing C, Wang M, Ajibade AA, et al. Microbiota regulate innate immune signaling and protective immunity against cancer. Cell Host Microbe. 2021;29(6):959-74.e7. doi:10.1016/j.chom.2021.03.016.

34. Parada Venegas D, De la Fuente MK, Landskron G, et al. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. Front Immunol. 2019;11;10:277. doi:10.3389/fimmu.2019.00277. Erratum in: Front Immunol. 2019;28;10:1486.

35. Blaschitz C, Raffatellu M. Th17 cytokines and the gut mucosal barrier. J Clin Immunol. 2010;30(2):196-203. doi:10.1007/s10875-010-9368-7.

36. Lee C, Hong SN, Paik NY, et al. CD1d Modulates Colonic Inflammation in NOD2-/- Mice by Altering the Intestinal Microbial Composition Comprising Acetatifactor muris. J Crohns Colitis. 2019;13(8):1081-91. doi:10.1093/ecco-jcc/jjz025.

37. Alexander M, Ang QY, Nayak RR, et al. Human gut bacterial metabolism drives Th17 activation and colitis. Cell Host Microbe. 2022;30(1):17-30.e9. doi:10.1016/j.chom.2021.11.001.

38. Hiippala K, Kainulainen V, Kalliomäki M, et al. Mucosal Prevalence and Interactions with the Epithelium Indicate Commensalism of Sutterella spp. Front Microbiol. 2016;7:1706. doi:10.3389/fmicb.2016.01706.

39. Kaakoush NO. Sutterella Species, IgA-degrading Bacteria in Ulcerative Colitis. Trends Microbiol. 2020;28(7):519-22. doi:10.1016/j.tim.2020.02.018.

40. Hansen IS, Baeten, DLP, den Dunnen J. The inflammatory function of human IgA. Cell. Mol. Life Sci. 2019;76**:**1041–55. doi:10.1007/s00018-018-2976-8.

41. Ligthart K, Belzer C, de Vos WM, Tytgat HLP. Bridging Bacteria and the Gut: Functional Aspects of Type IV Pili. Trends Microbiol. 2020;28(5):340-8. doi:10.1016/j.tim.2020.02.003.

42. Li YH and Tian X. Quorum sensing and bacterial social interactions in biofilms. Sensors (Basel). 2012;12(3):2519-38. doi:10.3390/s120302519.

43. Tytgat HLP, Nobrega FL, van der Oost J, de Vos WM. Bowel Biofilms: Tipping Points between a Healthy and Compromised Gut? Trends Microbiol. 2019;27(1):17-25. doi:10.1016/j.tim.2018.08.009.

44. Kim MH, Kim H. The Roles of Glutamine in the Intestine and Its Implication in Intestinal Diseases. Int J Mol Sci. 2017;18(5):1051. doi:10.3390/ijms18051051.

45. Liu Y, Hou Y, Wang G, et al. Gut Microbial Metabolites of Aromatic Amino Acids as Signals in Host-Microbe Interplay. Trends Endocrinol Metab. 2020;31(11):818-34. doi:10.1016/j.tem.2020.02.012.

46. Douglas GM, Bielawski JP, Langille MGI. Re-evaluating the relationship between missing heritability and the microbiome. Microbiome. 2020;8:87. doi:10.1186/s40168-020-00839-4.

47. Goodrich JK, Davenport ER, Clark AG, Ley RE. The Relationship Between the Human Genome and Microbiome Comes into View. Annu Rev Genet. 2017;51:413-33. doi:10.1146/annurev-genet-110711-155532

48. Grieneisen L, Dasari M, Gould TJ, et al. Gut microbiome heritability is nearly universal but environmentally contingent. Science. 2021;373(6551):181-86. doi:10.1126/science.aba5483.

49. Xie Y, Chen J, Wu B, et al. Dock2 affects the host susceptibility to Citrobacter rodentium infection through regulating gut microbiota. Gut Pathog. 2021;14;13(1):52. doi:10.1186/s13099-021-00449-x.

50. Dong TS, Gupta A. Influence of Early Life, Diet, and the Environment on the Microbiome. Clin Gastroenterol Hepatol. 2019;17(2):231-42. doi:10.1016/j.cgh.2018.08.067.

8. ADDITIONAL FILES

Additional files are available with the preprint at https://doi.org/10.21203/rs.3.rs-1597523/v1

Additional File 1

File format: .sh
Title: Additional file 1
Description: Full pipeline of the pre-processing raw FASTQ files

Additional File 2

File format: .sh
Title: Additional file 2
Description: Full pipeline of Illumina adapter removal and quality trimming of reads

Additional File 3

File format: .sh
Title: Additional file 3
Description: Full pipeline to run the SqueezeMeta software

Additional File 4

File format: .html
Title: Additional file 4
Description: Full pipeline to obtain the relevant KEGG IDs

Additional File 5

File format: .html
Title: Additional file 5
Description: Full pipeline to obtain the relevant COG IDs

Additional File 6

File format: .html
Title: Additional file 6
Description: Full pipeline to obtain the relevant taxa

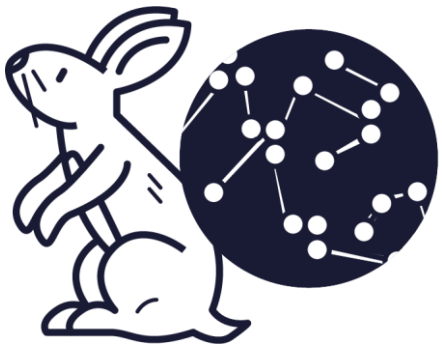
Additional File 7

File format: .html
Title: Additional file 7
Description: Full pipeline with the Bayesian statistical analysis

Additional File 8

File format: .html
Title: Additional file 8

Description: Full pipeline with the alpha- and beta-diversity
Title: Additional file 9
Description: Results of the Bayesian statistical analysis. File includes the name of the variables and the type (taxonomic ranks, KEGG or COG), posterior mean of the differences among the resilient and non-resilient populations (meanDiff), the probability of the difference being higher (if the difference is positive) or lower (if negative) than 0 (P0), the highest posterior interval density of 95% (HPD95), and the name and pathways of the genes for the KEGG and COG IDs.



CHAPTER 6: Dissimilarities in the gut metabolome of rabbits with genetic differences in their resilient potential

Authors

Cristina Casto-Rebollo[1], María José Argente[2], María Luz García[2], Agustín Blasco[1], Noelia Ibáñez-Escriche[1].

Institutional affiliations

[1]Institute for Animal Science and Technology, Universitat Politècnica de València, València, Spain

[2]Centro de Investigación e Innovación Agroalimentaria y Agroambiental (CIAGRO_UMH), Miguel Hernández University, Orihuela 03312, Spain

The content of this chapter is under review by Genetic Selection Evolution (Version 2). Submitted 23th September 2022. The preprint is available at https://doi.org/10.21203/rs.3.rs-1808216/v1

1. ABSTRACT

1.1. BACKGROUND

The gut metabolites are key actors in the host-microbiota crosstalk and can have either beneficial or detrimental effects on the host. In livestock, the study of the gut metabolome is emerging. Its study could help to understand their effects on key traits such as animal resilience and welfare. Furthermore, we could expand our knowledge about host-gut interactions. Animal resilience has now become a critical livestock trait because of the high demand for more sustainable productions. Environmental variance (V_E), the within-individual variance of traits due to environmental factors, is used as a measure of resilience. Studying the gut microbiome composition can reveal the mechanisms behind animal resilience due to its influence on host immunity. The aim of this study was to identify the gut metabolites underlying the differences in the resilience potential of animals from a divergent selection for V_E of litter size (LS). We performed an untargeted gut metabolome analysis in two rabbit populations divergently selected for low (n=13) and high (n=13) V_E of LS. Partial least square-discriminant analysis and Bayesian statistics were computed to determine dissimilarities in the gut metabolites of the rabbit populations.

1.2. RESULTS

We identified 15 metabolites that can discriminate between rabbit populations with a prediction performance of 99.18% for resilient and 90.42% for non-resilient populations. These metabolites were suggested to be biomarkers of animal resilience as they were the most reliable: glycerophosphoglycerol, equol, behenoylcarnitine, arachidoylcarnitine, ethyl beta-glucopyranoside, 3-(4-hydroxyphenyl)lactate, N6-acetyllysine, 5-aminovalerate, betaine, succinylcarnitine and palmitoyl dihydrosphingomyelin. Five of them were suggested to come from the microbiota metabolism, being indicators of dissimilarities in the microbiome composition between the rabbit populations. A low abundance of acylcarnitines and metabolites derived from the phenylalanine, tyrosine, and tryptophan metabolism were found in the resilient population. These pathways can impact the inflammatory response and the health status of animals.

1.3. CONCLUSIONS

This is the first study to identify gut metabolites that could act as potential resilience biomarkers. These results supported the differences in the resilience between the rabbit populations, that were generated by the selection for V_E of LS. Then, the selection for

environmental variance modified the gut metabolome which could be another factor to modulate animal resilience. However, further studies will be needed to determine the causal role of these metabolites in health and disease.

2. BACKGROUND

The gut metabolites are key actors in the host-microbiota crosstalk and can have either beneficial or detrimental effects on the host [1-3]. They can act in the gut or travel through the plasma to reach the host's tissues, influencing the functions of the liver, brain, and immune system [4]. Gut metabolites can be derived from (i) the microbiota, due to the conversion of non-digestible components from diet or to synthesis "de novo", (ii) the host, and (iii) the diet. Moreover, metabolites from the host could be modified by the gut microbiota (5-6). In livestock, the study of the gut metabolome is emerging. Recently, gut metabolome dissimilarities have been found in traits such as feed efficiency [7-8], and milk yield protein [9]. Thus, the study of the gut metabolome could help to expand the knowledge of host-gut interactions. Furthermore, it will allow to understand their effects on key traits such as methane emissions and animal resilience.

Animal resilience has become a critical trait in the livestock system because of the high demand for a more sustainable livestock system. Resilience is the ability of animals to maintain or quickly recover their production performance despite an environmental perturbation [10]. Environmental variance (V_E) was highly correlated with animal resilience, becoming an interesting trait to measure resilience [10-12]. Indeed, animals with low V_E for a given trait seem to cope better with environmental disturbances [11-12]. Quantitative genetics and genomic studies in different species underline the importance of the immune system in modulating animal resilience [13-16]. The gut metabolome is closely related to the modulation of the immune system [3, 17], so the study of the gut metabolome could insight into promising mechanisms underlying the animal's resilience.

A previous metagenomic study found differences in microbiome composition (genes and species) between two rabbit populations selected for high and low V_E of litter size (LS) [18]. These lines showed a notable genetic response to the selection performed, as well as correlated responses in resilience indicators [12,14]. The present study is an extension of the previous metagenome study [18] and aims to identify the gut metabolites related to the resilient potential of these rabbit populations. The gut microbiome is a complex ecosystem and is strongly influenced by environmental factors [19-20], and the metabolite origin [6]. The reduction of cofounded factors is necessary to correctly decipher the gut metabolome variability behind the complex traits. The rabbit

populations used in this study were coetaneous and were selected under the same environmental conditions and diet throughout 13 generations. Thus, they were an exceptional material to unravel the gut metabolites which can act as biomarkers of resilience.

3. METHODS

The rabbits used in this study belonged to the 13th generation of a divergent selection experiment for high and low V_E of LS carried out at the University Miguel Hernández of Elche, Spain [21]. Cecum samples were collected from 28 does (14 from both populations) sacrificed after their first parity by an intravenous injection of sodium thiopental at a dose of 50mg/kg of body weight (Thiobarbital, B. Braun Medical S. A., Barcelona, Spain). The samples were homogenized in 50 mL Falcon tubes and aliquoted in 2mL cryotubes for immediate snap-freezing in liquid nitrogen and storage at -80ºC until processed.

Untargeted metabolite analysis of the gut content was conducted on the Metabolon Discovery HD4 platform. The samples were prepared by the Hamilton Company's automated MicroLab STAR® system. Prior to extraction, several recovery standards were added for quality control (QC) purposes. The proteins were precipitated in methanol under vigorous shaking for two minutes followed by centrifugation to recover chemically diverse metabolites. The resulting extract was divided into five aliquots and the organic solvent was removed by a TurboVap® (Zymark).

The gut content metabolites were profiled by Ultrahigh Performance Liquid Chromatography (UPLC)-Tandem Mass Spectrometry (UPLC-MS/MS) methods with negative and positive ion mode electrospray ionization (ESI). All the methods used a Waters ACQUITY UPLC system (Waters, Milford, MA, USA) and a Q-Exactive high resolution/accurate mass spectrometer (Thermo Fisher Scientific) interfaced with a heated electrospray ionization (HESI-II) source and an Orbitrap mass analyser operated at 35,000 mass resolution. The aliquots were dried and reconstituted in solvents compatible with the method used and containing standards at fixed concentrations to ensure injection and chromatographic consistency. QC samples were injected throughout the platform run to remove artifacts and background noise and to distinguish biological variability from process/instrument variability.

Of a total of five aliquots, two were analysed by two separate reverse phases (RP)/UPLC-MS/MS methods with acidic positive ion mode ESI. One was chromatographically optimized for more hydrophilic compounds, and the other for more hydrophobic

compounds. To detect the former, the aliquot was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 μm) using water and methanol, containing 0.05% of perfluoropentanoic acid (PFPA) and 0.1% of formic acid (FA). To detect the latter, the aliquot was gradient eluted from the same C18 column but using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA, and was operated at an overall higher organic content. The third aliquot was analysed by RP/UPLC-MS/MS with basic negative ion mode ESI using a separate dedicated C18 column and eluted with methanol, water and 6.5mM of Ammonium Bicarbonate at pH 8. The fourth aliquot was analysed via negative ion mode ESI with a HILIC column (Waters UPLC BEH Amide 2.1x150mm, 1.7 μm) using a gradient of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The last aliquot was reserved for backup. Raw data files were obtained after a tandem mass spectrometry analysis, alternating between MS and data-dependent MS^n scans, using dynamic exclusion. The scan range varied slightly between methods but covered a 70-1000 mass to charge ratio (m/z).

Raw data were extracted, peaks were identified, and QC was processed on Metabolon hardware and software. A total of 765 metabolites were identified by a library that included three criteria of more than 3300 authenticated standard components: retention time/index (RI), mass to charge ratio (m/z), and chromatographic data including MS/MS spectral data. All three criteria can be used to distinguish and differentiate metabolites. Metabolite quantification was based on the area-under-the-curve of detected peaks.

All statistical analyses were done in R [22]. A principal component analysis was computed using 480 metabolites (with no missing values) from a total of 765. Of the 28 animals, 13 from the low V_E of LS (resilient) population and 13 from the high V_E (non-resilient) population remained in the datasets. Metabolites with more than 20% zeros [23] within each population or in total (without a relevant difference of zeros among populations higher than 50%) were removed. The remaining zeros were replaced by the half of minimum peak intensity identified by the UPLC-MS/MS method to which each metabolite belonged. A total of 654 metabolites from 26 samples (13 from resilient and 13 from non-resilient populations) were transformed using the additive log-ratio (ALR) transformation, due to the compositional nature of metabolomic data [24]. The reference variable (x_{ref}; Uracil nucleotide) was the one with the lowest coefficient of variation. The ALRs were calculated following:

$$ALR(\mathrm{j}|ref) = log\left(\frac{x_{\mathrm{j}}}{x_{ref}}\right) = \log(x_{\mathrm{j}}) - \log(x_{ref}) \quad (1)$$

where the number of total ALR is j-1, being j the total number of variables in the dataset. Procrustes correlation was performed to check lack of isometry in the transformed dataset [25]. ALR transformed data were auto-scaled with mean of 0 and standard deviation of 1.

A partial least square-discriminant analysis (PLS-DA) was performed to identify the most relevant metabolites to discriminate resilient from non-resilient populations. The PLS-DA model was computed on the mixOmics packages in R [26], using a categorical vector **y**, indicating the rabbit population of each sample, and a matrix **X** with the ALR of each metabolite for each sample. The optimal number of components were those with the lowest balance error rate (BER) for the Mahalanobis distance, computed by 4-fold cross-validation repeated 100 times. Metabolites with a contribution to the model prediction (VIP) lower than 1 were removed from the dataset, and a new PLS-DA model was computed [27]. PLS-DA model computing and variable selection were performed until the highest classification performance (lowest BER) was achieved.

The prediction performance of the final model was validated using a confusion matrix and a permuted-confusion matrix. The former was constructed by 4-fold cross-validation repeated 10,000 times using the Mahalanobis distance to predict the rabbit populations. The models' accuracy and precision were calculated considering the resilient population as the true positive value. We also computed a permuted-confusion matrix randomizing the categorical **y** vector of the rabbit populations to check whether the prediction performance of the final models was spurious, that is when the percentage of true positives in the permuted-confusion matrix was far from 50% (random probability of two categories).

Bayesian statistics were used to determine the relevance of the difference in the metabolite abundance between the two rabbit populations using a model with a single effect of line and flat priors for all unknowns. Marginal posterior distributions of the unknowns were estimated by MCMC (Gibbs sampling) using four chains with a length of 50,000 iterations, a lag of 10, and a burn-in of 1,000 iterations. The posterior mean of the differences in metabolite abundance was estimated as the mean of the marginal posterior distribution of differences between the resilient minus the non-resilient population. These differences were indicated as units of standard deviations (SD) of each metabolite (unit of SD). Moreover, the probability of the difference [28] being higher (if the difference is positive) or lower (if negative) than 0 (P_0) was calculated. We calculated

an approximation of the FDR from Storey (2003) [29] by using the cumulative posterior error probability (PEP), similar to the q-value, to establish the threshold for the identification of relevant metabolites. The PEP was calculated as $(1 - P_o)/0.5$. We assumed a cumulative PEP of 0.05, meaning that approximately 5% of the relevant metabolites identified were false positives. Later, an analysis for assigning the biological origin of each relevant metabolite was performed using the tool metOrigin [30].

4. RESULTS

The study included 26 rabbits from resilient (13) and non-resilient (13) populations, from whose gut contents a total of 765 untargeted metabolites were identified. The Bayesian statistical analysis identified 66 metabolites with relevant differences in abundance between both populations (with units of SD higher than 0.67) [see Additional File 2]. The PLS-DA model for these 66 metabolites showed a prediction performance of 71% for the non-resilient and 90% for the resilient populations. We found that from the 66 relevant metabolites 29% were associated with a co-metabolism because both host and microbiota can produce the metabolite, the12% were associated with the microbiota (*de novo* synthesis), and 27% with other sources (24% food related and 3% drug related). From the 66 relevant metabolites, 32% (21) of them could not be traced back to their origins (Fig. 1B). The most representative pathways were the long-chain fatty acids (LCFAs) acylcarnitines metabolism, the histidine metabolism, endocannabinoid metabolites, glycine, serine and threonine metabolism, and tryptophan metabolism (Fig. 1A) [see Additional file 2].

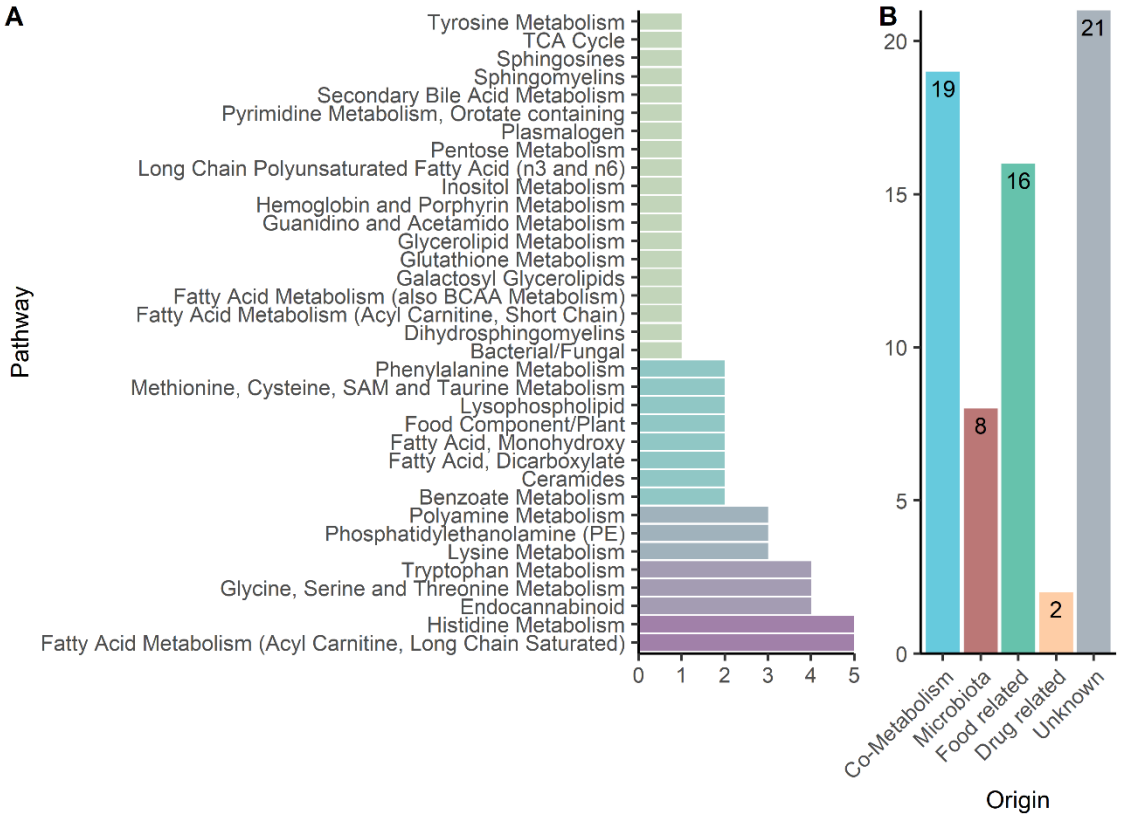


Figure 1. Pathway and biological origin of the metabolites with a relevant difference in abundance between the rabbit populations. A) Pathways of the 66 metabolites identified with relevant dissimilarities in their abundance between the resilient and non-resilient rabbit populations. B) Biological origin of the 66 metabolites with a relevant difference in abundance between the rabbit populations. "Co-metabolism" refers to the metabolites shared between the host and the microbiota. "Microbiota" are the microbiota-derived metabolites. "Food related" are the metabolites obtained from the diet. "Drug" refers to the drug-related metabolites. "Unknown" refers to the metabolites with unknown biological origin.

We also performed an optimized PLS-DA to obtain the most relevant metabolites for reaching the highest prediction performance. We identified 15 metabolites with a prediction performance of 99.18% for the resilient population and 90.42% for the non-resilient (Fig. 2.A). These metabolites were the behenoylcarnitine (C22), the arachidoylcarnitine, ethyl beta-glucopyranoside, 3-(4-hydroxyphenyl)lactate, 5-aminovalerate, glycerophosphoglycerol, succinylcarnitine, equol, cysteine s-sulfate, betaine, serine, palmitoyl dihydrosphingomyelin, thiamine, and aconitate. All of them are proposed as potential resilience biomarkers due to the low error for the prediction of each rabbit population achieved by the model. 13 of these 15 metabolites matched those identified in the Bayesian analysis as the relevant ones for the difference in abundance between the rabbit populations [see Additional file 2]. The two non-overlapping metabolites, the aconitate and thiamin, showed the lowest contribution to the optimized PLS-DA model (Fig. 2B) and a difference in abundance of 0.5 units of SD. The 13 most reliable metabolites seemed to be derived from the diet (behenoylcarnitine, arachidoylcarnitine, succinylcarnitie, betaine and palmitoyl dihihydrosphingomyelin),

the microbiota (3-(4-hydroxyphenyl)lactate, 5-aminovalerate and equol), and a co-metabolism between the host and the microbiota (N6-acetyllysine, serine). For the metabolites ethyl beta-glucopyranoside, glycerophosphoglycerol and cysteine s-sulfate no origin was determined.

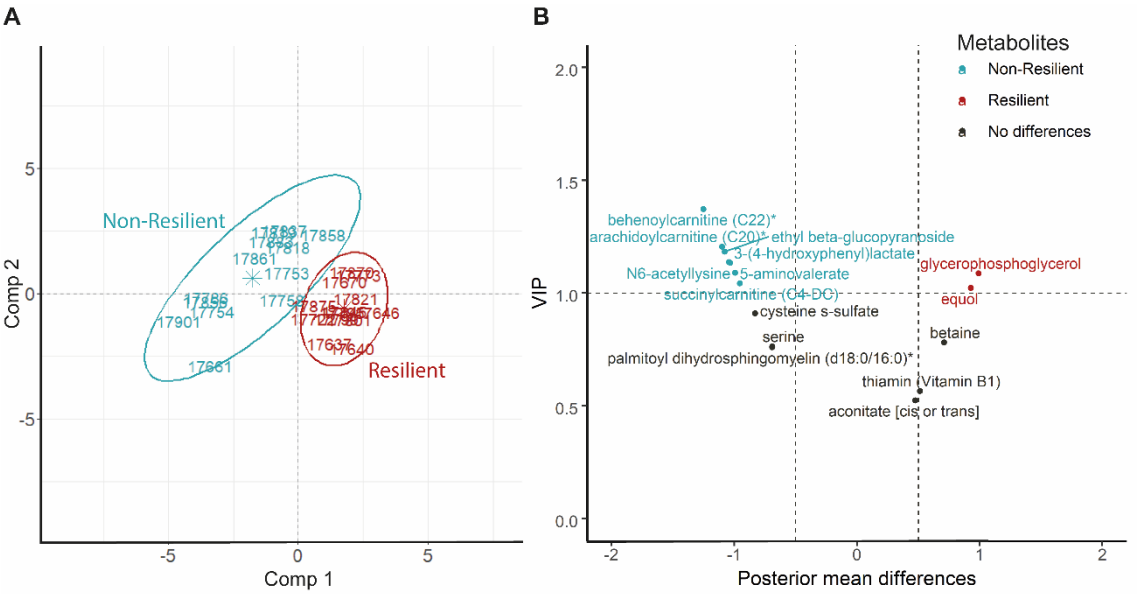


Figure 2. Final PLS-DA model. A) Representation of the first (Comp 1) and second (Comp 2) components of the PLS-DA used to discriminate rabbits belonging to resilient (red) and non-resilient (blue) populations. B) Representation of the posterior mean differences, given as units of standard deviation (SD), and the variable importance on prediction (VIP) of the metabolites included in the final PLS-DA model. Blue dots are relevant metabolites identified as more abundant in the non-resilient population. Red dots are the relevant metabolites identified with greater abundance in the resilient line. Black dots are metabolites included in the final PLS-DA model but did not overcome both thresholds: a VIP > 1 and a posterior mean of the differences higher than 0.5 SD.

5. DISCUSSION

The study of the gut metabolome could help to unravel its effects on key traits in livestock. In this study, we found differences in the metabolite profile [Additional file 2] of the rabbit populations with differences in resilience potential [12]. The Bayesian analysis identified 66 metabolites with dissimilarities in their abundances between the rabbit populations and good PLS-DA prediction performance. Nevertheless, 13 out of the 66 metabolites achieved the highest prediction performance to classify the resilient from non-resilient animals. Hence, these metabolites were proposed as potential biomarkers for resilience. Our study also showed that 27 of the 66 metabolites (5 of 13 candidates of resilience biomarkers [see Additional file 2]) were related to the microbiota (Fig.1B). These results suggested a different microbiome composition between the rabbit populations that agrees with a previous metagenomic study for genes and taxa using the same rabbit populations [18]. Additionally, we identified that 16 of 66 metabolites (5 of 15 candidates of resilience biomarkers [see Additional file 2]) showed a biological origin

related to the feeding (Fig. 1B). These results suggested differences in the use of dietary compounds between the rabbit populations because of the host itself or their microbiota.

Relevant functions were identified for four of the five resilience biomarkers related to microbiota-derived metabolites (equol, 3-(4-hydroxyphenyl)lactate, 5-aminovalerate, N6-acetyllisine and serine). Equol (0.93 units of SD) is from the daidzein metabolism. It can develop neuroprotective effects [31] and trigger the immune response as it enhances macrophages and protects from oxidative stress [32]. Our rabbits were sacrificed after their first parity. The high levels of equol in the resilient animals could have helped to reduce the inflammatory response triggered by the parity, a highly stressful event for the dam. Moreover, this stressful event may be increased the levels of the biomarker 3-(4-hydroxyphenyl)lactate (-1.04 units of SD) in the non-resilient population, a metabolite from tyrosine degradation that has been associated with non-alcoholic hepatic liver diseases [33]. This metabolite could be involved in a gut-liver crosstalk due to the differences found in plasma levels of cholesterol and triglycerides between these rabbit populations (after a challenge) [12,14]. 5-aminovalerate and N6-acetyllysine are products of lysine degradation (KEGG ID: C00431). The 5-aminovalerate is the precursor of 5-aminovaleric acid betaine (5-AVAB). This reaction used betaine (another resilience biomarker identified in this study; Fig. 2B) such as a methyl substrate donor to form the 5-AVAB [34]. It should be misidentified in some studies, so its role in health and disease is still unclear (see Haikonen et al. [34] for more information). Likewise, we did not find any evidence for the implication of the N6-acetyllysine in pathways related to animal resilience. On the other hand, the catabolism of the identified resilience biomarker serine (-0.69 unit of SD) was suggested to be related to the adaptation of pathogens during the inflammation process [35]. To support the relevance of this pathway, the glycine, an interconverted molecule to serine (KEGG ID: C00037), was identified with a difference of -0.68 units of SD between the rabbit populations. However, we cannot know how the serine levels are acting in the non-resilient population. In any case, it would be relevant to study its implication in animal resilience given its role in modulating bacterial pathogenesis.

Other metabolites derived from the AAA metabolisms (such as the abovementioned 3-(4-hydroxyphenyl) lactate) were identified in our study (Table 1). This is consistent with the dissimilarities in AAA biosynthesis genes (chorismite mutase and lyase) found in a previous metagenomic study using the same rabbit populations [18]. AAA metabolism can control health and disease [36], acting directly on the gut or on distal organs (i.e. liver, kidney or brain) [37]. Moreover, they can modulate inflammatory bowel diseases [37-39] as well as liver diseases such as hepatic inflammation and steatosis [37, 40]. Our

results showed high levels of kynurenine and anthranilate (Table 1) in the no-resilience rabbit population, the rabbit population that showed higher levels of CRP (an inflammation biomarker) [12]. High levels of these two metabolites were also found in individuals under stress with inflammation [41]. As the animals were sacrificed after their first parity, the overrepresentation of kynurenine and anthranilate in the non-resilient population could pinpoint a higher susceptibility to stress and inflammation in this population. Unexpectedly, the indole was found in low concentration in the resilient line (Table 1). This metabolite has protective functions in the gut as maintaining the intestinal barrier integrity and immune homeostasis to avoid dysbiosis during an inflammation response [42,43]. An in-depth study would be needed to really understand the interplay of this metabolites derived from the tryptophan on animal resilience.

Table 1. Metabolites from aromatic amino acids (AAAs) metabolism with relevant differences between the non-resilient and resilient populations.

Pathway	Metabolite	μ_{H-L}[(a)]	P_o[(b)]	HPD95[(c)]
Tyrosine metabolism				
	3-(4-hydroxyphenyl)lactate	-1.04	99.67	[-1.55, -0.22]
Phenylalanine metabolism				
	N-acetylphenylalanine	-0.82	98.17	[-1.57, -0.04]
	Phenyllactate	0.71	96.38	[-0.04, 1.52]
Tryptophan metabolism				
	Kynurenine	-0.79	97.80	[-1.55, -0.22]
	Anthranilate	-0.75	97.27	[-1.51, 0.04]
	Oxindolylalanine	-0.74	96.91	[-1.51, 0.03]
	Indole	-0.69	95.77	[-1.49, 0.10]

[(a)]Posterior mean of the differences between the non-resilient and the resilient populations.
[(b)] Probability of the difference being higher (if positive) or lower (if negative) than 0.
[(c)] High posterior density interval of 95%.

The metabolites behenoylcarnitine, arachidoylcarnitine, steroylcarnitine, palmitoylcarnitine, and formiminoglutamate are also highlighted. The behenoylcarnitine and the arachidoylcarnitine were identified as potential resilience biomarkers while the other ones only showed relevant differences in their abundance between the rabbit populations [Additional file 2]. The former four metabolites are long-chain fatty acylcarnitines (LCFA). Acylcarnitines are biomarkers of gut dysbiosis [44] and high levels of LCFA in the gut have been identified as a biomarker for inflammatory bowel diseases (IBD) due to mitochondrial dysfunction in the colonocytes [45]. A correct integrity and functionality of the intestinal epithelial barrier and colonocytes are essential to gut immunity homeostasis and pathogenesis [46-48]. These results could suggest differences in the assimilation by the gut of long-chain fatty acid for energy purposes that could influence the gut integrity and immunity. The latter, the formiminoglutamate, belongs to the histidine catabolism to L-glutamate. A low abundance of this metabolite was found in the resilient animals of this study [see

Additional file 2]. This result is in line with a previous metagenomic study that found an overrepresentation of the glutamate formiminotransferase in the resilient population [18]. Our findings suggested differences in the L-Glutamate metabolisms between the rabbit populations. Glutamate is an important neurotransmitter that can act in the gut, spinal cord, and brain, participating in the gut-brain axis and influencing the inflammatory response [49]. Lastly, for the metabolites palmitoyl dihydrosphingomyelin, ethyl beta glucopyranoside, glycerophosphoglycerol and cysteine-s-sulfate (potential resilient biomarkers; Fig. 2B) we can not hypothesize about the biological mechanisms affecting the animal resilience because their effects on the host are still unclear. We only can suggest that these metabolites are important to predict and classify between the rabbit populations used in this study.

6. CONCLUSIONS

This is the first study to identify gut metabolites which that could act as potential resilience biomarkers. The results supported the differences found in resilience potential between the rabbit populations, that were generated by the selection for V_E of LS. These differences could be due to the levels of acylcarnitines and metabolites derived from amino acid metabolism such as aromatic amino acids (tryptophan, phenylalanine and tyrosine), glycine, serine and glutamate. Moreover, relevant metabolites such as the 3-(4-hydroxyphenyl)lactate could be involved in a host-gut crosstalk. The selection for environmental variance modified the gut metabolome. Thus, the gut metabolome could be another factor to explain the differences in resilience between the rabbit populations. However, further studies will be needed to properly determine the origin and mode of action of each metabolite to unravel their causal role in health and disease, as well as in the host-gut crosstalk

7. DECLARATIONS

Ethics approval and consent to participate

All the experimental procedures were approved by the Committee of Ethics and Animal Welfare of the Miguel Hernández University, according to Council Directives 98/58/EC and 2010/63/EU (Ref. 2017/VSC/PEA/00212).

Consent for publication

Not applicable.

Availability of data and materials

Data are available upon request to the corresponding author.

Funding

This study was supported by Projects AGL2014-5592, C2-1-P and C2-2-P, AGL2017-86083, C2-1-P and C2-2-P, and PID2020-115558GB-C21, funded by the Spanish *Ministerio de Ciencia e Innovación (MCI)-Agencia Estatal de Investigación* (AEI) and the European Regional Development Funds (FEDER)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CCR analysed the data and wrote the manuscript. AB and MJA conceived and designed the study and contributed to the discussion of the results. MLG contributed to the study design and the discussion of the results. NIE conceived the study, contributed to the discussion of the results and edited the manuscript. All the authors have read and approved the final manuscript.

Acknowledgements

Cristina Casto-Rebollo acknowledges a FPU17/01196 scholarship from the Spanish Ministry of Science, Innovation and Universities.

8. REFERENCE

1. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. Gut. 2022;71(5):1020-1032. doi:10.1136/gutjnl-2021-326789

2. Manor O, Dai CL, Kornilov SA, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun. 2020;11(1):5206. doi:10.1038/s41467-020-18871-1.

3. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nat Rev Immunol. 2016;16(6):341-352. doi:10.1038/nri.2016.42

4. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30: 492–506. doi:10.1038/s41422-020-0332-7.

5. Krautkramer KA, Fan J, Bäckhed F. Gut microbial metabolites as multi-kingdom intermediates. Nat Rev Microbiol. 2021;19(2):77-94. doi:10.1038/s41579-020-0438-4

6. Postler TS and Ghosh S. Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. Cell Metabolism 2017;26: 110–30. doi:10.1016/j.cmet.2017.05.008

7. Liu Y, Liu C, Wu H, et al. Small Intestine Microbiome and Metabolome of High and Low Residual Feed Intake Angus Heifers. Front Microbiol. 2022;13:862151. doi:10.3389/fmicb.2022.862151

8. Wu J, Ye Y, Quan J, et al. Using nontargeted LC-MS metabolomics to identify the Association of Biomarkers in pig feces with feed efficiency. Porcine Health Manag. 2021;7(1):39. doi:10.1186/s40813-021-00219-w 8.

9. Xue MY, Sun HZ, Wu XH, et al. Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. Microbiome. 2020;8(1):64. doi:10.1186/s40168-020-00819-8

10. Berghof TVL, Poppe M, Mulder HA. Opportunities to improve resilience in animal breeding programs. Front Genet. 2019;9:692. doi:10.3389/fgene.2018.00692

11. Poppe M, Mulder HA, Veerkamp RF. Validation of resilience indicators by estimating genetic correlations among daughter groups and with yield responses to a heat wave and disturbances at herd level. J Dairy Sci. 2021;104(7):8094-8106. doi:10.3168/jds.2020-19817.

12. Argente MJ, García ML, Zbyňovská K, et al. Correlated response to selection for litter size environmental

variability in rabbits' resilience. Animal. 2019; 13:2348-55. doi:10.1017/S1751731119000302

13. Iung LHDS, Carvalheiro R, Neves HHDR, Mulder HA. Genetics and genomics of uniformity and resilience in livestock and aquaculture species: A review. J Anim Breed Genet. 2020;137:263–80. doi:10.1111/jbg.12454

14. Beloumi D, Blasco A, Muelas R, et al. Inflammatory Correlated Response in Two Populations of Rabbit Selected Divergently for Litter Size Environmental Variability. Animals. 2020;10:1540. doi:10.3390/ani10091540

15. Casto-Rebollo C, Argente MJ, García ML, et al. Identification of functional mutations associated with environmental variance of litter size in rabbits. Genet Sel Evol. 2020;52(1):22. doi:10.1186/s12711-020-00542-w.

16. Casto-Rebollo C, Argente MJ, García ML, et al. Selection for environmental variance of litter size in rabbits involves genes in pathways controlling animal resilience. Genet Sel Evol. 2021;13;53(1):59. doi:10.1186/s12711-021-00653-y.

17. Tilg H, Zmora N, Adolph TE, Elinav E. The intestinal microbiota fuelling metabolic inflammation. Nat Rev Immunol. 2020;20(1):40-54. doi:10.1038/s41577-019-0198-4

18. Casto-Rebollo C, Argente MJ, García ML, et al. Selection for Environmental Variance Shifted the Gut Microbiome Composition Driving Animal Resilience, 02 May 2022, PREPRINT (Version 1) available at Research Square doi:10.21203/rs.3.rs-1597523/v1

19. Rinninella E, Raoul P, Cintoni M, et al. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms. 2019;7(1):14. doi:10.3390/microorganisms7010014.

20. Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555(7695):210-215. doi:10.1038/nature25973.

21. Blasco A, Martínez-Álvaro M, García ML, et al. Selection for environmental variance of litter size in rabbit. Genet Sel Evol. 2017;49:48. doi:10.1186/s12711-017-0323-4

22. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

23. Bijlsma S, Bobeldijk I, Verheij ER, et al. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. Anal Chem. 2006;78(2):567-74. doi: 10.1021/ac051495j.

24. Milac TI, Randolph TW, Wang P. Analyzing LC-MS/MS data by spectral count and ion abundance: two case studies. Stat Interface. 2012;5(1):75-87. doi:10.4310/SII.2012.v5.n1.a7

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

26. Rohart F, Gautier B, Singh A, Le Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Computational Biology 2017;13(11):e1005752. doi: 10.1371/journal.pcbi.1005752

27. Galindo-Prieto B, Eriksson L, Trygg, J. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). Journal of Chemometrics. 2014;28:623– 632. doi:10.1002/cem.2627

28. Blasco, A. Bayesian Data Analysis for Animal Scientists: The Basics. 2017. Springer, Cham. doi:10.1007/978-3-319-54274-4

29. Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. Annals of Statistics, 2003;31:2013–2035. doi:10.1214/aos/1074290335

30. Yu G, Xu C, Zhang D, et al. MetOrigin: Discriminating the origins of microbial metabolites for integrative analysis of the gut microbiome and metabolome. iMeta. 2022;**1**:e10. doi:10.1002/imt2.10

31. Subedi L, Ji E, Shin D, et al. Equol, a Dietary Daidzein Gut Metabolite Attenuates Microglial Activation and Potentiates Neuroprotection In Vitro. Nutrients. 2017;9(3):207. doi:10.3390/nu9030207.

32. Gou Z, Jiang S, Zheng C, et al. Equol Inhibits LPS-Induced Oxidative Stress and Enhances the Immune Response in Chicken HD11 Macrophages. Cell Physiol Biochem. 2015;36(2):611-621. doi:10.1159/000430124

33. Caussy C, Hsu C, Lo MT, et al. Link between gut-microbiome derived metabolite and shared gene-effects with hepatic steatosis and fibrosis in NAFLD. Hepatology. 2018;68(3):918-932. doi: 10.1002/hep.29892

34. Haikonen R, Kärkkäinen O, Koistinen V, Hanhineva K. Diet- and microbiota-related metabolite, 5-aminovaleric acid betaine (5-AVAB), in health and disease. Trends Endocrinol Metab. 2022;33(7):463-480. doi:10.1016/j.tem.2022.04.004

35. Kitamoto S, Alteri CJ, Rodrigues M, et al. Dietary L-serine confers a competitive fitness advantage to Enterobacteriaceae in the inflamed gut. Nat Microbiol. 2020;5(1):116-125. doi:10.1038/s41564-019-0591-6

36. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. Nat Commun. 2018;9(1):3294. Published 2018 Aug 17. doi:10.1038/s41467-018-05470-4

37. Liu Y, Hou Y, Wang G, et al. Gut Microbial Metabolites of Aromatic Amino Acids as Signals in Host-Microbe Interplay. Trends Endocrinol Metab. 2020;31(11):818-834. doi:10.1016/j.tem.2020.02.012

38. Yachida S, Mizutani S, Shiroma H, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nat Med. 2019;25(6):968-976. doi:10.1038/s41591-019-0458-7

39. Nikolaus S, Schulte B, Al-Massad N, et al. Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases. Gastroenterology. 2017;153(6):1504-1516.e2. doi:10.1053/j.gastro.2017.08.028

40. Hoyles L, Fernández-Real JM, Federici M, *et al.* Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. Nat Med. 2018;**24:**1070–1080. doi:10.1038/s41591-018-0061-3.

41. Doney E, Cadoret A, Dion-Albert L, et al. Inflammation-driven brain and gut barrier dysfunction in stress and mood disorders. Eur J Neurosci. 2022;55(9-10):2851-2894. doi:10.1111/ejn.15239

42. Bansal T, Alaniz RC, Wood TK, Jayaraman A. The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. Proc Natl Acad Sci U S A. 2010;107(1):228-233. doi:10.1073/pnas.0906112107

43. Li X, Zhang B, Hu Y, Zhao Y. New Insights Into Gut-Bacteria-Derived Indole and Its Derivatives in Intestinal and Liver Diseases. Front Pharmacol. 2021;12:769501. doi: 10.3389/fphar.2021.769501.

44. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655-662. doi:10.1038/s41586-019-1237-9

45. Smith SA, Ogawa SA, Chau L, et al. Mitochondrial dysfunction in inflammatory bowel disease alters intestinal epithelial metabolism of hepatic acylcarnitines. J Clin Invest. 2021;131(1):e133371. doi:10.1172/JCI133371

46. Peterson L, Artis D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. Nat Rev Immunol 2014;**14:**141–153. doi:10.1038/nri3608

47. Zhou A, Yuan Y, Yang M, et al. Crosstalk Between the Gut Microbiota and Epithelial Cells Under Physiological and Infectious Conditions. Front Cell Infect Microbiol. 2022;12:832672. doi: 10.3389/fcimb.2022.832672

48. Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science. 2018;362(6418):eaat9076. doi:10.1126/science.aat9076

49. Baj A, Moro E, Bistoletti M, et al. Glutamatergic Signaling Along The Microbiota-Gut-Brain Axis. Int J Mol Sci. 2019;20(6):1482. doi:10.3390/ijms20061482

9. ADDITIONAL FILES

Additional files are available with the preprint at https://doi.org/10.21203/rs.3.rs-1808216/v1

Additional File 1

File format: .html

Title: Additional File 1

Description: Full pipeline with all the metabolomic analyses

Additional File 2

File format: .xlsx

Title: Additional File 2

Description: Results of the Bayesian statistical analysis. File includes (from left to right) the metabolite ID, posterior mean of the differences among the resilient and non-resilient populations (meanDiff), the probability of the difference being higher (if the difference is positive) or lower (if negative) than 0 (P0), the highest posterior interval density of 95% (HPD95), the chemical name of the metabolites and the general and specific pathways in which they are involved, the posterior error probability (PEP), the cumulative PEP, the metabolites identified by the Bayesian analysis (Bayes), the metabolites identified by the PLS-DA (PLS), and the biological origin determined for each metabolite.

CHAPTER 7: simuGMsel: A tool for simulating the genome and microbiome coevolution in animal breeding programs

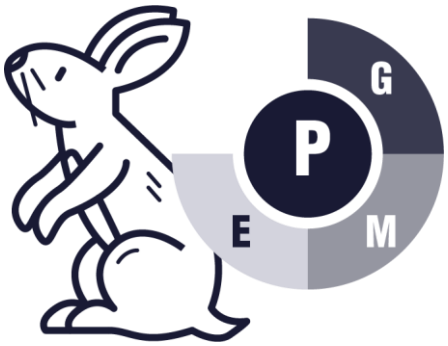
Authors

C. Casto-Rebollo[1], I. Pocrnic[2], G. Gorjanc[2], N. Ibáñez-Escriche[1]

Institutional affiliations

[1]Institute for Animal Science and Technology, Universitat Politècnica de València, C/Camino de Vera s/n, València, 46022, Spain

[2]The Roslin Institute, The University of Edinburgh, Easter Bush Campus, Edinburgh, EH25 9RG, UK



The content of this chapter will be submitted to Genetic Selection Evolution.

1. ABSTRACT

1.1. BACKGROUND

Animal and plant breeders study the microbiome as another source of phenotypic variability, affecting both the mean and the variance of traits. Differences in gut microbiome composition have been widely identified for key traits in livestock. Then, including the microbiome information in the breeding schemes could be useful in improving livestock populations. However, the complexity of the microbiome makes it difficult to model across species and traits. Moreover, the expensive sequence-based techniques limit the availability of empirical data to study microbiome composition at scale to decipher key animal breeding questions. Thus, in this study, we aimed to develop simuGMsel, a simulation tool to study the role of the microbiome in animal breeding.

1.2. RESULTS

simuGMsel simulates the inheritance of genome and microbiome between generations and their effect on livestock phenotypes in addition to the environmental effects. The microbiome is a function of the host genome effect and interaction among microbial species and the environment. Different scenarios can be simulated by varying several parameters to define the genome and the microbiome of founder populations. We tested simuGMsel by simulating an in-silico replica of a real rabbit divergent selection experiment across 13 generations of phenotypic selection. An expected divergent trend was observed for all the scenarios simulated when considered the microbiome effect. Microbial heritability affected the selection response when the microbial species had a lower fitness due to allow reducing their variability. Furthermore, scenarios that included the microbial species interactions showed the greatest phenotypic response.

1.3. CONCLUSION

simuGMsel enables simulation of genome and microbiome to study their coevolution and effect on phenotype in animal breeding programs. As such, it provides a research platform to test different hypotheses and scenarios related to the role of the microbiome in animal breeding. The presented simulation study showed that the selection based on phenotype generated changes in the genome and microbiome and highlighted the importance of microbial heritability, the symbiosis among species, and the fitness of microbial species as key variables impacting response to selection. simuGMsel will enable future studies on the role of the microbiome in animal breeding, including genome-assisted and microbiome-assisted selection.

2. BACKGROUND

The coevolution of both the genome and the microbiome drives the adaptation of plants and animals [1-3]. As a result, animal and plant breeders began to study the microbiome as another source of phenotypic variability which can affect both the mean and the variance of traits (see the review of Henry et al., 2021[4]). Differences in gut microbiome composition have been widely identified for key traits in livestock such as feed intake [5-6], growth performance, and carcass quality in pigs [7], feed intake in chickens [8], milk protein yield [9] and methane emissions in cattle [10], and intramuscular fat and environmental variance in rabbit [11-12]. Then, including the microbiome information in the breeding schemes must be useful to improve the progress of livestock populations [13].

The complexity of the microbiome makes difficult its modelling across species and traits. A stable microbiome is assembled by the acquisition and colonization of microorganisms. The former depends on the individual's exposure to different environmental sources such as their mother (pregnancy, birth, and suckling process) [14-16], the diet [17] and the housing environment [18]. Colonization is most dependent on the survival and growth potential of each microbial specie, as well as on the individual immune system and the tissue (gut, skin, mouth, vagina, etc.). Host genome could affect around 5-10% of the microbiome compositions variability [19-20], being the most variability due to environmental factors [21]. Although most microbial species seem to be influenced to a low or moderate degree by the host genome [22]. Furthermore, the lack of consensus on theoretical models and analytic methodologies constrains comparisons between experiments and learning how various microbiome-related factors impact the animal's phenotype.

The expensive sequence-based techniques limit the availability of empirical data to study microbiome composition at scale to decipher key questions for animal breeders such as how the selection response is influenced by: (i) the microbiome composition, (ii) the microbial heritability, (iii) the microbial interactions and (iv) the microbiome inheritance (dam or environment). Thus, in this study, we proposed a simulation approach to uncover the effect of the microbiome on the selection response of critical traits for livestock systems. This simulation is based on a phenotypic selection considering both the genome and the microbiome inheritance. Different scenarios were implemented to consider the contribution of the genome and/or the microbiome on the phenotype, as well as to model the microbial heritability. Furthermore, microbial interactions could be considered to know if the symbiosis among microorganisms is a

key factor to determine the overall effect of the microbiome. This is a tool under development, so further implementations and optimization will be included.

3. METHODS

The developed simulator allows users to simulate a host genome (G), the microbiome composition (M), and the animal's phenotype (P) as a function of G, M, and the environment (E). The simulator uses a combination of custom R-code available at github.com/HighlanderLab/ccastorebollo_metagenomics_sim.

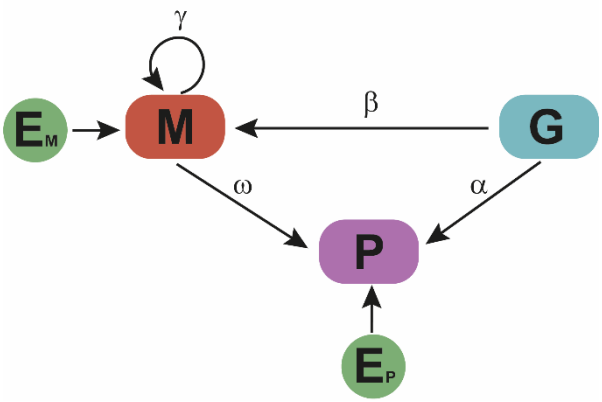


Figure 1. Simulation scheme. G: genome, M: microbiome, P: phenotype, E_M: environment over M, E_P: Environment over P. Arrows indicate causality. G has an effect α on P, determined by the QTLs, and β on M, determined by a proportion of the QTLs with an effect on P. M is also influenced by its species composition due to symbiosis interaction among them, being this effect determines by γ. P is a combination of G, M, and E_P.

3.1. GENOME SIMULATION

The simulation of the individual genome is performed with the R packages AlphaSimR [23]. Users can specify a huge number of parameters to simulate the haplotypes in the founder population such as the demographic history of the population under study, the number of segregating sites, the effective population size, and the number of chromosomes to simulate, among others. The quantitative trait is simulated under the additive model, indicating its mean (μ), phenotypic variance (σ_P^2) and heritability (h^2), and the number of QTLs per chromosome. The allele substitution effects of QTLs ($\boldsymbol{\alpha}$) are sampled from a gamma distribution $\boldsymbol{\alpha}$~Γ(k=0.2, θ=1) with the parameters suggested by Perez-Enciso et al. (2021) [24].

3.2. MICROBIOME SIMULATION

Four steps are implemented in the simulator to compute the microbiome composition of each animal: (i) microbial species distribution, (ii) microbial heritability, (iii) symbiosis or microbial interaction, and (iv) microbiome composition.

i. Microbial species distribution

Microbial species abundance is considered another complex trait. So, to simulate it, the microbial species distributions were first established. For that, a vector with the average abundances of a set of microbial species (μ) is computed with a negative binomial distribution, x_k~NB(r=2, q=0.0001) for 10^8 total cells. This initial vector is the distribution of the microbial species abundance in the environment (Environmental microbiome; EM) for the selection process, being EM constant throughout generations. To consider that the microbial species can grow differently in the individual with respect to the environment, the initial vector is sorted to establish the mean distribution of the microbial species abundance in the parental microbiome or PM ($\boldsymbol{\mu_{PM}}$). Microbial variability in the individual is modulated by defining a microbial fitness. This fitness account for the stability of the growth performance of each microbial species in the individual. For that, a coefficient of variation (CV) is assigned to each microbial species based on real data. This CV applies to the $\boldsymbol{\mu_{PM}}$ allowing to obtain the standard deviation of the abundance distribution for each microbial specie. This abundance distribution is the variability that we can observe for a determined microbial specie in the population. We considered that this variability is generated due to the contribution of the host genome, the symbiosis relationship among species, and the environment over M (E_M). Thus, our definition of fitness refers to the inverse of the coefficient of variation. For low values of CV, the fitness of the microbial species is higher.

ii. Microbial heritability

The simulation tool allows to define the effect of the host genome on the microbial species abundance, i.e., a microbial heritability effect or h_m^2. This effect follows a gamma distribution $\boldsymbol{\beta}$~Γ(k=0.2, θ=1) and establishes the percentage of the microbial species variability due to the host genome (σ_{hg}^2). Users can define the percentage of the individual QTLs that is affecting a determine number of microbial species. Default values are set to 10% for both QTLs and microbial species affected, with the restriction that 50% of the microbial species with an effect on the phenotypes must be influenced by the host genome. Moreover, the user can define the h_m^2, that is the degree of the host genetic variance contribution to the abundance variability. This h_m^2 is equal for all the microbial species to simplify the result interpretation. We can simulate simultaneously four different scenarios depending on if the h_m^2 has a low, medium, high, or null value. Default values are set to 0.1 for the low microbial heritability (LMH) scenario, 0.3 for the medium microbial heritability (MMH) scenario, and 0.6 for the high microbial heritability (HMH) scenario. Scenarios without microbial heritability (NMH) and only with the

contribution of the microbiome (M) or de genome (G) can also be simulated. For the M scenario, a microbial heritability of 0.2 is set by default.

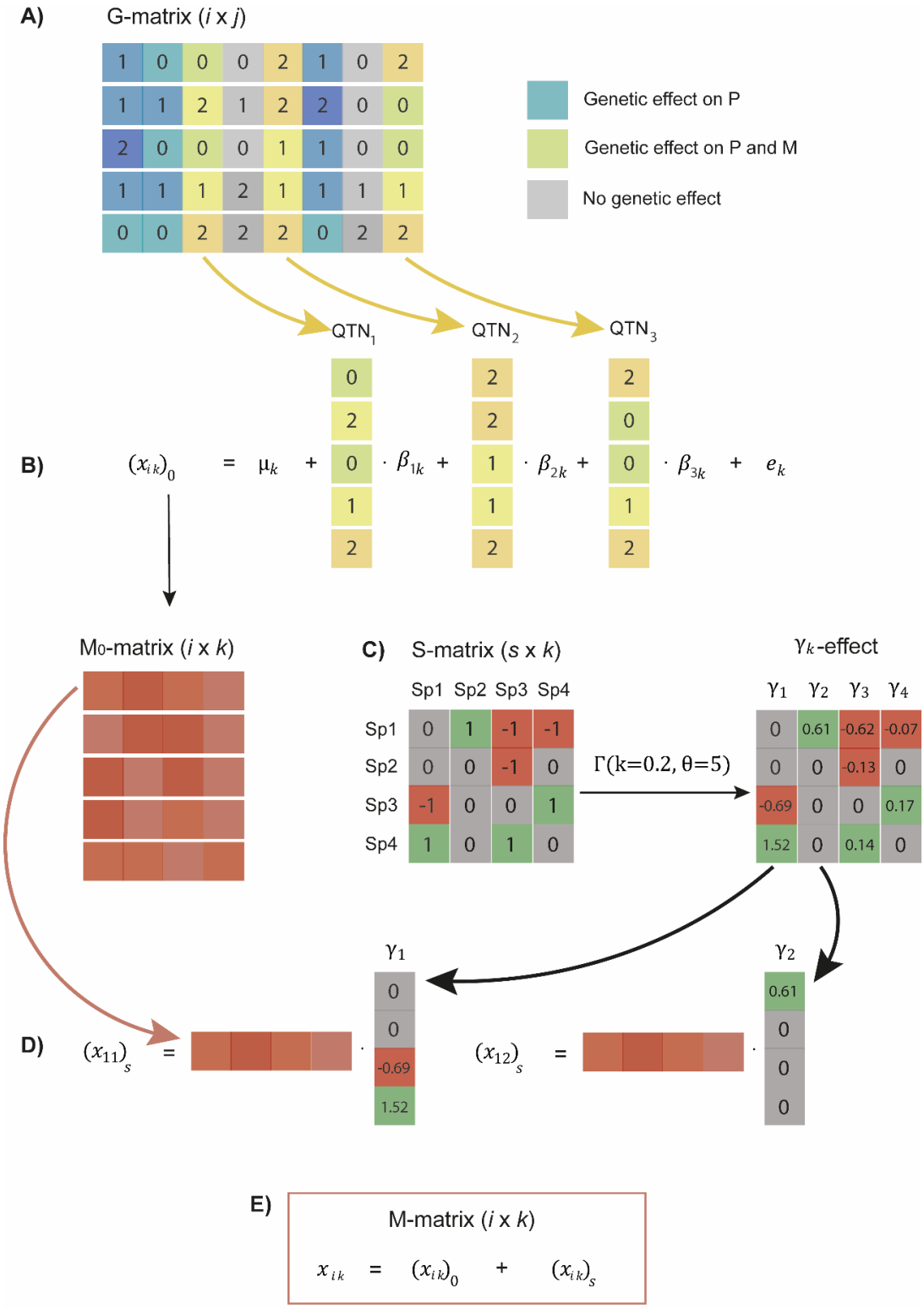


Figure 2. A hierarchical model for the calculation of the microbiome composition. A) Genotype of *i* individuals of *j* SNPs. Blue and yellow SNPs are quantitative trait loci (QTL) with an effect on the phenotype (blue) or on the phenotype and the microbiome (yellow). Grey SNPs do not have an effect on the phenotype. B) Stable microbial species abundances or $(x_{ik})_0$ for the microbial species *k* in individual *i*,

considering only the host-genetic and the environmental effect. M_0 is the stable microbiome composition in the entire population. C) Symbiosis matrix (S) harbouring the microbial species interactions. A gamma distribution Γ(k=0.2, θ=5) is applied to each column of the S-matrix to obtain the matrix with the interaction effects among species (γ_k). D) Calculation of the fluctuation in the microbial species abundance or $(x_{ik})_s$ due to the symbiosis effect over the stable microbiome of the individual *i* for the microbial species *k*. E) Total microbiome composition for the entire population (M). M is the combination of the stable microbiome plus the fluctuation due to the microbial species interaction.

iii. Symbiosis or microbial interactions

Symbiosis relationships are established in the simulation to consider that a part of the variability in M is due to the interaction among the microbial species. Users can model the interaction variance (σ_S^2) as a percentage of the environmental variance for the microbial species abundance (σ_{Em}^2). Default is established as the 20% of σ_{Em}^2. The types and subtypes of symbiosis relationships (Table 1) are based on Coyte and Rakoff-Nahoum (2019) [25]. These are (I) positive interactions, which include commensal and cooperative interactions; and (II) negative interactions, which include competitive, exploitation, and ammensalism interactions. Users can define the percentage of each one. By default, these are set to 20% for each subtype of microbial interaction. Most of the symbiosis relationships were suggested to be negative interactions [25]. In this line, we considered 60% of negative interactions, and 40% of positive interactions with this percentage assignment. The interaction is modelled as 0 (neutral interaction), 1 (positive interaction), and -1 (negative interaction). In this way, a symbiosis matrix with *s* x *k* interactions was constructed with the effect of the interaction between the species and *k* (Fig. 2C). The degree of the effect of all interactions for each microbial species *k* is randomly sampled from a gamma distribution, $\gamma_k \sim \Gamma(k=0.2, \theta=5)$.

Table 1. Types of symbiosis relationships among microbial species and their interaction effects when both species are in the individual.

Type	Subtype	Effect	
		Species A	Species B
I. Positive			
	I. Commensalism	0	+
	II. Cooperation	+	+
II. Negative			
	I. Competitive	-	-
	II. Exploitation	+	-
	III. Ammensalism	0	-
III. Neutral		0	0

Symbols +, – and 0 indicate the effect of the interaction on the abundance of the species, being + when increases, – when decreases, and 0 when do not change (0) its abundance in the presence of the other species.

iv. Microbiome composition

Microbial species abundance is treated as another quantitative trait, so their observed variability can be simplified as:

$$\boldsymbol{\sigma_m^2 = \sigma_{hg}^2 + \sigma_s^2 + \sigma_{Em}^2} \quad (1)$$

The abundance of each microbial species is computed independently and then, the symbiosis matrix is applied to take the microbial species interaction into account and generate a correlation between the microbial species. Therefore, we consider that each microbial species has a stable abundance determined by the host genome and the environmental effects. Therefore, the stable microbial species abundance is computed following the Pérez-Enciso et al. (2021) [24]:

$$(x_{ik})_0 = \mu_k + \sum_{j=1}^{J} z_{ij}\beta_{jk} + e_{ik} \quad (2)$$

where $(x_{ik})_0$ is the stable abundance of the specie k in the individual i; μ_k is the average abundance of the specie k in the base population; $\sum_{j=1}^{n} z_{ij}\beta_{jk}$ is the genetic value (gv_{ik}) of individual i for the abundance of specie k, where z_{ij} is the genotype of SNP j for individual i coded as 0, 1 or 2 according to the reference allele, and β_{jk} being the substitution allele effect of SNP j for the abundance of the species k; and e_{ik} is the residual of individual i for the species k distributed as N(0, $\sigma_{e_k}^2$), where $\sigma_{e_k}^2$ is residual variance for the species k. The gv_{ik} were distributed with mean 0 and variance $\sigma_{hg_k}^2$.

All stable microbial species abundance for the entire population is computed before applying the symbiosis matrix. The aim of the symbiosis matrix is to model a fluctuation in the microbiome composition due to the abundance observed in the stable microbiome for each microbial specie. This fluctuation in the stable microbial species abundance is computed as:

$$(x_{ik})_s = \sum_{k=1}^{K} (x_i)_0 \cdot \gamma_k \quad (3)$$

where $(x_{ik})_s$ is the abundance due to the microbial species interactions of specie k in the individual i, $(x_i)_0$ is the stable abundance in the individual i (Eq. 2), and γ_k is all the interaction effects for the species k. The $(x_{ik})_s$ is distributed with mean 0 and variance

$\sigma^2_{s\,k}$. Finally, the total microbiome composition is computed as the sum of the stable microbial species abundance plus the fluctuation over this one:

$$x_{ik} = (x_{ik})_0 + \ (x_{ik})_s \tag{4}$$

3.3. PHENOTYPE SIMULATION

The phenotype is simulated as:

$$y_i = \mu_P + \sum_{j=1}^{n} z_{ij}\alpha_j + \sum_{k=1}^{m} x_{ik}\omega_k + e_i \tag{5}$$

where y_i is the phenotype of individual i; μ_P is the phenotypic mean in the base population; $\sum_{j=1}^{n} z_{ij}\alpha_j$ is the genetic value of individual i (gv_i) , where z_{ij} is the genotype of individual i for the SNP j code as 0,1 or 2 according to the reference allele, and α_j is the allele substitution effect of the SNP j; $\sum_{k=1}^{m} x_{ik}\omega_k$ is the microbiome value of individual i (mv_i), where x_{ik} is the abundance of species k in individual i (see equation 4) and ω_k is the effect of species k on the host phenotype; and e_i is the residual of individual i distributed as N(0, σ^2_e), and σ^2_e is the residual variance. The contribution of both genome and microbiome on the phenotype is modelled using the established values for the heritability (h^2) and microbiability (m^2) to determine the genetic (σ^2_g) and microbiome variance (σ^2_m). By default, h^2 and m^2 are set to 0.15. The individual's gv are Normally distributed with mean 0 and variance σ^2_g and the gv with mean 0 and variance σ^2_m. Furthermore, users can establish the number of microbial species with ω_k, that is the effect of the microbial species on the phenotype. This effect follows a gamma distribution $\omega_k \sim \Gamma(k=0.2, \theta=5)$.

3.4. INHERITANCE

The simulator allows to implement a selection process base on the phenotype for one-way or for divergent selection. Mating and genome inheritance are simulated using the AlphaSimR package [23]. In the base population, the breeding females will be those individuals with the highest phenotypic performance. Breeding males will be full sibs of the best breeding females. Users can determine the number of breeding males and females. The default is set to 125 females and 25 males. Each breeding male is mated with the number of breeding females that the user defines (Default 5 females per male), avoiding a close relationship among them. In the following generations, the breeding females are selected as in the base population, and the breeding males will be males from the offspring of the best female for each sire (best mating). Users can define the number

of generations and the number of offspring per litter. By default, we set the number of generations to 13 and the litter size to 8.

Microbiome inheritance is modelled based on the relative abundance (RA) of each microbial species for the breeding females and the environmental microbiome. Breeding males do not contribute to the microbiome composition of the offspring under this assumption that is the common management in livestock system [13]. Offspring will inherit a percentage of the PM from their mother and another percentage from the EM. The microbial species that will be sampled from the PM and the EM will be determined by the RA of each specie. The RAs are used as a probability to random sample the microbial species from each microbiome source (PM or EM). The users can establish the percentage of species that we can sample from the PM and the EM. By default, we define that offspring would inherit 80% of the microbial species from the mother or PM and 20% from the EM. These percentages determine the number of microbial species present in each offspring that would grow according to equations 1 and 3.

3.5. PARAMETER ESTIMATIONS

The estimation of the h^2 and m^2 was performed to check the smooth running of the tool. Estimations were calculated using the genotype and the microbiome from the base population. The genotype included 200K SNPs and the microbiome all those microbial species present in at least one individual. Bayes B and Bayes C models were used to estimate the parameters using the R packages BGLR (CITA) and the Equation 5. We set the number of iterations to 10,500, the burn-in to 2500, and the lag to 5 (Default) for both methods. In the Bayes B model, we used the default parameters from BGLR package. In Bayes C model, we used as prior π for the probability of SNPs and microbial species being included in the model with effect and variance of the probability simulated. Thus, for SNPs π~Beta ($p_o = 5$, $\pi_o = 0.02$) and for microbial species π~Beta ($p_o = 5$, $\pi_o = 0.035$). We used the proposal from Pérez-Enciso et al. (2021) [26] to estimate the h^2 and the m^2. Estimates were calculated 50 times with each methodology.

4. RESULTS

Application of the simulation tool in a divergent selection process

The tool allowed to successfully simulate a divergent selection process considering different scenarios in which heritability (h^2), microbiability (m^2) and microbial interaction (h_m^2) can be modelled. Default values were used in this example. A clear divergent trend was observed for all the scenarios simulated with and without microbial interaction [Additional file 1]. The total number of microbial species shaping the

phenotype was 35. From them, a total of 22 microbial species were assigned to be a positive contribution and 13 a negative effect (Fig. 3). Of the 22 with a positive effect, 15 showed microbial heritability while only six of the 13 with a negative effect were influenced by the genotype. Lower fitness values were observed for 4 microbial species with a positive effect on the phenotype higher than 10 (Fig. 3). These four microbial species were influenced by the host genome. Any microbial species with a low fitness (a coefficient of variation higher than 0.25) and a high negative effect on the phenotype was found (Fig 3).

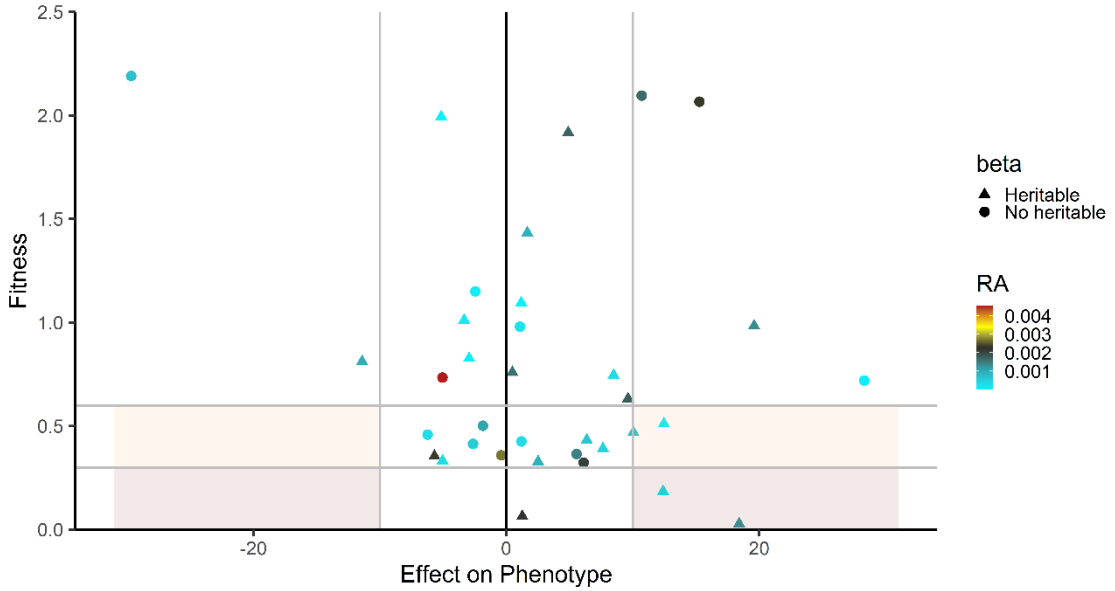


Figure 3. Microbial species effect on the phenotype (ω_k). The graph represents the effect of each microbial species on the individual phenotype with respect to the fitness of each one. Fitness is the base 10 logarithm of the inverse of the coefficient of variation assigned to each microbial specie. Triangle is microbial species affected by the host genome (with microbial heritability; beta: β) while dot is microbial species not affected by it. Colours are the mean for the relative abundance (RA) of each microbial species in the parental microbiome (PM), meaning a low abundance of the light blue colour and a high abundance of the red colour. The red rectangle is the section of values with an effect higher than ten (in absolute values) and a coefficient of variation higher than 0.5 (Fitness of 0.3) and the yellow rectangle is the section with a coefficient of variation between 0.5 and 0.25 (fitness of 0.6).

Microbiome abundance was considered as another quantitative trait following Pérez-Enciso et al. (2021) [24] but with some modifications (Eq. 4). This assumption allowed to simulate in a simpler way the microbial species abundance in each individual with or without a symbiosis effect (Table 1; Eq. 3). Furthermore, we used the relative abundance of each microbial species from the mother and a shared environment as the probability of being acquired by the offspring. The establishment of each microbial species in the offspring was determined by their microbial fitness which is constant throughout the simulation process. That means that the microbiome trends observed across the selection process [Additional file 1] are a response to the selection based on the phenotype by effective implementation of the microbiome inheritance. Furthermore, the estimation of the h^2 and m^2 were calculated at the base population 50 times using Bayes

B and Bayes C models (Fig. 4). Bayes C method was most precise in estimating the h^2 and m^2 because variable π was set to the exact number of QTLs and microbial species affecting the phenotype. In the G scenario, with Bayes B the mean of the estimations of h^2 was 0.19 and 0.03 for m^2 while with the Bayes C model was 0.17 for h^2 and 0.01 for m^2 (Fig. 4). In the M scenario, with Bayes B the h2 was 0.08 and the m2 was 0.18 while with Bayes C the h2 was 0.03 and the m2 was 0.14 (Fig. 4). The other scenarios all of them showed a mean of the estimations of 0.19 for h^2 and 0.18 for m^2 with Bayes B and 0.16 for h^2 and 0.14 for m^2 with Bayes C (Fig. 4). All estimates were quite close to the value set by default in the simulation that was 0.15 for both (Fig. a). That means that the developed R code was working properly with the parameters established.

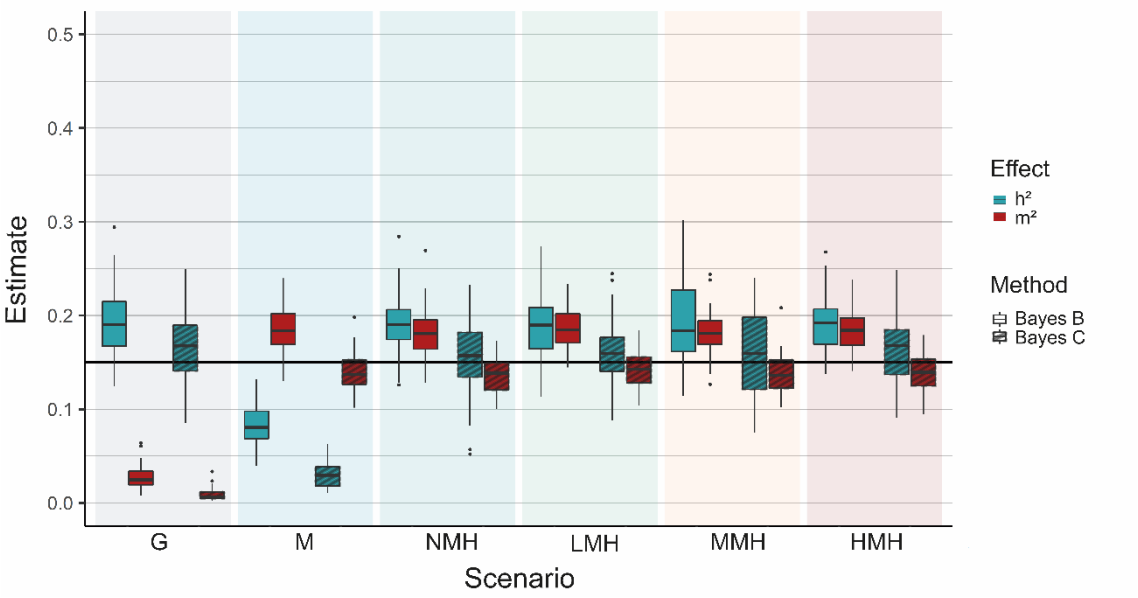


Figure 4. Boxplot of the estimates of h^2 and m^2 using Bayes B and Bayes C methods. The base population was simulated 50 times to obtain repeated estimations for each parameter. The microbiome was computed considering the symbiosis effect. Boxplots for h^2 values are in blue while boxplots for m^2 values are in red. The black horizontal line represented the simulated value for the h^2 and m^2. Bayes C and Bayes B models are represented with stripes or not in the boxplot, respectively. Background colours represented each scenario simulated, being G (grey): genome scenario; M (dark blue) the microbiome scenario; NMH (light blue): non microbial heritability; LMH (green); low microbial heritability; MMH (yellow): medium microbial heritability; and HMH (red): high microbial heritability.

Microbiome shaped in a different degree the phenotype depending on the selection performed

The trend of the phenotypic value [Additional 1] showed differences depending on the selection performed, i.e., to decrease or increase the trait performance (Fig.5). For the divergent line selected for decreasing the phenotypic value (Fig. 5A), the low line, the microbiome effect allowed to improve the phenotypic selection response, with a difference of 1.3 points in the mean between the G and M scenarios (without considering the symbiosis effect) (Fig. 5A). We simulated the phenotype of litter size in rabbit, so this 1.3 point meaning 1.3 kids more in the M scenario than in G scenario. For the divergent

line selected for increasing the phenotypic value (Fig. 5B), the high line, the G scenario overcame the M scenario (Fig. 5B) with a difference in mean of 1.1 kids. In the high line, when the microbiome is computed the phenotypic response was lower than when the genome was considered (Fig. 5B). Moreover, when microbial heritability was considered in the simulation, the phenotypic response increased with respect to the G and M scenario (G) for the high line (Fig. 5B). In general, the phenotypic response in the last generation observed in the low line was greater than in the high line since the differences in the phenotypic mean respect the base population was higher in this population (Fig. 5).

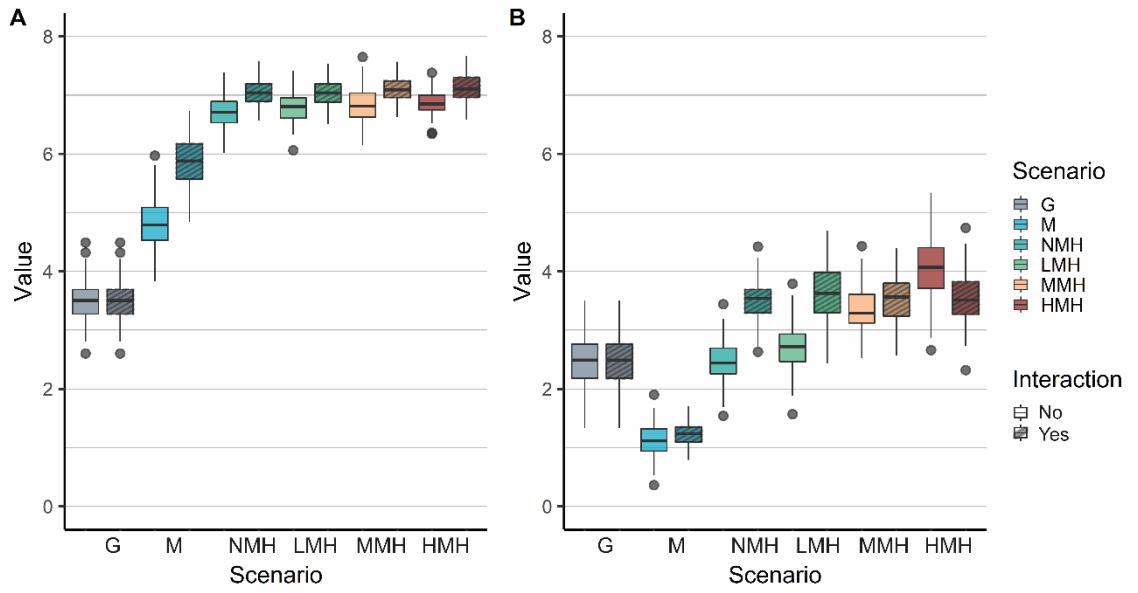


Figure 5. Boxplots of the phenotypic values on the last generation for all scenarios simulated. The values are the changes in the mean of the phenotypic values with respect to the base population after 13 generations of selection. The values were obtained after 100 simulations of each scenario. A) Absolute phenotypic values observed in the divergent line selected for decreasing the trait (Low line). Absolute values were computed in the low line to observe selection response in the same scale. B) Phenotypic values observed in the divergent line selected for increasing the trait (High line). Each colour represents a simulated scenario, being G: only genome effect on the phenotype (grey); M: only microbiome effect on the phenotype (Blue dark) and NMH (light blue), LMH (green), MMH (yellow), and HMH (red) are different scenarios with both the effect of the microbiome and the genome on the phenotype. Differences between these scenarios lay in the degree of the microbial heritability simulated, NMH: the scenario without microbial heritability; LMH; the scenario with a low microbial heritability; MMH; the scenario with medium microbial heritability; and HMH: the scenario with high microbial heritability. Interactions are indicated with stripes in the boxplot.

Differences in the phenotypic trends between the lines depend on the genetic and microbiome response observed in each one. The microbiome response was higher than the genetic response for the low line (Fig. 6A) while these responses were opposite in the high line (Fig 6B). That explains why the G scenario obtained higher phenotypic values than the M scenario at the last generation for the high line (Fig. 5B). The highest mean for the genetic value was 3.52 in the low line and 2.47 in the high line for the G scenario. The highest microbiome values were observed in the M scenario with interaction (6.25)

for the low line (Fig. 6A) and in the HMH scenario without interaction (1.85) for the high line (Fig. 6B). On the other hand, the low line showed an increase in the microbiome values and a decreasing in the genetic values when the symbiosis effect was modelled. In the high line, no differences were observed for the genetic values when the symbiosis was considered. However, for scenarios with both the G and M effects, the high line showed greater microbiome values at the last generation as the microbial heritability increased, when no microbial interaction was modelled. No differences in the microbiome values were observed for these scenarios when the symbiosis effect was considered.

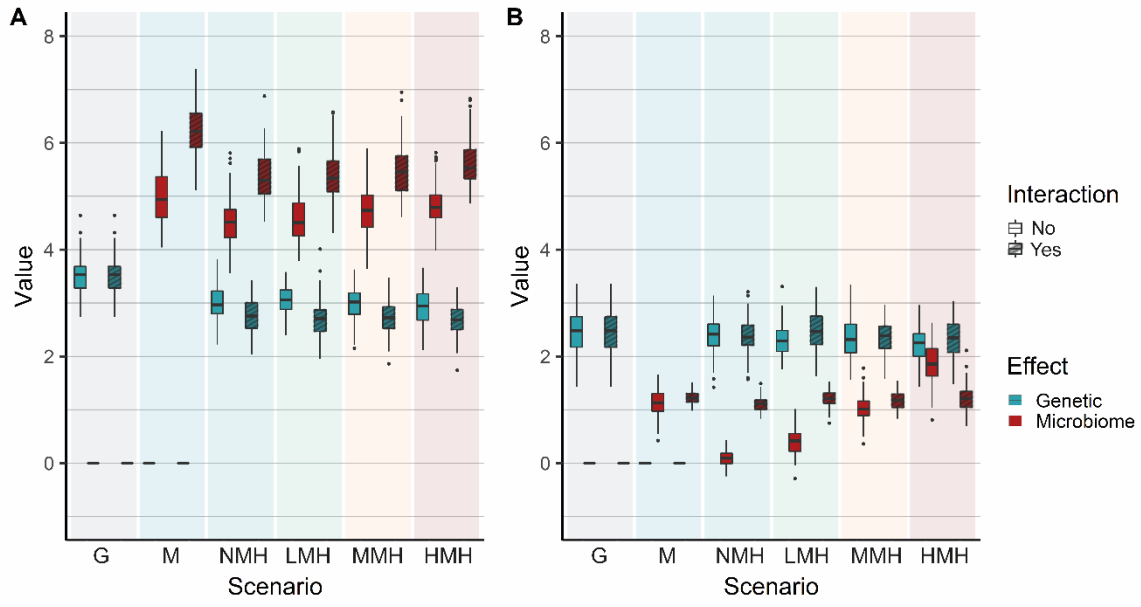


Figure 6. Boxplots of the genetic and microbiome value on the last generation for all scenarios simulated. The values were obtained after 100 simulations of each scenario. A) Absolute values observed in the divergent line selected for decreasing the trait (Low line). Absolute values were computed in the low line to observe the selection response on the same scale. B) Values observed in the divergent line selected for increasing the trait (High line). The genetic values are in blue, and the microbiome value is in red. Background colours represented each scenario simulated, being G (grey): genome scenario; M (dark blue) the microbiome scenario; NMH (light blue): non microbial heritability; LMH (green); low microbial heritability; MMH (yellow): medium microbial heritability; and HMH (red): high microbial heritability.

Symbiosis enhanced the selection response

Differences were observed in the scenarios when the interaction is modelled (Fig 5-6). Scenarios that included the microbial species interactions showed higher differences in their phenotypic values in the last generation with respect to the base population (Fig. 5). For the low line, a difference in the mean of the phenotypic values of one kid is observed between the M scenario with and without microbial interaction (Fig. 5A). When both, genome and microbiome influenced the phenotype, subtle differences were observed with the inclusion of the microbial interaction effect. Furthermore, no differences dependently of the microbial heritability were observed (Fig. 5A). A different effect of the symbiosis on the response to selection was observed in the high line (Fig.

5B). Dissimilarities in the phenotypic value with or without interaction was decreasing at the time that the microbial heritability increase for this line (Fig. 5B). Scenario with symbiosis showed higher phenotypic values until a strong microbial heritability was modelled (Fig 5B; HMH scenario), being 0.5 higher the HMH without symbiosis than with symbiosis in the high line.

5. DISCUSSION

In a traditional framework, the variability of the phenotype observed were assumed to be a consequence of the individual genome and the environment [27]. However, the microbiome can also shape phenotypes [1-4]. Thus, the coevolution between the microbiome and individual genome allows the adaptation and evolution of species [1-3], being of interest to animals and plant breeders. However, the microbiome is composed by living organisms influenced by a multitude of factors so deciphering it is still a challenge [28]. Complications of the microbiome and genome coevolution lie in the microbiome inheritance because it depends on the colonization of the microbial species in the new host [14-15,18, 29-31]. In this tool, we suggested a successful microbiome inheritance based on splitting the colonization into two actions: (i) the acquisition of microbial species and (ii) the establishment of these ones in the offspring. The (i) was modelled considering the relative abundance of the microbial species from the mother and a simulated constant environment as the probability of random sampling to the offspring. In our example, we considered that the mother does the major contribution to the microbiome of the offspring. However, it seemed that the mother only contributes to a few microbial species in the adult individual and that the microbiome composition is highly influenced by the diet and the housing environment [17-18]. The user can change the degree of acquisition of microbial species from the environment and the mother to explore these scenarios and observe their effects on the selection response.

The establishment of the mature microbiota depends on the education of the host immunity [32]. Then, (ii) was developed and assigned a fitness value to each microbial species that is a coefficient of variation to modulate their ability to survive in the new individual due to their own genes. For that, we used the approach suggested by Pérez-Enciso et al. (2021) [24] were each microbial species were considered another quantitative trait (Eq. 1). Moreover, we determine that some part of the abundance of the microbial species is influenced by the host genome [22] and the community of bacteria present in the host [25,33]. Microbial heritability and symbiosis effect were considered key factors to determine the final abundance of the microbiome (Fig. 2). Fitness was constant across all the simulations as well as the symbiosis effect. That means that the observed microbiome trend [Additional file 1] is a response to the selection based on the

phenotype. Hence, assumptions made in this simulation tool to model the microbiome inheritance showed promising results, considering both the contribution of a parental and environmental microbiome.

An asymmetry response was observed for the phenotypic, genetic, and microbiome values (Fig 5-6) in the divergent selection carried out. The G and M scenario showed that the selection response is more favourable to decrease than increase the trait (Fig 5). In the G scenario, the phenotypic response was higher in the low line (Fig 5.) suggesting that the effect of the QTLs was more favourable for decreasing the phenotype than for increasing it. For, the M scenario, the same trend was observed although the number of microbial species affecting positively the trait (22) was higher than those affecting negatively (13). This could be a consequence of the high fitness of the microbial species with a high negative effect on the phenotype (Fig. 3). Fitness is a variable that represents the stability for the growth of the microbial species, i.e., the fluctuation of their abundance from the mean. There are four microbial species with a high positive effect showing also low fitness values. This means that these microbial species showed high fluctuations from their mean abundance. So, the abundance distribution of these microbial species is highly variable, which could increase the noise in the individual phenotype avoiding selecting the best reproductive animals and decreasing the selection response. Hence, the number of stable microbial species seems to be more important than the number of species affecting the trait. This is supported by the low phenotypic value observed in the M scenario respected to the G scenario (Fig. 5B) which is in line with the challenge to select the best breeding animals when highly noisy microbial species are. influencing the trait. We also observed that with the inclusion of the symbiosis effect, the noise associated to the microbiome is reduced and the selection response is increased (Fig. 5) since the fluctuation of the abundance value of low fitness microbial species is also reduced (Fig. 5B). Furthermore, the negative effect of the fitness on the selection response seemed to be reduced by the microbial heritability (the host genome). In the high line which was more affected by fitness, higher phenotypic (Fig. 5B) and microbiome values (Fig. 6B) were observed when the microbial heritability increased. The four microbial species with low fitness were influenced by the host genome (Fig. 3) meaning that this component is key to reducing and controlling the microbiome fluctuations. Along the same line, symbiosis allowed to obtain better phenotypic values since part of the environmental variance is associated with the community of microbial species in the individual. Then, microbial species fitness, microbial heritability, and symbiosis must be carefully considered in the construction of strategies to properly improve the selection response using the microbiome information.

6. CONCLUSION

This is the first simulation tool to model the coevolution of the genome and the microbiome in a selection process based on the phenotype. The key to this tool is the implementation of microbiome inheritance. It is constructed in R and based on AlphaSimR so different scenarios could be implemented by the user. We performed a first approximation, showing as the selection based on the phenotype is successful and generated phenotypic, genetic, and microbiome responses. The results obtained highlight the importance of microbial heritability, symbiosis, and microbial species fitness as key actors to obtain a proper selection response of traits. Further updates will be implemented in this tool for performing genomic or microbiome selection to decipher their effect on the phenotypic response of key traits.

7. DECLARATIONS

Ethics approval and consent to participate

All the experimental procedures were approved by the Committee of Ethics and Animal Welfare of the Miguel Hernández University, according to Council Directives 98/58/EC and 2010/63/EU (Ref. 2017/VSC/PEA/00212).

Consent for publication

Not applicable.

Availability of data and materials

Data are available upon request to the corresponding author.

Funding

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CCR analysed the data and wrote the manuscript. AB and MJA conceived and designed the study and contributed to the discussion of the results. MLG contributed to the study design and the discussion of the results. NIE conceived the study, contributed to the discussion of the results and edited the manuscript. All the authors have read and approved the final manuscript.

Acknowledgements

Cristina Casto-Rebollo acknowledges a FPU17/01196 scholarship from the Spanish Ministry of Science, Innovation and Universities.

8. REFERENCE

1. Rosenberg E, Zilber-Rosenberg I. The hologenome concept of evolution after 10 years. Microbiome. 2018;6(1):78. doi:10.1186/s40168-018-0457-9

2. Shapira M. Gut Microbiotas and Host Evolution: Scaling Up Symbiosis. Trends Ecol Evol. 2016;31(7):539-549. doi:10.1016/j.tree.2016.03.006

3. Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev. 2008;32(5):723-735. doi:10.1111/j.1574-6976.2008.00123.x

4. Henry LP, Bruijning M, Forsberg SKG, Ayroles JF. The microbiome extends host evolutionary potential. Nat Commun. 2021;12(1):5141. doi:10.1038/s41467-021-25315-x.

5. Bergamaschi M, Tiezzi F, Howard J et al. Gut microbiome composition differences among breeds impact feed efficiency in swine. Microbiome. 2020; 8:110. doi:10.1186/s40168-020-00888-9

6. Déru V, Tiezzi F, Carillier-Jacquin C, et al. Gut microbiota and host genetics contribute to the phenotypic variation of digestive and feed efficiency traits in growing pigs fed a conventional and a high fiber diet. Genet Sel Evol. 2022;54(1):55. doi:10.1186/s12711-022-00742-6

7. Maltecca C, Dunn R, He Y, et al. Microbial composition differs between production systems and is associated with growth performance and carcass quality in pigs. Anim Microbiome. 2021;3(1):57. doi:10.1186/s42523-021-00118-z

8. Liu J, Stewart SN, Robinson K, et al. Linkage between the intestinal microbiota and residual feed intake in broiler chickens. J Anim Sci Biotechnol. 2021;12(1):22. doi:10.1186/s40104-020-00542-2

9. Xue MY, Sun HZ, Wu XH, et al. Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. Microbiome. 2020;8(1):64. doi:10.1186/s40168-020-00819-8

10. Martínez-Álvaro M, Auffret MD, Stewart RD, et al. Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. Front Microbiol. 2020;11:659. doi:10.3389/fmicb.2020.00659

11. Martínez-Álvaro M, Zubiri-Gaitán A, Hernández P, et al. Comprehensive functional core microbiome comparison in genetically obese and lean hosts under the same environment. Commun Biol. 2021;4(1):1246. doi:10.1038/s42003-021-02784-w

12. Casto-Rebollo C, Argente MJ, García ML, et al. Selection for Environmental Variance Shifted the Gut Microbiome Composition Driving Animal Resilience, 02 May 2022, PREPRINT (Version 1) available at Research Square doi:10.21203/rs.3.rs-1597523/v1

13. David I, Ricard A. A Unified Model for Inclusive Inheritance in Livestock Species. Genetics. 2019;212(4):1075-1099. doi:10.1534/genetics.119.302375

14. Coelho GDP, Ayres LFA, Barreto DS, et al. Acquisition of microbiota according to the type of birth: an integrative review. Rev Lat Am Enfermagem. 2021;29:e3446. doi:10.1590/1518.8345.4466.3446

15. Penders J, Thijs C, Vink C, et al. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics. 2006;118(2):511-521. doi:10.1542/peds.2005-2824

16. Senn V, Bassler D, Choudhury R, et al. Microbial Colonization From the Fetus to Early Childhood-A Comprehensive Review [published correction appears in Front Cell Infect Microbiol. 2021;11:715671]. Front Cell Infect Microbiol. 2020;10:573735. doi:10.3389/fcimb.2020.573735

17. Bian G, Ma S, Zhu Z, et al. Age, introduction of solid feed and weaning are more important determinants of gut bacterial succession in piglets than breed and nursing mother as revealed by a reciprocal cross-fostering model. Environ Microbiol. 2016;18(5):1566-77. doi:10.1111/1462-2920.13272

18. Mulder IE, Schmidt B, Stokes CR, et al. Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. BMC Biol. 2009;7:79. doi:10.1186/1741-7007-7-79

19. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG, Ley RE. Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe. 2016;19(5):731-43. doi: 10.1016/j.chom.2016.04.017

20. Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. Nat Rev Genet. 2017;18(11):690-699. doi:10.1038/nrg.2017.63

21. Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555(7695):210-215. doi:10.1038/nature25973

22. Grieneisen L, Dasari M, Gould TJ, et al. Gut microbiome heritability is nearly universal but environmentally contingent. Science. 2021;373(6551):181-186. doi:10.1126/science.aba5483

23. Gaynor RC, Gorjanc G, Hickey JM. AlphaSimR: an R package for breeding program simulations. G3 (Bethesda). 2021;11(2):jkaa017. doi:10.1093/g3journal/jkaa017

24. Pérez-Enciso M, Zingaretti LM, Ramayo-Caldas Y, de Los Campos G. Opportunities and limits of combining microbiome and genome data for complex trait prediction. Genet Sel Evol. 2021;53(1):65. doi:10.1186/s12711-021-00658-7. Erratum in: Genet Sel Evol. 2021 Dec 20;53(1):97

25. Coyte KZ, Rakoff-Nahoum S. Understanding Competition and Cooperation within the Mammalian Gut Microbiome. Curr Biol. 2019;29(11):R538-R544. doi:10.1016/j.cub.2019.04.017

26. Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. Genetics. 2014 Oct;198(2):483-95. doi:10.1534/genetics

27. Falconer D, Mackay TFC. (1996) Introduction to Quantitative Genetics, 4th ed. Edinburgh Gate, Harlow, U.K.Addison Wesley Longman Limited.

28. Martino C, Dilmore AH, Burcham ZM, Metcalf JL, Jeste D, Knight R. Microbiota succession throughout life from the cradle to the grave. Nat Rev Microbiol. 2022. doi:10.1038/s41579-022-00768-z

29. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A. 2010;107(26):11971-11975. doi:10.1073/pnas.1002601107

30. He Q, Kwok LY, Xi X, et al. The meconium microbiota shares more features with the amniotic fluid microbiota than the maternal fecal and vaginal microbiota. Gut Microbes. 2020;12(1):1794266. doi:10.1080/19490976.2020.1794266

31. Donnet-Hughes A, Perez PF, Doré J, et al. Potential role of the intestinal microbiota of the mother in neonatal immune education. Proc Nutr Soc. 2010;69(3):407-15. doi:10.1017/S0029665110001898

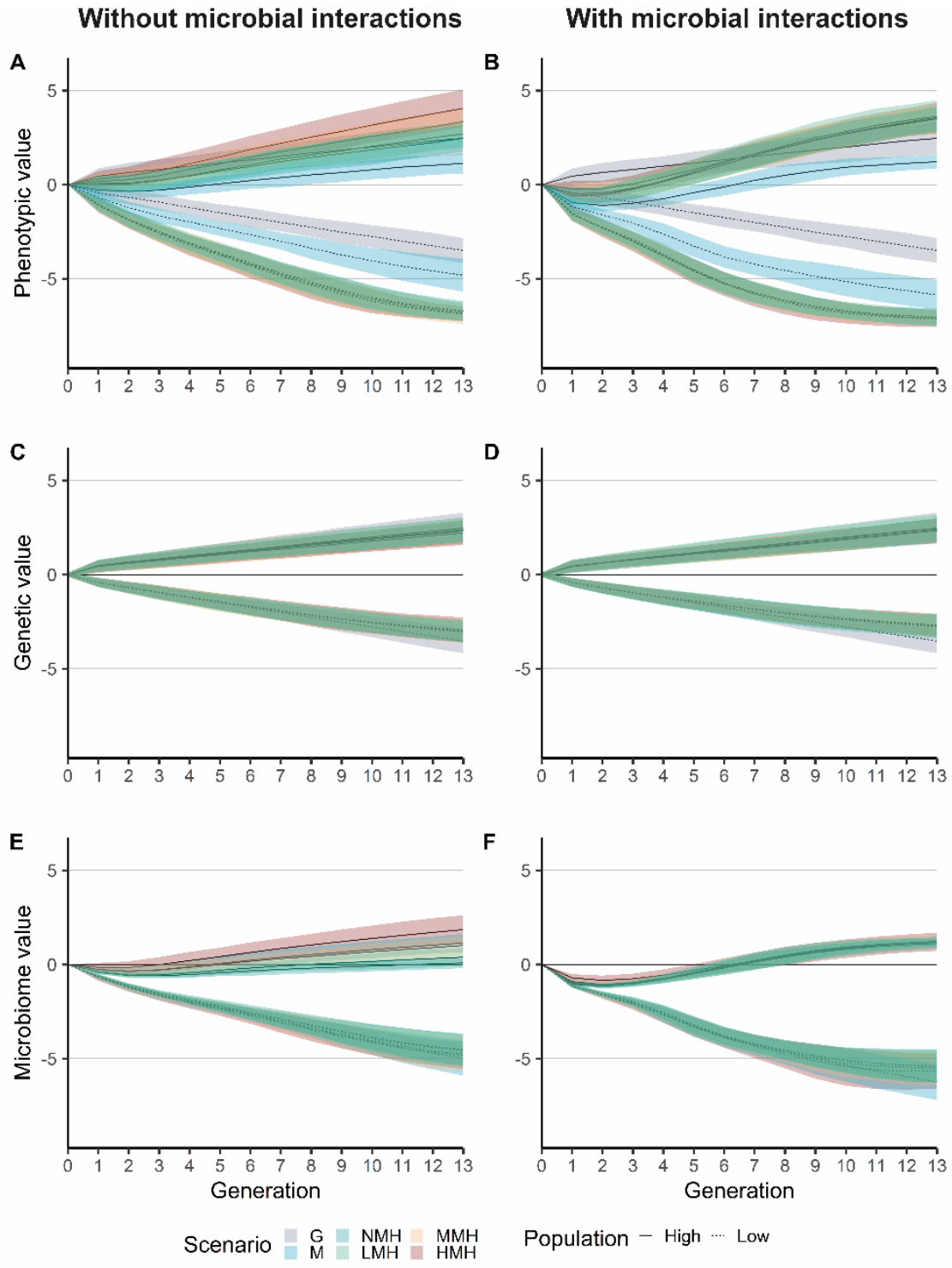
32. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30(6):492-506. doi:10.1038/s41422-020-0332-7

33. Liévin-Le Moal V, Servin AL. Anti-infective activities of lactobacillus strains in the human intestinal microbiota: from probiotics to gastrointestinal anti-infectious biotherapeutic agents. Clin Microbiol Rev. 2014;27(2):167-99. doi:10.1128/CMR.00080-13

9. ADDITIONAL FILES

Additional file 1

Selection trends with and without microbial interactions for all scenarios throughout 13 generations of divergent selection. A) and B) are the phenotypic trends for each scenario under the assumption of microbial interaction (B) or not (A). C) and D) are the genotypic trends with (D) and without (C) microbial interactions. E) and F) are the microbiome trends with (F) and without (E) interactions. Trends are plotted as the confidence interval of the values for 100 simulations of the divergent selection process for each scenario. Each line represents the mean for the values of each scenario for the high or low divergent population. Each colour represents a simulated scenario, G: only genome effect on the phenotype (grey); M: only microbiome effect on the phenotype (Blue dark) and NMH, LMH, MMH, and HMH are different scenarios with both the effect of the microbiome and the genome in the phenotype. Differences between these scenarios lie in the degree of microbial heritability simulated, being NMH: the scenario without microbial heritability; LMH; the scenario with a low microbial heritability; MMH; the scenario with medium microbial heritability; and HMH: the scenario with high microbial heritability.


Without microbial interactions
With microbial interactions
A
B
C
D
E
F
Phenotypic value
Genetic value
Microbiome value
5
0
-5
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Generation
Scenario
G
M
NMH
LMH
MMH
HMH
Population
High
Low

CHAPTER 8: GENERAL DISCUSSION

1. INTRODUCTION

The residual variance of traits, referred in this project as V_E, is under genetic control. That means that some of the dissimilarities among individuals observed in the V_E result from a genetic variation in response to microenvironmental factors [1-2]. Response to selection for V_E was observed in divergent selection experiments [3-5]. A low V_E was positively correlated with productive traits such as birth weight [6-7] and milk daily yield [8] and negatively correlated with health traits [9-12] and animal resilience [10-11]. Moreover, the inclusion of V_E as a breeding goal could homogenize production on farms by its reduction. V_E is a promising trait to improve animal resilience and welfare [10-13] to reach a more sustainable livestock system. This thesis used animals from a divergent selection experiment for high and low V_E of litter size (LS) carried out in the University of Miguel Hernández in Elche. They showed an outstanding genetic response to the selection for V_E of LS [3] as well as differences in their resilience potential [10], being an exceptional biological material to study animal resilience through the V_E. Thus, this thesis aimed to investigate the molecular basics of these rabbit populations to gain insight into the biological mechanisms of animal resilience. This thesis focused on studying the genome (*Chapter 3* and *4*) and the gut microbiome (*Chapter 5* and *6*) because both are important to shape phenotypes. Moreover, selection could act in both the genome and the microbiome allowing their coevolution. To study that, we developed a tool for simulating the host-genome and microbiome evolution throughout a phenotypic selection process (*Chapter 7*).

2. IDENTIFICATION OF CANDIDATE GENES FOR THE V_E

In *Chapter 3*, GWAS was performed to identify candidate genes that could influence the V_E and animal resilience. Two methodologies were implemented because SNPs or variants' size effect strongly depends on the model used [14]. We used single marker regression (SMR) and Bayesian multiple marker regression (BMMR) to perform the GWAS. The combination of both methodologies allowed us to obtain more reliable genomic regions associated with the V_E. The principal problem of SMR is the overestimation of the SNP effect because it fits one SNP in each model, considering all SNPs independently among them [15]. Moreover, there is no consensus on the threshold for controlling the number of false positives and negatives to correct the multiple testing performed by SMR. Thus, using another methodology could help to detect the true SNP associated with the trait. The BMMR methodologies also have limitations because they depend on the genomic architecture assumption [14]. For instance, we used Bayes B to

perform the BMMR analysis, considering that approximately only 100 SNPs influenced the V_E. Other methodologies allow estimating the number of SNPs affecting the trait to avoid making assumptions (i.e., Bayes Cpi). However, our sample size (367 animals) was insufficient to implement them.

We detected four genomic regions in the *Oryctolagus cuniculus* chromosome (OCU) 3, 7, 10, and 14, which only explained the 8.6% of the total genetic variance for the V_E. That's a common problem in GWAS studies for most complex traits because there are a vast number of variants with small effect sizes [16-18]. Thus, the genetic variance explained for each variant is rather small, making their identification difficult. Proper sample size and SNP density and distribution are necessary to correctly estimate the SNP effect and obtain adequate statistical power to detect most of the QTLs. Likewise, V_E has a low to moderate heritability (0 to 0.21) and its estimation depends on the animal species, the trait and the methodology used to calculate it [19-20]. In our rabbit population, the heritability estimation was from 0.05 to 0.11 (0.08 on average). Thus, there is a lot of noise due to environmental variability, which complicates the estimation of the SNP effect on the population. Inappropriate estimation of the SNP effect hinders the identification of the causal variants. For that, in *Chapter 4* we tried to map most of the QTLs for the V_E without calculating the effect size of SNPs. We used methodologies for identifying signatures of selection (SS) regardless the phenotype. SS allows using animals without phenotypic records. We used the same animals from the GWAS (from the base population and generation 11) and another 93 animals from generation 13 (without phenotype) to validate the results.

SS methodologies try to identify a disruption in the patterns of the inheritance of neutral alleles, which will indicate genomic regions harbouring causal variants under selection [21]. Low correlations among statistics were reported by Ma et al. (2015) [22], González-Rodríguez et al. (2016) [23], and Sosa-Madrid et al. (2020) [24]. Thus, we used intra- and inter-population analyses to explore most of the possible patterns generated by selection. We searched for the reduction in the local variability (Runs of homozygosity; ROH), variation in the linkage disequilibrium between populations (varLD), and differences in the allele frequencies between populations (F_{ST}). The combination of multiple methodologies with different assumptions was proposed to detect most SS [21, 25]. Thanks to this study (*Chapter 4*), we identified another 134 genomic regions under selection pressure for V_E of LS. These regions were validated on the animals from generation 13. An overlapping between these genomic regions and those identified by the GWAS (*Chapter 3*) was expected. However, no overlapping regions were identified. That could be due to the restrictive threshold used to determine true signatures of selection

(above or equal to the weighted F_{ST} value at percentile 99.9). With a modification of the threshold for the weighted F_{ST} values, we identified the genomic region on OCU3 (51-51.75Mb) as SS (*Chapter 4*) which was previously found in the GWAS (*Chapter 3*). For that, we considered a 98 percentile instead of the 99.9 percentile for the weighted F_{ST} values. Only a few studies showed overlapping results among GWAS and SS [24, 26-29]. That may be because these studies used high heritable traits (intramuscular fat) [24, 27] or a large selection process (40 generations) [28-29]. Hence, the overlapping region on OCU3 (*Chapter 3* and *Chapter 4*) could indicate that it is a segregating region relevant to the V_E.

All genomic region identified in *Chapter 3* and *Chapter 4* was considered genomic regions with candidate genes for the V_E because each methodology used different assumptions and algorithms, limiting the identification of QTLs for V_E. However, GWAS and SS methodologies do not allow the identification of causal variants due to the large LD generated by selection and the low effective population size in livestock populations [30-31]. Moreover, causal variants are not usually included in the SNP chips. For all that, both studies were complemented with a whole genome sequencing (WGS) analysis to discover the functional mutation in the genomic regions identified for the V_E. DNA pools with 27 males for each rabbit population were used to compare the whole genome and identify functional mutations in the transcription unit of the genes (UTR region, exon, and splicing site). These animals were all the sires from animals of generation 11 (used in GWAS and SS), so most of the genetic information of the population was sequenced. DNA pools give reliable estimates of the allele frequencies, although they confound low frequent variants with sequencing errors [32]. This complicates accurately identifying rare variants. However, it is a good approximation to search for the most critical functional variants at a more affordable price.

We expected that the candidate genes were involved in pathways related to the immune system since other GWAS for V_E suggested this (see the review of Iung et al., 2020 [20]). Moreover, Argente et al. (2019) [10] showed that the rabbit lines used in this study showed differences in their resilience potential, as well as in inflammatory biomarkers such as CRP, measured after the stressful condition of the parity [33]. These results suggested a link between V_E and animal resilience through the immune system. The immune system is essential to perceive and respond to environmental disturbances for modulating animal resilience [34-35]. The proposed candidate genes identified in *Chapter 3* and *Chapter 4* were related to important biological processes for animal resilience; the immune response (*DOCK2*, *HDAC9*, *ITGB8*, *HUNK*), stress response (*ENSOCUG00000021276*, *MC2R*), development of sensory structures (*FOXI1*, *FGF18*,

ECA1C), behaviour (*FBXL20*) and nervous system (*MC2R*, *SLC18A1*, *FBXL20*). Genes from the family of the heat-shock protein (HSP) were identified for the V_E by Sell-Kubiak et al. (2015) [36] in pigs, Iung et al. (2018) [37] in cattle, and Morgante et al. (2015) [38] in Drosophila Melanogaster. In *Chapter 3,* we identified the *HDAC9* gene that can control the expression of genes from the HSP family such as HSP70 [39], in line with these previous GWAS results for V_E. Then, it seems that independently of the resilience indicator used, the HSP protein is critical to control the V_E. HSP is related to the response to heat stress by boosting the immune response and countering cell damage [40-42], so it could be critical for modulating animal resilience. Furthermore, in *Chapter 4,* we found the *GATA3*, *FKBP10*, *KAT2A*, *CYP1B1*, *BRCA1*, *PGM3*, and *ACE* genes with pleiotropic function in the immune response, lipid and carbohydrates metabolism, and reproduction and embryo development. These genes are functional mutations whose alleles are fixed in these rabbit populations. Their pleiotropic functions could explain the differences between the rabbit populations generated by the V_E selection for immunological biomarkers [33], litter size [3, 43], reproductive traits [43-44], triglycerides and cholesterol plasma levels and animal resilience [10].

DOCK2, *INSYN2B,* and *FOXI1* are the only genes with functional mutation harboured in the genomic regions identified in both *Chapter 3* and *Chapter 4* (with a less restrictive threshold for F_{ST} values). *DOCK2* is important for developing the inflammatory response [45] since its expressed in most hematopoietic cells such as monocytes, macrophages, and lymphocytes [46-47]. The functional mutation identified by WGS was found in the 3'UTR, the splicing site, and on an exon that could greatly impact the gen's functionality. The alternative allele is absent in the resilient population and is segregating in the non-resilient population. These results may indicate differences in the expression and functionality of the *DOCK2* gene. The *FOXI1* gene was identified as relevant to develop sensory structures such as the inner ear [48]. We did not find much information about the *INSYN2B* gene in the literature. However, it seemed that its paralog, the *INSYN2A* gene, is important to modulate a correct nervous system (GeneCards [49]), so it could develop a similar function. These genes are functional mutations on the 3'UTR (*FOXI1)* and the 5'UTR (*INSYN2B*), which could affect the stability and expression of their mRNA. Functions of all these three genes are related to the biological mechanisms which may influence animal resilience: immune system, sensory perception, and nervous system. Further studies are necessary to know their real implications in modulating animal resilience. There are the most reliable genes found in this study because they were identified as associated with the V_E (*Chapter 3*) and with differences in their allele frequency (*Chapter 4*; F_{ST} value) between the rabbit populations.

Another important gene identified in this study (*Chapter 3*) was the gen *ENSOCUG00000021276,* an orthologue to the human *MRAP* (melanocortin receptor). The *MRAP* gene can modulate the expression and function of the *MC2R* gene [50-51] found with signatures of positive selection in *Chapter 4*. MRAP forms a complex with MC2R to respond to the adrenocorticotropin hormone or ACTH [50] for stimulating cortisol production, a biomarker of the stress response [52]. Different levels of cortisol were shown in the rabbit populations [10], which is in line with the function of the MRAP/MC2R complex. The genes found in *Chapter 3* and *Chapter 4* were not previously identified for V_E or animal resilience. However, genes with similar functions were recently identified in studies for heat stress in cattle [53], pig [54] and sheep [55], and disease resilience in pigs [56-57]. Further studies are necessary to really know their implications on the control of the immune and stress response and therefore modulate animal resilience.

3. DISSIMILARITIES IN THE MICROBIOME COMPOSITION

The results from *Chapter 5* and *Chapter 6* showed that the selection for V_E modified the microbiome composition between the rabbit populations. We expected differences in alpha- and beta-diversity between the rabbit lines. However, we observed that the microbiome composition was generally the same for both lines, as alpha- and beta-diversity did not show relevant differences in diversity and richness (*Chapter 5*). High microbial diversity is related to a healthy and resilient gut because it allows individuals to cope better with gut disturbances [58-59]. However, there is functional redundancy among microbial species [60-62]. The relevance lies in the functionality (genes) of the microbiome instead of the microbial species itself because its function is most conserved [63]. Unrelated taxa can have a similar subset of genes, and the acquisition and colonization of each one depends on a huge number of factors [64-65]. Selection will act on the microbial species genes, selecting functionality that affects the phenotype. Indeed, the acquisition of microbial species with healthy properties should be most important than the diversity because they can directly control the growth of harmful bacteria [66]. This implies bacterial cell communications to determine microbial interactions.

Differences in bacterial cell communication can be supported by differences in the abundance of genes related to quorum sensing and ABC transporters (*Chapter 5*). Both are important to develop proper bacterial cell communication to control gene expression and pathogenesis [67-68]. Moreover, these genes are important to the development of biofilms. A biofilm is a community of associated bacteria assembled in a matrix and

adhered to a surface [69-70]. The formation of biofilms allows the optimization and coordination of activities and functions [69-70]. The simulation developed in Chapter 7 showed that microbial interactions are critical in shaping phenotypes and controlling the microbiome composition. Moreover, we showed that the phenotypic selection response was the greatest with the inclusion of microbial interaction in the model (Chapter 7). This could explain the outstanding selection response observed in these rabbit lines [3]. On the other hand, biofilms are a source of bacteria to protect against high-stress levels [71]. Selection for V_E generated differences in the susceptibility to stress between the rabbit populations [10, 33], which could be influenced by microbial interactions. A proper understanding of microbial interaction is needed to accurately determine the overall effect of the microbiome on the phenotype. The results from *Chapter 6* and *Chapter 7* suggested that microbial interactions are important to modulate the V_E and animal resilience and boost the phenotypic selection response.

Diet [72] and housing environment [73] are major factors in determining the gut microbiome. These two factors were the same for both rabbit populations, so microbial species acquisition should have been the same through the selection process. Moreover, they belonged to the same base population. The differences found in the microbiome composition could be explained by (i) an influence of the host genome on the colonization and persistence of some microbial species or by (ii) the selection of those species with effect on the V_E. *Chapter 7* shows a good tool for studying the implications of the microbiome composition in the phenotype. We observed that the strength of the trends for the microbiome value was influenced by the microbial species heritability. Microbial heritability seemed critical for microbial species with high variability or low fitness values (*Chapter 7*). Microbial heritability may be important to maintain the stability of the microbiome by reducing their abundance fluctuations among individuals. Hence, increasing the microbiome stability may allow reducing the environmental noise to select the breeding animals and thus increasing the selection response (*Chapter 7*). However, the microbiome strongly also depends on the relationship or symbiosis between the bacteria species.

i. Influence of the rabbit genome

Searching for genes that could modify the microbiome composition, we found that *DOCK2* and ACE genes could be involved. The *DOCK2* gene located in the genomic region identified in *Chapter 3* and *Chapter 4* was suggested to modulate the microbiome composition and the susceptibility to pathogen infections like *Citrobacter rodentium* [74]. A reduction in the abundance of the genus *Lactobacillus* was found in the knockout

DOCK2 mice, indicating higher susceptibility to *C. Rodentium*. In our study, the genus *Lactobacillus* showed an increased abundance in the resilience line (low V_E), supporting their low susceptibility to stress conditions. *Lactobacillus* could activate the T_{reg} cells and control gut inflammation after pathogen infection [75]. The same applies to the *ACE* gene. This one produces the substrate (Angiotensin II) for its homologous *ACE2* to produce Angiotensin I. Mutant mice for *ACE2* showed different microbiome composition, which could be a consequence of the dysregulation of tryptophan homeostasis, affecting the gut integrity and immunity [76]. An incorrect substrate bioavailability for *ACE2* generated by *ACE* action could influence the gut microbiota via the tryptophan metabolism. The findings in *Chapter 6* support this result showing differences in the levels of metabolites from tryptophan metabolism between the rabbit populations. Hence, both *DOCK2* and *ACE* could modulate the microbiome composition and consequently animal resilience. We suggested that the rabbit genome could be influencing the abundance of the microbial species in the gut. However, direct empirical analyses are needed to validate this hypothesis. Moreover, the implications of the other candidate genes in the gut microbiome composition must be studied.

ii. Microbial species affecting the V_E

Selection for V_E could have modified the microbiome composition by selecting microbial species affecting the V_E. We showed that animal resilience and V_E are highly related to the immune system (*Chapter* 3 and *Chapter 4*) [10, 33], so we expected differences in beneficial and harmful bacteria between the rabbit populations, which could influence the health status of the animals, affecting their susceptibility to stress conditions. In chapter 5, we found in the resilient population an overrepresentation of beneficial bacteria belonging to the genus *Rikenella*, *Lactobacillus*, *Alistipes,* and *Odoribacter*, being the most representative species *Alistipes prutedinis*, *Alistipes shahii*, *Odoribacter splanchnicus* and *Limosilactobacillus fermentum*. Moreover, harmful species such as *Acetatifactor muris* and *Eggerthella sp.* were more abundant in non-resilience animals. Likewise, the study of the gut metabolome (*Chapter* 6) revealed gut-derived metabolites such as equol, 3-(4-hydroxyphenyl)lactate, 5-aminovalerate, N6-acetyllisine and serine, supporting the differences in the microbial species composition between the rabbit populations. The genus *Lactobacillus* seems to be a promising probiotic due to its beneficial properties for maintaining the gut epithelial barrier, modulating gut immunity, and controlling pathogen infections [75]. Likewise, a recent study suggested that this genus could also contribute to develop a protective mechanism to cope with heat stress [77]. Likewise, tryptophan metabolism has been related to the *Lactobacillus* genus [78], so we expected an overrepresentation of their metabolites in the resilient line.

However, in *Chapter 6,* we observed that the tryptophan metabolites indole, kynurenine, and anthranilate, showed a high abundance in the non-resilient population. The degradation of tryptophan to metabolites such as kynurenic acid and aryl indole seems important in mediating the inflammation process [78-79]. Elevated levels of kynurenine and anthranilate are found in individuals under stress and inflammation [80]. Rabbits were under high-stress conditions given they were slaughtered after their first parity. The overrepresentation of kynurenine and anthranilate in the non-resilient population could pinpoint a higher susceptibility to stress and inflammation in this non-resilient population. This is in line with the higher levels of CRP (an inflammation biomarker) found in this population [33], which could activate the tryptophan metabolism to counter its effects [78, 81-82]. Unexpectedly, the indole was found in low concentration in the resilient line (Table 1). This metabolite has protective functions in the gut as maintaining the intestinal barrier integrity and immune homeostasis to avoid dysbiosis during an inflammation response [83-84]. An in-depth study would be needed to understand the interplay of these metabolites derived from tryptophan on animal resilience. Moreover, the *HDAC9* gene identified in *Chapter 3*, which could modulate the expression of HSP70 and the functions of Treg cells, modulate the inflammation in the gut. Thus, functional mutation of this gene could increase the inflammation levels in the non-resilient line. This is also supported by the high levels of long-chain fatty acylcarnitines shown in non-resilient animals, biomarkers of inflammation in the gut [85].

We identified other derived metabolites from the aromatic amino acid (AAA) metabolism (phenylalanine, tyrosine, and tryptophan). This is consistent with the dissimilarities in AAA biosynthesis genes (chorismite mutase and lyase) found in *Chapter* 5. Results from *Chapter 5* and *Chapter 6* highlighted the importance of this pathway to the V_E. However, the implications of AAAs metabolism remain unclear because these depend on the active metabolites, which determine the beneficial or harmful effect for the host [86]. Moreover, gut-derived metabolites could not act directly in the gut but in other organs and tissues, inducing crosstalk with the host. For instance, the 3-(4-hydroxyphenyl)lactate, an end-product of tyrosine metabolism, is a biomarker of liver disease [87] and was overrepresented in the non-resilient line. This could suggest that gut-derived metabolites in the resilience animal could affect liver functions. This is supported by the differences between these rabbit lines in plasma cholesterol and triglyceride levels found by Argente et al. (2019) [10]. Likewise, overlapping results from *Chapter 5* and *Chapter 6* were found for the synthesis of formiminoglutamate as differences in its enzyme, the glutamate formiminotransferase, were found among the rabbit populations. This could also suggest the importance of histidine degradation and the synthesis of L-GLutamate for V_E and animal resilience. However, further studies are

needed to properly determine each metabolite's origin and mode of action to unravel their causal role in health and disease, as well as in the host-gut crosstalk.

4. REFERENCES

1. Falconer D, Mackay TFC. (1996) Introduction to Quantitative Genetics, 4th ed. Edinburgh Gate, Harlow, U.K.Addison Wesley Longman Limited.

2. Lynch M, and Walsh B. (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland.

3. Blasco A, Martínez-Álvaro M, García ML, et al. Selection for environmental variance of litter size in rabbit. Genet Sel Evol. 2017;49:48. doi:10.1186/s12711-017-0323-4

4. Formoso-Rafferty N, Cervantes I, Ibáñez-Escriche N, Gutiérrez JP. Correlated genetic trends for production and welfare traits in a mouse population divergently selected for birth weight environmental variability. Animal. 2016;10(11):1770-1777. doi:10.1017/S1751731116000860.

5. Garreau H, Bolet G, Larzul C, et al. Results of four generations of a canalising selection for rabbit birth weight. Livest Sci, 2008;119(1–3):55–62. doi:10.1016/j.livsci.2008.02.009

6. Sell-Kubiak E, Bijma P, Knol EF, Mulder HA. Comparison of methods to study uniformity of traits: Application to birth weight in pigs. J Anim Sci, 2015;93(3):900–911. doi:10.2527/jas.2014-8313

7. Sell-Kubiak E, Wang S, Knol EF, Mulder HA. Genetic analysis of within-litter variation in piglets' birth weight using genomic or pedigree relationship matrices. J Anim Sci. 2015;93(4):1471-1480. doi:10.2527/jas.2014-8674

8. Poppe M, Bonekamp G, van Pelt ML, Mulder HA. Genetic analysis of resilience indicators based on milk yield records in different lactations and at different lactation stages. J Dairy Sci. 2021;104(2):1967-1981. doi:10.3168/jds.2020-19245

9. Elgersma GG, de Jong G, van der Linde R, Mulder HA. Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. J Dairy Sci. 2018;101(2):1240-1250. doi:10.3168/jds.2017-13270

10. Argente MJ, García ML, Zbyňovská K, et al. Correlated response to selection for litter size environmental variability in rabbits' resilience. Animal. 2019; 13:2348-55. doi:10.1017/S1751731119000302

11. Poppe M, Veerkamp RF, van Pelt ML, Mulder HA. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. J Dairy Sci. 2020;103(2):1667-1684. doi:10.3168/jds.2019-17290

12. Formoso-Rafferty N, Cervantes I, Ibáñez-Escriche N, Gutiérrez JP. Correlated genetic trends for production and welfare traits in a mouse population divergently selected for birth weight environmental variability. Animal. 2016;10(11):1770-1777. doi:10.1017/S1751731116000860.

13. Berghof TVL, Poppe M, Mulder HA. Opportunities to improve resilience in animal breeding programs. Front Genet. 2019;9:692. doi: 10.3389/fgene.2018.00692

14. López de Maturana E, Ibáñez-Escriche N, González-Recio Ó, et al. Next generation modeling in GWAS: comparing different genetic architectures. Hum Genet. 2014;133(10):1235-1253. doi:10.1007/s00439-014-1461-1

15. Gondro C, Werf JVD, Hayes, B. (2013) Genome-Wide Association Studies and Genomic Prediction, in: Methods in Molecular Biology. Media Springer Science+ Business LLC. doi:10.1007/978-1-62703-447-0

16. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1918;52:339-433.

17. Holland D, Frei O, Desikan R, et al. Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. PLoS Genet. 2020;16(5):e1008612. doi:10.1371/journal.pgen.1008612

18. McCartney DL, Min JL, Richmond RC, et al. Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. Genome Biol. 2021;22(1):194. doi:10.1186/s13059-021-02398-9

19. Hill WG, Mulder HA. Genetic analysis of environmental variation. Genet Res (Camb). 2010;92(5-6):381-395. doi:10.1017/S0016672310000546

20. Iung LHS, Carvalheiro R, Neves HHR, Mulder HA. Genetics and genomics of uniformity and resilience in livestock and aquaculture species: A review. J Anim Breed Genet. 2020;137(3):263-280. doi:10.1111/jbg.12454

21. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. Livest Sci. 2014;166:133-143. doi:10.1016/j.livsci.2014.05.003

22. Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different selection signature statistics and a new strategy for combining them. Heredity (Edinb). 2015;115(5):426-436. doi:10.1038/hdy.2015.42

23. González-Rodríguez A, Munilla S, Mouresan EF, et al. On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. Genet Sel Evol. 2016;48(1):81. doi:10.1186/s12711-016-0258-1

24. Sosa-Madrid BS, Varona L, Blasco A, et al. The effect of divergent selection for intramuscular fat on the domestic rabbit genome. Animal. 2020;14(11):2225-2235. doi:10.1017/S1751731120001263

25. Staubach F, Lorenc A, Messer PW, et al. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (Mus musculus). PLoS Genet. 2012;8(8):e1002891. doi:10.1371/journal.pgen.1002891

26. Igoshin AV, Yurchenko AA, Belonogova NM, et al. Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. *BMC Genet*. 2019;20(Suppl 1):26. doi:10.1186/s12863-019-0725-0

27. Kim ES, Ros-Freixedes R, Pena RN, et al, Rothschild MF. Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations. J Anim Sci. 2015;93(7):3292-302. doi:10.2527/jas.2015-8879

28. Lillie M, Sheng Z, Honaker CF, et al. Genome-wide standing variation facilitates long-term response to bidirectional selection for antibody response in chickens. BMC Genomics. 2017;18(1):99. doi:10.1186/s12864-016-3414-7

29. Johansson AM, Pettersson ME, Siegel PB, Carlborg O. Genome-wide effects of long-term divergent selection. PLoS Genet. 2010;6(11):e1001188. doi:10.1371/journal.pgen.1001188

30. Qanbari S. On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. Front Genet. 2020;10:1304. doi:10.3389/fgene.2019.01304

31. MacLeod IM, Hayes BJ, Goddard ME. The effects of demography and long-term selection on the accuracy of genomic prediction with sequence data. Genetics. 2014;198(4):1671-84. doi:10.1534/genetics.114.168344

32. Anand S, Mangano E, Barizzone N, et al. Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. Sci Rep. 2016;6:33735. doi:10.1038/srep33735

33. Beloumi D, Blasco A, Muelas R, et al. Inflammatory Correlated Response in Two Populations of Rabbit Selected Divergently for Litter Size Environmental Variability. Animals. 2020;10:1540. doi:10.3390/ani10091540

34. Colditz IG, Hine BC. Resilience in farm animals: biology, management, breeding and implications for animal welfare. Anim Prod Sci. 2016;56:1961–83. doi:10.1071/AN15297

35. Dantzer R, Cohen S, Russo SJ, Dinan TG. Resilience and immunity. Brain Behav Immun. 2018;74:28-42. doi:10.1016/j.bbi.2018.08.010

36. Sell-Kubiak E, Duijvesteijn N, Lopes MS, et al. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. BMC Genomics. 2015;16:1049. doi:10.1186/s12864-015-2273-y

37. Iung LHS, Mulder HA, Neves HHR et al. Genomic regions underlying uniformity of yearling weight in Nellore cattle evaluated under different response variables. BMC Genomics 2018;19:619. doi:10.1186/s12864-018-5003-4

38. Morgante F, Sørensen P, Sorensen DA, et al. Genetic Architecture of Micro-Environmental Plasticity in Drosophila melanogaster. Sci Rep. 2015;5:9785. doi:10.1038/srep09785

39. de Zoeten EF, Wang L, Sai H, et al. Inhibition of HDAC9 increases T regulatory cell function and prevents colitis in mice. Gastroenterology. 2010;138(2):583-94. doi:10.1053/j.gastro.2009.10.037

40. Gibson OR, Tuttle JA, Watt PW, et al. Hsp72 and Hsp90α mRNA transcription is characterised by large, sustained changes in core temperature during heat acclimation. Cell Stress Chaperones. 2016;21(6):1021-1035. doi:10.1007/s12192-016-0726-0

41. Nava R, Zuhl MN. Heat acclimation-induced intracellular HSP70 in humans: a meta-analysis. Cell Stress Chaperones. 2020;25(1):35-45. doi:10.1007/s12192-019-01059-y

42. Sareh H, Tulapurkar ME, Shah NG, et al. Response of mice to continuous 5-day passive hyperthermia resembles human heat acclimation. Cell Stress Chaperones. 2011;16(3):297-307. doi:10.1007/s12192-010-0240-8

43. Argente MJ, Calle EW, García ML, Blasco A. Correlated response in litter size components in rabbits selected for litter size variability. J Anim Breed Genet. 2017;134(6):505-511. doi:10.1111/jbg.12283

44. Calle EW, García ML, Blasco A, Argente MJ. Correlated response in early embryonic development in rabbits selected for litter size variability. World Rabbit Sci. 2017;25:323-7.

45. Chen Y, Meng F, Wang B, et al. Dock2 in the development of inflammation and cancer. Eur J Immunol. 2018;48(6):915-922. doi:10.1002/eji.201747157

46. Nishihara H. (2012) DOCK2; Dedicator of Cytokinesis 2. In: Choi, S. (eds) Encyclopedia of Signaling Molecules. Springer, New York, NY. doi:10.1007/978-1-4419-0461-4_518

47. Kikuchi T, Kubonishi S, Shibakura M, et al. Dock2 participates in bone marrow lympho-hematopoiesis. Biochem Biophys Res Commun. 2008;367(1):90-6. doi:10.1016/j.bbrc.2007.12.093

48. Enerbäck S, Nilsson D, Edwards N, et al. Acidosis and Deafness in Patients with Recessive Mutations in FOXI1. J Am Soc Nephrol. 2018;29(3):1041-1048. doi:10.1681/ASN.2017080840

49. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54:1.30.1-1.30.33. doi:10.1002/cpbi.5

50. Clark AJL, Chan L. Stability and Turnover of the ACTH Receptor Complex. *Front Endocrinol (Lausanne)*. 2019;10:491. doi:10.3389/fendo.2019.00491

51. Metherell, L., Chapple, J., Cooray, S. *et al.* Mutations in *MRAP*, encoding a new interacting partner of the ACTH receptor, cause familial glucocorticoid deficiency type 2. *Nat Genet*. 2005;**37**:166–170. doi:10.1038/ng1501

52. Goldsby RA, Kindt TJ, Osborne BA, Kuby J. (2003) Immunology, 5th ed. W.H Freemand and Company.

53. Cheruiyot EK, Haile-Mariam M, Cocks BG, et al. New loci and neuronal pathways for resilience to heat stress in cattle. Sci Rep. 2021;11(1):16619. doi:10.1038/s41598-021-95816-8.

54. Joy A, Dunshea FR, Leury BJ, et al. Resilience of Small Ruminants to Climate Change and Increased Environmental Temperature: A Review. Animals (Basel). 2020;10(5):867. doi:10.3390/ani10050867

55. Tsartsianidou V, Sánchez-Molano E, Kapsona VV et al. A comprehensive genome-wide scan detects genomic regions related to local adaptation and climate resilience in Mediterranean domestic sheep. Genet Sel Evol. 2021:53, 90. doi:10.1186/s12711-021-00682-7

56.Cheng J, Fernando R, Cheng H, et al. Genome-wide association study of disease resilience traits from a natural polymicrobial disease challenge model in pigs identifies the importance of the major histocompatibility complex region. G3 (Bethesda). 2022;12(3):jkab441. doi:10.1093/g3journal/jkab441

57. Laghouaouta H, Fraile L, Suárez-Mesa R, et al. A genome-wide screen for resilient responses in growing pigs. Genet Sel Evol. 2022;54(1):50. doi:10.1186/s12711-022-00739-1

58. Mosca A, Leclerc M, Hugot JP. Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? Front Microbiol. 2016;7:455. doi:10.3389/fmicb.2016.00455

59. Griffiths B, Ritz K, Bardgett R D, et al. Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity–ecosystem function relationship. Oikos. 2003;90:279–294. doi:10.1034/j.1600-0706.2000.900208.x

60. Allison SD, Martiny JB. Colloquium paper: resistance, resilience, and redundancy in microbial communities. Proc Natl Acad Sci U S A. 2008;105 Suppl 1(Suppl 1):11512-9. doi:10.1073/pnas.0801925105

61. Sommer F, Anderson JM, Bharti R, et al. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017;15(10):630-638. doi:10.1038/nrmicro.2017.58

62. Moya A, Ferrer M. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends Microbiol. 2016 May;24(5):402-413. doi:10.1016/j.tim.2016.02.002

63. Tian L, Wang XW, Wu AK, et al. Deciphering functional redundancy in the human microbiome. Nat Commun. 2020;11(1):6217. doi:10.1038/s41467-020-19940-1

64. Coelho GDP, Ayres LFA, Barreto DS, et al. Acquisition of microbiota according to the type of birth: an integrative review. Rev Lat Am Enfermagem. 2021;29:e3446. doi:10.1590/1518.8345.4466.3446

65. Martino C, Dilmore AH, Burcham ZM, et al. Microbiota succession throughout life from the cradle to the grave. Nat Rev Microbiol. 2022. doi:10.1038/s41579-022-00768-z

66. Liévin-Le Moal V, Servin AL. Anti-infective activities of lactobacillus strains in the human intestinal microbiota: from probiotics to gastrointestinal anti-infectious biotherapeutic agents. Clin Microbiol Rev. 2014;27(2):167-99. doi:10.1128/CMR.00080-13

67. Rutherford ST, Bassler BL. Bacterial quorum sensing: its role in virulence and possibilities for its control. Cold Spring Harb Perspect Med. 2012;2(11):a012427. doi:10.1101/cshperspect.a012427

68. de Kievit TR, Iglewski BH. Bacterial quorum sensing in pathogenic relationships. Infect Immun. 2000;68(9):4839-49. doi:10.1128/IAI.68.9.4839-4849.2000

69. Davey ME, O'toole GA. Microbial biofilms: from ecology to molecular genetics. Microbiol Mol Biol Rev. 2000;64(4):847-67. doi:10.1128/MMBR.64.4.847-867.2000

70. Watnick P, Kolter R. Biofilm, city of microbes. J Bacteriol. 2000;182(10):2675-9. doi:10.1128/JB.182.10.2675-2679.2000

71. Rode DKH, Singh PK, Drescher K. Multicellular and unicellular responses of microbial biofilms to stress. *Biol Chem*. 2020;401(12):1365-1374. doi:10.1515/hsz-2020-0213

72. Bian G, Ma S, Zhu Z, et al. Age, introduction of solid feed and weaning are more important determinants of gut bacterial succession in piglets than breed and nursing mother as revealed by a reciprocal cross-fostering model. Environ Microbiol. 2016;18(5):1566-77. doi:10.1111/1462-2920.13272

73. Mulder IE, Schmidt B, Stokes CR, et al. Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. BMC Biol. 2009;7:79. doi:10.1186/1741-7007-7-79

74. Xie Y, Chen J, Wu B, et al. Dock2 affects the host susceptibility to Citrobacter rodentium infection through regulating gut microbiota. Gut Pathog. 2021;14;13(1):52. doi:10.1186/s13099-021-00449-x.

75. Heeney DD, Gareau MG, Marco ML. Intestinal Lactobacillus in health and disease, a driver or just along for the ride? Curr Opin Biotechnol. 2018;49:140-147. doi:10.1016/j.copbio.2017.08.004

76. Hashimoto T, Perlot T, Rehman A, et al. ACE2 links amino acid malnutrition to microbial ecology and intestinal inflammation. *Nature*. 2012;487(7408):477-481. doi:10.1038/nature11228

77. Cao Y, Liu Y, Dong Q, et al. Alterations in the gut microbiome and metabolic profile in rats acclimated to high environmental temperature. Microb Biotechnol. 2022;15(1):276-288. doi:10.1111/1751-7915.13772

78. Zelante T, Iannitti RG, Cunha C, et al. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity*. 2013;39(2):372-385. doi:10.1016/j.immuni.2013.08.003

79. Puccetti P, Grohmann U. IDO and regulatory T cells: a role for reverse signalling and non-canonical NF-kappaB activation. Nat Rev Immunol. 2007;7(10):817-23. doi:10.1038/nri2163

80. Doney E, Cadoret A, Dion-Albert L, et al. Inflammation-driven brain and gut barrier dysfunction in stress and mood disorders. Eur J Neurosci. 2022;55(9-10):2851-2894. doi:10.1111/ejn.15239

81. Kennedy PJ, Cryan JF, Dinan TG, Clarke G. Kynurenine pathway metabolism and the microbiota-gut-brain axis. Neuropharmacology. 2017;112(Pt B):399-412. doi:10.1016/j.neuropharm.2016.07.002

82. Gao K, Mu CL, Farzi A, Zhu WY. Tryptophan Metabolism: A Link Between the Gut Microbiota and Brain. Adv Nutr. 2020;11(3):709-723. doi:10.1093/advances/nmz127

83. Bansal T, Alaniz RC, Wood TK, Jayaraman A. The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. Proc Natl Acad Sci U S A. 2010;107(1):228-233. doi:10.1073/pnas.0906112107

84. Li X, Zhang B, Hu Y, Zhao Y. New Insights Into Gut-Bacteria-Derived Indole and Its Derivatives in Intestinal and Liver Diseases. Front Pharmacol. 2021;12:769501. doi: 10.3389/fphar.2021.769501.

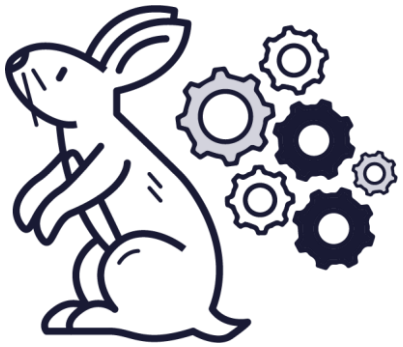
85. Smith SA, Ogawa SA, Chau L, et al. Mitochondrial dysfunction in inflammatory bowel disease alters intestinal epithelial metabolism of hepatic acylcarnitines. J Clin Invest. 2021;131(1):e133371. doi:10.1172/JCI133371

86. Liu Y, Hou Y, Wang G, et al. Gut Microbial Metabolites of Aromatic Amino Acids as Signals in Host-Microbe Interplay. Trends Endocrinol Metab. 2020;31(11):818-834. doi:10.1016/j.tem.2020.02.012

87. Caussy C, Hsu C, Lo MT, et al. Genetics of NAFLD in Twins Consortium. Link between gut-microbiome derived metabolite and shared gene-effects with hepatic steatosis and fibrosis in NAFLD. Hepatology. 2018;68(3):918-932. doi:10.1002/hep.29892

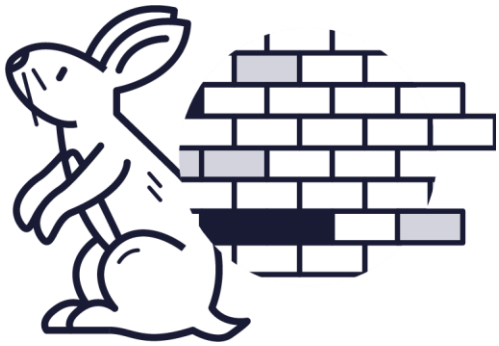
CHAPTER 9: GENERAL CONCLUSIONS

This is the first thesis to unravel the biological mechanisms underlying animal resilience using the V_E of LS. In this thesis, we found relevant genes, taxa, and gut metabolites involved in controlling stress and immune response. The results could explain the differences in resilience potential between the rabbit populations, which were generated by selection for V_E of LS. Moreover, relevant genes for the development of the neural system and sensory receptors such as olfactory receptors were identified, which is also important to determine animal resilience. This thesis suggested that the microbiome composition changed due to (i) an effect of the host genome and (ii) the selection of relevant microbial species for the V_E. The candidate genes *DOCK2, ACE,* and *HDAC9* could modify the gut microbiome composition. Beneficial species such as *Rikenella*, *Lactobacillus*, *Alistipes prutedinis*, *Alistipes shahii*, *Odoribacter splanchnicus,* and *Limosilactobacillus fermentum* were selected in the resilient lines, increasing their abundance. In contrast, harmful species such as *Acetatifactor muris* and *Eggerthella sp*. were decreased. Furthermore, we found overlapping results between all the analyses performed that suggest a relevant implication of the aromatic amino acid metabolism and the L-glutamate synthesis in the phenotype of the V_E. These metabolisms could be influencing the susceptibility of animals to stress conditions which supported the differences in the resilience potential between the rabbit lines. Further studies are needed to validate all these results and determine their real implications in V_E and animal resilience. Moreover, this is the first thesis developing a tool (simuGMsel) for simulating a selection process with the inheritance of both the genome and the microbiome. These results agree with those found in the microbial study highlighting the importance of microbial interactions and microbial heritability for shaping the V_E and modulating animal resilience. simuGMsel enables simulation of genome and microbiome to study their coevolution and effect on phenotype in animal breeding programs.



CHAPTER 10: IMPLICATIONS

The real implication of this thesis is the contribution to generate knowledge about the mechanisms underlying animal resilience, highlighting the importance of the immune system. This is the first step to develop strategies and future works to improve animal resilience. Genes and causal variants identified could be used for genomic selection, as well as the bacteria identified for a selection based on the microbiome. A selection combining genomic and metagenomic information can improve the selection response, as we showed in the results generated by simulation. Moreover, gut-derived metabolites with evidence of crosstalk can be used as biomarkers to identify resilient animals in plasma, avoiding the extraction of faecal samples to determine the microbiome composition. However, further studies are necessary to validate their presence in the animals' plasma. In the future, gene editing, faecal microbial transfer, and probiotics could be developed to test their effects on animal resilience following the knowledge starting in this thesis. If success is obtained, animal resilience could be improved following these strategies with the aim of search a more sustainable livestock system. Lastly, the simulation tool developed could help to study microbiome composition at scale to decipher key animal breeding questions.



CHAPTER 11: REMARKABLE LIMITATIONS

i. Sample size

For genomic analyses, the moderate sample size used is sufficient to detect the most relevant genomic regions for the V_E. However, due to the low heritability of the V_E, a large sample size should be used to have good statistical power and better estimate the SNP effect. For metagenomic, the cost of this technique complicates the use of a considerable number of animals. This study was based on category assignation (at the population level), so a low sample size could be used. Proper sample size will allow for separating the true effect from the environmental noise, reducing the false positives and negatives in the analyses.

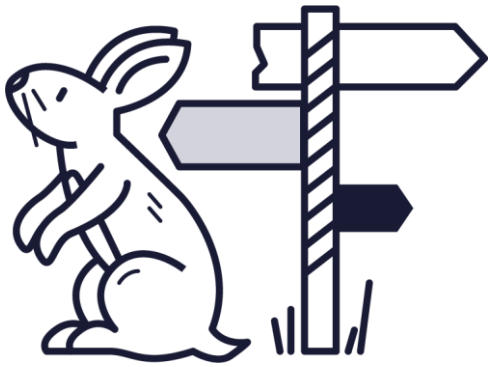
ii. Genome annotation

Oryctolagus cuniculus genome is poorly annotated. Recent updates to the Ensembl database considerably changed the genes harbouring the relevant genomic regions. The Ensembl Version from *Chapter 2* and *Chapter 3* changed markedly. Essential genes such as *DOCK2* were reviewed using the genome version from *Chapter 3*. An ID changed its name from Ensembl, but its description determines that this gene is orthologous to the *DOCK2* gene. The information did not change in this case, but all the other genomic regions must be reviewed.

iii. Inferences

Effects of the genes, microbiome, and metabolites were based on functions reported in the literature. So, no direct effects were validated for all the inferences and hypotheses made in this project. Moreover, the selection trait was the V_E, and we assumed that the impact of the candidate genes and the microbiota must also influence the animal resilience. Experimental analyses must be needed to validate all the hypotheses made in this project. Furthermore, we could lose promising genes or microbiota because of their

effects on the V_E and resilience have not been described previously in the literature. There is a lack of information. For instance, the microbiome is an emerging field, and there are no previous VE studies. We based all our inferences on the implications of the gut microbiome on the health status of the individual.



CHAPTER 12: FURTHER PERSPECTIVES

i. Transcriptomic analysis

Transcriptomic analysis for V_E is complicated because we do not know the tissue to analyse. Transcriptomic analyses of the epithelial gut barrier could be a good approximation due to its direct implications for gut immunity and microbiome composition. Moreover, some candidate genes such as *HDAC9* were found to be a protective effect on the gut due to the expression of HSP, a protein related to the V_E in several studies.

ii. Target metabolites in the gut

Target metabolites could help validate the project's results and the implications of the tryptophan metabolism for the V_E. Moreover, we could not identify other candidate metabolites in this project, such as SCFAs, which could be targeted with a higher-performance technique.

iii. Genomic and metagenomic interaction

Analyses to establish the link between the genome and microbiome must be needed, but a large sample size is necessary. Likewise, the taxa and gut-derived metabolite relationship must be studied.

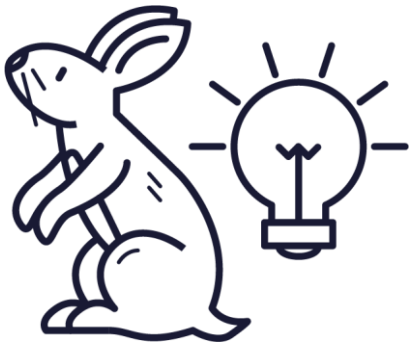
iv. Untargeted metabolite from blood samples

This must be useful to know the bioavailability of gut-derived metabolites in the plasma, which could be an indicator for gut-host crosstalk. Moreover, we can identify other molecules directly influencing V_E and resilience.

v. Update the simulation tool

Made a new version with the implementation of a microbial gene model to study the functional redundancy of the microbiome. Made a version to implement a genomic and microbiome selection.

All of them are proposals for further studies that could complement and validate the results of this project. Any of them, except point 5, will be really developed due to technical issues.



CHAPTER 13: CONSIDERATIONS

V_E was correlated negatively with health and reproductive traits and positively with performance traits such as average milk yield and litter size. So, including V_E as a selection criterion for improving animal resilience must be carefully evaluated. Other studies are necessaries to know the relation between the mean and the variance of the trait used. Moreover, if the improvement of animal resilience involves a detriment of key traits, a balance must be found to improve animal resilience without drastically worsening production.

This study was supported by:

Projects AGL2014-5592, C2-1-P and C2-2-P, and AGL2017-86083, C2-1-P and C2-2-P, funded by the Spanish *Ministerio de Ciencia e Innovación (MIC)-Agencia Estatal de Investigación* (AEI) and the European Regional Development Fund (FEDER).

Projects PID2020-115558GB-C21, funded by the Spanish *Ministerio de Ciencia e Innovación (MIC)-Agencia Estatal de Investigación* (AEI) and the European Regional Development Funds (FEDER)

FPU17/01196 scholarship from the Spanish Ministry of Science, Innovation and Universities.