Universitat Autònoma de Barcelona

Exploring the molecular basis of dairy production and viability traits in Murciano-Granadina goats with genomic, metagenomic and transcriptomic approaches

Maria Gracia Luigi Sierra

EXPLORING THE MOLECULAR BASIS OF DAIRY PRODUCTION AND VIABILITY TRAITS IN MURCIANO-GRANADINA GOATS WITH GENOMIC, METAGENOMIC AND TRANSCRIPTOMIC APPROACHES

DOCTORAL THESIS TO OBTAIN THE PHD DEGREE IN ANIMAL PRODUCTION OF THE UNIVERSITAT AUTÒNOMA DE BARCELONA, SEPTEMBER 2022

by

Maria Gracia Luigi Sierra

SUPERVISOR

Marcel Amills Eras

El Dr. Marcel Amills Eras, professor agregat del Departament de Ciència Animal i dels Aliments de la Universitat Autònoma de Barcelona,

fa constar:

que el treball de recerca i la redacció de la memòria de la tesi doctoral titulada:

"Exploring the molecular basis of dairy production and fitness traits in Murciano-Granadina goats with genomic, metagenomic and transcriptomic approaches"

han estat realitzats sota la seva direcció per

Maria Gracia Luigi Sierra

i certifica:

que aquest treball s'ha dut a terme al Departament de Ciència Animal i del Aliments de la Facultat de Veterinària de la Universitat Autònoma de Barcelona i al Departament de Genètica Animal del Centre de Recerca en Agrigenòmica (CRAG), i es considera que la memòria resultant és apta per optar al grau de Doctor en Producció Animal per la Universitat Autònoma de Barcelona.

I perquè en quedi constància, signen aquest document a Setembre del 2022.

Dr. Marcel Amills Maria Gracia Luigi Sierra

This research was funded by the European Regional Development Fund (FEDER)/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación/Project Reference grants: AGL2016-76108-R and also by project PID2019-105805RB-I00 funded by MCIN/AEI/10.13039/501100011033. We also acknowledge the support of the CERCA programme of the Generalitat de Catalunya and the Center of Excellence Severo Ochoa 2020–2023 (CEX2019–000902-S) grant funded by MCIN/AEI /10.13039/501100011033 and awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain).

Maria Luigi-Sierra was funded with a PhD fellowship Formación de Personal Investigador BES-C-2017-079709 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future".

CONTENTS

SUMMARY

In this thesis, we aimed to investigate the genetic basis of milk and morphological traits in Murciano-Granadina goats as well as to explore the microbial composition of several types of caprine cheeses. Moreover, we wanted to assess the consequences of inbreeding on dairy performance and to study the molecular basis of traits related with the biological viability of goats. In **study 1**, we have identified several genetic determinants of milk yield and composition traits in Murciano-Granadina goats with records for three lactations by performing three independent GWAS (one for each lactation) and a longitudinal GWAS. Both approaches consistently revealed a genome-wide significant QTL for protein percentage on a chromosome 6 (74.8 – 94.6 Mb) region which harbours the four casein genes. Additionally, a QTL on chromosome 2 (129.77 – 131.01 Mb) for lactose percentage was also consistently detected. More QTL were identified in the longitudinal GWAS than in the three independent GWAS, possibly because the former has far more statistical power than the latter. In **study 2**, we also used a GWAS approach to identify genomic regions associated with the variation of morphological traits in Murciano-Granadina goats. In this analysis, few significant associations were found and positional coincidence with morphology QTL detected in other breeds was low. These results are compatible with a highly polygenic determinism of morphological traits in goats. Since dairy production in goats is focused to the elaboration of cheese and the technological and organoleptic properties of this food are largely determined by its microbiota, in **study 3**, we have characterized the microbiota of six different goat cheeses by sequencing hyper-variable regions of the bacterial 16S rRNA gene. We have identified lactic acid bacteria as the

predominant microbial community in all six cheeses. Besides, we have detected an abundance of psychrophilic bacteria, which are common post-pasteurisation contaminants of milk, in fresh cheeses.

With regard to viability traits, in **study 4**, we detected low levels of inbreeding in the Murciano-Granadina population under study (media $F_{ROH} = 0.053 \pm 0.04$). Despite this, a significant inbreeding depression on the logarithm of the somatic cell score (lnSCC), an important trait for udder health and milk quality, was detected and its determinants were finely mapped to a few chromosome 8 and 25 regions containing immunity genes. To ascertain genomic regions harbouring potentially harmful mutations (that are not expected to show a Mendelian segregation in the offspring of carrier individuals), in **study 5** we carried out a scan to detected transmission ratio distortion (TRD) in Murciano-Granadina goats at a genome-wide scale. From the 36 SNPs displaying significant TRD, 25 had low GenTrain scores (< 0.8) indicating poor assignment of the genotypes. This result implies that most of the TRD signals detected by us are artifacts attributable to defective genotyping and stress the importance of filtering SNPs according to their GenTrain scores when carrying out TRD scans. As maternal care has an important influence on offspring viability, in **study 6**, we characterized the mRNA expression profiles of 12 brain tissues in 7 goats, of which 3 were 1 month-pregnant. Principal component analysis of the data revealed that most tissues tend to cluster according to their neural vesicle of origin, suggesting that embryonic development leaves a durable footprint that affects gene expression in the adult goat. Exceptions to this trend were cerebellum and glandular tissues, possibly because of their highly specialized functions. Important gene expression changes associated with 1 month gestation were observed in the adenohypophysis, frontal neocortex, hippocampus, pineal gland, pons and particularly in the olfactory bulb. Interestingly, many of the genes

differentially expressed in the olfactory bulb are related to human behavior, and there is evidence that this anatomical structure has a key role in the development of maternal care.

RESUM

En aquesta tesi, hem volgut investigar les bases genètiques dels caràcters lleters i morfològics de les cabres de la raça Murciano-Granadina, així com explorar la composició microbiana de diversos tipus de formatges caprins. A més, hem volgut avaluar les conseqüències de la consanguinitat sobre la producció i la composició de la llet de cabra i estudiar la base molecular de diversos caràcters relacionats amb la viabilitat biològica. En l'**estudi 1**, hem identificat diversos marcadors genètics associats a la producció i composició de la llet en la raça Murciano-Granadina emprant registres de tres lactacions i duent a terme tres GWAS independents (un per cada lactació) i un GWAS longitudinal. Totes dues aproximacions van revelar un QTL significatiu per al percentatge de proteïnes en una regió del cromosoma 6 (74, 8 – 94, 6 Mb) que alberga els quatre gens de les caseïnes. A més, també es va detectar un QTL al cromosoma 2 (129,77 – 131,01 Mb) per al percentatge de la lactosa. Es van identificar més QTL al GWAS longitudinal que als tres GWAS independents, possiblement perquè el GWAS longitudinal té molt més poder estadístic. A l'**estudi 2**, també hem utilitzat una aproximació GWAS per identificar regions genòmiques associades a la variació de caràcters morfològics en cabres de la raça Murciano-Granadina. En aquesta anàlisi, es van trobar poques associacions significatives i la coincidència posicional amb QTL morfològics detectats en altres races va ser baixa. Aquests resultats són compatibles amb un determinisme altament poligènic dels caràcters morfològics en cabrum. Atès que la producció de llet en cabra està enfocada a l'elaboració de formatges i les propietats tecnològiques i organolèptiques d'aquest aliment estan en gran part determinades per la seva microbiota, a l'**estudi 3**, hem caracteritzat la microbiota de sis formatges de cabra mitjançant la seqüenciació de regions hipervariables del gen bacterià 16S ARNr. Els bacteris acido-làctics són els més predominants en els sis formatges. A més,

en els formatges frescos hem detectat una certa abundància de bacteris psicròfils, que són contaminants habituals de la llet després de la pasteurització,.

Pel que fa als trets de viabilitat, a l'**estudi 4** s'han detectat nivells baixos d'endogàmia a la població Murciano-Granadina objecte d'estudi (mitjana $F_{ROH} = 0.053 \pm 0.04$). Malgrat això, es va detectar una depressió endogàmica significativa pel logaritme del recompte de cèl·lules somàtiques (lnSCC), un tret important per a la salut de la glàndula mamària i la qualitat de la llet, i es va mapejar finament els determinants d'aquesta depressió a diverses regions dels cromosomes 8 i 25 que contenen gens immunitaris. Per determinar les regions genòmiques que contenen mutacions potencialment adverses sobre la viabilitat (i que per tant, s'espera que no mostrin una segregació Mendeliana en la descendència d'individus portadors), a l'**estudi 5** vam investigar l'existència de distorsions de la ratio de transmissió al·lèlica (TRD) en cabres Murciano-Granadines. Dels 36 SNP que mostraven una TRD significativa, 25 tenien puntuacions GenTrain baixes (< 0, 8), la qual cosa indica una mala assignació dels genotips. Aquest resultat implica que la majoria dels senyals TRD detectats en el nostre estudi són artefactes atribuïbles a un genotipat defectuós i subratllen la importància de filtrar els SNP segons les seves puntuacions GenTrain quan es realitzen anàlisis TRD. Com que el comportament matern té una influència important en la viabilitat de la descendència, a l'**estudi 6**, vam caracteritzar els perfils d'expressió d'ARNm de 12 teixits cerebrals en 7 cabres, de les quals 3 estaven prenyades d'un mes. L'anàlisi de components principals va revelar que la majoria dels teixits tendeixen a agrupar-se segons la seva vesícula neural d'origen, cosa que suggereix que el desenvolupament embrionari deixa una empremta duradora que afecta l'expressió gènica a la cabra adulta. El cerebel i els teixits glandulars constitueixen excepcions a aquesta tendència, possiblement a causa de les seves funcions altament especialitzades. Igualment, es van observar notables canvis d'expressió gènica associats a la gestació a l'adenohipòfisi, neocòrtex frontal, hipocamp,

glàndula pineal, pont i, particularment, al bulb olfactiu. Curiosament, molts dels gens expressats diferencialment al bulb olfactiu estan relacionats amb el comportament humà, i hi ha proves que aquesta estructura anatòmica té un paper clau en el desenvolupament del comportament matern.

RESUMEN

En esta tesis, nuestro objetivo fue investigar la base genética de caracteres de producción de leche y morfológicos en cabras Murciano-Granadina, así como explorar la composición microbiana de varios tipos de quesos caprinos. Además, queríamos evaluar las consecuencias de la consanguinidad en el rendimiento lechero y estudiar la base molecular de los caracteres relacionados con la viabilidad biológica de las cabras. En el **estudio 1** hemos identificado varios determinantes genéticos de la producción y composición de la leche en cabras Murciano-Granadina con registros de tres lactancias mediante la realización de tres GWAS independientes (uno para cada lactancia) y un GWAS longitudinal. Ambos enfoques revelaron consistentemente un QTL significativo en todo el genoma para el porcentaje de proteína en leche en una región del cromosoma 6 (74,8 - 94,6 Mb) que alberga los cuatro genes de la caseína. Además, también se detectó consistentemente un QTL en el cromosoma 2 (129,77 – 131,01 Mb) para el porcentaje de lactosa. Se identificaron más QTL en el GWAS longitudinal que en los tres GWAS independientes, posiblemente porque el primero tiene mucho más poder estadístico que el segundo. En **el estudio 2**, también utilizamos un enfoque GWAS para identificar regiones genómicas asociadas con la variación de caracteres morfológicos en cabras Murciano-Granadina. En este análisis, se encontraron pocas asociaciones significativas y la coincidencia posicional con la morfología QTL detectada en otras razas fue baja. Estos resultados son compatibles con un determinismo altamente poligénico de los rasgos morfológicos en cabras. Dado que la producción láctea en cabras está enfocada a la elaboración de queso y que las propiedades tecnológicas y organolépticas de este alimento están determinadas en gran medida por su microbiota, en el **estudio 3** hemos caracterizado la microbiota de seis quesos de cabra diferentes mediante la secuenciación de regiones hipervariables de la gen bacteriano 16S

rRNA. Hemos identificado bacterias del ácido láctico como la comunidad microbiana predominante en los seis quesos. Además, hemos detectado una gran cantidad de bacterias psicrófilas, que son contaminantes comunes de la leche después de la pasteurización, en los quesos frescos.

En cuanto a los rasgos de viabilidad, en el **estudio 4** detectamos bajos niveles de consanguinidad en la población murciano-granadina objeto de estudio (F_{ROH} medio = 0,053 ± 0,04). A pesar de esto, se detectó una depresión consanguínea significativa en el logaritmo de la puntuación de células somáticas (lnSCC), un rasgo importante para la salud de la ubre y la calidad de la leche, y sus determinantes se mapearon finamente en algunas regiones del cromosoma 8 y 25 que contienen genes de inmunidad. Para determinar las regiones genómicas que albergan mutaciones potencialmente dañinas (que no se espera que muestren una segregación mendeliana en la descendencia de individuos portadores), en el **estudio 5** llevamos a cabo una exploración para detectar la distorsión de la relación de transmisión (TRD) en cabras Murciano-Granadina. De los 36 SNP que mostraban un TRD significativo, 25 tenían puntuaciones GenTrain bajas (< 0,8), lo que indicaba una asignación deficiente de los genotipos. Este resultado implica que la mayoría de las señales TRD detectadas por nosotros son artefactos atribuibles a un genotipado defectuoso y enfatizan la importancia de filtrar los SNP de acuerdo con sus puntajes GenTrain al realizar escaneos TRD. Dado que el cuidado materno tiene una influencia importante en la viabilidad de las crías, en el **estudio 6** caracterizamos los perfiles de expresión de ARNm de 12 tejidos cerebrales en 7 cabras, de las cuales 3 tenían 1 mes de embarazo. El análisis de componentes principales de los datos reveló que la mayoría de los tejidos tienden a agruparse según su vesícula neural de origen, lo que sugiere que el desarrollo embrionario deja una huella duradera que afecta la expresión génica en la cabra adulta. Las excepciones a esta tendencia fueron el cerebelo y los tejidos glandulares, posiblemente debido a sus

funciones altamente especializadas. Se observaron importantes cambios en la expresión génica asociados con 1 mes de gestación en la adenohipófisis, el neocórtex frontal, el hipocampo, la glándula pineal, el puente y particularmente en el bulbo olfatorio. Curiosamente, muchos de los genes expresados diferencialmente en el bulbo olfatorio están relacionados con el comportamiento humano y existe evidencia de que esta estructura anatómica tiene un papel clave en el desarrollo del cuidado materno.

LIST OF PUBLICATIONS

The present tesis is based on the work contained in the following articles:

Paper I: **Luigi-Sierra, MG**., Martínez, A., Macri, M., Delgado, JV., Castelló, A., Fernádez Alvarez, J., Such, X., Jordana, J., Amills., M., 2022. Single and longitudinal genome-wide association studies for dairy traits available in goats with three recorded lactations. (in preparation).

Paper II: **Luigi-Sierra, MG**., Landi, V., Guan, D., Delgado, JV., Castelló, A., Cabrera, B., Mármol-Sánchez, E., Fernádez Alvarez, J., Gómez-Carpio, M., Martínez, A., Such, X., Jordana, J., Amills, M, 2020. A genome-wide association analysis for body, udder and leg conformation traits recorded in Murciano-Granadina goats. J Dairy Sci. 103, 11605–11617. doi: 10.3168/jds.2020-18461.

Paper III: **Luigi-Sierra, MG**., Ramayo-Caldas, Y., Guan, D., Amills, M., 2022. Microbiome composition of six commercial goat cheeses based on targeted 16S rRNA sequencing. Food Microbiol. (submitted).

Paper IV: **Luigi-Sierra, MG**., Fernández, A., Martínez, A., Guan, D., Delgado, JV., Fernádez Alvarez, J., Landi, V., Such, X., Jordana, J., Saura, M., Amills., M, 2022. Genomic patterns of homozygosity and inbreeding depression in Murciano-Granadina goats. J Anim Sci Biotechnol. 13, 35. doi: 10.1186/s40104-022-00684-5.

Paper V: **Luigi-Sierra, MG**., Casellas, J., Martínez, A., Delgado, JV., Fernádez Alvarez, J., Such, X., Jordana, J., Amills, M., 2021. Impact of SNP calling quality

on the detection of transmission ratio distortion in goats. bioRxiv. doi: 10.1101/2021.06.09.447792.

Paper VI: **Luigi-Sierra, MG**., Guan, D., López-Béjar, M., Casas, E., Olvera, S., Gardela, J., Palomo, MJ., Osuagwuh, UI., Ohaneje, UI., Mármol-Sánchez, E., Amills, M., 2022. A protein-coding gene expression atlas from the brain of pregnant and non-pregnant goats. (in preparation).

LIST OF PUBLICATIONS NOT INCLUDED IN THIS THESIS

Mármol-Sánchez, E., **Luigi-Sierra, M. G.**, Castelló, A., Guan, D., Quintanilla, R., Tonda, R., Amills, M., 2021). Variability in porcine microRNA genes and its association with mRNA expression and lipid phenotypes. Genetics Selection Evolution, 53(1). https://doi.org/10.1186/s12711-021-00632-3

Cardoso, T. F., **Luigi-Sierra, M. G.**, Castelló, A., Cabrera, B., Noce, A., Mármol-Sánchez, E., García-González, R., Fernádez-Arias, A., Alabart, JL., López-Olvera, JR., Mentaberre, G., Granados-Torres, JE., Cardells-Peris, J., Molina, A., Sànchez, A., Clop, A., Amills, M., 2021. Assessing the levels of intraspecific admixture and interspecific hybridization in Iberian wild goats (*Capra pyrenaica*). Evol Appl. 14, 2618–2634. https://doi.org/10.1111/eva.13299

Guan, D., Martinez, A., **Luigi-Sierra, M. G.**, Delgado, J. V, Landi, V., Castelló, A., Fernández-Alvarez, J., Such, X., Jordana, J., Amills, M., 2021. Detecting the footprint of selection on the genomes of Murciano-Granadina goats. Anim Genet. 52, 683–693. https://doi.org/10.1111/age.13113

Rovelli, G., **Luigi-Sierra, M. G.,** Guan, D., Sbarra, F., Quaglia, A., Sarti, F. M., Amills, M., Lasagna, E., 2021. Evolution of inbreeding: a gaze into five Italian beef cattle breeds history. PeerJ, 9, e12049. https://doi.org/10.7717/peerj.12049

Luigi-Sierra, M. G., Mármol-Sánchez, E., & Amills, M., 2020. Comparing the diversity of the casein genes in the Asian mouflon and domestic sheep. Anim Genet, 51, 470–475. https://doi.org/10.1111/age.12937

Guan, D., Landi, V., **Luigi-Sierra, M. G.**, Delgado, J. V., Such, X., Castelló, A., Cabrera, B., Mármol-Sánchez, E., Fernádez Alvarez, J., Ruiza de la Torre Casañas, J.L., Martínez, A., Jordana, J.,Amills, M., 2020. Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. J Anim Sci Biotechnol, 11, 35. https://doi.org/10.1186/s40104-020-00435-4

Guan, D., Martinez, A., Castelló, A., Landi, V., **Luigi-Sierra, M. G.**, Fernández-Álvarez, J., Cabrera, B., Delgado, J. V., Such, X., Jordana, J., Amills, M., 2020. A genome-wide analysis of copy number variation in Murciano-Granadina goats. Genet Sel Evol. 52, 44. https://doi.org/10.1186/s12711-020-00564-4

Bâlteanu, V. A., Cardoso, T. F., Amills, M., **Luigi-Sierra, M. G.**, Egerszegi, I., Anton, I., Zsolnai, A., 2020. Red and blond Mangalitza pigs display a signature of divergent directional selection in the *SLC45A2* gene. Anim Genet. 52, 66–77. https://doi.org/10.1111/age.13031

Mármol-Sánchez, E., **Luigi-Sierra, M. G.**, Quintanilla, R., Amills, M., 2019. Detection of homozygous genotypes for a putatively lethal recessive mutation in the porcine argininosuccinate synthase 1 ASS1 gene. Anim Genet 51,106-110. https://doi.org/10.1111/age.12877

1 INTRODUCTION

1.1 AN INTRODUCTION TO GOAT BREEDING IN EUROPE.

1.1.1 Goat production in Europe and Spain

Goats were domesticated about 10,000 years ago in multiple locations in the Fertile Crescent (Zeder 2008, Daly et al. 2018) and subsequently dispersed across Europe by following two main routes: the Mediterranean corridor, through which several waves of seafaring colonists established coastal farming enclaves scattered across the Mediterranean coast, and the Danubian corridor that traversed central Europe until reaching the British Isles and Scandinavia (Zeder, 2008; Porter et al., 2016). Initially, goats and other ruminants were bred for their meat, but soon milk, and to a less extent fermented products, became important food resources due to their high nutritional quality and the possibility of obtaining them without the need of slaughtering (Evershed et al., 2008). In Europe, evidence of milk consumption and dairy food production, obtained from pottery with milk residuals and dental calculi in humans, dates back to 7,500 years ago (Salque et al., 2013; McClure et al., 2018; Charlton et al., 2019)

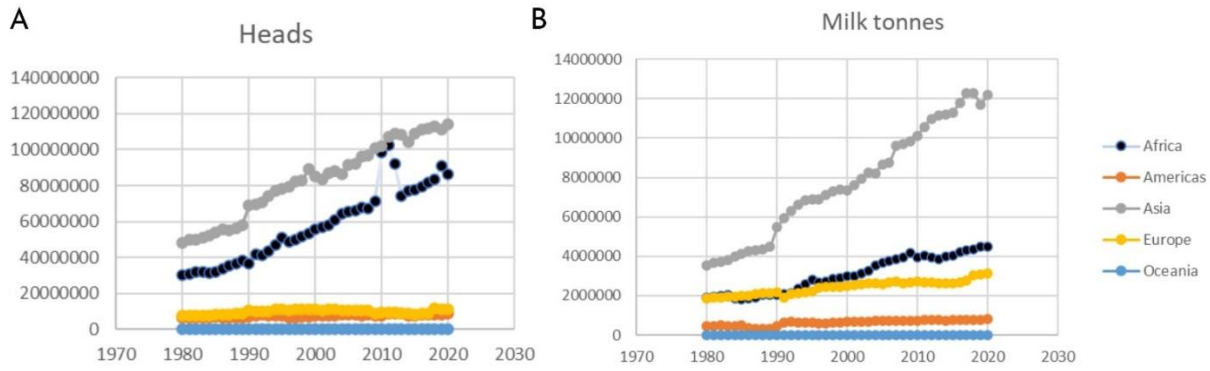
Traditionally, goats have been bred as marginal livestock in subsistence agricultural schemes due to their high rusticity and excellent adaptation to harsh environments (de Asís Ruiz Morales et al., 2019). During the past 20 years, goat census has remained quite stable in Europe while milk production has experienced a moderate increase (**Figure 1.1**), whereas in Africa and, above all, in Asia the number of heads and total milk production have increased dramatically (de Asís Ruiz Morales et al., 2019). Unlike cattle, the majority of goat milk is dedicated to the production of cheese and other (mainly fermented) products, except in a few countries, like China, that produce large amounts of powdered milk from goats. The United States of America and other western countries

commercialise goat milk for human consumption since the digestibility of caprine milk is better than that of cows (Pulina et al., 2018).

As depicted in **Figure 1.1**, Europe harbours 5.17% of the global goat census and produces over 15% and 35% of goat milk and cheese worldwide, respectively (FAOSTAT). Indeed, European countries have established increasingly intensive goat breeding programs that have substantially augmented the efficiency of the production system compared to other regions of the world (de Asís Ruiz Morales et al., 2019). Countries like France, The Netherlands and Spain have improved their goat breeding practices by incorporating technologies, like artificial insemination, computerised recording, optimised nutrition plans and efficient genetic selection, which have dramatically improved productivity per animal, mainly in terms of milk yield (de Asís Ruiz Morales et al., 2019). Noteworthy, European goats have an average milk production of 290 L/head per lactation, which is considerably higher than the 48.9, 76.2 and 93.4 L/head produced in Africa, Asia and America, respectively (Pulina et al., 2018). Spain occupies the second place in the ranking of goat milk production in Europe, right after France, with a mean of 500.000 tons of milk produced per year during the past decade (FAOSTAT).

There is a great diversity of goat breeds in Europe compared to other regions of the world (Cañón et al., 2006). Even more, several breeds of European origin with extensive census and popularity, like Alpine and Saanen, have become cosmopolitan due to their excellent dairy performance (Dubeuf and Boyazoglu, 2009). From the 1.8 million goats bred in Spain, over 70% are dairy goats, and around 6% belong to the Murciano-Granadina breed (MAPA).

Figure 1.1. Evolution of the (A) census and (B) total milk production (measured in tonnes) in Africa, America, Asia, Europe and Oceania from the late '70s until 2021. Information was retrieved from FAOSTAT.



1.2 THE MURCIANO-GRANADINA BREED: AN OVERVIEW.

The Murciano-Granadina breed, with a census of 119,354 heads and 194 farms registered in the herd book in Spain (and over 500,000 individuals considering those not included in the herd book), is the most popular goat breed in the country (MAPA, https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-razas/caprino/murciano-granadina/iframe-ejemplo-arca.aspx). Although the majority of the farms are located in Andalusia (40.21 %), Murcia (18.04 %) and Castilla la Mancha (9.79 %), the Murciano-Granadina breed is distributed in most of Spain (MAPA). Moreover, its high capacity of adaptation to arid and dry environments with an extreme climate and its excellent milk production performance (586 kg milk per lactation) have increased the interest in this breed nationally and worldwide (Delgado et al., 2018). By these reasons, Murciano-Granadina might be considered as a cosmopolitan breed increasingly used in countries such as Portugal and France (MAPA).

1.2.1 Origins of the Murciano-Granadina breed.

This breed was officially established in 1975 by admixing two different Murciano and Granadina goat populations within the framework of a government program that aimed to implement a new breeding program (Delgado et al., 2018). After the creation of the herd book in 1979, a genetic-functional scheme for evaluating bucks was successfully developed (Delgado et al., 2018). In 1999, two different associations approved and implemented the official selection scheme of sires (CAPRIGRAN and ACRIMUR). Finally, in 2011, these two associations were merged into one single federation (MURCIGRAN) to control and monitor the genetic progress more effectively and optimise the use of the available genetic resources (Delgado et al., 2018).

1.2.2 The breeding program of the Murciano-Granadina breed.

The Murciano-Granadina breed is subjected to one of the most innovative and well developed caprine breeding programs in Spain (Delgado et al., 2018). The selection criteria aim to improve productivity in terms of milk production, body morphology, and profitability (https://www.mapa.gob.es/en/ganaderia/temas/zootecnia/report_resolucionaprobacionprogramadecriamurciano-granadina_tcm38-579711.pdf). In 2021, longevity was included as a selection goal in the breeding program. The three main selection goals and criteria that are taken into account in the Murciano-Granadina breed program are described below:

Goal 1. To improve milk production in terms of quantity and quality. To do so, milk controls are carried out in the registered females by measuring milk yield and quality at official laboratories. Additionally, sires are systematically genotyped to ascertain if they

carry favourable alleles for the casein genes. The selection criteria taken into consideration are:

- Total milk yield at 210 and 240 days of lactation.
- Protein content at 210 and 240 days of lactation.
- Fat content at 210 and 240 days of lactation.
- Casein genotype.

Goal 2. Improvement of body conformation ("Dairy type"). Morphological traits are categorised according to a linear score ranging from 1 to 9, and intermediate scores are usually optimal. The specific selection criteria are:

- Linear classification of body structure and capacity.
- Linear classification of mammary system.
- Linear classification of feet.
- Linear classification of "dairy type".

Goal 3. Improvement of the longevity of sires and dams by taking into account the following criteria:

- Lifetime milk production.
- Lifetime protein production.
- Lifetime fat production.
- Productive longevity.

The breeding program, which is applied in several steps (**Figure 1.2**), only takes place at the selection nucleus, which is integrated by herds that apply artificial insemination (AI) with semen from males with genetic superiority. The several phases of the breeding program are described below and summarised in **Figure 1.2**.

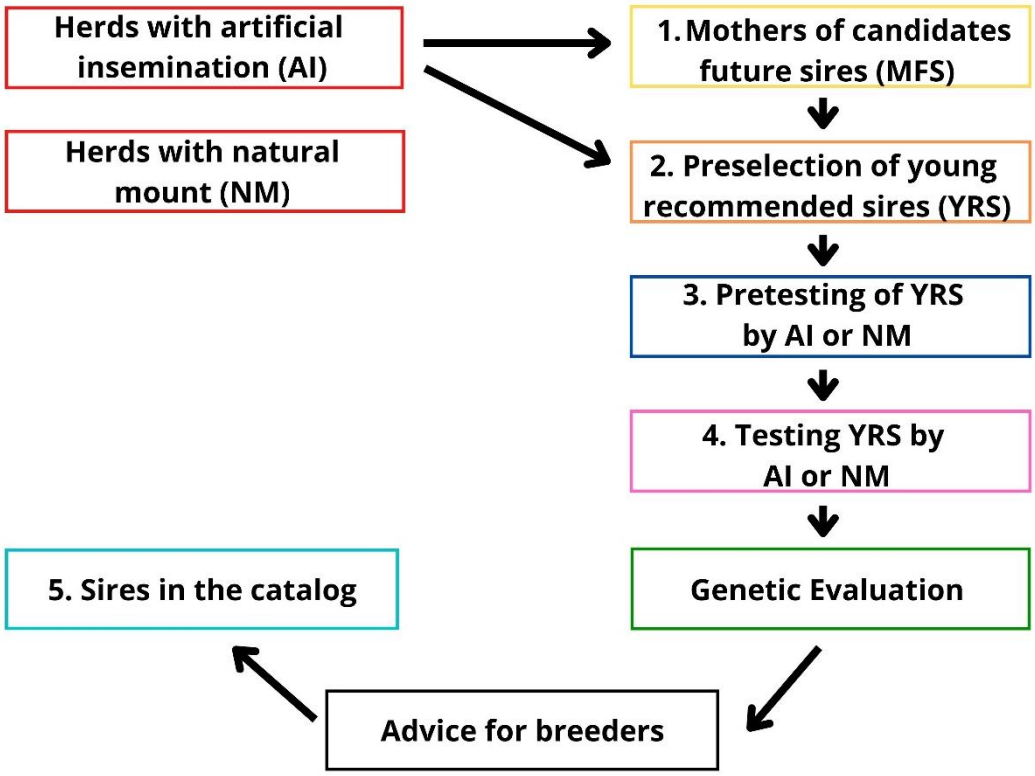
1. **The selection of the mothers of the future sires (MFS).** This implies the selection of female goats from the "selection nucleus" based on their genetic indices for dairy and conformation traits as well as on the granting of awards.
2. **Preselection of young recommended sires (YRS).** In this step, the male offspring of MFS are tested for paternity and morphology at a very young age. A subset of the offspring is subjected to testing in reproduction centres to be used for AI, and the rest are tested for natural service in their own herd.
3. **Pretesting of YRS by AI.** Selected males that are sent to the reproductive centre are trained in semen collection with an artificial vagina.
4. **Testing of YRS by AI.** Selected males are used to perform a maximum of 400 inseminations to females randomly selected from the selection nucleus, with a minimum of 5 inseminations per herd to maximise connections between herds. Female offspring from bucks under evaluation are tested for genealogy and subjected to morphological classification and milk controls.
5. **Genetic evaluation.** An annual genetic evaluation of all animals from the nucleus is applied using a BLUP animal model and all the available information from the individuals to be evaluated and their relatives. In the case of males, productive traits limited by gender are retrieved from the female offspring and collateral relatives. The model includes as fixed effect the herd-year interaction, month of parturition and the litter size, and the age at the parturition, and the individual additive breeding value as covariates (Delgado et al., 2018).

The breeding program also considers a periodical evaluation of the inbreeding and kinship coefficients of the herds, as well as a control of the genetic flow between selection nucleus and other herds to minimise the increment of inbreeding

https://www.mapa.gob.es/en/ganaderia/temas/zootecnia/report_resolucionaprobacionprogramadecriamurciano-granadina_tcm38-579711.pdf.

Since the implementation of the breeding program, there has been a sustained improvement of several traits of economic interest. An increase of 892 g of milk/year was estimated as the response to selection between 2000-2014, making a total gain of 12.5 kg of milk per individual (Delgado et al., 2018). An increment in the milk solid content was also reported, i.e. 38.5 g of fat/year and 36.6 g of protein/year (Delgado et al., 2018).

Figure 1.2. Diagram of the process for selecting sires of Murciano-Granadina goat breed in Spain. Figure adapted from Murciano-Granadina breeding program available at https://www.mapa.gob.es/fr/ganaderia/temas/zootecnia/report_resolucionaprobacionprogramadecriamurciano-granadina_tcm36-579711.pdf.



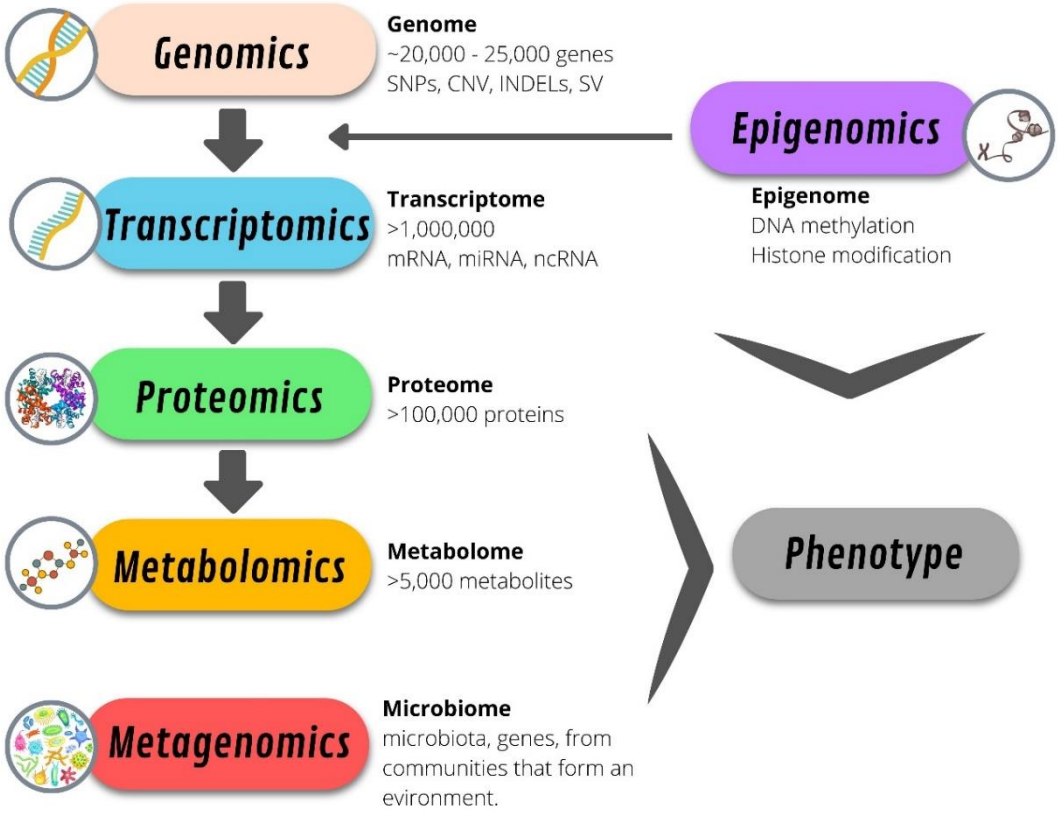
1.3 STRUCTURAL AND FUNCTIONAL ANALYSES OF CAPRINE GENOMES AND TRANSCRIPTOMES.

Nowadays, with the availability of multiple “omics” technologies related to genomics, transcriptomics, epigenomics, metagenomics and metabolomics, it has been possible to study the underlying biological basis of multiple livestock traits that are economically relevant (Chakraborty et al., 2022). In this section, we will review these “omic” technologies and their impact on the elucidation of the structure and function of the goat genome.

Genomics and transcriptomics are disciplines that study the structure and function of genomes and transcriptomes (whole set of RNAs expressed in the cells of an organism), respectively (**Figure 1.3**). They contemplate the whole molecular machinery from the DNA molecule to the RNA transcript, considering all the steps in between, with the aim of understanding and characterising the biological background of cells, tissues and molecular processes of the organisms under study (Kenny and McCarthy, 2017; Vailati-Riboni et al., 2017). Other -omics disciplines are epigenomics, which focuses on the epigenetic modifications of the genome, proteomics, which analyses the set of proteins expressed in the cells of an organism, metagenomics, which studies the genomes of organisms recovered from a particular environment, and metabolomics, which elucidates chemical processes involving metabolites. The effective integration of these multiple sources of information leads to the generation of a holistic view of the biological mechanisms involved in the determination of phenotypes (Kenny and McCarthy, 2017). Methodological advances in data generation, like high throughput technologies, and computation, have been determinant to study molecules with unprecedented resolution and at a very large scale (Tebani et al., 2016). This section includes a revision of the

different -omic technologies and their implementation in livestock production and research.

Figure 1.3. Dataflow and subject of study of different omic-sciences, genomics, transcriptomics, proteomics, metabolomics and metagenomics.



1.3.1 Genomics.

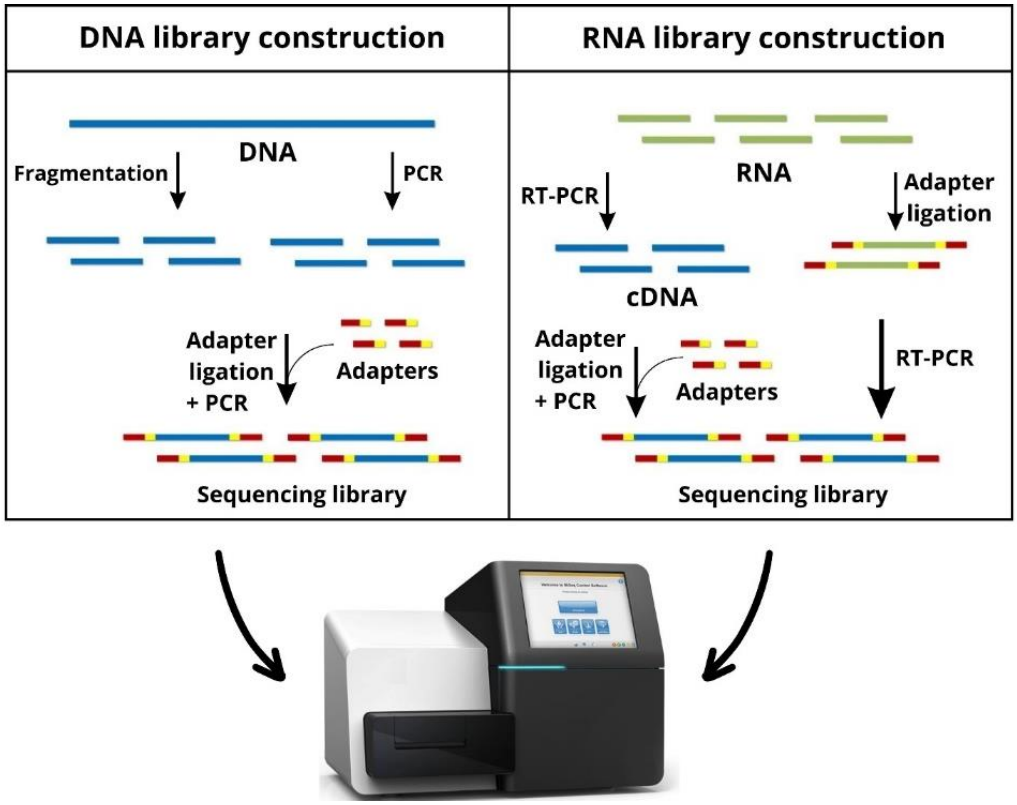
Genomics is the study of the whole DNA sequence contained in the chromosome set of an individual (Kenny and McCarthy, 2017). This discipline initially started in 1920, with the creation of the "genome" concept by Hans Wrinkler, who considered the whole set of chromosomes contained in the nucleus of eukaryotic cells as a unit (Noguera-Solano et al., 2013). This view changed the paradigm of the study of genetics from "genes" to a combination of genes and chromosomes (Noguera-Solano et al., 2013). The first approaches devoted to characterising all the genes in a given genome included the

identification of mutations and the construction of linkage maps using mutants (Klug et al., 2011). These approaches were broadly applied to several model organisms such as maize, *Drosophila*, yeast and bacteria (Klug et al., 2011).

The emergence of the first genome sequencing methods in the late 1970s by Sanger and colleagues (Sanger et al., 1977; Sanger and Coulson, 1975) provided solid grounds for the establishment of the "genomics" field. Since then, sequencing technologies have evolved dramatically into faster, automated, massive and cheaper analyses, moved by the demand for increasing amounts of genomic data from complex organisms (Giani et al., 2020). Nowadays, next-generation sequencing technologies (NGS) are the gold standard for sequencing whole genomes. NGS generate massive amounts of data quickly, combining chemistry with computerised automated processes and parallelising reactions and processes in an efficient way (Giani et al., 2020). Over 90% of genomic data obtained to date have been generated by Illumina platforms mostly represented by highly efficient machines from the HiSeq and NovaSeq lines for short-read sequencing (Giani et al., 2020). NGS usually involves three steps: library preparation, sequencing and data analysis. During library preparation, nucleic acids are fragmented and specific adapters, the nature of which depends on the technology used, are attached to both ends of each fragment. Adapters bind to the flow cell and nucleic acids are amplified and purified. In the sequencing step, libraries are amplified and subjected to sequencing by synthesis. In this last step, nucleotides tagged with different fluorophores bind to the nucleic acid template and the type of fluorescence allows inferring the series of nucleotides that are sequentially incorporated to the elongating complementary chain. Finally, the software of the sequencer performs a base calling procedure, identifying each nucleotide and registering the accuracy of each detection event (https://www.illumina.com/science/technology/next-generation-

sequencing/beginners/ngs-workflow.html). The whole process is summarised in **Figure 1.4**. During the past decade, efforts have been focused on the generation of technologies that allow the sequencing of single molecules with longer reads, with PacBio and Nanopore as leaders in the market. These technologies now offer reads in a range of 10–20 kbp with reliability similar to Illumina short-reads (Giani et al., 2020).

Figure 1.4. Next generation sequencing workflow from the DNA (left) or RNA (right) molecules to the generation of the sequences files.



1.3.1.1 Generation of a goat reference genome.

In the genomic era, it becomes essential to have reference genomes to identify genomic variants, regulatory regions, and any other functional element of interest. The availability of reference genomes is especially useful because it makes possible to compare the

discoveries of different researchers worldwide working on the same species (Ballouz et al., 2019). With regard to goats, Dong et al. (2012) characterised the genome of a female Yunnan black goat by combining Illumina sequencing with optical mapping, a technique by which intact large single DNA molecules are labelled, stretched out and imaged using a fluorescence microscope (Dong et al., 2012). Information provided by optical mapping substantially facilitated the assembly of contigs (Coulson et al., 1986). The goat assembly generated through this approach made possible to annotate 22,175 protein-coding genes (Dong et al., 2012), thus representing the first goat reference genome available for the research community. Five years later, Bickhart et al. (2017) released the ARS1 goat genome assembly, which is currently used as reference, by combining several approaches that included single-molecule long-reads sequencing, optical mapping and Hi-C-based chromatin interaction maps to enhance contig assembly, and short-reads for validation. The ARS1 genome has a total length of 2.92 Mb, with 28,921 genes annotated, 20,606 of which are protein-coding genes (as annotated in the NCBI, https://www.ncbi.nlm.nih.gov/). New efforts are being made to improve the goat reference genome by increasing the sequencing resolution and fixing errors and gaps present in ARS1 (Li et al., 2021). More specifically, in the Saanen_v1 genome, eight assembly errors of ARS1 were corrected, the number of gaps was strongly reduced, and a first goat Y chromosome scaffold was obtained (Li et al., 2021). Importantly, a reference genome from a single individual does not represent the whole variation segregating in a given species, and this is why there is an impetus to construct pangenomes, which model the genomic totality of a taxon of interest (Vernikos, 2020). The pangenome comprises a core genome, the set of genes shared by all individuals from a given species, and the dispensable genome, which contains genes only present in a subset of the individuals (Vernikos, 2020). In the case of goats, the construction of the first “pangenome” has

demonstrated an improvement in variant identification and transcriptomic mapping rate compared to the current ARS1 reference genome (Li et al., 2019). Currently, there are 4,158 whole genome sequences of *Capra hircus* individuals uploaded in the Sequence Read Archive (SRA) database from the NCBI (https://www.ncbi.nlm.nih.gov/sra).

In livestock, whole-genome sequencing data facilitate the analysis of the genomic basis of diseases and complex traits and genomic data are often used in animal breeding and genomic selection (Sharma et al., 2017). Variable sites on the genomes are used as genetic markers in association analyses with traits of interest, making it possible to identify causal variants or variants in high linkage disequilibrium with the causal variant for multiple phenotypes (Uffelmann et al., 2021). Variants highly associated with productive traits, including sustainability, resilience and disease resistance, can be used in the selection schemes of livestock species (Hayes et al., 2013).

The most common sources of genomic variation are single nucleotide polymorphisms (SNPs), small insertions and deletions (INDELs), copy number variations (CNVs) and other structural variants (SVs). Single nucleotide polymorphisms, which are the most frequent genomic variants in all species, are substitutions of a single nucleotide by another one carrying a different nitrogenous base (Eichler, 2019). The effect of the SNP on biological function will depend on the genomic region where it is located. Synonymous SNPs, for instance, map to a protein-coding region but do not involve any change in amino acid sequence, so they are less likely to affect protein function than non-synonymous substitutions, which always imply the replacement of an amino acid by another one (Eichler, 2019). In mammals, there is approximately one SNP every 300-1000 bp on average (Aitken et al., 2004). In cattle, for instance, the frequency of SNPs is one every 700 bp in *Bos taurus* and every 300 bp in *Bos indicus* (The Bovine HapMap

Consortium 2009), meaning that they are widely distributed and well represented along the taurine and indicine genomes.

1.3.1.2 Development of high-throughput genotyping technologies in goats and their practical applications.

In order to make available a tool for high throughput genotyping in goats, Tosser-Klopp et al. (2014) designed a genotypic array that contains about 54,000 SNPs segregating in multiple breeds, evenly spaced along the genome, and covering all chromosomes. The identification of the SNPs was performed by obtaining whole-genome sequences from 97 individuals from six breeds (Alpine, Boer, Creole, Katjang, Saanen and Savanna), and the selected SNPs were validated in a population of 285 individuals from 10 breeds. Currently, the chip is commercially available as GoatSNP50 Illumina BeadChip (http://snp.toulouse.inra.fr/~sigenae/50K_goat_snp_chip/index.html) and a new version containing 59K SNPs has been launched recently (http://www.goatgenome.org/projects.html). A new user-friendly database of goat genomic variation (GGVD, http://animal.nwsuaf.edu.cn/GoatVar) was created by Fu et al. (2021) by retrieving 41.4 million SNPs, 5.14 million indels and a large amount of selected or introgressed regions detected in a sample of 360 individuals comprising modern and ancient goats, wild ibex and bezoar.

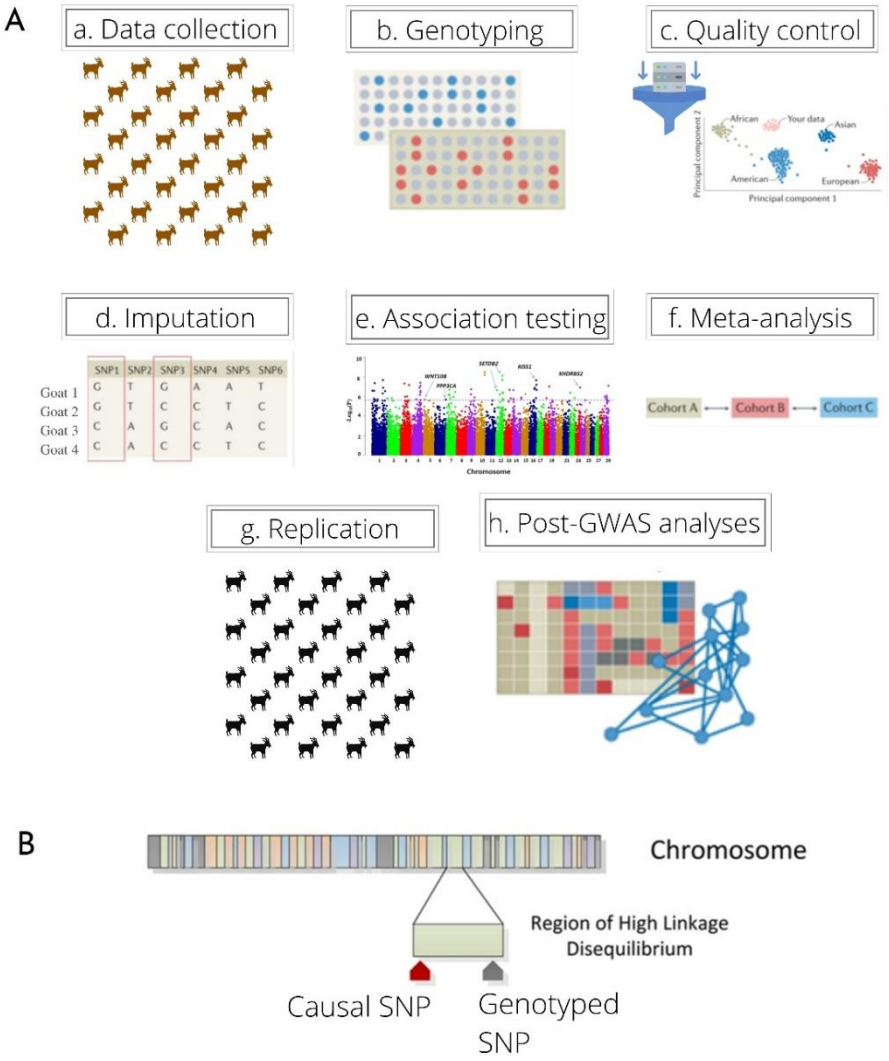
Amongst other things, the availability of high throughput SNP genotyping arrays makes it possible to carry out genome-wide association studies (GWAS) in order to detect quantitative trait loci (QTL) for phenotypes of interest (Sharma et al., 2015). In a GWAS, individuals from a population are phenotyped for a number of traits and also genotyped, either with an SNP array or by whole-genome sequencing, and then genotypic means are contrasted for each marker by using linear or logistic regression (this depends on the distribution of the studied trait) to investigate whether significant differences exist. To

ensure that the detected associations are not artifacts produced by confounding effects, the statistical model includes covariates like sex, age, farm, ancestry, and other factors that might contribute to phenotypic variation (Uffelmann et al., 2021). These analyses assume that the causal variant is physically close and linked to any of the genetic markers integrated into the genotyping platform (Bush and Moore, 2012). The study design of a GWAS analysis is explained in **Figure 1.5**.

In goats, until April 2022, 128 QTL have been reported for traits of economic interest, such as udder morphology and milk composition, in the AnimalQTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/), a curated QTL database for livestock species (Hu et al., 2022). Compared with other ruminant species such as sheep (4,207 QTL) and cattle (192,925 QTL), few QTL have been detected in goats. GWAS for multiple phenotypes have been performed in goats during the last 5 years, including those related with milk yield and quality (Mucha et al., 2018; Guan et al., 2020; Scholtens et al., 2020), reproduction (Talouarn et al., 2020), pigmentation (Nazari-Ghadikolaei et al., 2018), growth (Zhang et al., 2021; Moaeen-ud-Din et al., 2022) and morphology (Rahmatalla et al., 2018; Luigi-Sierra et al., 2020; Guo et al., 2021; Zhang et al., 2022). A total of 857 genomic variants are described as significantly associated with 51 traits in the GWAS atlas (https://ngdc.cncb.ac.cn/gwas/).

Figure 1.5. (A) Steps for conducting a genome-wide association analysis: (a) phenotyping and (b) genotyping of the target population, (c) quality control of the phenotypes and genomic markers, (d) if imputation from whole-genome sequence data if available, (e) application of statistical association models (if multiple groups or populations are analysed a meta-analysis can be applied joining all groups together, as in f), (g) replication of the analysis in other individuals from the same population to assess the reliability of the results obtained in the first analysis, (h) additional down-stream

analyses using external data to identify associated genes, metabolic pathways, etc. This figure has been adapted from Uffelmann et al. (2021). (B) Graphical explanation of the GWAS basis, which assumes that the causal variant is in high linkage disequilibrium with the significantly associated genotyped SNP. Image adapted from Bush and Moore, (2012).



1.3.2 Transcriptomics

The transcriptome is the total set of RNAs, including coding (mRNA) and non-coding (e.g. rRNA, tRNA, lncRNA, miRNA) RNAs expressed by a specific cell or tissue. The characterisation of the transcriptome by sequencing unveils the identities of genes

expressed in single cells or bulk tissues as well as their expression levels at a given time or condition (Kenny and McCarthy, 2017). High throughput transcriptomic analyses started with the development of microarrays (Schena et al., 1995). Nowadays, this technology has been replaced by the massive sequencing of RNAs through RNA-Seq (Conesa et al., 2016). The RNA-Seq procedure is based on the same principles and technological platforms used in DNA sequencing, but requires an additional process of retrotranscription of the RNA transcripts to cDNA before libraries are prepared as depicted in **Figure 1.4** (Conesa et al., 2016). Once the RNA sequences are generated, they are usually stored in FASTQ files, and downstream bioinformatic analyses are implemented to map the sequencing reads to the reference genome and quantify the expression of genes and transcripts (Conesa et al., 2016).

In general, pipelines to process and analyse RNA-Seq data comprise three different steps:

1. Quality control of the raw reads: it is performed to evaluate sequence quality, checking for the presence of adapters (which need to be removed), depth, GC content, duplication levels, and artifacts, among others. Software packages like FastQC are used to perform the quality control of the sequencing data (Andrews, 2010), and others like Trimmomatic and Cutadapt are employed to filter out low-quality reads and remove adapters (Bolger et al., 2014; Martin, 2011).
2. Read alignment of the trimmed reads: reads are mapped to the reference genome, if available, and an additional quality control step is performed to check for the percentage of reads mapping to the genome. From 70 to 90% of the reads are expected to be aligned to the reference genome in good-quality RNA-seq experiments (Conesa et al., 2016). There are multiple software available for read mapping, but some of the most popular are HISAT2 (Kim et al., 2019), STAR (Dobin et al., 2013) and BWA (Li and Durbin, 2009). When the reference genome

is unavailable or incomplete for the species under study, a *de novo* assembly process is performed. Software like Trinity (Grabherr et al., 2011), SPAdes (Bankevich et al., 2012), Oases (Schulz et al., 2012), SOAP-Trans (Xie et al., 2014) and Bridger (Chang et al., 2015), are widely used for this task because of their excellent performances in the *de novo* reconstruction of transcripts based on short-reads (Hölzer and Marz, 2019).

3. Transcript identification and quantification: After reads are aligned to the reference genome, the level of expression of each gene is measured by counting the number of reads mapping to it. Gene expression can be measured as raw read counts, which can be estimated by using software such as HTSeq (Anders et al., 2015) and Featurecounts (Liao et al., 2014). An alternative approach is to correct raw read counts for different factors, including library size and feature length (Conesa et al., 2016), in order to obtain normalised metrics such as RPKM (reads per kilobase per million), FPKM (fragments per kilobase per million), TPM (transcripts per million), TMM (trimmed mean M-values) and CPM (counts per million). Importantly, the choice of one metric or another depends on the context: if the analysis aims to perform comparisons across samples for the same gene, correction for gene length is not necessary. However, for gene ranking based on gene expression this is absolutely required since longer genes harbour more reads than the shorter ones (Conesa et al., 2016). It is also relevant to emphasise that TPM and TMM have been described to preserve the biological signal more faithfully and have a better performance than other metrics, and they are more comparable between samples of different origins (Abrams et al., 2019). Several software are available to normalise sequencing data and estimate transcript

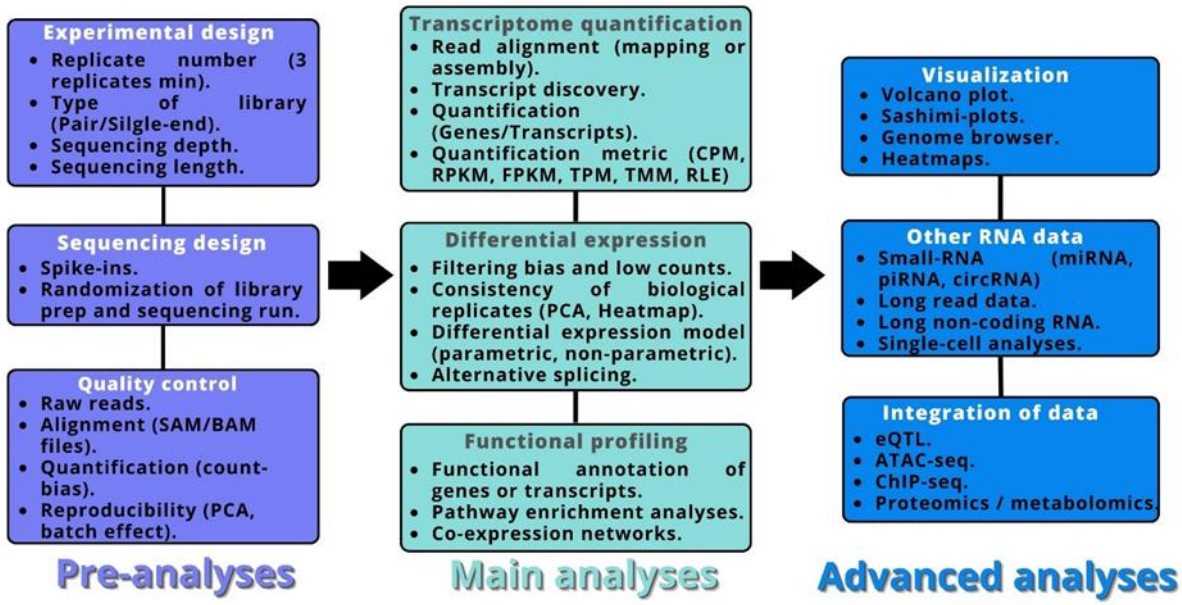
expression, i.e. StringTie (Pertea et al., 2015), Cufflinks (Trapnell et al., 2010), RSEM (Li and Dewey, 2011) and Kallisto (Bray et al., 2016).

From the matrix of counts or normalised metrics different transcriptomic analyses can be performed, including the characterisation of expression patterns, identification of alternative splicing isoforms, identification of non-coding RNAs and differential expression analyses (Conesa et al., 2016).

Differential expression analysis and the characterisation of gene expression have been used in the current thesis, so they will be further discussed. In the differential expression analysis, the transcriptomes of two or more groups of biological replicates (at least three per group) are sequenced to characterise gene expression levels. Subsequently, transcriptomes from different groups are compared to identify genes with statistically significant differences in terms of averaged (across replicates within groups) expression levels between groups (Conesa et al., 2016; Schurch et al., 2016). To perform this analysis, the quantification of the expression per gene or transcript is inferred by considering specific mathematical distributions e.g. the negative binomial distribution in the case of edgeR (Robinson et al., 2009) and DESeq2 (Love et al., 2014). Other software, such as SAMseq (Li and Tibshirani, 2013) use non-parametrical models to infer the distribution of gene expression from the experimental data, but such approach requires a large number of replicates (Anders and Huber, 2010; Soneson and Delorenzi, 2013; Conesa et al., 2016). Since differential expression analysis considers thousands of genes, statistical significance needs to be corrected for multiple testing by procedures, such as the false discovery rate method (Benjamini and Hochberg, 1995), aiming to reduce the occurrence of false positive results. The set of differentially expressed genes can be explored with tools like DAVID (Huang et al., 2009), enrichR (Chen et al., 2013) and Uniprot (Bateman, 2019), to annotate genes and to identify biological pathways that are

enriched in the list of differentially expressed genes. In **Figure 1.6**, we summarise the whole workflow, from sequencing to result analyses and data visualisation, associated with the analysis of RNA-seq data.

Figure 1.6. RNA-seq workflow and analyses commonly performed from transcriptomic sequences. This figure was adapted from Conesa et al. (2016).



1.3.2.1 Studies on the goat transcriptome.

A mini-atlas of gene expression has been constructed in goats by sequencing the transcriptomes of 17 different tissues and three cell types in seven goats (Muriuki et al., 2019). By doing so, 18,528 unique protein-coding genes expressed in at least one tissue were identified (Muriuki et al., 2019). A network clustering analysis based on the expression pattern of each sample showed that tissue is the main factor governing sample clustering. Moreover, this research reported a high transcriptional activity in neural tissues, i.e. brain cortex and cerebellum. Indeed 1,795 genes were exclusively expressed in these tissues and mainly associated with cognitive and neurotransmission functions (Muriuki et al., 2019). This study also showed a high number of genes with a restricted

single tissue expression e.g. myosin and transcription factors as well as genes that regulate muscle growth like *MSTN, MYH1*, *MYL1*, *MYOG* and *MYOD1* were highly and specifically expressed in the muscle. Muriuki et al. (2019) also described the expression patterns of genes that are assumed to be important in the determinism of productive traits. For instance, the *DGAT1* gene, which strongly influences milk fat content in goats (Martin et al., 2017), displays its highest expression levels in the colon and liver. Muriuki et al. (2019) also contrasted the expression of genes in alveolar macrophages of the udder exposed to bacterial lipopolysaccharide in sheep and goats. They found similar patterns of gene activation in both species, with high expression of genes involved in immune functions such as interleukins (i.e. *IL34, IL1RN, IL33, IL1B*) and cytokines (*CCDC80*).

Besides the mini-atlas constructed by Muriuki et al. (2019), several tissues related with traits of economic interest have been analysed with transcriptomic tools in goats. For instance, Cremonesi et al. (2012) detected the upregulation of mRNAs involved in the immune and inflammatory response (*NFKB1*, *BASP1*, *BATF3* and *TNFAIP6*) and innate response to pathogens (*PLEK*) when sequencing the RNA from mammary tissue exposed to *Staphylococcus aureus* infection. With regard to milk production, the transcriptomic profiling of the goat mammary gland at different time points of the productive cycle evidenced a significant mRNA upregulation (in lactating goats) of genes involved in fatty acid and lipid synthesis and transport, including those related with the PPAR signalling pathway, e.g. *FASN*, *ACACA*, *SCD*, *FAB3* and *4*, *GPAT2* and *4*, *LPL* (Guan et al., 2020; Li et al., 2020). In lactating goats, Guan et al. (2020) also reported an increased mRNA expression of genes encoding caseins, the main proteins of milk, and of α-lactalbumin which is essential for lactose synthesis. In contrast, the ceasing of lactation entailed the upregulation of several genes involved in apoptosis and tissue remodelling of the mammary gland (Guan et al., 2020).

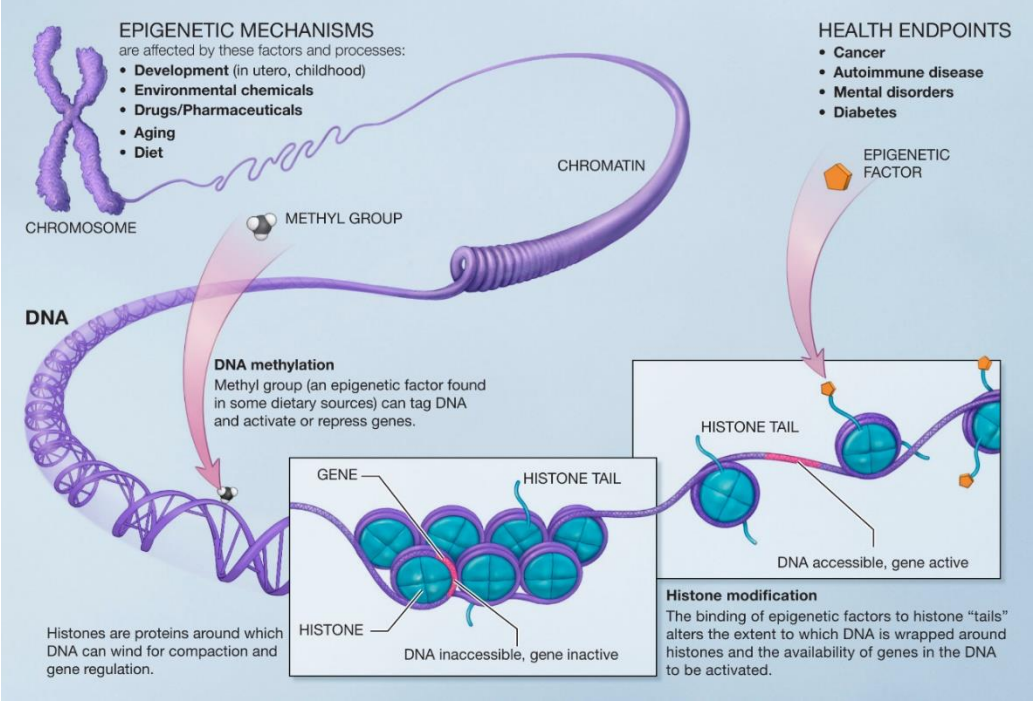
1.3.3 Epigenomics.

This discipline is devoted to the characterisation of the factors that regulate chromatin remodelling and gene expression without altering the DNA sequence. Epigenetic changes are commonly produced by DNA methylation (usually in CpG islands located in promoter regions) and also by post-translational modifications (acetylation, phosphorylation, methylation etc.) of histones that result in modifications of chromatin condensation and the accessibility of it to transcription factors, as depicted in **Figure 1.7** (Jaenisch and Bird, 2003; Agarwal et al., 2020). Non-coding RNAs are also important mediators of epigenetic changes (Kumar et al., 2020). The whole set of epigenetic marks present in a specific genome are called "epigenome" (Jaenisch and Bird, 2003). Epigenetic modifications play an essential role in gene expression regulation, switching "in" and "off" gene transcription (Agarwal et al., 2020). For instance, DNA methylation usually has a repressive effect on gene expression while acetylation stimulates it (Agarwal et al., 2020). Epigenetic changes probably account for part of the "missing heritability" reported for complex phenotypic traits, as chromatin modifications and methylation patterns might display intergenerational transmission (Trerotola et al., 2015; Khatib, 2021). Nowadays, several methods have been developed to analyse the epigenome of any organism.

For instance, methylation can be investigated by treating DNA with sodium bisulfite, which transforms unmethylated cytosines into uracils (recognised as thymines in subsequent PCR amplification and sequencing stages) while 5-methylcytosine remains unaffected by such treatment (Li and Tollefsbol, 2011). Enzymatic methyl sequencing (EM-seq), which is based on the oxidation of methylated cytosines to protect them from deamination and the subsequent use of APOBEC which deaminates non-modified,

cytosines, is a powerful alternative to bisulfite sequencing because such enzymatic treatment does not affect the integrity of the DNA (Vaisvila et al., 2021).

Figure 1.7. Epigenetic mechanisms and variations on the DNA, comprising histone modifications and DNA methylation. Figure obtained from http://commonfund.nih.gov/epigenomics/.



Histone modifications can be analysed using chromatin immunoprecipitation (ChIP-seq). This method uses an antibody that targets a specific histone modification or a DNA-binding protein site (Park, 2009). In brief, cells are treated with formaldehyde and chromatin is sheared into fragments from 200 to 600 bp. Subsequently, an antibody specific to the protein of interest is added to immunoprecipitate the DNA-protein complex (Park, 2009). Finally, libraries are prepared from the immunoprecipitated DNA and fully sequenced (Park, 2009).

The landscape of chromatin accessibility can be explored with ATAC-Seq, which stands for "assay for transposase accessible chromatin sequencing". In this method, a

hyperactive transposase 5 (Tn5) that cuts and ligates adapters to regions of increased chromatin accessibility is used (Buenrostro et al., 2015). High-throughput sequencing of these fragments is subsequently performed using next-generation sequencing techniques. One of the main advantages of ATAC-Seq is that even a low input sample is, in principle, enough to generate a successful assay (Buenrostro et al., 2015).

The bioinformatic pipelines to analyse methylation, CHIP-Seq and ATAC-seq data are similar to those reported for whole-genome sequencing and RNA-seq. In brief, reads are aligned to the reference genome using algorithms like BWA (Li and Durbin 2009, 2010), and chemically modified or accessible sites are recognised along the genome with specific software. For methylation, Bismark (Krueger and Andrews, 2011) and Methyldackel (https://github.com/dpryan79/MethylDackel) are commonly used, while for ATAC-seq MACS2 (Gaspar, 2018) and Genrich (https://github.com/jsh58/Genrich) have been widely employed. Once the methylation, histone modification or chromatin accessibility patterns are determined, edgeR (Robinson et al., 2009) and DESeq2 (Love et al., 2014) can be used to identify genomic regions displaying differential methylation, accessibility or histone modification patterns.

1.3.3.1 Insights about the goat epigenome.

The epigenome of goats and other species has been described for multiple tissues, and, in the framework of projects like FAANG (Clark et al., 2020), there is a continuous effort to generate new insights about epigenetic regulation in domestic animals. The analysis of transcriptome and chromatin accessibility of 4 tissues in 4 species (chicken, pig, cattle and goat) revealed a correlation between chromatin accessibility and gene expression, and made it possible to elucidate complex mechanisms regulating gene expression (Foissac et al., 2019). Several of these regulatory processes were conserved among the

four domestic species under study, evidencing the existence of shared selection pressures over regulatory elements present in these four genomes. In goats, methylation patterns associated with phenotypes, like hair morphogenesis, female reproduction and hormonal cycles as well as adaptation to climate with low and high temperatures, have been described (Wang et al., 2020; Xiao et al., 2020; Denoyelle et al., 2021; Kang et al., 2022). In cattle, studies have reported methylation patterns associated with lactation and milk production, and changes of these patterns during different phases of the lactation have been characterised in depth (Singh et al., 2012; Osorio et al., 2016). Information about methylation and chromatin accessibility related with milk synthesis in the mammary gland are not yet available in goats, despite the fact that they could provide valuable clues about the genetic regulation of lactation.

1.4 THE INHERITANCE OF DAIRY AND MORPHOLOGY TRAITS IN GOATS.

As commented in section 1.1, goat breeding is an important component of dairy production. Most traits of economic interest are related with milk production and morphology (in terms of dairyness and longevity) and all of them are amenable to be modified by selection because they have an additive genetic component (Zonaed Siddiki et al., 2020).

1.4.1 Genetic parameters of morphological and milk traits in goats

In dairy goats, moderate heritabilities have been estimated for milk yield and milk composition traits, with values ranging from 0.20 to 0.69 (**Table 1.1**). In contrast, the

heritabilities of morphological traits in goats are quite variable (**Table 1.1**), with figures that go from 0.02 for feet and hock traits to 0.52 for udder morphology traits (Manfredi et al., 2001; Carillier et al., 2014; McLaren et al., 2016).

Miranda et al. (2019) estimated genetic parameters for peak milk yield, milk yield and persistency of the lactation in a population of 122,883 Murciano-Granadina goats. Different models were adjusted (i.e. uni-, bi- and multi-trait), and all of them evidenced the existence of low levels of heritability i.e. from 0.005 for lactation persistency to 0.015 for peak milk yield (Miranda et al., 2019). These results would indicate that the additive genetic basis of these traits is weak, maybe because the environmental component is important. In contrast, morphological traits displayed heritabilities that go from 0.12 for the fore udder attachment trait to 0.28 for the medial suspensory ligament of the udder trait (Gómez-Carpio et al., 2012). By analysing the production of 4,967 goats using single traits models, Analla et al. (1996) reported heritabilities of 0.18, 0.16 and 0.25, for milk yield, fat and protein content in the Murciano-granadina breed, respectively. Many of these traits display moderate to high genetic correlations between each other, implying the existence of a common genetic background with pleiotropic effects (Falconer and Mackay, 1989). In Murciano-Granadina goats, milk yield and composition traits are negatively correlated i.e Analla et al. (1996) reported genetic correlations of r_g = –0.89 (milk yield vs fat content) and –0.65 (milk yield vs protein content), while the r_g between milk fat and protein contents was 0.93 (Analla et al., 1996). Udder morphology in goats shows some degree of correlation with milk production and composition (Rupp et al., 2011). In Saanen and Alpine goats, udder floor position displays a negative genetic correlation with somatic cell score (r_g of 0.24 and -0.19, respectively) as reported by Rupp et al. (2011). In the same study, Rupp et al. (2011) described moderate genetic correlations between somatic cell scores and teat length, width and shape (r_g of 0.29, 0.34

and –0.27, respectively). Morphological traits are also relevant for longevity, a trait of increasing importance in goat production. For instance, Castañeda-Bustos et al. (2017) detected moderate correlations between rump width and udder morphology, which influences productive life duration, highlighting the importance of using these traits as selection criteria.

1.4.2 Understanding the genomic basis of milk production and morphology traits in goats.

The genomic basis of dairy traits has been explored in goats by using multiple approaches to understand the molecular events that determine such phenotypes as well as to detect genetic variants with significant effects on traits of economic interest. In the next two sections, we will summarise the main findings reported for goats in this field.

1.4.2.1 Genomic basis of milk yield and milk composition in goats.

Milk composition traits are essential for dairy goat production as milk is mainly devoted to the manufacturing of cheese. In the cosmopolitan breeds Alpine and Saanen, a QTL on a goat chromosome (CHI) 6 region which coincides with the casein gene cluster (CHI 6:82–86 Mb) has been reported to have strong effects on milk protein content (Martin et al., 2017; Tilahun et al., 2020). The same association was described by Guan et al. (2020) in Murciano-Granadina goats (with effects on both protein and dry matter contents), evidencing that variability in the casein genes is one of the most important genetic determinants of milk composition in goats.

Table 1.1. Heritabilities (h2) estimated in Alpine and/or Saanen dairy breeds for milk and morphological traits.

Trait	h^2	Population	N individuals	Reference
Milk yield	0.26	Alpine	32,256	Arnal et al. 2019
	0.31	Alpine	~1.5 M	Carillier et al. 2014
	0.34	Alpine	161,121	Rupp et al. 2011
	0.29	Saanen	30,186	Arnal et al. 2019
	0.26	Saanen	~1.1 M	Carillier et al. 2014
	0.34	Saanen	121,411	Rupp et al. 2011
Protein Yield	0.25	Alpine	32,256	Arnal et al. 2019
	0.31	Alpine	~1.5 M	Carillier et al. 2014
	0.61	Alpine	161,121	Rupp et al. 2011
	0.3	Saanen	30,186	Arnal et al. 2019
	0.25	Saanen	~1.1 M	Carillier et al. 2014
	0.3	Saanen	121,411	Rupp et al. 2011
Fat yield	0.26	Alpine	32,256	Arnal et al. 2019
	0.28	Alpine	~1.5 M	Carillier et al. 2014
	0.35	Alpine	161,121	Rupp et al. 2011
	0.33	Saanen	30,186	Arnal et al. 2019
	0.25	Saanen	~1.1 M	Carillier et al. 2014
	0.6	Saanen	121,411	Rupp et al. 2011
Protein content	0.66	Alpine	32,256	Arnal et al. 2019
	0.67	Alpine	161,121	Rupp et al. 2011
	0.62	Saanen	30,186	Arnal et al. 2019
	0.62	Saanen	121,411	Rupp et al. 2011
Fat content	0.65	Alpine	32,256	Arnal et al. 2019
	0.31	Alpine	161,121	Rupp et al. 2011
	0.6	Saanen	30,186	Arnal et al. 2019
	0.32	Saanen	121,411	Rupp et al. 2011
Somatic cell score	0.2	Alpine	~1.5 M	Carillier et al. 2014
	0.24	Alpine	161,121	Rupp et al. 2011
	0.16	Saanen	~1.1 M	Carillier et al. 2014
	0.2	Saanen	121,411	Rupp et al. 2011
Udder shape	0.4	Alpine	~1.5 M	Carillier et al. 2014
	0.47	Saanen	~1.1 M	Carillier et al. 2014
Fore udder	0.44	Alpine	~1.5 M	Carillier et al. 2014
	0.42	Saanen	~1.1 M	Carillier et al. 2014
Rear udder attachment	0.47	Alpine	~1.5 M	Carillier et al. 2014
	0.52	Saanen	~1.1 M	Carillier et al. 2014
Teat angle	0.42	Alpine	~1.5 M	Carillier et al. 2014
	0.45	Saanen	~1.1 M	Carillier et al. 2014

Before the advent of GWAS, very few QTL had been identified in goats (due to a lack of well characterized microsatellite markers) but key candidate gene studies, mostly focused

on the casein loci, had been carried out (reviewed in Amills et al., 2012). The four casein genes, i.e. casein α_{S1} (*CSN1S1*), casein α_{S2} (*CSN1S2*), casein β (*CSN2*) and casein κ (*CSN3*), map to a 250 kb region on caprine chromosome 6 and they encode about 80% of the proteins in milk (Martin et al., 2002). Since caseins are insoluble, they form multi-molecular aggregates, called micelles, that might precipitate at certain conditions of temperature and acidity (Jenness, 1980). In Saanen and Alpine goats, SNPs on the casein cluster explain 39.1% of the variance in protein content, out of which 24 to 38% is attributed to the polymorphism of the *CSN1S1* gene (Martin et al., 2017). Different alleles have been reported for each casein gene based on genomic variations and their effect on protein content in milk (Marletta et al., 2007). The alleles of the goat *CSN1S1* gene are classified in four categories according to their levels of CSN1S1 synthesis (Moioli et al., 2007):

- High content alleles: A, B_1, B_2, B_3, B_4, C, H, L, M) producing almost 3.5 g/L of CSN1S1 per allele.
- Medium content alleles: E and I, 1.1 g/L of CSN1S1 per allele.
- Low content alleles: F and G, 0.45 g/L of CSN1S1 per allele.
- Null content alleles: O1, O2 and N, 0 g/L of CSN1S1 per allele.

In Murciano-Granadina goats, Caravaca et al. (2008) described that individuals carrying the *CSN1S1* BB genotype show increased levels of CSN1S1 in milk (8.50 ± 0.60 g/l) when compared to other genotypes. In a subsequent study, Caravaca et al. (2011) demonstrated that the curdling rate of the milk from EE goats was significantly higher than that of BB individuals. Moreover, cheese yield was not significantly different in BB, EE and EF goats, a finding that does not match what has been published in Alpine and Saanen goats in which the *CSN1S1* genotype has strong effects on milk and fat protein

content, technological parameters and cheese yield and organoleptic properties (Grosclaude et al., 1987; Barbieri et al., 1995; Manfredi et al., 1995; revised in Trujillo et al., 1998). Another relevant finding of the study of Caravaca et al. (2011) was that the *CSN3* genotype was significantly associated with the rennet coagulation time but not with cheese yield. Currently, the *CSN1S1* and *CSN3* genotypes are taken into account in the selection of Murciano-Granadina goats, a procedure that makes a lot of sense in the light of the GWAS results presented by Guan et al. (2020) and also in this thesis.

Besides caseins, the whey proteins lactalbumin α (*LALBA*) and lactoglobulin β (*BLG*) are important components of the milk proteome (Selvaggi et al., 2014). The two genes encoding these proteins have been much less characterised than the casein loci and a few polymorphisms have been reported, the majority located in non-coding regions (Pena et al., 2000; Cosenza et al., 2003; Ballester et al., 2005; Zidi et al., 2014; Dettori et al., 2015; Cardona et al., 2016). Interestingly, Cardona et al. (2016) described higher protein and fat contents in milk from goats with AA *BLG* genotypes than in that from individuals with AB and BB genotypes across different stages of lactation.

Regarding fat content, the *DGAT1* gene (CHI 14:81,329,989 – 81,338,811 Mb), which plays a key role in the synthesis of triglycerides, contains polymorphisms associated with milk fat content. More specifically, two mutations mapping to this gene (p.Arg251Leu and p.Arg396Trp) cause a decrease in milk fat content (Martin et al. 2017). Additionally, small insertions and deletions in non-coding regions of the *DGAT1* gene are associated with changes in the fat composition of Xinong Saanen and Guanzhong goats (An et al., 2012).

In terms of milk yield, Scholtens et al. (2020) reported a QTL on CHI 19:26-27 Mb with effects on milk protein, fat and somatic cell content in New Zealand Alpine, Saanen,

Nubian, Toggenburg and crossbreed goats. This genomic region has similar effects on milk traits in French Alpine and Saanen goats (Martin et al. 2017). Notably, this locus has also been associated with longevity and morphology traits in French and New Zealand goat populations (Martin et al. 2018; Jiang et al. 2022), indicating a pleiotropic effect of the QTL on milk production and longevity, two traits that are correlated (Castañeda-Bustos et al. 2014).

1.4.2.2 Genomic basis of morphology traits in goats.

Considering the correlations between the morphology of the udder, feet and body capacity with milk production and longevity, multiple studies have focused on identifying the genomic basis of the variation of these traits. As commented in the previous section, in French and New Zealand goats a QTL with pleiotropic effects on CHI 19:24-29 Mb affects udder morphology traits, i.e. fore and rear udder attachment, udder depth, udder floor position (Martin et al. 2018; Jiang et al. 2022). Martin et al. (2018) reported a certain degree of breed-specificity of this QTL, as significant effects were found in Saanen individuals but not in the Alpine ones. In a population of mixed-breed dairy goats (Saanen×Toggenburg×Alpine), significant associations between this chromosome 19 region and udder attachment, udder depth and front legs were detected (Mucha et al., 2018). This chromosome 19 region has a high density of genes per Mb (22.5 genes/Mb), making it challenging to identify potential candidate genes. To address this issue, Jiang et al. (2022) obtained whole-genome sequences from 302 individuals, including 48 from the GWAS analysis, and these sequences were screened to identify SNPs in high linkage disequilibrium ($R^2 > 0.8$) with the significant SNPs identified in the GWAS. They found two candidate non-synonymous SNPs mapping to the proteasome 20S subunit β6 (*PSMB6*) and sex hormone binding globulin (*SHBG)*, and one in-frame deletion the SUMO specific peptidase 3 (*SENP3*) gene. The *SHBG* gene is a steroid transporter that

regulates the bioavailability of these to target cells and mediates the uptake of steroids into cells, stimulating mitogenesis and cellular division in the mammary gland (Denholm et al., 2018). Besides, the *PSMB6* locus is associated with human lipid metabolism (Klarin et al., 2018), while *SENP3* encodes a SUMO-specific protease involved in ribosome biogenesis (Finkbeiner et al., 2011).

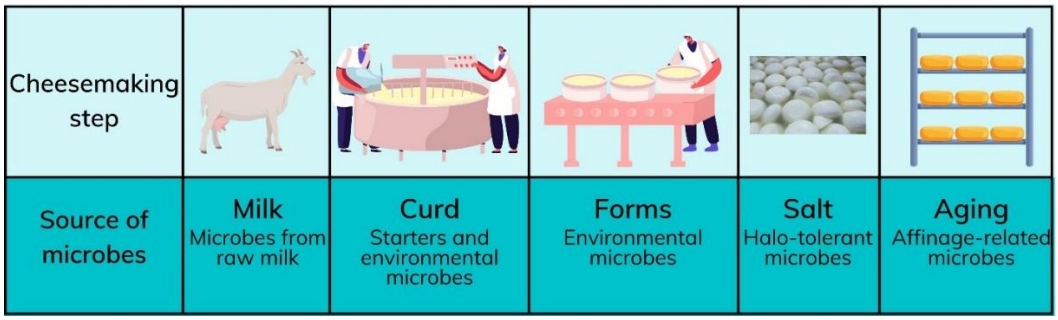
In general, there is a lack of knowledge about the molecular basis of morphological traits of goats and other species. The inherent complexity of morphological traits, which are determined by multiple pathways related with highly heterogeneous biological and developmental processes, poses a considerable challenge to elucidate the molecular basis of morphology. Interestingly, the presence of wattles in goats has been associated with the *FMN1/GREM1* region on chromosome 10 (Reber et al., 2015).

1.5 THE MICROBIOME AS AN IMPORTANT DETERMINANT OF CHEESE TECHNOLOGICAL AN ORGANOLEPTIC PROPERTIES.

As previously said, goat milk is mostly devoted to the elaboration of cheese, and, in this context, a high proportion of solids in milk is desirable. Besides the genetic background of goats, the technological and organoleptic properties of cheese are strongly determined by its microbiome (Yeluri Jonnala et al., 2018). A complex network of interactions between enzymes, microorganisms and milk constituents, i.e. proteins, lipids, and acids, induces multiple physicochemical changes and results in the production of volatile and non-volatile compounds that determine the aroma and flavour of cheese (Fox et al., 1995). Certain bacterial strains are used as primary starters in cheesemaking, particularly lactic-acid bacteria (LAB), including *Lactococcus lactis*, *Leuconostoc sp*, and *Streptococcus thermophilus*, amongst others (Fox et al., 2017; Parente et al., 2017). The main function

of starters is the production of lactic acid from lactose during the first stages of cheesemaking. This results in the acidification of milk and the induction of proteolysis and lipolysis events as well as the triggering of casein precipitation to produce the curd (Fox et al., 2017; Parente et al., 2017). The lowering of the milk pH also prevents the growth of pathogens like *Listeria* (Coelho et al., 2014). The selection of the microbial strain to be used as starter is determined by the type of cheese and the manufacturing technique employed in its elaboration (Parente et al., 2017). Cured and hard cheeses commonly use *Streptococcus,* which is a thermophilic strain and works at high temperatures, while in the making of fresh cheese bacteria from the *Lactococcus* and *Lactobacillus* genera, that are activated at lower temperatures, are commonly used (Parente et al., 2017). Apart from the starter cultures, other LAB and non-LAB genera are added to the milk or to the curd to induce the production of specific flavours, textures and aromas, including *Hafnia alvei*, several strains of *Lactobacillus*, *Brevibacterium*, *Brachybacterium*, and many others (Fox et al., 2017). Besides, bacteria present in milk (especially if unpasteurised) or in the surfaces of materials in contact with milk (e.g. the milking machine, milk receivers and pipelines, the bulk tank, equipments in the dairy plant etc.) can contaminate the milk or the curd anytime (Fox et al., 2017). In **Figure 1.8** the sources of microbes at different steps of cheesemaking are displayed. The growth of the microorganisms in cheese depends to a high degree on water activity, salt concentration, pH, temperature and oxidation-reduction potential, which are conditioned by the manufacturing techniques and the type of cheese (Fox et al., 2017).

Figure 1.8. Sources of microbes at different stages of the cheese manufacturing process.



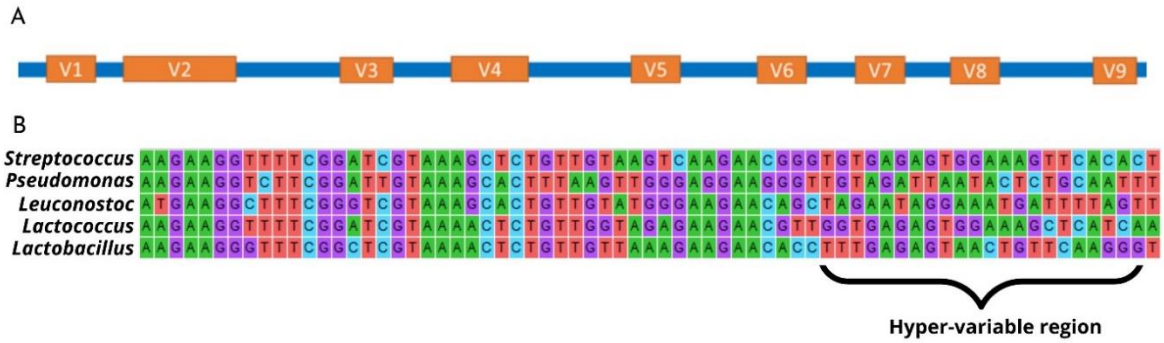
For the cheesemaking, microbiological information is useful to optimise technological processes and to ensure the safety and shelf-life of the products (Yeluri Jonnala et al., 2018). Traditionally, the identification of microorganisms in food has been performed through the use of appropriate culture media, immunoassays and PCR (Beresford et al., 2001). These techniques have some limitations. For instance, culture media cannot identify unculturable bacteria, PCR requires the design of specific primers for each taxon to be detected, and immunoassays are often taxa-specific (Miller et al., 2013). Nowadays, with the development of high throughput sequencing techniques, it is possible to characterise with unprecedented resolution the whole genome, transcriptome or specific marker genes of any microorganism present in cheese (Yeluri Jonnala et al., 2018). In this thesis, we will focus on the sequencing of marker genes to provide a first picture about the composition of the cheese microbiome in goats.

1.5.1 Marker gene16S rRNA for phylogenetical analyses.

Marker genes contain highly variable regions, flanked by conserved regions, that vary across species and that can be used to identify the taxonomy of the sequenced individuals (Knight et al., 2018). In bacteria and archaea, the most widely used marker locus for taxonomy assignation is the 16S rRNA gene, which encodes the RNA component of the 30S subunit of the prokaryote ribosome (Chakravorty et al., 2007). As depicted in **Figure**

1.9, this gene contains nine hypervariable regions, with a length between 30 and 100 bp (Chakravorty et al., 2007). Each region of the 16S rRNA gene has a different resolution in identifying taxa levels, being V2 and V3 the most effective ones in identifying most bacteria up to the genus level (Chakravorty et al., 2007) while V4 displays a very good performance for clustering of communities using short reads (Caporaso et al., 2011).

Figure 1.9. Alignment of hyper-variable regions in the 16S rRNA gene from different microorganisms.



The analysis of bacterial marker genes requires (1) The extraction of the bacterial DNA from the samples, (2) The partial amplification of the 16S rRNA gene using specific primers that hybridise to conserved regions and amplify the targeted hypervariable regions, (3) The preparation of the libraries by attaching sequencing adapters to amplicons, and (4) The sequencing of the libraries with high throughput technologies (Knight et al., 2018). The amplification step of this methodology allows for detecting the contamination of samples with host DNA and bacterial DNA from the environment (Knight et al., 2018). Some of the most used sequencing platforms for metagenomics are Illumina MiSeq, Ion Torrent PGM and Roche 454 GS FLX+ (Allali et al., 2017).

The bioinformatics analyses of the sequence data require a quality control step to remove adapters and sequencing errors (Knight et al., 2018). Subsequently, sequences are clustered into features, like operational taxonomic units (OTUs) or amplicon sequence

variants (ASVs), depending on the algorithm used. Algorithms like Deblur (Amir et al., 2017) and DADA2 (Callahan et al., 2016) are commonly used for this purpose, as they assign sequences to exact features using error profiles in order to discriminate closely related but distinct taxa (Knight et al., 2018). Subsequently, taxonomy is assigned to the features using machine learning classifier processes and databases with reference sequences and taxonomy. Popular software to perform these analyses are QIIME (Bolyen et al., 2019) and Mothur (Schloss et al., 2009).

Besides microbiome composition, the microbial diversity of each sampled environment can be studied. Common methods are alpha diversity, which measures the diversity within an environment, and beta diversity which measures the changes in taxa diversity across environments (Whittaker, 1972). Alpha diversity can be expressed in terms of richness, the total number of observed species in a sample like Chao1 (Chao, 1984), evenness, which combines richness and species abundances (Shannon, 1948) and measures that include phylogenetic distances across species (Faith, 1992). On the other hand, dissimilarities across samples (beta diversity) are commonly measured using Bray-Curtis metrics for compositional data (Bray and Curtis, 1957) and metrics that include phylogenetic distances, like Unifrac (Lozupone and Knight, 2005).

1.5.2 Sequencing of the cheese microbiome.

A considerable effort has been made to sequence the microbiomes of a wide variety of cheeses, mainly in cattle (Yeluri Jonnala et al., 2018). As said LAB are overwhelmingly predominant in the majority of cheeses, but high-depth sequencing has made it possible to detect many bacterial genera that were overlooked in culture-based studies (e.g. D*ehalobacter*, *Desulfohalobium*, *Halomonas*, *Thermohalobacter*, *Haloquadratum*, *Prevotella*, *Faecalibacterium* etc.), thus paving the way to characterise their effects, if

any, on cheese quality. The analysis of cheeses has also facilitated the identification of minority halotolerant or halophilic genera (e.g. *Marinilactibacillus*, *Idiomarina*, *Halomonas*, *Pseudoalteromona*, *Oceanobacillus* etc.) that are probably incorporated to the cheese during the brining process (Yeluri Jonnala et al., 2018; Kothe et al., 2021). The presence of contaminating bacteria, such as *Pseudomonas*, has also been detected in cheese samples and could be explained, in some instances, by the ability of psychrophilic genera to grow at cold temperatures (Ribeiro Júnior et al., 2018). The detection of contaminating bacteria is of outmost importance because they can cause the spoilage of cheese or even food poisoning due to the ingestion of metabolites such as biogenic amines (Yeluri Jonnala et al., 2018). There are also evidences that the composition of the microbiome is affected by spatial factors, with the rind microbiome being more diverse than the core one due to its decreased acidity, augmented oxygen availability and close interaction with the environment (Montel et al., 2014).

Few studies have analysed the microbiomes of cheeses of caprine origin. In a meta-analysis of 55 artisanal cheeses, comprising cheeses from cow, goat and sheep origins and 107 publicly available data, Walsh et al. (2020) analysed the whole metagenome sequences and the volatilome of cheese samples. They reported a high microbiome variability across cheese types and found relations between the microbiome and the volatile profile. However, the microbiome of goat cheeses was not the main focus of such research. Salazar et al. (2018) characterised the microbiota of pasteurised and unpasteurised Gouda cheeses from cattle and goats. They described a higher microbiome diversity in goat pasteurised cheeses than in the unpasteurised cow ones. Predominance of bacteria from the *Bacillacea* family was reported in goat cheeses (~50% of the total abundance). Besides, *Staphylococcus* microorganisms were found in high abundance (5–25%) in goat cheese samples (Walsh et al., 2020). Penland et al. (2021) characterised the

microbiota of Pelardon goat cheese and reported substantial changes during the ripening phase, and they related such changes with the development of the organoleptic properties of cheese. Noteworthy, Pelardon cheese harboured *Lactococcus lactis*, associated with cheese acidification, *L. mesenteroides* and *geotrichum candidum* associated with amino acid catabolism, and *L. paracasei*, *Enterococcus faecalis* and *Penicillium commune*, that were highly abundant during ripening and might have an environmental origin (Penland et al., 2021). Despite these initial efforts, there is a lot of work to do in the characterisation of the microbiome of goat cheeses. In this regard, implementing meta-transcriptomics and meta-genomics approaches to the microbiological analysis of caprine cheese might help to better understand the dynamics of microorganisms in diverse cheese environments and their impact on cheese features and properties.

1.6 FACTORS INFLUENCING BIOLOGICAL VIABILITY IN GOATS AND OTHER LIVESTOCK.

Artificial selection applied within livestock breeding programs traditionally has been focused on increasing productive traits, not devoting so much attention to traits associated with welfare, adaptation, viability, health and disease resistance, among others (Goddard 2009). Often, the intensity of selection leads to the usage, as sires and dams, of few individuals with superior phenotypic performance. Nowadays, livestock breeding programs are giving high importance to functional traits, including those related to health, longevity, and survival, as a strategy to make animal production more sustainable. Indeed, fitness and viability traits have a high impact on reproductive performance and, in consequence, on the economic profit of farms (Mellado et al., 2006). Viability traits are affected by the genetics of animals as well as by non-genetic determinants including nutrition, health and management amongst others.

1.6.1 The genetics of viability traits.

Viability traits reflect the reproductive and survival rate of individuals (Goddard, 2009). Mucha et al. (2022) retrieved data from different studies reporting genetic parameters of viability traits in sheep and goats. Moderate heritabilities were estimated in both species with regard to somatic cell score (SCS), with pooled heritabilities of 0.21 ± 0.01 in goats and 0.13 ± 0.02 in sheep. The increment of SCS is an indicator of the presence of mastitis and it has a negative impact on the welfare of the animals while causing economic losses (Pérez-Méndez et al., 2020). The abundance of somatic cells is highly correlated with the functional longevity of dairy animals and with culling in cases of high SCS (Sewalem et al., 2006). Negative correlations between SCS and milk composition (-0.19 ± 0.01 for fat content and -0.06 ± 0.05 for protein content) were reported in dairy goats. Other traits involved in longevity and resilience are fecal egg count (h^2 = 0.07 ± 0.01 in goats and h^2 = 0.29 ± 0.14 in sheep), parasitism antibodies (h^2 = 0.18 ± 0.07 in sheep) and parasitism immunoglobulins (h^2 = 0.36 ± 0.06 in sheep). These heritabilities were estimated in the meta-analysis performed by Mucha et al. (2022), which had access to a higher amount of data from sheep than from goats and considered mainly cosmopolitan dairy goat breeds. In Creole goats from Guadeloupe, heritabilities for other viability and resilience traits have been reported, i.e. h^2 of 0.13 ± 0.05 for the packed cell volume (proportion of red cells in blood, a proxy to anaemia produced by haemotrophics parasites) and of 0.18 ± 0.04 for faecal egg count (Gunia et al., 2011). With regard to reproductive traits, Creole goats displayed heritabilities for litter size of 0.11 ± 0.02 (Gunia et al., 2011). The age at kidding has a reported h^2 of 0.08 ± 0.02, and this trait displays negative genetic correlations with milk yield in cosmopolitan dairy goats (Desire et al., 2018). Other traits analysed by Desire et al. (2018) are reproduction out of season (h^2 = 0.11 ± 0.02) and pseudopregnancy (h^2 = 0.11 ± 0.02).

1.6.2 The measurement of inbreeding and inbreeding depression.

Inbreeding, defined as the mating of close relatives, creates an excess of homozygosity in the offspring population that can have adverse consequences on any trait, but mostly on phenotypes related with reproduction and viability (Agarwal and Whitlock, 2012). The increase of inbreeding augments the frequency, in the population, of genotypes homozygous for deleterious alleles that segregate at low frequencies (Charlesworth and Willis, 2009). Deleterious mutations can disrupt gene function, thus causing diseases or even the death of the individual at any developmental stage (Henn et al., 2015). In the case of haploinsufficient genes, heterozygous individuals are likely to suffer common or complex diseases related with the function of the affected gene (Henn et al., 2015).

Breeding programs have increased the selection pressure over traits of economic interest, like milk traits in dairy animals. Dairy cattle herds subjected to artificial selection display an increment of inbreeding across time, as shown in reports made by the Canadian Dairy Network (CDN, https://lactanet.ca/en/inbreeding-update-august-2021/). Artificial selection also influences the genomic patterns of homozygosity (e.g. frequency of regions of homozygosity and their genomic location), as Kim et al. (2013) reported when comparing selected and unselected cattle.

The availability of high throughput genotyping and sequencing methods allows the characterization of homozygosity patterns along the genome and the estimation of inbreeding coefficients using SNP data independently of the availability of pedigree data (Ferenčaković et al., 2013). The inbreeding coefficient can be defined as the probability that two randomly selected alleles at a given locus are identical by descent or IBD (Wright, 1922). Traditionally, inbreeding coefficients were estimated from pedigree data

(F_{PED}), thus requiring reliable records and a genealogical depth including as many ancestral generations as possible to accurately infer coefficients (Keller et al., 2011). Notably, pedigree-based methods assume no relationship between individuals of the base population and do not consider the effects of selection and demographic events that affect population diversity (Keller et al., 2011). Molecular methods, in contrast, can estimate inbreeding coefficients on a SNP by SNP basis, measured from deviations from Hardy-Weinberg proportions, i.e. F_{YAN} (Li and Horvitz, 1953) or from the diagonal elements of the kinship matrix estimated from genomic markers, i.e. F_{VR}, $F_{L\&H}$ and F_{NEJ} (Nejati-Javaremi et al., 1997; VanRaden, 2008; Yang et al., 2011). Runs of homozygosity (ROH) are genomic regions with homozygous genotypes in all nucleotide positions (Gibson et al., 2006). Thus, molecular inbreeding coefficients can be also estimated by measuring the fraction of the genome covered by ROH, i.e. F_{ROH} (McQuillan et al., 2008). The relationship between F_{ROH} and F_{PED} has been analysed in studies carried out in different domestic species, finding a moderate to high correlation (Peripolli et al., 2017). Currently, the F_{ROH} coefficient is frequently used as an indicator of inbreeding. It is considered a more accurate tool than the F_{PED} (Keller et al., 2011), and this is because F_{ROH} has the advantage of not being affected by pedigree depth. However, its magnitude can be affected by the density of markers (Ceballos et al., 2018) and its measurement involves a great economic cost compared to the estimation of F_{PED} from pedigrees (Zhang et al., 2015b). Molecular inbreeding coefficients have been widely used in livestock to assess the inbreeding levels of the populations and study the effect of inbreeding on productive traits (Purfield et al., 2012, 2017; Saura et al., 2015; Bertolini et al., 2018; Peripolli et al., 2017; Caballero et al., 2021).

In goat dairy farms, a long productive life of the milked goats is desirable to maximise economic profit (Ithurbide et al., 2022). Indeed, the cost of maintaining individuals with

low production performances in the herd can be economically unsustainable for the farmer. An analysis of 86 herds of dairy goats in France evidenced that the main causes of culling are low production followed by udder health problems, high incidences of abortion and locomotor problems (Malher et al., 2001). A meta-analysis of multiple livestock species considering different traits indicated the existence of a negative impact of inbreeding depression on productive traits, diminishing 0.35% of the production for multiple traits (e.g. milk yield, somatic cells count, protein yield, fat yield, litter weight) with regard to the population mean per each 1% increment on inbreeding (Leroy, 2014). The effects of inbreeding reported by Leroy et al. (2014) on reproduction/survival, conformation and weight/growth traits are -0.22, -0.09 and -0.24 % of the average per each 1% increment of inbreeding. In dairy cattle, it has been reported that a 1% increase in F_{ROH} is associated with a 36.3 kg decrease in 305-day milk yield, a 0.48 day increase in calving interval and a 0.86 unit increase in somatic cell score for day 150 through to 400. The magnitude of these changes were equivalent to − 0.45, 0.12 and 0.05% of the trait means, respectively (Doekes et al., 2019). From this study, it was also evident that recent inbreeding was more harmful than ancient inbreeding (Doekes et al., 2019). In another study, Ercanbrack and Knight (1991) analysed the consequences of inbreeding on the reproduction and wool production of Rambouillet, Targhee, and Columbia sheep. They found that the mean loss of profitability per ewe in value of production was $17 for average inbreeding and as high as $36 for inbreeding approaching 50%. These results highlight the importance of minimising the incidence of inbreeding in livestock in order to reduce economic losses and improve animal welfare.

Inbreeding depression has been characterised in Murciano-Granadina goats for production traits using pedigree information (Deroide et al., 2016). A decrease in total milk production was detected when inbreeding levels reached 10.59%, while no effect

was reported for milk composition traits. Deroide et al. (2016) reported a low number of inbred animals and low inbreeding levels in Murciano-Granadina goats. In Saanen goats, an increment of 1% of inbreeding represented a reduction of 2.31 kg of milk at 305 days of lactation (Paiva et al., 2020).

1.6.3 Detection of deleterious mutations through the analysis of transmission ratio distortion.

In diploid organisms, allelic transmission from one generation to another is expected to be Mendelian, meaning that when an individual carries two different alleles at a given locus both are expected to be equally transmitted to the offspring. A deviation from this ratio is produced when one of the two alleles is preferably inherited by the offspring, and this phenomenon is called transmission ratio distortion or TRD (Pardo-Manuel de Villena Fand Sapienz, 2001; Fishman and Mcintosh, 2019). An allele might not be efficiently transmitted to the next generation at any reproductive or developmental stage including gametogenesis, meiosis, fertilisation and zygote and fetal development (Huang et al., 2013). The lack or lowered transmission of a given allele is usually due to the fact that it harbours a deleterious mutation affecting the function of an essential gene (Huang et al., 2013). Deleterious mutations can be targeted by gamete killers that disable male sperm carrying deleterious mutations, like in the case of the t-haplotype in mice (Fishman and Mcintosh, 2019). Other sources of TRD are embryo lethality and mother-fetal incompatibility (Huang et al., 2013). The detection of TRD is usually performed in trios or extended family analyses, where allelic transmission ratios from parents and offspring are measured to identify deviations from the Mendelian ratio (Labbe et al., 2013). Transmission ratio distortion has been studied mostly in humans and model species like mice and *Drosophila melanogaster* (Huang et al., 2013), while in livestock little has been

reported so far (Casellas et al., 2017; Abdalla et al., 2020; Lahoucine et al., 2020; Vázquez-Gómez et al., 2020). In cattle, several genes close to TRD signals are related with postnatal (*TNS3*) and embryonic lethality (*HUS1*) as well as with mitochondrial architecture maintenance (*GTPBP10*) (Casellas et al., 2017).

Interestingly, Casellas et al. (2014) developed a flexible Bayesian algorithm for testing the occurrence of transmission ratio distortion at the genome-wide level, which was subsequently refined to analyse haplotypes rather than single SNPs (Id-Lahoucine et al., 2019) and to discriminate between allele and genotype-specific transmission ratio distortion as well (Casellas et al., 2020). This methodology has been successfully used to detect TRD in multiple species such as cattle, pigs and turkey (Casellas et al., 2017; Abdalla et al., 2020; Gòdia et al., 2020; Lahoucine et al., 2020; Vázquez-Gómez et al., 2020).

1.6.4 Maternal care has an important influence on offspring viability.

Maternal care has a strong influence on the biological viability of the offspring. The contribution of the parents to the offspring takes place in two dimensions, encompassing the genetic load (the transmission of two states of paternal and maternal alleles) and the environment provided fundamentally by the mother. This maternal environment is determined to a certain degree by genetic factors influencing parental behaviour, nutrition, and any other aspects fundamental for offspring survival (Reinhold, 2002; Lévy and Keller, 2009).

Parental care is a major determinant of offspring survival, and in ruminants it is provided exclusively by the mother (Dwyer, 2014). In natural populations, maternal care has a

direct effect on offspring viability, through milking, and an indirect effect on survival through the mother-kid bond that provides protection against predators, shelter and knowledge of survival strategies (Théoret-Gosselin et al., 2015). In domestic goats and sheep, maternal care after birth comprises licking and grooming, vocalisation and assistance to help the kid finding the udder and feeding as well as acceptance of the own offspring while rejecting others to keep the milk for its offspring (Bordi et al., 1994; Dwyer, 2014). In summary, maternal care is a key factor determining the viability of the offspring in terms of immunological defences, growth and social behaviour (Merlot et al., 2013).

Maternal behaviour in mammals develops before the parturition through changes in physiology and sensory perception (Poindron et al., 2007a). In the framework of sensory perception, the olfactory function plays an important role (Romeyer et al., 1994). Studies focused on sheep and goats report that olfactory stimuli from amniotic fluid at the moment of parturition trigger a series of chemical responses at the neural and endocrine levels that lead to maternal responsiveness and selectivity (Poindron et al., 2007a). In goats, olfactory inputs from parturition and offspring odour are perceived by the vomeronasal organ during licking and transmitted to the olfactory bulb, where a series of neurochemical events elicit substantial behavioural changes related with maternity (Poindron et al., 2007a, 2007b). As reported during pregnancy and after parturition in mice (Navarro-Moreno et al., 2020), the olfactory bulb displays high plasticity and continuous neurogenesis related with the development of olfactory memory, odour discrimination and processing of olfactory cues (Belnoue et al., 2016). In sheep, the damage of noradrenergic projections of olfactory bulb neurons hinders offspring recognition and facilitates the acceptance of alien kids (Pissonnier et al., 1985), while in mice, the removal of the olfactory bulb in females induces a lack of maternal behaviour

and rejection plus cannibalism of the pups (Gandelman et al., 1971). These results highlight the role of olfaction, and consequently of the olfactory bulb, in maternal behaviour and offspring survival.

Besides the olfactory bulb, several other brain regions display changes during pregnancy. In humans, a reduction of grey matter, associated with social behavioural changes, occurs during pregnancy and lasts up to two years after parturition (Hoekzema et al., 2017). The main cerebral regions displaying changes activated by pregnancy are the cerebral cortex and hippocampus (Hoekzema et al., 2017). Ray et al. (2016) described the transcriptomic changes in the hippocampus, hypothalamus, cerebellum and neocortex of virgin, pregnant and post-partum mice. In this study, the neocortex presented the most remarkable changes in gene expression when comparing individuals from different stages. The gene expression patterns of hippocampus and hypothalamus were also affected by pregnancy, while the cerebellum remained stable (in terms of gene expression) in pregnant and non-pregnant mice (Ray et al., 2016). These studies highlight that gestation has an important impact on the mRNA expression profile of different brain regions.

Although maternal behaviour is probably explained by changes in brain gene expression, very little information is available about the expression profiles of different brain regions in goats or any other livestock species. The lack of this fundamental information makes it difficult to dissect the molecular basis of maternal behaviour. In sheep, goats and cattle, a few brain tissues have been characterised when building atlases of gene expression at multiple organs (Harhay et al., 2010; Clark et al., 2017; Muriuki et al., 2019). A more systematic study was undertaken by Sjöstedt et al. (2020) in pigs. In this work, the cerebellum was the tissue with the most divergent profile of gene expression, and the olfactory bulbs from pigs and mice grouped closely together, while the olfactory bulb

from humans behaved as an outlier and showed similarities with the cerebrum regions of humans (Sjöstedt et al., 2020).

2 OBJECTIVES

This PhD thesis was performed within the framework of project "CAPRAMUR" PID2019-105805RB-I funded by the European Regional Development Fund (FEDER)/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación. This project is divided in two main work packages. One of them is dedicated to explore the genomic basis of dairy and morphological traits in Murciano-Granadina goats as well as to characterise the caprine cheese microbiome. The second workpackage investigates the genomic basis of traits related with biological viability and maternal behaviour in goats.

The specific goals of this thesis are:

- To identify genomic regions associated with milk and morphology traits in Murciano-Granadina goats.
- To describe the microbiota of a representative set of goat cheeses.
- To characterise the levels of inbreeding of Murciano-Granadina goats by using molecular coefficients and to assess the existence of inbreeding depression for phenotypes of economic interest.
- To map polymorphisms with potentially harmful effects on viability by identifying genomic regions displaying transmission ratio distortions in sire-offspring families.
- To determine the effect of early pregnancy on the mRNA expression profile of 12 brain regions by performing a differential expression analysis in pregnant and non-pregnant Murciano-Granadina goats.

3 PAPERS AND STUDIES

RUNNING HEAD: GWAS for dairy traits in goats

Single and longitudinal genome-wide association studies for dairy traits available in goats with three recorded lactations

Maria Gracia Luigi-Sierra[1], Amparo Martínez[2], Martina Macri[2], Juan Vicente Delgado[2], Anna Castelló[1,3], Javier Fernández Alvarez[4], Xavier Such[5], Jordi Jordana[3], Marcel Amills[1,3*]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.
[2]Departamento de Genética, Universidad de Córdoba, Córdoba 14071, Spain;
[3]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.
[4]Asociación Nacional de Criadores de Caprino de Raza Murciano-Granadina (CAPRIGRAN), 18340 Granada, Spain.
[5]Group of Research in Ruminants (G2R), Department of Animal and Food Science, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona 08193, Spain.
[7]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.

Corresponding author: *

In preparation

Abstract

Background: Milk yield and composition phenotypes are systematically recorded across several lactations in goats, but the majority of GWAS performed so far have rather ignored the longitudinal nature of such data. Moreover, comparisons of GWAS results across different lactations have not been carried out in goats despite the fact that milk traits might be affected by genetic determinants with temporal patterns of expression. Here, we have used two different GWAS approaches to analyse data from three lactations recorded in Murciano-Granadina goats. In the first one (Analysis 1), independent GWAS have been carried out for each trait and lactation, while a single longitudinal GWAS jointly considering all data has been performed in the second one (Analysis 2).

Results: In both analyses, genome-wide significant QTL for lactose percentage on chromosome 2 (129.77 – 131.01 Mb) and for milk protein percentage on the chromosome 6 (74.8 – 94.6 Mb) casein gene region were detected. In Analysis 1, a substantial number of QTL were not replicated in all three lactations due to the existence of lactation-specific genetic determinants. In Analysis 2, we identified several genome-wide significant QTL related to milk yield and protein content that were not uncovered in Analysis 1.

Conclusion: The increased discovery of QTL in Analysis 2 indicates that longitudinal GWAS is particularly well suited for the genetic analysis of dairy traits. Moreover, our data confirm that the polymorphism of the casein complex is the main genetic determinant of milk protein content in Murciano-Granadina goats.

Keywords: GWAS, milk yield and composition traits, Murciano-Granadina breed

Main text

The performance of genome-wide association studies (GWAS) has been helpful to identify genomic regions associated with milk traits in several caprine breeds including New Zealand goats [1], French Alpine and Saanen [2, 3] and a composite breed of Saanen, Toggenburg, and Alpine goats [4]. One of the most consistently identified regions associated with milk yield maps to chromosome 19 at 26 Mb, while another chromosome 6 region encompassing the casein cluster has been associated with milk protein content [1, 3, 4]. In a previous study, Guan et al. [5] performed a GWAS for dairy traits recorded in 822 Murciano-Granadina goats during a single lactation, and they found three genome-wide significant associations. The location of the most significant QTL detected by Guan et al. [5] was coincident with the cluster of the four genes encoding casein α_{s1} (*CSN1S1*), casein α_{s2} (*CSN1S2*), casein β (*CSN2*) and casein κ (*CSN3*). The goal of the current study was to expand the reach of the GWAS carried out by Guan et al. [5] to dairy traits measured in three consecutive lactations. First, we aimed to compare the positional concordance of the QTL identified for each specific lactation by performing independent GWAS for each lactation (Analysis 1), and second we wanted to carry out a longitudinal GWAS (Analysis 2) and compare its results with those obtained in Analysis 1.

Milk production and composition phenotypes for three lactations were recorded in Murciano-Granadina goats distributed on 15 farms in Andalusia (Spain). A total of 917 individuals had phenotypes available for the first lactation, while 805 and 660 had records for the second and third lactations, respectively. The list of measured phenotypes included milk yield in kilograms at 210 days (MY210), 240 days (MY240) and 305 days (MY305) of lactation as well as milk fat (Fat), protein (Protein), lactose (Lactose) and dry matter (DM) percentages and the natural logarithm of the somatic cell count divided by 1000 (SCS). Milk composition traits were standardised to a lactation of 210 days. The

distribution of each phenotype was depicted in histograms (**Supplementary Figures 1, 2 and 3)** and normality was tested with the Shapiro test [6]. Traits deviating significantly from normality, i.e. protein, fat, dry matter and lactose percentages, were rank-based transformed using GenABEL in R [7]. Summary statistics for each trait and lactation are displayed in **Supplementary Table 1**. Pearson phenotypic correlations (r_p) between traits were estimated with the R software [8]. Positive and significant correlations were observed between the same trait measured in different lactations (**Supplementary Figure 4 and Supplementary Table 2**). As expected (**Supplementary Figure 4**), milk yield and composition traits showed low to moderate negative correlations (r = -0.05 to -0.2).

Blood samples were collected using vacuum tubes with EDTA K_3 and stored at –20ºC until processing. Genomic DNA was isolated using a modified salting-out protocol as explained in Guan et al. [5]. Goats were genotyped with the Goat SNP50 BeadChip [9] which features 53 347 SNP probes distributed across the whole goat genome, following the instructions of the manufacturer (Illumina Inc., San Diego, CA, USA). Genomic positions and single nucleotide polymorphism (SNP) identifiers were updated using PLINK 1.9 [10] and the ARS1 goat genome was used as reference [11]. The filtering of the data was performed using PLINK 1.9 [10] by removing (1) SNPs with missing genotypes in more than 10 % of the samples, (2) SNPs with a minor allele frequency below 0.01, (3) SNPs deviating significantly from the Hardy-Weinberg equilibrium (*P*-value $\leq 1x10^{-5}$). Besides, individuals with a SNP missing rate over 10% were removed from the dataset. After quality control and filtering of the data, a total of 48,785 SNPs and 917 (lactation 1), 805 (lactation 2) and 660 (lactation 3) goats were retained for downstream analyses.

To identify SNPs significantly associated with the analysed traits, we adjusted a univariate linear mixed model (LMM) using two approaches, (1) Independent GWAS

were carried out for each lactation, and (2) A single GWAS combining data from the three lactations was performed (longitudinal analysis).

Analysis 1 (three lactations considered independently): here, we carried out three independent GWAS, one for each lactation, by using GEMMA [12]. The model was defined by the formula:

$$y = W\propto + X\beta + \mu + \epsilon$$

where *y* is a vector of phenotypic records from 917 (lactation 1), 805 (lactation 2), and 660 (lactation 3) goats; W = (w1, …, wc) is an n × c matrix of three fixed effects (farm, with 16 levels for lactation 1 and 15 levels for lactations 2 and 3; year of birth, with 10 levels; litter size, with 5 levels); α is a c-vector of the corresponding fixed effects including the intercept; X is a n-vector of marker genotypes; β is the effect size of the marker (allele substitution effect); u is a n-vector of random individual genetic effects with a normal distribution $u \sim N(0, \lambda \tau^{-1} K)$, where τ^{-1} is the variance of the residual error, λ is the ratio between the 2 variance components, and K is the relatedness matrix derived from SNP genotypes. Finally, ϵ is an n-vector of errors. We corrected the results for multiple testing using the false discovery rate (FDR) method [13], and the significance threshold was set at a *q*-value ≤ 0.05. Results were visualised using custom scripts implemented in the ggplot2 package [14] on R [8].

This approach made it possible to identify several associations that reached the genome-wide significance level (**Table 1 and Figure 1**). The most relevant result corresponded to one region on chromosome 6 (74.8 – 94 Mb) which was consistently associated (*q*-value < 0.05) with milk protein percentage in the three lactations (**Figure 1; Table 1**). In contrast, several associations were not replicated in different lactations. When analysing lactation 1 data, for instance, we found several genetic markers significantly associated with lactose percentage on chromosome 2 (125.96, 129.77 – 131.01 Mb) that did not yield

genome-wide significant associations with such trait in lactations 2 and 3. Similarly, when using data from lactation 2 we identified an association between markers from the chromosome 6 (81.08 – 87.41 Mb) region and dry matter. At the chromosome-wide level of significance, this lack of replicability was even more evident. In the case of fat percentage, for instance, a QTL on chromosome 18 (15.92 – 15.97 Mb) was identified when using data from lactation 1. For the same trait, a QTL on chromosome 25 (36.46 Mb) and three QTL on chromosomes 2 (112.5 Mb), 11 (45.34 Mb) and 19 (34.33 –34.41; 53.67 Mb) were identified when analysing data from lactations 2 and 3, respectively **(Supplementary Table 3**). A non-mutually exclusive explanation for the lack of consistency across lactations would be the existence of genetic determinants with temporal patterns of expression. Indeed, Cho et al. [15] performed a GWAS for estimated breeding values for milk production traits recorded in Holstein cattle from the 1st to 4th lactations and identified many associations in lactation 1 that could not be replicated in subsequent lactations.

Analysis 2 (longitudinal analysis): In this second analysis, we did a longitudinal GWAS taking into account the joint phenotypic data from the three lactations. Herewith, we employed the rGLS function from the RepeatABEL package [16] included in the GenABEL suite [7]. The model is very similar to the one implemented in GEMMA [11], and it is defined as follows:

$$y = W\mu + x_{snp}\beta_{snp} + Zg + \epsilon$$

where y is a vector of phenotypic records from 660 goats with three lactations; W is a matrix of three fixed effects (farm, with 15 levels; year of birth, with 10 levels; litter size, with 5 levels); μ is a vector of the corresponding fixed effects including the intercept; x_{snp} is a vector of the genotype dosage for each genomic position (0, 1 or 2); β_{snp} is the effect size of the marker (allele substitution effect); Z is an incidence matrix relating the

individuals to their observed values; g is a vector of random individual genetic effects with a normal distribution $g \sim N(0, \sigma^2_g K_n)$, where K is the relatedness matrix derived from SNP genotypes and n the number of individuals. Finally, ϵ is an n-vector of errors defined as $\epsilon \sim N(0, \sigma^2 \epsilon I_N)$, where I is the identity matrix and N the total number of observations. The false discovery rate method [13] was used to correct for multiple testing and ggplot2 [14] was employed to visualise the results of the GWAS.

This longitudinal analysis yielded genome-wide significant associations between one chromosome 6 (78.51-93.50 Mb) region containing the casein genes and milk protein and dry matter contents as well as between one chromosome 2 region (129.80-130.75 Mb) and lactose percentage, thus confirming the results obtained in Analysis 1 (**Table 1; Supplementary Table 4 and Figure 1**). However, with the longitudinal analysis it was possible to identify a larger number of SNPs significantly associated with milk traits e.g. one chromosome 6 region (17.02 Mb) was associated with milk yield at 210 and 240 days, and two regions on chromosomes 4 and 9 were associated with protein percentage. Detailed information about the effect of each SNP within the QTL regions can be found in **Supplementary Table 4**. This increased capacity of the longitudinal analysis to detect QTL is probably due to the fact that the repeated measurements model implemented in RepeatABEL has more statistical power than a model using single or averaged measurements [16]. At the chromosome-wide level of significance, the longitudinal analysis made it possible to detect QTL for fat percentage on chromosomes 2 (112.55 Mb), 20 (31.78 and 53.61 Mb) and 27 (19.79 Mb). Besides, the protein content QTL on chromosome 6 mapping to the casein cluster was significantly associated with somatic cell score (**Supplementary Table 5**).

Figure 1. Negative $\log_{10}$ q-values (y-axis) of the associations between SNPs and milk production and composition traits are plotted against the genomic location of each SNP marker (x-axis). Markers on different chromosomes are indicated with different colors. Two analyses have been performed: (**A**) GWAS are independently done for each of the three lactations; and (**B**) A single longitudinal GWAS jointly considering the three lactations is carried out. Only traits with genome-wide significant results are plotted. The dashed red line indicates the genome-wide level of significance and it corresponds to -$\log_{10}$(q-value = 0.05) = 1.30.

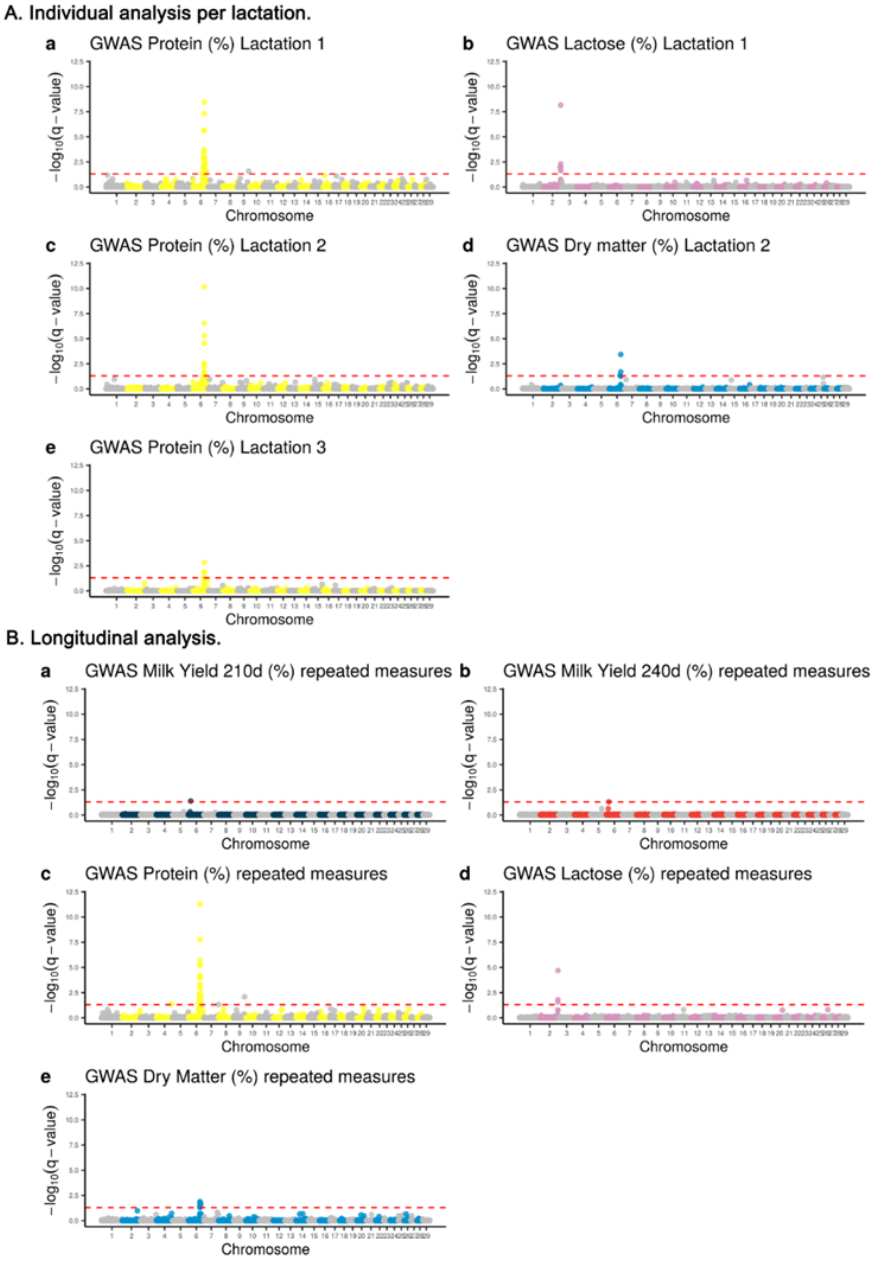


Table 1. Genome-wide significant SNPs associated with milk traits recorded in Murciano-Granadina goats with three available lactations (allele frequencies, substitution effects and statistical significance correspond to the lead SNP).

Univariate analysis (3 lactations considered independently)										
Trait	**Lead SNP**	**Chr[1]**	**Position, Mb**	**# SNPs[2]**	**AF[3]**	**A1[4]**	**A0[5]**	**β ± SE[6]**	***p*-value**	***q*-value**
Lactation 1										
Protein, %	rs268290908	6	74.8 – 94.6	40	0.42	A	G	-0.36 ± 0.05	7.47E-14	3.67E-09
	rs268234071	6	99.76	1	0.29	G	A	-0.22 ± 0.05	1.72E-05	0.026
	rs268258054	6	103.34	1	0.13	G	A	-0.29 ± 0.07	1.76E-05	0.026
	rs268268932	9	82.43	1	0.21	A	G	-0.26 ± 0.06	1.67E-05	0.026
Lactose, %	rs268253126	2	125.96	1	0.27	A	G	-0.27 ± 0.05	8.30E-07	0.01
	rs268253425	2	129.77 – 131.01	5	0.21	A	G	-0.44 ± 0.06	1.42E-13	7E-0.9
Lactation 2										
Protein, %	rs268290908	6	85.57 – 87.85	11	0.43	A	G	-0.41 ± 0.05	1.40E-15	6.87E-11
Dry matter, %	rs268260283	6	81.08	1	0.28	A	G	0.28 ± 0.06	2.87E-06	0.04
	rs268290908	6	86.85 – 86.90	2	0.43	A	G	-0.3 ± 0.05	7.50E-09	0.0003
Lactation 3										
Protein, %	rs268290908	6	83.2 – 86.9	5	0.43	A	G	-0.3 ± 0.05	8.17E-08	0.004
Longitudinal analysis (3 lactations considered jointly)										
Trait	**Lead SNP**	**Chr[1]**	**Position, Mb**	**# SNPs[2]**	**AF[3]**	**A1[4]**	**A0[5]**	**β ± SE[6]**	***p*-value**	***q*-value**
Protein, %	rs268288251	4	113.17	1	0.38	A	C	0.21 ± 0.05	2.93E-05	0.044
Protein, %	rs268290909	6	78.51 – 93.50	32	0.06	A	G	-0.37 ± 0.04	1.11E-16	5.46E-12
Protein, %	rs268235611	7	107.74	1	0.34	A	G	0.18 ± 0.04	3.39E-05	0.048
Protein, %	rs268268930	9	82.51	1	0.34	A	G	-0.24 ± 0.05	3.29E-06	0.008
MY210, Kg	rs268284580	6	17.02	1	0.2	A	G	41.28 ± 8.38	8.40E-07	0.041
MY240, Kg	rs268284580	6	17.02	1	0.2	A	G	48.13 ± 9.83	9.94E-07	0.048
Dry matter, %	rs268290909	6	86.20 – 86.94	6	0.06	A	G	-0.21 ± 0.04	2.68E-07	0.013
Dry matter, %	rs268273385	6	92.85	1	0.42	A	G	-0.28 ± 0.04	1.36E-06	0.022
Lactose, %	rs268253426	2	129.80 - 130.75	3	0.29	G	A	-0.30 ± 0.05	4.11E-10	2.02E-05

[1]**Chr**, chromosome; [2]**#SNPs**, number of SNPs within a QTL region significantly associated with a specific dairy trait; [3]**AF**, alternative allele frequency; [4]**A1**, alternative allele; [5]**A0**, reference alelle; [6]**β ± SE**, allele substitution effect ± standard error.

Based on the linkage disequilibrium estimates obtained by Guan et al. [5] in the same goat population, we retrieved all protein-coding genes mapping within a 1 Mb window around the significantly associated SNPs taking as a reference the NCBI ARS1 reference genome (GCF_001704415.1). Genes were functionally annotated using David bioinformatics tools [17] with goat as reference database (containing specie-specific data from NCBI, Uniprot, Ensembl, Gene Ontology, KEGG, among others). Both analyses 1 and 2 provided very strong evidence about the key role of casein gene variability on the determinism of milk protein content in Murciano-Granadina goats. Caseins are the majority proteins in milk and several polymorphisms in the *CSN1S1*, *CSN1S2*, *CSN2* and *CSN3* genes have been reported to be associated, often causally, with milk protein content as well as with many other dairy and cheese traits (reviewed by Amills et al. [18]). Our results agree well with what was published by Guan et al. [5] in the same breed as well as by Martin et al. in Alpine and Saanen goats [3]. Regarding the chromosome 2 QTL for lactose percentage, it is worth mentioning that the very same region was identified by Costa et al. [19] as associated with milk lactose content in Fleckvieh cattle. This region contains the ORMDL sphingolipid biosynthesis regulator 1 (*ORMDL1*) gene, which is involved in the negative regulation of the synthesis of ceramides which are necessary for the production of sphingolipids [20]. Noteworthy, galactose can be a component of sphingolipids [21] and it is also a key precursor in the synthesis of lactose. Another locus of interest mapping to the chromosome 2 QTL for lactose percentage is the inositol polyphosphate-1-phosphatase (*INPP1*) gene, which is involved in phosphate and phospholipid metabolism and described, in cattle, as

highly associated with lactose content and with effects on the protein and mineral composition of milk [22]. The hydroxyacyl-CoA dehydrogenase (*HADH*, chromosome 6: 17,529,001–17,573,045) gene, mapping close to QTL (chromosome 6:17.02 Mb) associated with milk yield at 210 and 240 days of lactation, and the 3-hydroxyisobutyryl-CoA hydrolase (*HIBCH*, chromosome 2: 130,422,688–130,535,003) gene mapping close to the QTL (chromosome 2:129.8–130.7 Mb) associated with lactose percentage, are both involved in valine degradation. In dairy cattle, diets rich in valine significantly increase milk yield [22, 23].

Conclusions

Independent GWAS based on data from lactations 1, 2 and 3 made it possible to detect a common chromosome 6 QTL for milk protein content mapping to the casein cluster, while several additional QTL were not replicated across lactations (e.g. the QTL for lactose percentage on chromosome 2 only reached genome-wide significance in lactation 1). The implementation of a longitudinal GWAS integrating data from the three lactations confirmed the QTL for protein content on chromosome 6 as well as a QTL for lactose on chromosome 2, but it also uncovered several QTL not identified in the three separate GWAS. This finding is consistent with the increased statistical power of longitudinal GWAS (when compared to the non-longitudinal ones) and supports its widespread use in the genetic analysis of dairy traits. Finally, the strong consistency of the milk protein content QTL located on the chromosome 6 region that harbours the casein cluster evidences the

fundamental role of casein polymorphism on the determination of milk protein and dry matter contents in Murciano-Granadina goats.

Acknowledgements

Many thanks to CAPRIGRAN for carrying out phenotype recording and blood sample collection in Murciano-Granadina goats.

Funding

This research was funded by the European Regional Development Fund (FEDER)/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación/Project Reference grant: PID2019-105805RB-I00 and by the CERCA Programme/Generalitat de Catalunya. We also acknowledge the support of the Spanish Ministerio de Ciencia e Innovación for the Center of Excellence Severo Ochoa 2020–2023 (CEX2019–000902-S) grant awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). We also acknowledge the support of the CERCA programme of the Generalitat de Catalunya. Maria Luigi-Sierra was funded with a PhD fellowship Formación de Personal Investigador (BES-C-2017-079709) awarded by the Ministerio de Economía y Competitividad (Spain).

Contributions

MA, JJ, XS and JVD designed the study. JFA, JVD and AM coordinated all tasks involved in phenotype recording. AM carried out the DNA extractions. AC did

genotyping tasks. MGLS carried out all data analyses. MGLS and MA wrote the first draft of the paper. All authors read and approved the content of the paper.

Ethics declarations

Ethics approval and consent to participate

Blood collection is a routine procedure performed by CAPRIGRAN, so it does not require approval by the Ethics Committee on Animal and Human Experimentation of the Universitat Autònoma de Barcelona.

Consent for publication

Not applicable.

Competing interest

The authors declare that they have no competing interests.

References

1. Scholtens M, Jiang A, Smith A, Littlejohn M, Lehnert K, Snell R, et al. Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. J Anim Sci Biotechnol. BioMed Central Ltd.; 2020;11:1–14.

2. Talouarn E, Bardou P, Palhière I, Oget C, Clément V, Tosser-Klopp G, et al. Genome wide association analysis on semen volume and milk yield using different strategies of imputation to whole genome sequence in French dairy goats. BMC Genet [Internet]. BioMed Central Ltd; 2020 [cited 2022 Jul 22];21:1–13. Available from: https://bmcgenomdata.biomedcentral.com/articles/10.1186/s12863-020-0826-9

3. Martin P, Palhière I, Maroteau C, Bardou P, Canale-Tabet K, Sarry J, et al. A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. Sci Reports 2017 71. Nature Publishing Group; 2017;7:1–13.

4. Mucha S, Mrode R, Coffey M, Kizilaslan M, Desire S, Conington J. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. J Dairy Sci. 2018;101:2213–25.

5. Guan D, Landi V, Luigi-Sierra MG, Delgado JV, Such X, Castelló A, et al. Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. J Anim Sci Biotechnol. 2020;

6. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika [Internet]. Oxford Academic; 1965 [cited 2022 Jul 29];52:591–611. Available from: https://academic.oup.com/biomet/article/52/3-4/591/336553

7. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: An R library for genome-wide association analysis. Bioinformatics. 2007;

8. R Core team. R Core Team. R A Lang Environ Stat Comput R Found Stat Comput , Vienna, Austria ISBN 3-900051-07-0, URL http//wwwR-project.org/ [Internet]. 2015;55:275–86. Available from: http://www.mendeley.com/research/r-language-environment-statistical-computing-96/%5Cnpapers2://publication/uuid/A1207DAB-22D3-4A04-82FB-D4DD5AD57C28

9. Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and characterization of a 52K SNP chip for goats. PLoS One. 2014;9.

10. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience. 2015;4.

11. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet. 2017;49:643–50.

12. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4.

13. Benjamini, Yoav ; Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995.pdf. J R Stat Soc Ser B [Internet]. 1995;57:289–300. Available from: http://www.jstor.org/stable/2346101

14. Wilkinson L. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. Biometrics. 2011;67.

15. Cho KH, Oh JD, Kim HB, Park K Do, Lee JH. Genome Wide Association Studies Using Multiple-lactation Breeding Value in Holsteins. Asian-Australasian J Anim Sci [Internet]. Asian-Australas J Anim Sci; 2015 [cited 2022 Aug 9];28:328–33. Available from: https://pubmed.ncbi.nlm.nih.gov/25656182/

16. Rönnegård L, McFarlane SE, Husby A, Kawakami T, Ellegren H, Qvarnström A. Increasing the power of genome wide association studies in natural populations using repeated measures - evaluation and implementation. Methods Ecol Evol [Internet]. Methods Ecol Evol; 2016 [cited 2022 Jul 28];7:792–9. Available from: https://pubmed.ncbi.nlm.nih.gov/27478587/

17. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.

18. Amills M, Jordana J, Zidi A, Serradilla JM. Genetic Factors that Regulate Milk Protein and Lipid Composition in Goats. Milk Prod - Adv Genet Trait Cell Mech Anim Manag Heal [Internet]. IntechOpen; 2012 [cited 2022 Aug 2]; Available from: undefined/state.item.id

19. Costa A, Schwarzenbacher H, Mészáros G, Fuerst-Waltl B, Fuerst C, Sölkner J, et al. On the genomic regions associated with milk lactose in Fleckvieh cattle. J

Dairy Sci [Internet]. Elsevier; 2019 [cited 2022 Jul 29];102:10088–99. Available from: http://www.journalofdairyscience.org/article/S0022030219307301/fulltext

20. Green CD, Weigel C, Oyeniran C, James BN, Davis D, Mahawar U, et al. CRISPR/Cas9 deletion of ORMDLs reveals complexity in sphingolipid metabolism. J Lipid Res [Internet]. American Society for Biochemistry and Molecular Biology Inc.; 2021 [cited 2022 Aug 4];62:100082. Available from: http://www.jlr.org/article/S002222752100064X/fulltext

21. Quinville BM, Deschenes NM, Ryckman AE, Walia JS. A Comprehensive Review: Sphingolipid Metabolism and Implications of Disruption in Sphingolipid Homeostasis. Int J Mol Sci [Internet]. Int J Mol Sci; 2021 [cited 2022 Aug 4];22. Available from: https://pubmed.ncbi.nlm.nih.gov/34071409/

22. Sanchez MP, Ramayo-Caldas Y, Wolf V, Laithier C, El Jabri M, Michenet A, et al. Sequence-based GWAS, network and pathway analyses reveal genes co-associated with milk cheese-making properties and milk composition in Montbéliarde cows. Genet Sel Evol [Internet]. BioMed Central Ltd.; 2019 [cited 2022 Jul 29];51:1–19. Available from: https://gsejournal.biomedcentral.com/articles/10.1186/s12711-019-0473-7

23. Park JK, Yeo JM, Bae GS, Kim EJ, Kim CH. Effects of supplementing limiting amino acids on milk production in dairy cows consuming a corn grain and soybean meal-based diet. J Anim Sci Technol [Internet]. J Anim Sci Technol; 2020 [cited 2022 Aug 4];62:485–94. Available from: https://pubmed.ncbi.nlm.nih.gov/32803181/

24. Hultquist KM, Casper DP. Effects of feeding rumen-degradable valine on milk production in late-lactating dairy cows. J Dairy Sci [Internet]. Elsevier; 2016 [cited 2022 Aug 4];99:1201–15. Available from: http://www.journalofdairyscience.org/article/S0022030215009297/fulltext

Supplementary tables

Supplementary Table 1. Summary statistics of milk production and composition traits recorded in 917, 805 and 660 Murciano-Granadina goats during lactations 1, 2 and 3, respectively.

Supplementary Table 2. Pearson correlations between milk production and composition traits measured in 660 Murciano-Granadina goats with three available lactations.

Supplementary Table 3. SNPs associated at the chromosome-wide level of significance with milk traits recorded in Murciano-Granadina goats with three available lactations.

Supplementary Table 4. (**A**) SNPs identified in the longitudinal GWAS as associated at the genome-wide level of significance with milk protein percentage recorded in Murciano-Granadina goats with three available lactations (**B**). Genes located within a ± 500 kb window around each SNP displaying significant associations with protein percentage indicated in Table 4A (**C**) SNPs identified in the longitudinal GWAS as associated at the genome-wide level of significance with milk yield in kg at 210 and 240 days of lactation. (**D**). Genes located within a ± 500 kb window around each SNP displaying significant associations with milk yield at 210 and 240 lactation days indicated in Table 4C (**E**) SNPs identified in the longitudinal GWAS as associated at the genome-wide level of significance with milk lactose percentage recorded in Murciano-Granadina goats with three available lactations. (**F**) Genes located within a ± 500 kb window around each SNP displaying significant associations with lactose percentage indicated in Table 4E. (**G**). SNPs identified in the longitudinal GWAS as associated at the genome-wide level of significance with milk dry matter percentage recorded in Murciano-Granadina goats with three available lactations. (**H**) Genes located within a ± 500 kb window around each SNP displaying significant associations with dry matter percentage indicated in Table 4G.

Supplementary Table 5. SNPs identified in a longitudinal GWAS as associated at the chromosome-wide level of significance with milk traits recorded in Murciano-Granadina goats with three available lactations.

Supplementary figures

Supplementary Figure 1. Distribution of the raw (1A) and corrected (1B, rank-based transformation implemented in GenABEL) milk production and composition measurements recorded during the first lactation of 917 Murciano-Granadina goats. We only corrected data that were not normally distributed.

Supplementary Figure 2. Distribution of the raw (2A) and corrected (2B, rank-based transformation implemented in GenABEL) milk production and composition measurements recorded during the second lactation of 805 Murciano-Granadina goats. We only corrected data that were not normally distributed.

Supplementary Figure 3. Distribution of the raw (3A) and corrected (3B, rank-based transformation implemented in GenABEL) milk production and composition measurements recorded during the third lactation of 660 Murciano-Granadina goats. We only corrected data that were not normally distributed.

Supplementary figure 4. Pearson correlation across phenotypic records of milk production and composition measured during three lactations of 660 Murciano-Granadina goats.

Journal of Dairy Science
Official Publication of the American Dairy Science Association

RUNNING HEAD: GWAS FOR MORPHOLOGY TRAITS IN GOATS

A genome-wide association analysis for body, udder and leg conformation traits recorded in Murciano-Granadina goats

Maria Gracia Luigi-Sierra[1], Vincenzo Landi[2,3], Dailu Guan[1], Juan Vicente Delgado[2], Anna Castelló[1,4], Betlem Cabrera[1,4], Emilio Mármol-Sánchez[1], Javier Fernández Alvarez[5], Mayra Gómez-Carpio[2], Amparo Martínez[2], Xavier Such[6], Jordi Jordana[4], Marcel Amills[1,4,*]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain.
[2]Departamento de Genética, Universidad de Córdoba, Córdoba 14071, Spain.
[3]Department of Veterinary Medicine, University of Bari "Aldo Moro", SP. 62 per Casamassima km. 3, 70010 Valenzano (BA), Italy.
[4]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain,
[5] Asociación Nacional de Criadores de Caprino de Raza Murciano-Granadina (CAPRIGRAN), 18340 Granada, Spain.
[6]Group of Research in Ruminants (G2R), Department of Animal and Food Science, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain

*Corresponding author

J Dairy Sci. 2020 Dec;103(12):11605-11617. doi: 10.3168/jds.2020-18461.

Abstract

Morphological traits are of great importance to dairy goat production given their effect on phenotypes of economic interest. However, their underlying genomic architecture has not yet been extensively characterized. Herein, we aimed to identify genomic regions associated with body, udder, and leg conformation traits recorded in 825 Murciano-Granadina goats. We genotyped this resource population using the GoatSNP50 BeadChip (Illumina Inc., San Diego, CA) and performed genome-wide association analyses using the GEMMA software. We found 2 genome-wide significant associations between markers rs268273468 [Capra hircus (CHI) 16:69617700] and rs268249346 (CHI 28:18321523) and medial suspensory ligament. In contrast, we did not detect any genome-wide significant associations for body and leg traits. Moreover, we found 12, 19, and 7 chromosome-wide significant associations for udder, body, and leg traits, respectively. Comparison of our data with previous studies revealed a low level of positional concordance between regions associated with morphological traits. In addition to technical factors, this lack of concordance could be due to a substantial level of genetic heterogeneity among breeds or to the strong polygenic background of morphological traits, which makes it difficult to detect genetic factors that have small phenotypic effects.

Key words: genome-wide association study (GWAS), goat, Murciano-Granadina, morphological trait.

Introduction

Since its establishment in 1975 (Delgado et al., 2018), the Murciano-Granadina breed has become the most important dairy goat breed in Spain, reaching 112,417 heads in 2019 (https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas). Murciano-Granadina goats show great ability to adapt to harsh environments, yielding 530 kg of milk per lactation (250 d) with fat and protein contents of 5.6 and 3.6%, respectively (https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas).

The genetic improvement program of the Murciano-Granadina breed is mostly focused on increasing milk yield and quality and on optimizing body and dairy conformation traits. The association of Murciano-Granadina breeders (Caprigran) systematically records information about milk yield and composition and performs linear scoring of 17 morphological traits (Delgado et al., 2018). The inclusion of morphological traits as selection criteria is motivated by their association with mammary health and longevity (Shelton, 1978; Manfredi et al., 2001; Montaldo and Manfredi, 2002). Studies performed in cattle (Seykora and McDaniel, 1985; Rogers et al., 1991; Boettcher et al., 1998; Rupp and Boichard, 1999; Miglior et al., 2017) support a relationship between udder morphology and health. Indeed, higher, nonpendulous, and more tightly attached udders are less susceptible to mastitis (Seykora and McDaniel, 1985; Rupp and Boichard, 2003). Moreover, flat, disk, or inverted teat ends are associated with an increased risk of suffering mastitis, whereas funnel-shaped teats seem to be less prone to mastitis (Seykora and

McDaniel, 1985). Numerous reports associate body traits with fertility and longevity (Bastin and Gengler, 2013; Miglior et al., 2017). For instance, foot and leg conformation traits could be considered indicators of claw health, which, after reproduction and mastitis, is one of the main determinants for culling animals (Egger-Danner et al., 2015).

Few studies about the genetics of morphological traits have been performed in goats. According to Rupp et al., 2011, udder floor position showed negative genetic correlations (r_g) with SCS in both Alpine ($r_g = -0.24$) and Saanen goats ($r_g = -0.19$). The same authors demonstrated that, in the Saanen breed, SCS was correlated with teat length ($r_g = 0.29$), teat width ($r_g = 0.34$), and teat form ($r_g = -0.27$). These results suggest that a reduction in SCC could be achieved by selection, while still improving milk production and udder type and teat traits. In Tinerfeña goats, moderate to high and mostly positive phenotypic correlations have been detected between udder traits and milk yield (Capote et al., 2006).

Manfredi et al., 2001 found a high heritability ($h^2 > 0.4$) for thorax perimeter in Alpine and Saanen goats, whereas other body traits such as rump angle, feet angle, and hock distance showed low heritabilities ($h^2 = 0.03–0.16$). In contrast, heritabilities for udder and teat scores were around 0.3 for most traits, with teat angle displaying the lowest value ($h^2 = 0.15$ in Saanen). Genetic correlations among teat dimension traits and between udder floor and rear udder attachment ($r_g > 0.7$ in Alpine and Saanen) were generally high, whereas the majority of genetic correlations between body and udder scores were <0.3 (Manfredi et al., 2001).

In a more recent study, McLaren et al., 2016 described low to moderate heritabilities (from 0.02 to 0.38) for conformation traits recorded in mixed-breed dairy goats: although udder and teat traits had the highest heritabilities (h^2 ~0.28, from 0.15 to 0.38), feet and leg traits showed lower values (h^2 ~0.13, from 0.02 to 0.25). Although most of the correlations estimated between milk yield and udder and teat traits were negative, their magnitude and sign fluctuated across the first lactation. For instance, estimates of the genetic correlation between udder furrow and milk yield ranged from −0.42 to 0.18 depending on the time point in first lactation when they were calculated (McLaren et al., 2016). Castañeda-Bustos et al., 2017 described a high genetic correlation between the productive life of dairy goats (i.e., total days in production until 72 mo of age) and final score (appraisal of the general conformation of the animal), fore udder attachment, and rump width. These findings demonstrate that considering conformation and udder traits in selection schemes would be expected to increase productivity without compromising the viability of the animals (Castañeda-Bustos et al., 2017).

Several investigations have been carried out to identify genomic regions associated with body conformation and udder traits in goats. Mucha et al., 2018 performed a GWAS for morphology traits in mixed-breed dairy goats and found a significant association between a region on chromosome 19 and udder attachment, udder depth, and front leg morphology. Moreover, Martin et al., 2018 detected 37 genome-wide significant QTL for type and SCS phenotypes with linkage analyses, whereas a much larger number of QTL were identified by association mapping. These authors concluded that the inheritance of body and udder conformation traits

is markedly polygenic and that genetic determinants are often breed-specific. In the current work, we aimed to identify genomic regions associated with the phenotypic variation of 17 morphological traits in Murciano-Granadina goats.

Materials and methods

Phenotypic Recording

A total of 825 female goats distributed in 13 farms, with an average herd size of 500 individuals, were scored for 17 morphological traits included in the breeding program of the Murciano-Granadina goat breed. The scoring is performed only once in the lifetime of the animal. Most of the goats in our study were scored during their first lactation, although a group of 89 animals was scored between the second and sixth lactations. All traits were scored by the same specialist, using a personal digital device equipped with the Kalifadroid app (Fernández Alvarez, 2017) for carrying out scoring tasks. The following phenotypes were evaluated, with linear scores ranging from 1 to 9 according to the criteria established in Sánchez-Rodríguez, 2012.

Udder Traits

Seven udder traits were scored:

Fore udder attachment (**FUA**)—Corresponds to the angle formed by the line of the udder insertion and the abdominal wall. Scores 1 and 9 correspond to angles of 45° and 120°, respectively.

Rear udder height (**RUH**)—Scored by measuring the distance between the bottom of the vulva and the top of the secretory tissue of the mammary gland. Scores 1 and 9 correspond to distances of 11 and 3 cm, respectively.

Udder depth (**UD**)—Distance from the lowest part of the udder floor to the hock joint (tibiotarsal joint). A linear score of 1 corresponds to an udder with its deepest part 10 cm over the hock, whereas a score of 9 would define an udder with its deepest part 10 cm down the hock joint.

Medial suspensory ligament (**MSL**)—Depth of the udder cleft measured at the base of the rear udder. Scores 1 and 9 correspond to 1 and 9 cm (or more) deep udder clefts, respectively.

Udder width (**UW**)—Measured at the crease where the udder meets the leg. Scores of 1 and 9 correspond to measurements of 3 and 11 cm (or more), respectively.

Teat diameter (**TD**)—Diameter of the teat at its base, when it meets the udder. The measurement was performed for each teat individually and the average score was used as final score. Score 1 = diameter of 0.5 cm and score 9 = diameter of 4.5 cm or more.

Teat placement (**TP**)—Defines the position of the teats on the udder half. Teats located on the outside third of the udder half are scored as 1, whereas teats located very close to the medial suspensory ligament, that almost touch each other, are scored as 9. The measurement was performed for each teat individually and the average score was used as final score. Teats on the center of the udder half, with an intermediate placement, are considered desirable and are scored as 5.

Body Conformation Traits

Seven body conformation traits were scored:

Height (**HT**)—Measures the distance from ground level to the top of the withers. Goats with a height of 62 cm or less received a score of 1, and those over 78 cm received a score of 9.

Chest width (**CW**)—Measured from the inside surface of the chest between the top of the front legs. Score 1 = 15 cm or less; score 9 = 23 cm or more.

Body depth (**BD**)—Distance between the top of the spine and bottom of the body at the beginning of the last rib. Score 1 = low depth, if the beginning of the last rib is located above the elbow joint, and score 9 = high depth, if the beginning of the last rib is located below the elbow joint.

Rump width (**RW**)—Distance between the most posterior points of the pin bones (ischial tuberosities). Score 1 = 13 cm (or less); score 9 = 21 cm (or more).

Rump angle (**RA**)—Angle between the hook (coxal tuberosity) and pin (ischial tuberosity) bones. Scores 1 and 9 correspond to angles of approximately 55° and 31°, respectively.

Angularity or dairyness (**ANG**)—Angle and openness of the ribs. Score 1 defines an animal extremely coarse for this trait; score 9 is assigned to goats that are very angular.

Bone quality (**BQ**)—Appraisal of the thickness and width of the bone structure, assessed by examining the rear leg from the rear and from the side. Score 1

corresponds to goats with thick and round bones; score 9 is assigned to goats with flat and sharp bones.

Feet Structure

Three feet structure traits were scored:

Rear legs rear view (**RLR**)—Direction of rear feet when viewed from the rear. Score 1: extreme toe-out feet; score 9: parallel feet.

Rear legs side view (**RLS**)—Curvature of the hock viewed laterally. Score 1: straight legs; score 9: very curved legs.

Mobility (**MOB**)—Evaluates the locomotion patterns, including the length and direction of the step. Score 1: bad locomotion, with severe abduction and short steps; score 9: harmonic, long and uniform locomotion.

Isolation of Genomic DNA from Blood and Genotyping with the Goat SNP50 BeadChip

Blood samples from the 825 Murciano-Granadina goats with morphology records were collected in vacuum tubes coated with K3-EDTA anticoagulant and stored at −20°C until processing. Genomic DNA was purified using a modified salting-out procedure (Miller et al., 1988). In brief, we combined 3 mL of whole blood plus 4 volumes of Red Cell Lysis Solution (10 mM Tris-HCl, pH 6.5; 2 mM EDTA; 1% Tween 20) and then this mixture was centrifuged at 2,000 × g for 2 min. The resulting cell pellet was resuspended in 3 mL of lysis buffer (200 mM Tris-HCl, pH 8, 30 mM EDTA, 1% SDS, 250 mM NaCl) and 100 µL of proteinase K (20 mg/mL) and incubated for 3 h at 55°C. The lysate was chilled, and 1 mL of 10 M ammonium

acetate was added. This mixture was centrifuged at 2,000 × g for 10 min, and the supernatant (~4 mL) was transferred to a new tube with 3 mL of 96% isopropanol. Subsequently, samples were centrifuged at 2,000 × g for 3 min. The resulting DNA pellet was washed with 3 mL of ethanol 70% followed by an additional centrifugation step at 2,000 × g for 1 min. The DNA pellet was left at room temperature until it dried, and then was resuspended in 1 mL of Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA, pH = 8). All 825 goats were genotyped with the Illumina Goat SNP50 BeadChip (Illumina Inc., San Diego, CA), which contains 54,241 SNP, following the instructions of the manufacturer. The genomic location of the SNPs was obtained using the goat ARS1 genome (Bickhart et al., 2017) as reference, and the position and the name of each SNP was updated using the software PLINK v 1.9 (Chang et al., 2015). The genotypic information was filtered using PLINK v 1.9 (Chang et al., 2015). Only individuals with <5% missing genotypes were taken into consideration. With regard to SNPs, only those meeting the following requirements were used in the GWAS: (1) mapping to autosomes, (2) displaying a minor allele frequency of ≥0.05, (3) not deviating very significantly (P > 0.001) from Hardy-Weinberg expectation, and (4) with a genotype call rate >90%. After applying these filtering criteria, 47,880 SNPs and 811 animals were selected to perform subsequent analyses. Population structure was assessed with a principal component analysis (PCA) implemented in PLINK v 1.9 (Chang et al., 2015). The visualization of the PCA results was based on the first 2 components of the PCA.

Statistical Analyses

We calculated summary statistics for each of the morphological traits using R (R Core Team, 2017) as well as Pearson correlations (r_p) between conformation traits and milk composition and yield records; that is, total milk yield, milk yield at 210 d, milk yield at 240 d, milk yield at 305 d, SCS, fat percentage, protein percentage, and lactose percentage. The correlations were estimated and heatmap plots constructed with R software (R Core Team, 2017) to visualize the correlation matrix and the P-values of each correlation.

The software GEMMA (Zhou and Stephens, 2012) was used to carry out the GWAS. This method corrects population structure by taking into account the relatedness matrix, which is built by considering all genome-wide SNPs as a random effect. Morphological phenotypes were rank-based transformed using the package GenAbel from R (Aulchenko et al., 2007) because we assessed, with our data, that this transformation yields residuals that are normally distributed. A univariate linear mixed model was fit for each trait as follows:

$$y = W\alpha + x\beta + u + \varepsilon,$$

where y is a vector of corrected scores for morphological traits recorded in 811 individuals; W = (w1, …, wc) is an n × c matrix of 3 fixed effects (farm, with 13 levels; year of birth, with 10 levels; number of lactations, with 6 levels) and 1 covariate (days producing milk); α is a c-vector of the corresponding fixed effects including the intercept; x is a n-vector of marker genotypes; β is the effect size of the marker (allele substitution effect); u is a n-vector of random individual genetic effects with a normal distribution $u \sim N(0, \lambda \tau^{-1} K)$, where τ^{-1} is the variance of

the residual error, λ is the ratio between the 2 variance components, and K is the relatedness matrix derived from SNP genotypes. Finally, ε is a n-vector of errors.

A false discovery rate (FDR) approach was applied to correct for multiple testing, setting the significance level to a q-value of 0.05 (Benjamini and Hochberg, 1995). Graphical visualization of the results of the GWAS was achieved by using the R software (R Core Team, 2017).

The proportion of the phenotypic variance explained by the significant SNPs (PVE) was estimated using the formula reported by Shim et al., 2015:

$$PVE = \frac{2\,\beta^2\,MAF\,(1-MAF)}{2\beta^2\,MAF\,(1-MAF) + [se(\beta)]^2\,2N\,MAF\,(1-MAF)}$$

where β is the effect size of the SNP variant estimated from the association analysis; MAF is the minor allele frequency of the SNP, se is the standard error, and N is the sample size. Lambda inflation factors were calculated with the median method (1 df) implemented in GenABEL (Aulchenko et al., 2007), whereas quantile-quantile (Q-Q) plots were built with the gg_qqplot() function (https://www.rdocumentation.org/packages/lindia/versions/0.9/topics/gg_qqplot).

To retrieve candidate genes mapping close to significant SNP, we considered an interval of ±1 Mb based on data previously reported by Guan et al., 2020 for the same population. Genes mapping within these defined boundaries were listed by using the Biomart tool from Ensembl (Kinsella et al., 2011) and subsequently analyzed with Uniprot (UniProt Consortium, 2019) and David Bioinformatic Resources (Huang et al., 2009) to annotate their function.

Results and discussion

Population Structure and Analysis of Morphological Traits

The first 2 components (C1 and C2) of the principal component analysis accounted for 42.64% and 21.76% of the genetic variance, respectively (Supplemental Figure S1; https://doi.org/10.3168/jds.2020-18461). Several samples were grouped according to their farm of origin, but we did not find an obvious within-population substructure in this sample of Murciano-Granadina goats. Descriptive statistics of the raw conformation scores are reported in Supplemental Table S1 and Supplemental Figures S2 and S3 (https://doi.org/10.3168/jds.2020-18461). The estimated phenotypic correlations and their significances are depicted in **Supplemental Table S2** (https://doi.org/10.3168/jds.2020-18461) and **Figure 1**. We classified phenotypic correlations as low ($r_p < 0.2$), moderate ($r_p = 0.2–0.4$), or high ($r_p > 0.4$). Phenotypic correlations between udder traits were generally low and positive, except for the correlation between MSL and UD that was moderate ($r_p = 0.37$, $P < 0.001$). In the study of McLaren et al., 2016, genetic correlations between udder traits ranged from 0.12 to 0.77, whereas those between teat traits were in the range of −0.10 to 0.69. Udder and teat traits were positively correlated, a result coincident with our findings, and also with those reported by Manfredi et al., 2001. We found low and positive correlations between SCS and MSL ($r_p = 0.14$, $P < 0.001$), UD ($r_p = 0.19$, $P < 0.001$), and TD ($r_p = 0.13$, $P < 0.001$). Udder depth ($r_g = 0.10$) and teat size ($r_g = 0.29$) also showed positive genetic correlations with SCS in Latxa sheep (Legarra and Ugarte, 2005). Pendulous udders are closer to the floor, exposing the mammary gland to direct contact with fecal and other environmental

contaminants (Pugh and Baird, 2012). Moreover, teat diameter is highly correlated with the diameter of the internal cistern (Guarín et al., 2017), and it is expected that teat sphincters in animals with wider cisterns do not close completely, leaving a channel open for pathogens, thus increasing the risk of suffering mastitis (Seykora and McDaniel, 1985). In contrast, UW ($r_p = -0.11$, $P < 0.001$) and TP ($r_p = -0.13$, $P < 0.001$) were negatively correlated with SCS, but these 2 values were low, and these 2 traits in other studies do not show strong genetic correlations with SCS (Legarra and Ugarte, 2005; Pérez-Cabal et al., 2013). Notably, UW ($r_p = 0.20–0.21$, $P < 0.001$) and UD ($r_p = 0.29–0.32$, $P < 0.001$) showed moderate positive correlations with milk yield. Pérez-Cabal et al., 2013 found moderate to high and positive phenotypic correlations between milk yield and udder width ($r_p = 0.29$) and udder depth ($r_p = 0.47$) in Spanish Assaf sheep, and Legarra and Ugarte, 2005 reported a strong positive genetic correlation between milk yield and udder depth ($r_g = 0.43$). It is reasonable to infer that goats with wider and deeper udders produce more milk.

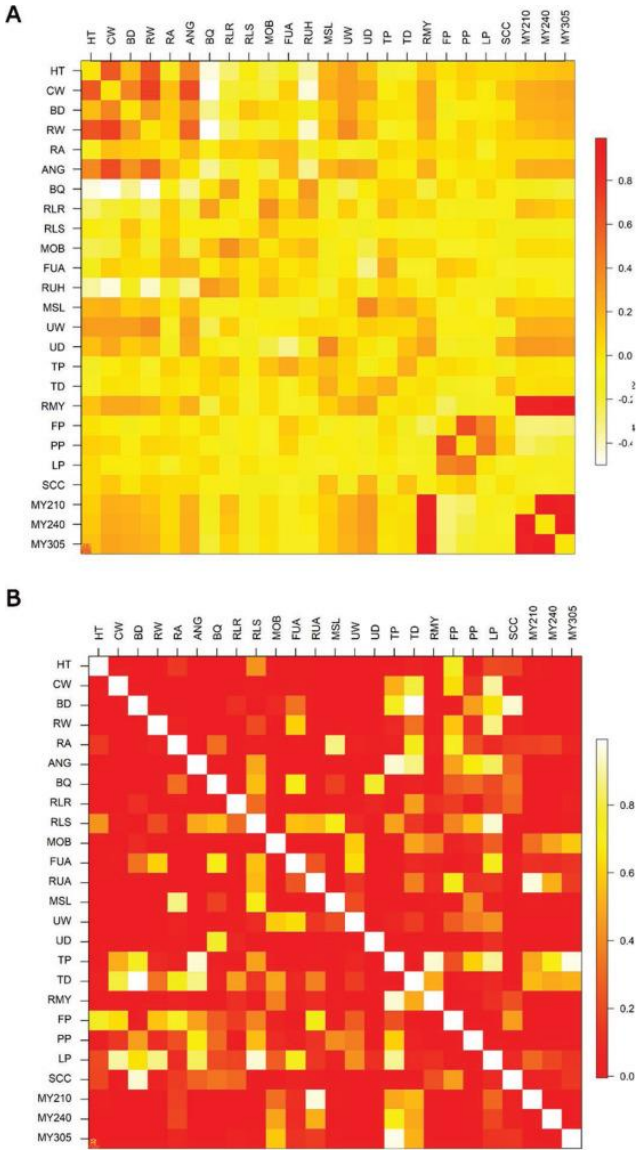
Phenotypic correlations between body conformation traits were generally high and positive. For instance, HT was correlated with CW ($r_p = 0.65$, $P < 0.001$), RW ($r_p = 0.67$, $P < 0.001$), and ANG ($r_p = 0.42$, $P < 0.001$); and CW showed positive correlations with BD ($r_p = 0.41$, $P < 0.001$), RW ($r_p = 0.74$, $P < 0.001$), and ANG ($r_p = 0.68$, $P < 0.001$). Zujovic et al., 2011 also observed high phenotypic correlations between body traits measured in Balkan goats, and Chacón et al., 2011 observed a similar trend in Cuban goats. In general, taller goats are also bigger and have a wider chest and rump. We observed moderate positive correlations between

CW and BD and milk yield (rp = 0.2–0.3, P < 0.001) and between RW and milk yield (r_p = 0.15–0.2, P < 0.001). In cows, individuals with a wider rump and increased space across their hooks and pins are reported to accommodate a higher and wider udder (Campbell and Marshall, 2016). These results suggest that the increase in udder capacity leads to a higher milk yield while decreasing the percentages of solid milk components (% lactose, % fat, and % protein). There are also indications that cows with a short and round body often lack dairy character and udder capacity (Campbell and Marshall, 2016), supporting the positive correlation between ANG and milk yield observed in our Murciano-Granadina population.

With regard to leg traits, MOB displayed a positive phenotypic correlation with RLR (r_p = 0.42, P < 0.001) and RLS (r_p = 0.23, P < 0.001). Interestingly, RLR and RLS showed low positive and low negative correlations with milk yield, respectively. McLaren et al., 2016 observed a correlation of 0.33 between back legs and milk yield at 305 d, whereas de la Fuente et al., 2011 estimated a small genetic correlation of −0.09 between the back legs and milk yield in Churra ewes. Our interpretation is that leg morphology could be associated with the predisposition of goats to lameness, a pathology that results in decreased milk production and often in the culling of the affected animal (Archer et al., 2010).

Figure 1(A) Heatmap depicting Pearson correlations between morphological and milk yield phenotypes recorded in 811 Murciano-Granadina goats. The intensity of the color indicates the magnitude of the correlation. **(B)** Heatmap displaying the significance of the Pearson correlations shown in panel A. The intensity of the color

indicates the significance of the association. Traits: HT, height; CW, chest width; BD, body depth; RW, rump width; RA, rump angle; ANG, angularity; BQ, bone quality; RLR, rear legs rear view; RLS, rear legs side view; MOB, mobility; FUA, fore udder attachment; RUH, rear udder height; MSL, medial suspensory ligament; UW, udder width; UD, udder depth; TP, teat placement; TD, teat diameter; RMY, total milk yield; MY210, milk yield corrected at 210 d of milking; MY240, milk yield corrected at 240 d of milking; MY305, milk yield corrected at 305 d of milking.



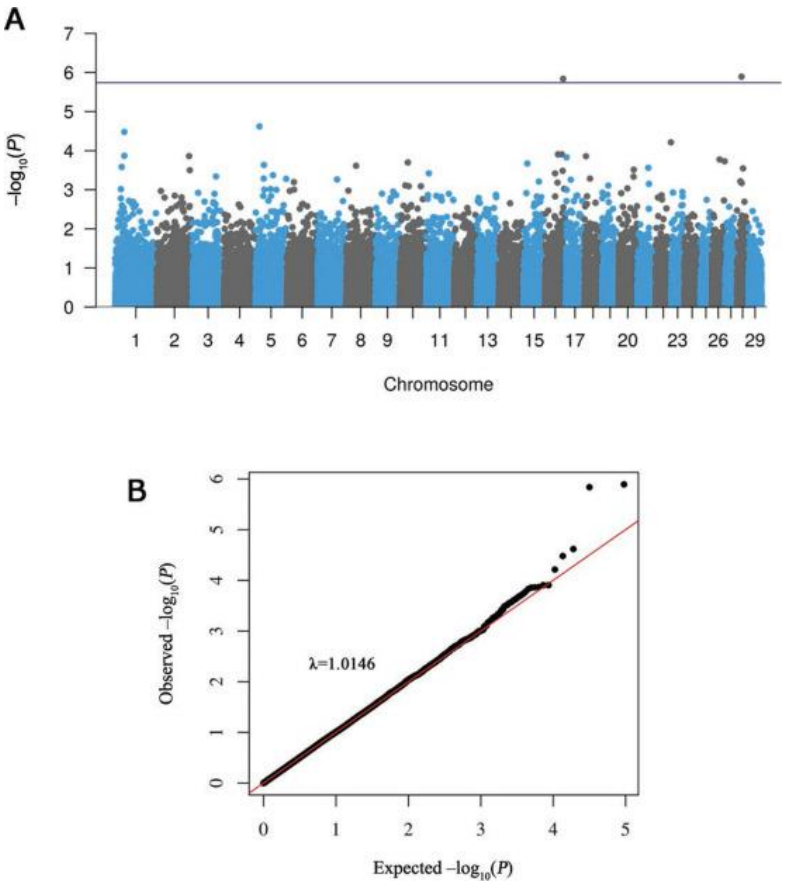
Identification of Genetic Determinants for Udder Traits

After performing the GWAS, we detected 2 genome-wide significant associations for the trait MSL (**Table 1** and **Figures 2A** and **2B**). We also found chromosome-wide associations for the traits MSL (2 SNP), UW (1 SNP), UD (3 SNP), TP (4 SNP), and TD (2 SNP), as shown in **Table 1** and **Supplemental Figures S4** (Manhattan plots) and S5 (Q-Q plots; https://doi.org/10.3168/jds.2020-18461). No significant SNP was found for FUA or RUH.

The rs268273468 (CHI 16: 69617700) marker, which was significantly associated with MSL at the genome-wide level of significance (**Table 1** and **Figure 2**), is located less than 1 Mb from the lysophosphatidylglycerol acyltransferase 1 (*LPGAT1*) gene (**Supplemental Table S3**; https://doi.org/10.3168/jds.2020-18461). This gene encodes an enzyme involved in conversion of lysophosphatidylglycerol into phosphatidylglycerol, a membrane phospholipid that is a key precursor in the biosynthesis of cardiolipin (Yang et al., 2004). Interestingly, cardiolipin is located in the inner mitochondrial membrane and plays a fundamental role in maintaining mitochondrial membrane stability and dynamics as well as in regulating apoptosis (Paradies et al., 2014). Proper mitochondrial function, in turn, is essential to ensure the integrity of tendons and other connective tissues (Lowes et al., 2009; Thankam et al., 2018). With regard to rs268249346 (CHI 28: 18321523), the other genome-wide significant SNP for MSL, we found that it maps close to the ADAM metallopeptidase with thrombospondin type 1 motif 14 (*ADAMTS14*) locus (**Supplemental Table S3**). This gene has been reported to encode a procollagen N-proteinase that cleaves the amino-propeptide of

procollagen to allow the assembly of elongated and cylindrical collagen fibrils (Bekhouche and Colige, 2015), although one recent study indicates that the main role of this molecule is the regulation of the immune response (Dupont et al., 2018).

Figure 2(A) Negative log10 P-values (y-axis) of the associations between SNPs and the medial suspensory ligament phenotype are plotted against the genomic location of each SNP marker (x-axis). Markers on different chromosomes are denoted by different colors. The blue line indicates the −log10(P-value) equivalent to a q-value of 0.05. **(B)** Quantile-quantile plot corresponding to the genome-wide corrected P-values of the genome-wide association study for the trait medial suspensory ligament and its lambda (λ) inflation factor.



We identified 12 SNP showing significant associations with udder traits at the chromosome-wide level (**Table 1**, **Supplemental Figures S4** and **S5**). The activating transcription factor 3 (*ATF3*) gene is located close to the SNP rs268273468 (CHI 16: 69617700) associated with MSL (**Supplemental Table S3**). Interestingly, this gene modulates the synthesis of collagen I and III (Zhou et al., 2011) and regulates matrix metalloproteinases, which are fundamental in the development, renewal, and remodeling of tendons (Guenzle et al., 2017). Another interesting association is that between the rs268288193 marker (CHI 19: 38362152) and TP. The region containing this SNP was associated with SCS in Saanen goats (Martin et al., 2018). Of note, Lund et al., 1994 reported that cows with bad teat placement tend to be more susceptible to mastitis. Less than 1 Mb from this SNP, we identified the suppressor of cytokine signaling 7 (*SOCS7*) gene (**Supplemental Table S3**), which inhibits prolactin, growth hormone, and leptin signaling (Martens et al., 2005). Members of this gene family are regulators of mammary gland physiology. For instance, in mice, *SOCS1* and *SOCS2* attenuate prolactin signaling, thus preventing premature lactation (Sutherland et al., 2007). In dairy sheep, a missense polymorphism (R96C) in *SOCS2* is associated with mastitis susceptibility, while having a positive effect on milk production and BW (Rupp et al., 2015). This antagonistic relationship could be due to co-selection of alleles influencing milk yield and susceptibility to mastitis, due to a hitchhiking effect, or to pleiotropy (Oget et al., 2019). Moreover, *SOCS3* is a key regulator of mammary gland involution (Sutherland et al., 2007). In dairy cattle, polymorphisms in *SOCS7* have been associated with different milk traits such as protein yield and percentage and milk yield (Arun et al., 2015).

The collagen type XIV α1 chain (*COL14A1*) gene is located 0.1 Mb away from rs268281312 (CHI 14:868624), which was significantly associated with UD (**Supplemental Table S3**). This gene encodes a fibril-associated collagen that regulates fibrillogenesis (Lindholm et al., 2019). In the human mammary gland, large amounts of type XIV collagen have been found in interlobular stroma, which, compared with intralobular stroma, contains densely packed collagen (Atherton et al., 1998). This differential distribution of type XIV collagen might have an important effect on the architecture of the mammary connective tissue.

Identification of Genetic Determinants for Body Conformation Traits

Although we did not detect any genome-wide significant associations for body conformation traits, we found 19 SNPs at the chromosome-wide level significantly associated with ANG (4 SNP), RW (2 SNPs), RA (1 SNP), CW (5 SNP), HT (3 SNP), BD (3 SNP), and BQ (1 SNP). These results are displayed in **Table 2** and in **Supplemental Figures S6** (Manhattan plots) and **S7** (Q-Q plots; https://doi.org/10.3168/jds.2020-18461).

Several of the aforementioned SNP map close to genes involved in bone homeostasis and skeletal development (Supplemental Table S3). For instance, rs268245664, which is associated with ANG, maps to the parathyroid hormone 1 receptor (*PTH1R*) locus (**Table 2**, and **Supplemental Table S3**). This gene encodes a protein that acts as a receptor for parathyroid hormone (PTH) and parathyroid-related peptide (PTHrP), 2 factors regulating mineral ion homeostasis (Mannstadt et al., 1999). The dysfunction of *PTH1R* is associated with diseases that affect skeletal development and calcium homeostasis (Mannstadt et al., 1999). With

regard to the association between ANG and rs268265191 (CHI 1: 35829812), it is worth mentioning that this marker maps near to the ephrin receptor A3 gene (*EPHA3*; **Supplemental Table S3**). Ephrin receptors and their associated ligands are essential modulators of bone remodeling and they ensure adequate coupling between bone resorption and formation (Edwards and Mundy, 2008). Interestingly, the inactivation of the ephrin-B1 gene causes perinatal lethality, abdominal wall closure defects, and skeletal abnormalities, especially of the thoracic cage (Compagni et al., 2003). Another interesting gene is that encoding the CGG triplet repeat binding protein 1 (*CGGBP1*), which is located 1 Mb from rs268265191 (CHI 1: 35829812; **Supplemental Table S3**). Polymorphisms in this gene are associated with several carcass traits in cattle (Calonge, 2004), having considerable effects on growth (Sevane et al., 2014). Moreover, Sevane et al., 2014 described a nonsynonymous mutation, rs477676137 (c.206A>G, 1: 36060631), in *CGGBP1* associated with an increase in pelvis width and withers height measured at 9 mo. In contrast, rs268255133 (CHI 18: 34112104), which was associated with RW, maps close to the cadherin 11 gene (*CDH11*; **Supplemental Table S3**), which modulates postnatal bone growth and osteoblast differentiation (Di Benedetto et al., 2010). Furthermore, rs268262472 (CHI 27: 38084358), which is associated with RA, lies near the spermatogenesis-associated 4 gene (*SPATA4*), which is also involved in promoting osteoblast differentiation (Wang et al., 2011).

Another marker of interest is rs268285858 (CHI 22: 45075519). This SNP was associated with BD and is located close to the Wnt family member 5A (*WNT5A*) gene (**Supplemental Table S3**) that regulates planar cell polarity signaling during

embryonic development (Qian et al., 2007). Loss of this gene results in a shortened and widened cochlea in knockout mice embryos and severe shortening of the anterior-posterior axis and limb truncations due to abnormal convergent extension (Yamaguchi et al., 1999; Qian et al., 2007; Andre et al., 2015). Finally, rs268249930 (CHI 16: 26773653), which is associated with HT, co-localizes with *DNAH14* (**Supplemental Table S3**), which encodes an axonemal dynein heavy chain. Mutations in dynein genes can cause skeletal ciliopathies characterized by thoracic narrowing, short long bones, and pelvis dysplasia (Yildiz, 2018).

Table 1. Genome-wide and chromosome-wide significant associations between SNPs and udder traits (MSL, medial suspensory ligament; TD, teat diameter; TP, teat placement; UD, udder depth; UW, udder width) recorded in 811 Murciano-Granadina goats.

	Trait	Chr[1]	rs[2]	Pos[3]	A1[4]	MAF[5]	β ± SE[6]	*P-value*[7]	*q-value*[8]	*PVE*[9]
Genome Wide	MSL	16	rs268273468	69,617,700	C	0.469	-0.220 ± 0.045	1.45E-06	0.034	0.010
		28	rs268249346	18,321,523	A	0.485	-0.241 ± 0.049	1.28E-06	0.034	0.008
Chromosome Wide	MSL	16	rs268273468	69,617,700	C	0.469	-0.220 ± 0.045	1.45E-06	0.002	0.010
		28	rs268249346	18,321,523	A	0.485	-0.241 ± 0.049	1.28E-06	0.001	0.008
	TD	28	rs268248647	9,579,785	G	0.460	-0.214 ± 0.054	6.93E-05	0.042	0.007
		28	rs268243765	10,586,645	G	0.404	-0.215 ± 0.055	9.45E-05	0.042	0.007
	TP	9	rs268282545	20,855,427	G	0.498	-0.208 ± 0.047	1.09E-05	0.020	0.009
		19	rs268288193	38,362,152	G	0.181	0.243 ± 0.059	4.05E-05	0.047	0.006
		25	rs268246864	40,499,453	G	0.391	-0.205 ± 0.047	1.32E-05	0.011	0.009
		26	rs268291440	37,970,238	A	0.316	-0.238 ± 0.052	5.33E-06	0.005	0.008
	UD	14	rs268281312	868,624	A	0.263	0.216 ± 0.051	2.83E-05	0.041	0.008
		14	rs268276674	90,209,603	A	0.480	0.193 ± 0.047	4.48E-05	0.041	0.009
		15	rs268268821	47,457,448	A	0.430	-0.207 ± 0.050	3.14E-05	0.049	0.008
	UW	27	rs268251218	10,045,474	G	0.444	0.277 ± 0.062	1.00E-05	0.009	0.005

[1]**Chr**, chromosome; [2]**rs**, identifier code of the SNP according to the RefSNP database; [3]**Pos**, position in base pairs; [4]**A1,** minority allele; [5]**MAF**, allele frequency; [6]**β ± SE**, allelic substitution effect ± standard error; [7]***P-value***, raw *P*-values; [8]***q-value***, *P*-values corrected for multiple testing using a false discovery rate approach; [9]***PVE,*** percentage of proportion of variance in phenotype explained by a given SNP.

Table 2. Chromosome-wide significant associations between SNPs and body traits (ANG, angularity; BD, body depth; BQ, bone quality; CW, chest width; HT, height; RA, rump angle; RW, rump width) recorded in 811 Murciano-Granadina goats.

	Trait	Chr[1]	rs[2]	Pos[3]	A1[4]	MAF[5]	β ± SE[6]	*P-value*[7]	*q-value*[8]	*PVE*[9]
Chromosome Wide	ANG	1	rs268265191	35,829,812	G	0.496	-0.199 ± 0.045	0.032	4.980	0.010
		1	rs268280713	135,255,901	A	0.352	-0.200 ± 0.048	0.049	4.504	0.009
		22	rs268245662	52,518,237	A	0.074	-0.358 ± 0.083	0.010	4.759	0.003
		22	rs268245664	52,649,718	G	0.043	-0.512 ± 0.105	0.001	5.900	0.002
	BD	9	rs268236956	24,849,991	G	0.281	0.224 ± 0.050	0.012	5.161	0.008
		13	rs268290607	55,121,535	A	0.394	-0.190 ± 0.045	0.044	4.553	0.010
		22	rs268285858	45,075,519	G	0.48	-0.180 ± 0.044	0.045	4.393	0.011
	BQ	15	rs268289470	13,689,629	A	0.394	0.217 ± 0.052	0.012	5.109	0.008
	CW	3	rs268243320	53,360,997	C	0.139	-0.265 ± 0.062	0.044	4.715	0.005
		7	rs268240258	51,615,185	A	0.311	-0.214 ± 0.045	0.006	5.548	0.010
		17	rs268248977	26,865,599	A	0.454	0.174 ± 0.040	0.021	4.824	0.013
		17	rs268248975	26,942,955	G	0.252	0.193 ± 0.049	0.044	4.022	0.008
		17	rs268264934	58,534,119	G	0.494	0.171 ± 0.041	0.027	4.411	0.012
	HT	16	rs268249920	26,376,843	A	0.419	0.172 ± 0.043	0.031	4.213	0.011
		16	rs268249930	26,773,653	G	0.348	-0.207 ± 0.044	0.005	5.477	0.011
		16	rs268236696	59,779,757	A	0.126	-0.238 ± 0.059	0.031	4.262	0.006
	RA	27	rs268262472	38,084,358	G	0.207	-0.255 ± 0.063	0.044	4.299	0.005
	RW	17	rs268264930	58,698,058	G	0.359	0.182 ± 0.040	0.009	5.190	0.013
		18	rs268255133	34,112,104	G	0.134	-0.259 ± 0.055	0.003	5.553	0.007

[1]**Chr**, chromosome; [2]**rs**, identifier code of the SNP according to the RefSNP database; [3]**Pos**, position in base pairs; [4]**A1,** minority allele; [5]**MAF**, allele frequency; [6]**β ± SE**, allelic substitution effect ± standard error; [7]***P-value***, raw *P*-values; [8]***q-value***, *P*-values corrected for multiple testing using a false discovery rate approach; [9]***PVE,*** percentage of proportion of variance in phenotype explained by a given SNP.

Identification of Genetic Determinants for Leg Structure Traits

No genome-wide significant SNP was found for leg structure traits, but 7 SNP showed significant associations at the chromosome-wide level. These findings are reported in **Table 3** as well as in **Supplemental Figures S8** (Manhattan plot) and **S9** (Q-Q plot; https://doi.org/10.3168/jds.2020-18461). For the MOB trait, an association was found with rs268236663 (CHI 2:16211260), located 600 kb from the endothelin converting enzyme like 1 (*ECEL1*) gene (**Supplemental Table S3**), which encodes an endopeptidase member of the M13 family involved in the regulation of neuropeptide and peptide hormone activity. This molecule has an important function in the development of the neuromuscular junctions of the limbs in mice (Nagata et al., 2016). Mice lacking this gene display poor arborization of the neuromotor nerves and a significant reduction of the number of neuromuscular junctions (Nagata et al., 2016). In humans, digital arthrogryposis is caused by mutations in *ECEL1*, and affected individuals show limited flexion of the knee and fingers as well as muscular atrophy (Dieterich et al., 2013).

Another interesting association was that between RLR and rs268286224 (CHI 24:42536581), which is positioned near the gene encoding piezo type mechanosensitive ion channel component 2 (*PIEZO2*; **Supplemental Table S3**), which is also involved in the etiology of digital arthrogryposis (Delle Vedove et al., 2016). Homozygous individuals for mutations inactivating *PIEZO2* suffer from arthrogryposis and scoliosis (Haliloglu et al., 2017), while carriers of gain-of-function mutations can present multiple congenital contractures of limbs and

variable absence of cruciate knee ligaments, among other symptoms (Coste et al., 2013).

The Genomic Architecture of Morphological Traits

Comparing our results with those obtained by Martin et al., 2018 and Mucha et al., 2018, we can state that there is a general lack of positional coincidence between the genomic regions associated with conformation traits in the aforementioned studies. Furthermore, in the study of Martin et al., 2018, different regions were identified in the Alpine and Saanen breeds as being associated with body phenotypes, suggesting that this lack of concordance is not the result of technical factors. Indeed, we found only 2 SNP displaying genome-wide significant associations with morphological traits. Such limited results are probably due to the fact that inheritance of morphological traits in goats is highly polygenic, with many genetic variants having small effects determining phenotypic variation. In other words, the success of GWAS largely depends on the genomic architecture of the trait rather than on the magnitude of its heritability. Stature is a good example of this because it is a highly heritable and, at the same time, a highly polygenic trait. In cattle, 163 genomic regions associated with stature have been detected but they explain only 13.8% of the phenotypic variance (Bouwman et al., 2018). Similarly, in humans, at least 180 genetic markers mostly segregating in populations of European descent explain ∼10% of the variance in height, a phenotype that has high heritability close to ∼0.8 (Yang et al., 2010). This remarkable gap between genealogical heritability and the percentage of the phenotypic variance explained by the SNPs is caused, at least in part, by the existence of hundreds or thousands of genetic determinants with small

phenotypic effects on stature. Similar reasoning can probably be made for the majority of body, leg, and udder morphological traits, because with the population sizes often used in bovine, ovine or caprine GWAS, the number of significant hits is generally very low (Schmid and Bennewitz, 2017).

Currently, we do not know whether the lack of positional concordance between GWAS for morphology traits recorded in goats is due to the existence of a substantial genetic heterogeneity across populations or to the modest size of the populations used in GWAS, a circumstance that limits our ability to detect variants with small effects that explain the majority of the phenotypic variance of morphological traits. Indeed, large GWAS performed in humans have demonstrated that many of the genetic variants that are associated with height are shared between individuals of European and African descent (N'Diaye et al., 2011), reinforcing the idea that the magnitude of genetic heterogeneity between populations is greatly reduced when large sample sizes are used in GWAS.

Table 3. Chromosome-wide significant associations between SNPs and leg traits (MOB, mobility; RLR, rear legs rear view; RLS, rear legs side view) recorded in 811 Murciano-Granadina goats.

	Trait	Chr[1]	rs[2]	Pos[3]	A1[4]	MAF[5]	β ± SE[6]	*P-value*[7]	*q-value*[8]	*PVE*[9]
Chromosome Wide	MOB	2	rs268236663	16,211,260	G	0.237	0.325 ± 0.072	8.13E-06	0.022	0.004
	RLR	17	rs268258221	21,284,278	G	0.367	-0.255 ± 0.061	2.98E-05	0.041	0.006
		19	rs268288174	39,252,784	G	0.188	0.340 ± 0.074	4.69E-06	0.005	0.004
		24	rs268259571	38,761,188	A	0.422	-0.278 ± 0.065	1.78E-05	0.023	0.005
		24	rs268286224	42,536,581	A	0.117	-0.376 ± 0.092	4.26E-05	0.027	0.002
	RLS	3	rs268254620	8,818,355	G	0.356	-0.188 ± 0.046	4.41E-05	0.050	0.010
		3	rs268285963	43,496,497	A	0.3	0.211 ± 0.048	1.20E-05	0.027	0.009

[1]**Chr**, chromosome; [2]**rs**, identifier code of the SNP according to the RefSNP database; [3]**Pos**, position in base pairs; [4]**A1,** minority allele; [5]**MAF**, allele frequency; [6]**β ± SE**, allelic substitution effect ± standard error; [7]***P-value***, raw *P*-values; [8]***q-value***, *P*-values corrected for multiple testing using a false discovery rate approach; [9]***PVE,*** proportion of variance in phenotype explained by a given SNP.

Conclusions

The number of hits detected in our GWAS for body conformation, udder, and leg traits was quite limited, a result that agrees well with previous studies. This outcome probably reflects the highly polygenic nature of morphological traits in ruminants. In comparing our results with previous reports, we also detected low positional concordance. This could be the consequence of genetic heterogeneity in the genetic determinism of morphological traits or because the GWAS carried out to date are underpowered to reliably detect the genetic determinants of such phenotypes. Despite these limitations, we identified several genes related to collagen synthesis (*ATF3*, *ADAMTS14*, and *COL14A1*), growth (*CGGBP1*), development (*WNT5A* and *DNAH14*), bone homeostasis and remodeling (*PTH1R*, *CDH11*, *SPATA4*, and *EPHA3*), limb development (*ECEL1* and *PIEZO2*), and mammary physiology (*SOCS7*) mapping close to GWAS hits. Such information, combined with candidate gene sets generated in other GWAS, if possible with much larger reference populations, could provide valuable clues about the identity of the loci shaping the body, udder, and leg morphology of goats.

Acknowledgments

The authors are indebted to the Asociación Nacional de Criadores de Caprino de Raza Murciano-Granadina (Caprigran) for their cooperation in this study, specifically to Miguel García García and Teresa Novo Díaz, who collected all blood samples, and to Antonio Martín Ramírez who recorded all linear score measurements. This research was funded by the European Fund for Regional Development/Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal

de Investigación/ Project Reference (AGL2016-76108-R). We acknowledge the financial support from the Spanish Ministry of Economy and Competitiveness, through the “Severo Ochoa Programme for Centres of Excellence in R&D” 2016-2019 (SEV-2015-0533), and from the CERCA programme of the Generalitat de Catalunya. Emilio Mármol-Sánchez was funded with an FPU PhD grant awarded by the Spanish Ministry of Education (FPU15/01733). Maria Gracia Luigi-Sierra was funded with an FPI PhD grant from the Spanish Ministry of Economy and Competitivity (BES-2017-079709). The authors have not stated any conflicts of interest.

References

Andre, P., H. Song, W. Kim, A. Kispert, and Y. Yang. 2015. WNT5A and WNT11 regulate mammalian anterior-posterior axis elongation. Development 142:1516–1527. https://doi.org/10.1242/dev.119065.

Archer, S. C., M. J. Green, and J. N. Huxley. 2010. Association between milk yield and serial locomotion score assessments in UK dairy cows. J. Dairy Sci. 93:4045–4053. https://doi.org/10.3168/jds.2010-3062.

Arun, S. J., P. C. Thomson, P. A. Sheehy, M. S. Khatkar, H. W. Raadsma, and P. Williamson. 2015. Targeted analysis reveals an important role of JAK-STAT-SOCS genes for milk production traits in Australian dairy cattle. Front. Genet. 6:342. https://doi.org/10.3389/fgene.2015.00342.

Atherton, A. J., M. J. Warburton, M. J. O’Hare, P. Monaghan, D. Schuppan, and B. A. Gusterson. 1998. Differential expression of type XIV collagen/undulin by human mammary gland intralobular and interlobular fibroblasts. Cell Tissue Res. 291:507–511.https://doi.org/10.1007/s004410051020.

Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn. 2007. GenABEL: An R library for genome-wide association analysis. Bioinformatics 23:1294–1296. https://doi.org/10.1093/bioinformatics/btm108.

Bastin, C., and N. Gengler. 2013. Genetics of body condition score as an indicator of dairy cattle fertility: A review. Biotechnol. Agron. Soc. Environ. 17:64–75.

Bekhouche, M., and A. Colige. 2015. The procollagen N-proteinases ADAMTS2, 3 and 14 in pathophysiology. Matrix Biol. 44–46:46–53. https://doi.org/10.1016/j.matbio.2015.04.001.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate—A practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bickhart, D. M., B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie, S. Chan, J. Lee, E. T. Lam, I. Liachko, S. T. Sullivan, J. N. Burton, H. J. Huson, J. C. Nystrom, C. M. Kelley, J. L. Hutchison, Y. Zhou, J. Sun, A. Crisà, F. A. Ponce De León, J. C. Schwartz, J. A. Hammond, G. C. Waldbieser, S. G. Schroeder, G. E. Liu, M. J. Dunham, J. Shendure, T. S. Sonstegard, A. M. Phillippy, C. P. Van Tassell, and T. P. L. Smith. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat. Genet. 49:643–650. https://doi.org/10.1038/ng.3802.

Boettcher, P. J., J. C. M. Dekkers, and B. W. Kolstad. 1998. Development of an udder health index for sire selection based on somatic cell score, udder conformation, and milking speed. J. Dairy Sci. 81:1157–1168. https://doi.org/10.3168/jds.S0022-0302(98)75678 -4.

Bouwman, A. C., H. D. Daetwyler, A. J. Chamberlain, C. H. Ponce, M. Sargolzaei, F. S. Schenkel, G. Sahana, A. Govignon-Gion, S. Boitard, M. Dolezal, H. Pausch, R. F. Brøndum, P. J. Bowman, B. Thomsen, B. Guldbrandtsen, M. S. Lund, B. Servin, D. J. Garrick, J. Reecy, J. Vilkki, A. Bagnato, M. Wang, J. L. Hoff, R. D. Schnabel, J. F. Taylor, A. A. E. Vinkhuyzen, F. Panitz, C. Bendixen, L. E. Holm, B. Gredler, C. Hozé, M. Boussaha, M. P. Sanchez, D. Rocha, A. Capitan, T. Tribout, A. Barbat, P. Croiseau, C. Drögemüller, V. Jagannathan, C. Vander Jagt, J. J. Crowley, A. Bieber, D. C. Purfield, D. P. Berry, R. Emmerling, K. U. Götz, M. Frischknecht, I. Russ, J. Sölkner, C. P. Van Tassell, R. Fries, P. Stothard, R. F. Veerkamp, D. Boichard, M. E. Goddard, and B. J. Hayes. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that

regulate body size in mammals. Nat. Genet. 50:362–367. https://doi.org/10.1038/ s41588-018-0056-5.

Calonge, M. E. 2004. Identificación de genes con expresión diferencial en tejido muscular de bovinos pertenecientes a los tres genotipos de la miostatina (mutación nt821(del11)). PhD Thesis. Universidad Complutense de Madrid, Spain.

Campbell, J. R., and R. T. Marshall. 2016. Dairy Production and Processing: The Science of Milk and Milk Products. 1st ed. Waveland Press Inc., Long Grove, IL.

Capote, J., A. Argüello, N. Castro, J. L. López, and G. Caja. 2006. Short communication: Correlations between udder morphology, milk yield, and milking ability with different milking frequencies in dairy goats. J. Dairy Sci. 89:2076–2079. https://doi.org/10.3168/ jds.S0022-0302(06)72276-7.

Castañeda-Bustos, V. J., H. H. Montaldo, M. Valencia-Posadas, L. Shepard, S. Pérez-Elizalde, O. Hernández-Mendo, and G. TorresHernández. 2017. Linear and nonlinear genetic relationships between type traits and productive life in US dairy goats. J. Dairy Sci. 100:1232–1245. https://doi.org/10.3168/jds.2016-11313.

Chacón, E., F. Macedo, F. Velázquez, S. R. Paiva, E. Pineda, and C. McManus. 2011. Morphological measurements and body indices for Cuban Creole goats and their crossbreds. Rev. Bras. Zootec. 40:1671. https://doi.org/10.1590/S1516-35982011000800007.

Chang, C. C., C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. https: //doi.org/10.1186/s13742-015-0047-8.

Compagni, A., M. Logan, R. Klein, and R. H. Adams. 2003. Control of skeletal patterning by EphrinB1-EphB interactions. Dev. Cell 5:217–230. https://doi.org/10.1016/S1534-5807(03)00198-9.

Coste, B., G. Houge, M. F. Murray, N. Stitziel, M. Bandell, M. A. Giovanni, A. Philippakis, A. Hoischen, G. Riemer, U. Steen, V. M. Steen, J. Mathur, J. Cox, M. Lebo, H. Rehm, S. T. Weiss, J. N. Wood, R. L. Maas, S. R. Sunyaev, and A. Patapoutian. 2013. Gain-of-function mutations in the mechanically activated ion channel PIEZO2 cause a subtype of distal arthrogryposis. Proc. Natl. Acad. Sci. USA 110:4667–4672. https://doi.org/10.1073/ pnas.1221400110.

de la Fuente, L. F., C. Gonzalo, J. P. Sánchez, R. Rodríguez, J. A. Carriedo, and F. S. Primitivo. 2011. Genetic parameters of the linear body conformation traits and genetic correlations with udder traits, milk yield and composition, and somatic cell count in dairy ewes. Can. J. Anim. Sci. 91:585–591. https://doi.org/10.4141/cjas2010-031.

Delgado, J. V., V. Landi, C. J. Barba, J. Fernández, M. M. Gómez, M. E. Camacho, M. A. Martínez, F. J. Navas, and J. M. León. 2018. Murciano-Granadina goat: A Spanish local breed ready for the challenges of the twenty-first century. Pages 205–219 in Sustainable Goat Production in Adverse Environments. Vol. 2. J. Simões and C. Gutiérrez, ed. Springer International Publishing, Heidelberg, Germany.

Delle Vedove, A., M. Storbeck, R. Heller, I. Hölker, M. Hebbar, A. Shukla, O. Magnusson, S. Cirak, K. M. Girisha, M. O'Driscoll, B. Loeys, and B. Wirth. 2016. Biallelic loss of proprioceptionrelated PIEZO2 causes muscular atrophy with perinatal respiratory distress, arthrogryposis, and scoliosis. Am. J. Hum. Genet. 99:1206–1216. https://doi.org/10.1016/j.ajhg.2016.09.019.

Di Benedetto, A., M. Watkins, S. Grimston, V. Salazar, C. Donsante, G. Mbalaviele, G. L. Radice, and R. Civitelli. 2010. N-cadherin and cadherin 11 modulate postnatal bone growth and osteoblast differentiation by distinct mechanisms. J. Cell Sci. 123:2640–2648. https://doi.org/10.1242/jcs.067777.

Dieterich, K., S. Quijano-Roy, N. Monnier, J. Zhou, J. Fauré, D. A. Smirnow, R. Carlier, C. Laroche, P. Marcorelles, S. Mercier, A. Mégarbané, S. Odent, N. Romero, D. Sternberg, I. Marty, B. Estournet, P. S. Jouk, J. Melki, and J. Lunardi. 2013. The neuronal endopeptidase ECEL1 is associated with a distinct form of recessive distal arthrogryposis. Hum. Mol. Genet. 22:1483–1492. https://doi.org/10.1093/hmg/dds514.

Dupont, L., G. Ehx, M. Chantry, C. Monseur, C. Leduc, L. Janssen, D. Cataldo, M. Thiry, C. Jerome, J. M. Thomassin, B. Nusgens, J. Dubail, F. Baron, and A. Colige. 2018. Spontaneous atopic dermatitis due to immune dysregulation in mice lacking ADAMTS2 and 14. Matrix Biol. 70:140–157. https://doi.org/10.1016/j.matbio.2018.04.002.

Edwards, C. M., and G. R. Mundy. 2008. Eph receptors and ephrin signaling pathways: A role in bone homeostasis. Int. J. Med. Sci. 5:263–272. https://doi.org/10.7150/ijms.5.263.

Egger-Danner, C., J. B. Cole, J. E. Pryce, N. Gengler, B. Heringstad, A. Bradley, and K. F. Stock. 2015. Invited review: Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal 9:191–207. https://doi.org/10.1017/ S1751731114002614.

Fernández Álvarez, J. 2017. Herramientas Informáticas para el Control y Gestión de Rebaños: Eskardillo, Kalifa, Inteka y Siamelk. VIII Foro Nacional del Caprino, Carmona (Sevilla), Spain. https: //www.cabrandalucia.com/inicio/foro-nacional-caprino/viii-foro -nacional-del-caprino/.

Guan, D., V. Landi, M. G. Luigi-Sierra, J. V. Delgado, X. Such, A. Castelló, B. Cabrera, E. Mármol-Sánchez, J. Fernández-Alvarez, J. L. R. de la Torre Casañas, A. Martínez, J. Jordana, and M. Amills. 2020. Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. J. Anim. Sci. Biotechnol. 11:35. https://doi.org/10.1186/s40104-020-00435-4.

Guarín, J. F., M. G. Paixão, and P. L. Ruegg. 2017. Association of anatomical characteristics of teats with quarter-level somatic cell count. J. Dairy Sci. 100:643–652. https://doi.org/10.3168/jds.2016 -11459.

Guenzle, J., L. J. Wolf, N. W. C. Garrelfs, J. M. Goeldner, N. Osterberg, C. R. Schindler, J. E. Saavedra, and A. Weyerbrock. 2017. ATF3 reduces migration capacity by regulation of matrix metalloproteinases via NFκB and STAT3 inhibition in glioblastoma. Cell Death Discov. 3:17006. https://doi.org/10.1038/cddiscovery.2017 .6.

Haliloglu, G., K. Becker, C. Temucin, B. Talim, N. Küçüksąhin, M. Pergande, S. Motameny, P. Nürnberg, U. Aydingoz, H. Topaloglu, and S. Cirak. 2017. Recessive PIEZO2 stop mutation causes distal arthrogryposis with distal muscle weakness, scoliosis and proprioception defects. J. Hum. Genet. 62:497–501. https://doi.org/10 .1038/jhg.2016.153.

Huang, W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44–57. https://doi.org/10.1038/nprot .2008.211.

Kinsella, R. J., A. Kähäri, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek. 2011. Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database (Oxford) 2011:bar030. https://doi.org/10.1093/database/bar030.

Legarra, A., and E. Ugarte. 2005. Genetic parameters of udder traits, somatic cell score, and milk yield in Latxa sheep. J. Dairy Sci. 88:2238–2245. https://doi.org/10.3168/jds.S0022-0302(05)72899 -X.

Lindholm, M., T. Manon-Jensen, and M. A. Karsdal. 2019. Type XIV collagen. Pages 121–125 in Biochemistry of Collagens, Laminins and Elastin. Vol. 2. M. A. Karsdal, ed. Academic Press, London, UK.

Lowes, D. A., C. Wallace, M. P. Murphy, N. R. Webster, and H. F. Galley. 2009. The mitochondria targeted antioxidant MitoQ protects against fluoroquinolone-induced oxidative stress and mitochondrial membrane damage in human achilles tendon cells. Free Radic. Res. 43:323–328. https://doi.org/10.1080/10715760902736275.

Lund, T., F. Miglior, J. C. M. Dekkers, and E. B. Burnside. 1994. Genetic relationships between clinical mastitis, somatic cell count, and udder conformation in Danish Holsteins. Livest. Prod. Sci. 39:243–251. https://doi.org/10.1016/0301-6226(94)90203-8.

Manfredi, E., A. Piacere, P. Lahaye, and V. Ducrocq. 2001. Genetic parameters of type appraisal in Saanen and Alpine goats. Livest. Prod. Sci. 70:183–189. https://doi.org/10.1016/S0301 -6226(01)00180-4.

Mannstadt, M., H. Jüppner, and T. J. Gardella. 1999. Receptors for PTH and PTHrP: Their biological importance and functional properties. Am. J. Physiol. 277:F665–F675. https://doi.org/10 .1152/ajprenal.1999.277.5.F665.

Martens, N., G. Uzan, M. Wery, R. Hooghe, E. L. Hooghe-Peters, and A. Gertler. 2005. Suppressor of cytokine signaling 7 inhibits prolactin, growth hormone, and leptin signaling by interacting with STAT5 or STAT3 and attenuating their nuclear

translocation. J. Biol. Chem. 280:13817–13823. https://doi.org/10.1074/jbc .M411596200.

Martin, P., I. Palhière, C. Maroteau, V. Clément, I. David, G. T. Klopp, and R. Rupp. 2018. Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. J. Dairy Sci. 101:5214–5226. https://doi.org/10.3168/jds.2017-13625.

McLaren, A., S. Mucha, R. Mrode, M. Coffey, and J. Conington. 2016. Genetic parameters of linear conformation type traits and their relationship with milk yield throughout lactation in mixed-breed dairy goats. J. Dairy Sci. 99:5516–5525. https://doi.org/10.3168/ jds.2015-10269.

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

Miller, S. A., D. D. Dykes, and H. F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16:1215. https://doi.org/10.1093/nar/16.3 .1215.

Montaldo, H. H., and E. Manfredi. 2002. Organisation of selection programmes for dairy goats. No. 01–35 in 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France.

Mucha, S., R. Mrode, M. Coffey, M. Kizilaslan, S. Desire, and J. Conington. 2018. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. J. Dairy Sci. 101:2213–2225. https://doi.org/10.3168/jds.2017-12919.

N'Diaye, A., G. K. Chen, C. D. Palmer, B. Ge, B. Tayo, R. A. Mathias, J. Ding, M. A. Nalls, A. Adeyemo, V. Adoue, C. B. Ambrosone, L. Atwood, E. V. Bandera, L. C. Becker, S. I. Berndt, L. Bernstein, W. J. Blot, E. Boerwinkle, A. Britton, G. Casey, S. J. Chanock, E. Demerath, S. L. Deming, W. R. Diver, C. Fox, T. B. Harris, D. G. Hernandez, J. J. Hu, S. A. Ingles, E. M. John, C. Johnson, B. Keating, R. A. Kittles, L. N. Kolonel, S. B. Kritchevsky, L. Le Marchand, K. Lohman, J. Liu, R. C. Millikan, A. Murphy, S. Musani, C. Neslund-Dudas, K. E. North, S. Nyante, A.

Ogunniyi, E. A. Ostrander, G. Papanicolaou, S. Patel, C. A. Pettaway, M. F. Press, S. Redline, J. L. Rodriguez-Gil, C. Rotimi, B. A. Rybicki, B. Salako, P. J. Schreiner, L. B. Signorello, A. B. Singleton, J. L. Stanford, A. H. Stram, D. O. Stram, S. S. Strom, B. Suktitipat, M. J. Thun, J. S. Witte, L. R. Yanek, R. G. Ziegler, W. Zheng, X. Zhu, J. M. Zmuda, A. B. Zonderman, M. K. Evans, Y. Liu, D. M. Becker, R. S. Cooper, T. Pastinen, B. E. Henderson, J. N. Hirschhorn, G. Lettre, and C. A. Haiman. 2011. Identification, replication, and fine-mapping of loci associated with adult height in individuals of African ancestry. PLoS Genet. 7:e1002298. https: //doi.org/10.1371/journal.pgen.1002298.

Nagata, K., S. Kiryu-Seo, H. Tamada, F. Okuyama-Uchimura, H. Kiyama, and T. C. Saido. 2016. ECEL1 mutation implicates impaired axonal arborization of motor nerves in the pathogenesis of distal arthrogryposis. Acta Neuropathol. 132:111–126. https://doi .org/10.1007/s00401-016-1554-0.

Oget, C., G. Tosser-Klopp, and R. Rupp. 2019. Genetic and genomic studies in ovine mastitis. Small Rumin. Res. 176:55–64. https:// doi.org/10.1016/j.smallrumres.2019.05.011.

Paradies, G., V. Paradies, V. De Benedictis, F. M. Ruggiero, and G. Petrosillo. 2014. Functional role of cardiolipin in mitochondrial bioenergetics. Biochim. Biophys. Acta 1837:408–417. https://doi .org/10.1016/j.bbabio.2013.10.006.

Pérez-Cabal, M. A., E. Legaz, I. Cervantes, L. F. de la Fuente, R. Martínez, F. Goyache, and J. P. Gutiérrez. 2013. Association between body and udder morphological traits and dairy performance in Spanish Assaf sheep. Arch. Tierzucht 56:430–442. https://doi.org/ 10.7482/0003-9438-56-042.

Pugh, D. G., and A. N. Baird. 2012. Sheep and Goat Medicine. 2nd ed. Elsevier, St. Louis, MO. Qian, D., C. Jones, A. Rzadzinska, S. Mark, X. Zhang, K. P. Steel, X. Dai, and P. Chen. 2007. WNT5A functions in planar cell polarity regulation in mice. Dev. Biol. 306:121–133. https://doi.org/10 .1016/j.ydbio.2007.03.011.

R Core Team. 2017. R: A Language and environment for statistical computing. R Found. Stat. Comput., Vienna, Austria. http//www .R-project.org/.

Rogers, G. W., G. L. Hargrove, T. J. Lawlor Jr., and J. L. Ebersole. 1991. Correlations among linear type traits and somatic cell counts. J. Dairy Sci. 74:1087–1091. https://doi.org/10.3168/jds .S0022-0302(91)78259-3.

Rupp, R., and D. Boichard. 1999. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. J. Dairy Sci. 82:2198–2204. https: //doi.org/10.3168/jds.S0022-0302(99)75465-2.

Rupp, R., and D. Boichard. 2003. Genetics of resistance to mastitis in dairy cattle. Vet. Res. 34:671–688. https://doi.org/10.1051/vetres: 2003020.

Rupp, R., V. Clément, A. Piacere, C. Robert-Granié, and E. Manfredi. 2011. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. J. Dairy Sci. 94:3629–3634. https://doi .org/10.3168/jds.2010-3694.

Rupp, R., P. Senin, J. Sarry, C. Allain, C. Tasca, L. Ligat, D. Portes, F. Woloszyn, O. Bouchez, G. Tabouret, M. Lebastard, C. Caubet, G. Foucras, and G. Tosser-Klopp. 2015. A point mutation in suppressor of cytokine signalling 2 (SOCS2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. PLoS Genet. 11:e1005629. https://doi.org/10.1371/journal .pgen.1005629.

Sánchez-Rodríguez, M. 2012. Valoración morfológica del ganado caprino lechero. Juzgamiento y calificación. 2nd ed. Editorial Servet, Zaragoza, Spain.

Schmid, M., and J. Bennewitz. 2017. Invited review: Genome-wide association analysis for quantitative traits in livestock—A selective review of statistical models and experimental designs. Arch. Tierzucht 60:335–346. https://doi.org/10.5194/aab-60-335-2017.

Sevane, N., E. Armstrong, P. Wiener, R. Pong Wong, S. Dunner, V. Amarger, D. Delourme, H. Levéziel, S. Boitard, B. Mangin, J. Cañón, M. L. Checa, D. García, M. E. Miranda, R. Pérez, M. Christensen, P. Ertbjerg, A. Crisá, C. Marchitelli, A. Valentini, S. Failla, S. Gigli, J. F. Hocquette, G. Nute, I. Richardson, J. L. Olleta, B. Panea, C. Sañudo, N. Razzaq, G. Renand, and J. L. Williams. 2014. Polymorphisms in twelve candidate genes are associated with growth, muscle lipid

profile and meat quality traits in eleven European cattle breeds. Mol. Biol. Rep. 41:4721–4731. https://doi.org/10.1007/s11033-014-3343-y.

Seykora, A. J., and B. T. McDaniel. 1985. Udder and teat morphology related to mastitis resistance: A review. J. Dairy Sci. 68:2087– 2093. https://doi.org/10.3168/jds.S0022-0302(85)81072-9.

Shelton, M. 1978. Reproduction and breeding of goats. J. Dairy Sci. 61:994–1010. https://doi.org/10.3168/jds.S0022-0302(78)83680-7.

Shim, H., D. I. Chasman, J. D. Smith, S. Mora, P. M. Ridker, D. A. Nickerson, R. M. Krauss, and M. Stephens. 2015. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 caucasians. PLoS One 10:e0120758. https://doi.org/10.1371/journal.pone.0120758.

Sutherland, K. D., G. J. Lindeman, and J. E. Visvader. 2007. Knocking off SOCS genes in the mammary gland. Cell Cycle 6:799–803. https://doi.org/10.4161/cc.6.7.4037.

Thankam, F. G., I. S. Chandra, A. N. Kovilam, C. G. Diaz, B. T. Volberding, M. F. Dilisio, M. M. Radwan, R. M. Gross, and D. K. Agrawal. 2018. Amplification of mitochondrial activity in the healing response following rotator cuff tendon injury. Sci. Rep. 8:17027. https://doi.org/10.1038/s41598-018-35391-7.

UniProt Consortium. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47(D1):D506–D515. https://doi .org/10.1093/nar/gky1049.

Wang, X., K. Harimoto, J. Liu, J. Guo, S. Hinshaw, Z. Chang, and Z. Wang. 2011. SPATA4 promotes osteoblast differentiation through Erk-activated Runx2 pathway. J. Bone Miner. Res. 26:1964–1973. https://doi.org/10.1002/jbmr.394.

Yamaguchi, T. P., A. Bradley, A. P. McMahon, and S. Jones. 1999. A WNT5A pathway underlies outgrowth of multiple structures in the vertebrate embryo. Development 126:1211–1223.

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

Yang, Y., J. Cao, and Y. Shi. 2004. Identification and characterization of a gene encoding human LPGAT1, an endoplasmic reticulumassociated lysophosphatidylglycerol acyltransferase. J. Biol. Chem. 279:55866–55874. https://doi.org/10.1074/jbc.M406710200.

Yildiz, A. 2018. Single-molecule dynein motor mechanics in vitro. Pages 113–135 in Dynein Mechanics, Dysfunction, and Disease. 2nd ed. S. M. King, ed. Academic Press, London, UK.

Zhou, H., D. F. Shen, Z. Y. Bian, J. Zong, W. Deng, Y. Zhang, Y. Y. Guo, H. Li, and Q. Z. Tang. 2011. Activating transcription factor 3 deficiency promotes cardiac hypertrophy, dysfunction, and fibrosis induced by pressure overload. PLoS One 6:e26744. https://doi .org/10.1371/journal.pone.0026744.

Zhou, X., and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44:821–824. https:// doi.org/10.1038/ng.2310.

Zujovic, M., N. Memisi, V. Bogdanovic, and Z. Tomic. 2011. Correlation between body measurements and milk production of goats in different lactations. Biotechnol. Anim. Husb. 27:217–225. https:// doi.org/10.2298/BAH1102217Z.

Supplementary Tables

Supplemental Table S1. Descriptive statistics of body, udder and leg phenotypes recorded in 811 Murciano-Granadina goats and expressed as linear scores.

Supplemental Table S2. Phenotypic Pearson correlations matrix among morphological and milk yield traits recorded in 811 Murciano-Granadina goats. The lower diagonal shows the correlation between each pair of traits, the upper diagonal indicates the p-value of each correlation.

Supplemental Table S3. Distance in base pairs between SNPs showing significant associations with morphological traits and the closest functional and positional candidate gene identified in the goat ARS1 reference genome.

Supplementary Figures

Supplemental Figure S1. Principal component analysis based on the Illumina Goat SNP50 BeadChip (Illumina inc, San Diego, CA) genotypes of 811 Murciano-Granadina distributed in 13 farms (each one encoded by a different color) and used in the current GWAS.

Supplemental Figure S2. Histograms depicting the number of Murciano-Granadina goats corresponding to each one of the linear score categories defining udder traits. The following abbreviations have been used, **FUA**, fore udder attachment; **RUH**, rear udder height; **MSL**, medial suspensory ligament; **UW**, udder width; **UD**, udder depth; **TP**, teat placement; **TD**, teat diameter.

Supplemental Figure S3. Histograms depicting the number of Murciano-Granadina goats corresponding to each one of the linear score categories defining body (A) and leg (B) traits. The following abbreviations have been used, **HT**, height; **CW**, chest width; **BD**, body depth; **RW**, rump width; **RA**, rump angle; **ANG**, angularity; **BQ**, bone.

Supplemental Figure S4. Manhattan plots depicting negative $\log_{10}$ *P*-values (*y*-axis) of the associations at the chromosome-wide level between SNPs and udder morphology traits plotted against the genomic location of each SNP marker (*x*-axis). Only chromosomes hosting significant associations are shown. The dotted line indicates the $-\log_{10}$(*P-value*) equivalent to a *q*-value of 0.05. The following

abbreviations have been used, **MSL**, medial suspensory ligament; **TD**, teat diameter; **TP**, teat placement; **UD**, udder depth; **UW**, udder width.

Supplemental Figure S5. Quantile-Quantile plots showing the expected distribution of the -log10 *P*-values (*x*-axis) compared to the observed -log10 *P*-values (*y*-axis) in the GWAS performed at the chromosome wide level for udder traits.The mean lambda (λ) inflation factor was 1.060 ± 0.201. The following abbreviations have been used, **MSL**, medial suspensory ligament; **TD**, teat diameter; **TP**, teat placement; **UD**, udder depth; **UW**, udder width.

Supplemental Figure S6. Manhattan plots depicting negative log10 *P*-values (*y*-axis) of the associations at the chromosome-wide level between SNPs and body morphology traits plotted against the genomic location of each SNP marker (*x*-axis). Only chromosome hosting significant associations are shown. The dotted line indicates the -log10(*P-value*) equivalent to a *q*-value of 0.05. The following abbreviations have been used, **ANG**, angularity; **BD**, body depth; **BQ**, bone quality; **CW**, chest width; **HT**, height; **RA**, rump angle; **RW**, rump width.

Supplemental Figure S7. Quantile-Quantile plots showing the expected distribution of the -log10 *P*-values (*x*-axis) compared to the observed -log10 *P*-values (*y*-axis) in the GWAS performed at the chromosome wide level for body traits. The mean lambda (λ) inflation factor was 1.090 ± 0.113. The following abbreviations have been used, **ANG**, angularity; **BD**, body depth; **BQ**, bone quality; **CW**, chest width; **HT**, height; **RA**, rump angle; **RW**, rump width.

Supplemental Figure S8. Manhattan plots depicting negative log10 *P-values* (*y*-axis) of the associations at the chromosome-wide level between SNPs and leg traits plotted against the genomic location of each SNP marker (*x*-axis). Only chromosomes hosting significant associations are shown. The dotted line indicates the -log10(*P-value*) equivalent to a *q*-value of 0.05. The following abbreviations have been used, **MOB**, mobility; **RLR**, rear legs rear view; **RLS**, rear legs side view.

Supplemental Figure S9. Quantile-Quantile plots showing the expected distribution of the -log10 *P*-values (*x*-axis) compared to the observed -log10 *P*-values (*y*-axis) in the GWAS performed at the chromosome wide level for leg traits goats.

The mean lambda (λ) inflation factor was 1.098 ± 0.272. The following abbreviations have been used, **MOB**, mobility; **RLR**, rear legs rear view; **RLS**, rear legs side view.

RUNNING HEAD: Goat cheese microbiome

Microbiome composition of six commercial goat cheeses based on targeted 16S rRNA sequencing

M. G. Luigi-Sierra[1], Y. Ramayo-Caldas[2], D. Guan[3], M. Amills[1,4*]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain.
[2]Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, España.
[3]Department of Animal Science, University of California Davis, 95616, Davis, USA.
[4]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.

Corresponding author: *

Submitted Food Microbiology

Abstract

In this study, we have characterised the microbiota of six soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses by sequencing the V3-V4 ultravariable region of the 16S rRNA gene. Our data show that lactic acid bacteria from the genera *Lactococcus*, *Lactobacillus*, *Streptococcus*, and *Leuconostoc* are predominant in all six goat cheeses. Furthermore, we have identified several psychrophilic taxa, like *Pseudomonas*, *Pseudoalteromonas* and *Shewanella*, that are highly abundant in fresh cheeses (H and M). The presence of such microorganisms is often explained by the post-pasteurisation contamination of refrigerated milk. We have also compared α-diversity indices between the core, middle part and rind microbiomes of all six goat cheeses and such analysis did not reveal any significant difference, thus challenging the notion that the core microbiome is less diverse than that of the rind due to increased acidity and anaerobiosis. Finally, the construction of a dendrogram based on Aitchison distances revealed that cheese samples cluster according to their manufacturing characteristics, with a clear distinction between fresh vs semi-ripened or aged cheeses. This finding is consistent with the key impact of the ripening process on the composition of the cheese microbiome.

Keywords: Microbiome, Metagenomics, Goat cheese, Bacteria, 16S rRNA, Cheese.

Introduction

The complex network of interactions between enzymes, microorganisms and milk components determines, to a remarkable extent, the technological and organoleptic properties of cheese (Fox et al., 1995). Lactic acid bacteria (LAB), including mesophilic (e.g. *Lactococcus lactis* and *Leuconostoc*) and thermophilic (e.g. *Streptococcus thermophilus*) microorganisms, are regularly used as starters in cheese manufacturing to increase the acidity of milk to the desired point and to enhance the production of volatile and non-volatile compounds through proteolysis and lipolysis (Fox et al., 2017; Parente et al., 2017). Other microorganisms not added during the cheesemaking process can also have a strong impact on the physico-chemical and sensorial features of cheese (Fox et al., 2017).

Besides the raw ingredients of milk and the use of specific starter cultures, the composition of the cheese microbiome is also affected by technological factors associated with the manufacturing process. For instance, milk pasteurisation decreases the diversity of the cheese microbiome since only thermoduric bacteria survive exposure to high temperatures (Salazar et al., 2018). Poor hygienic conditions in the dairy plant or insufficient post-pasteurisation refrigeration (> 7-12ºC) might favour the growth of psychrotrophic bacteria, such as *Enterobacteriaceae* and *Pseudomonadaceae*, associated with spoilage (Sameli et al., 2021). The composition of cheese microbiome also changes during brining, rind development and ageing (Kamilari et al., 2019). For instance, cheese moisture and nutrient availability might become remarkably reduced during ripening, causing the

autolysis of certain bacteria (Salazar et al. 2018). There is also evidence that microbial diversity in the core part of cheese might be lower than in the rind, probably because the surface of the cheese is less acidic and more exposed to environmental influences (Montel et al., 2014).

Sequencing of marker genes, such as the 16S rRNA locus, and whole metagenome sequencing have made it possible to uncover, with unprecedented resolution, the main features of the cheese microbiome. For instance, the meta-analysis of the microbiomes of 184 cheese samples from cattle, sheep and goats revealed that bacteria are overwhelmingly predominant (78%), followed by viruses (20%) and, to a much lower extent (2%), eukaryotes (Walsh et al., 2020). Lactic acid bacteria from the order *Lactobacillales* were the most abundant ones, being *Lactococcus lactis* (78%), *Streptococcus thermophilus* (43.5%) and *Lactobacillus helveticus* (37%) particularly prevalent (Walsh et al., 2020). High throughput sequencing techniques have also demonstrated the presence, in cheese, of microbes hard to detect with traditional culture methods and present in small proportions such as *Pseudoalteromonas, Facklamia, Halomonas, Arthrobacter, Brachybacterium*, and *Vibrio*, to mention a few (Kamilari et al., 2019).

The majority of what is currently known about the cheese microbiome comes from the sequencing analysis of cow products (Afshari et al., 2020; Camargo et al., 2021; Dolci et al., 2014; Guidone et al., 2016; Salazar et al., 2018). In strong contrast, cheese microbiomes from goats have been much less characterised despite the fact that most of caprine milk is devoted to the production of cheese. The meta-analysis

of cheese microbiomes performed by Walsh et al. (2020) included a few goat cheese samples, although they were not the main focus of such full-scale metagenomics study, and the microbiomes of Gouda (Salazar et al., 2018) and Pélardon (Penland et al., 2021) cheeses have been reported. However, a thorough description and comparison of the microbial communities of a broader diversity of caprine cheese types is still lacking. In the current work, we have characterised the microbiomes of six Spanish commercial goat fresh and ripened cheeses with different textures (soft, semi-hard and hard) with the aim of providing a more comprehensive picture of the microbial diversity of this dairy product.

Materials and methods

Sample collection and DNA extraction.

We acquired six commercial Spanish goat cheeses (**Table 1)** that were stored at 4ºC until processing. Three samples were retrieved from the rind, middle part and the core of each cheese by using sterilised instruments in a UV-irradiated cabinet to avoid external microbial contamination. Cheese sample homogenisation and DNA extraction were performed with the DNeasy PowerFood Microbial Kit (Qiagen, Redwood) in accordance with the instructions of the manufacturer with a slight modification i.e. the lysate was heated at 56ºC before vortex homogenisation.

Library construction.

To construct the libraries for sequencing, an initial PCR was performed by using the KAPA HiFi PCR Mix (Kapa Biosystems, Cape Town) to amplify the

hypervariable V3-V4 region of the bacterial 16S rRNA gene with the set of universal primers 5'-CCTACGGGNGGCWGCAG-3' and 5'-GACTACHVGGGTATCTAATCC-3'. Amplicons were purified with the AMPure XP beads (Beckman Coulter, Indianapolis). Afterwards, a second PCR was carried out to add Nextera adapters, with barcodes for multiplexed sequencing, to amplicons. Subsequently, the concentration of the PCR product was normalised and purified with the SequalPrep kit (ThermoFisher, Barcelona, Spain). The sequencing of the libraries was performed on an Illumina MiSeq equipment in order to generate 2 × 301 bp reads. All library construction and sequencing tasks were performed at the Centre de Regulació Genòmica (CRG, Barcelona, Spain).

Table 1. Main features of the six goat cheeses analysed in this study.

Cheese ID	Texture	Rind	Milk treatment	Elaboration	Fat content	Production type
H	Soft	-	Pasteurization	Fresh	21 %	Industrial
M	Soft	-	Pasteurization	Fresh	23 %	Artisanal
C	Soft	Mouldy	Raw	Semi-ripened 21 – 45 days	?	Artisanal
P	Soft	Mouldy	Pasteurization	Semi-ripened 21 – 45 days	55 %	Artisanal
B	Hard	Bloomy	Pasteurization	Semi-ripened 30 – 42 days	55 %	Artisanal
G	Semi-hard	Natural & thin coat of wax	Pasteurization	Aged 180 days	35 %	Artisanal

Quality control of the data.

The quality control of the data was conducted with the QIIME2 version 2021.4 software (Bolyen et al., 2019). Demultiplexed sequences were retrieved as fastq files. Subsequently, primers were removed with the *cutadapt* plugin in QIIME2 (Martin, 2011). Singletons and sequences with length and quality below 260 bp and 20, respectively, were removed from the dataset. Reads were denoised and grouped into amplicon sequence variants (ASVs) using the DADA2 protocol (https://benjjneb.github.io/dada2/tutorial.html), and chimaeras were removed with the consensus method (Callahan et al., 2016). Filtered reads were aligned with MAFFT (Katoh and Standley, 2013) to generate phylogenetic trees.

Taxonomical analysis.

A taxonomical classifier was created using as reference DAIRYdb 2.0 (Meola et al., 2019), a manually curated database optimised for dairy product environments. The classifier was trained exclusively on the target reads generated with primers V3-V4 to improve its accuracy (Werner et al., 2012). Taxonomy was assigned to the identified ASVs using the feature-classifier *classify-sklearn* command of QIIME2 (Bolyen et al., 2019; Pedregosa et al., 2012). The unassigned features and those categorised as eukaryotes, archaea, mitochondria or bacteria with a taxonomy depth inferior to family, were removed. Barplots depicting the relative abundance of bacterial families and genera were built with R (R Core team, 2015). Bacterial taxa with relative abundances below 1% were catalogued as “Other”.

Microbial diversity analyses.

The feature table, phylogenetic tree and metadata were exported to R as a *phyloseq* object using the QIIME2R package (Bolyen et al., 2019). Sequences were rarefied to a depth of 60,000 to avoid library size bias (**Supplementary Figure 1**). Chao1, Shannon and Faith's phylogenetic diversity (Faith's PD) α-diversity indices were estimated with the R packages Microbiome (Lahti and Shetty, 2019) and Picante (Kembel et al., 2010). The non-parametric Kruskal-Wallis test (Kruskal and Wallis, 1952) was used to compare α-diversities between core, middle part and rind of all six cheeses. Bray-Curtis, unweighted Unifrac and weighted Unifrac β-diversity indices were estimated with the Vegan R package (Oksanen et al., 2022).

Results and discussion

Predominance of lactic acid bacteria (LAB) in goat cheeses.

A total of 3,831,251 reads were obtained by sequencing 18 cheese samples (212,847.28 reads per sample on average). Sequencing data can be accessed at https://doi.org/10.6084/m9.figshare.20179727.v1. After applying quality control procedures and removing undesired taxa, we retained 1,639,816 reads assigned to 283 ASVs for downstream analyses. Cheese microbiota was characterised at the genus level, because the sequencing of marker gene regions (such as 16S rRNA) allows a good taxonomical resolution up to such a level (Knight et al., 2018).

Similar to data reported in cow and ewe (Choi et al., 2020b; Walsh et al., 2020), the main genera identified in all goat cheese samples corresponded to LAB commonly

used as starters in cheesemaking e.g. *Lactococcus*, *Lactobacillus*, *Leuconostoc* and *Streptococcus* (**Figures 1** and **2**). Lactic acid bacteria can catabolise lactose to lactic acid, a metabolite that decreases the pH of milk, improves curd formation and, at the same time, prevents the growth of pathogenic bacteria such as *Listeria* (Beresford et al., 2001; Coelho et al., 2014). Depending on the technological treatment of the cheese, LAB can originate from external addition as starter or come from the raw milk microbiota and behave as an endogenous starter (Quigley et al., 2013).

In a meta-analysis of cheese microbiomes, Walsh et al. (2020) showed that three bacterial species were the most prevalent in a wide repertoire of cheese samples i.e. *Lactococcus lactis* (78%), *Streptococcus thermophilus* (43.5%) and *Lactobacillus helveticus* (37%). Consistently, in our survey *Streptococcus* was very abundant in soft (cheese P, 48.38 – 63.48%) and hard (cheese B, 40.48 – 49.53%) semi-ripened samples as well as in semi-hard aged cheese G (40.90 – 48.23%). Thermophilic starters based on *Streptococcus* are commonly used to enhance flavour development because they hydrolyse α_{s1} and β caseins (Gomez et al., 1998), thus releasing amino acids and small peptides which are subsequently metabolised into flavour compounds e.g. esters, alcohols and aldehydes (Smit et al., 2005).

The genus *Lactococcus*, a mesophilic starter, dominated the microbiota of soft fresh cheese M (relative abundance of 33.73 – 39.81%) and hard-paste semi-ripened cheese B (relative abundance of 43.95 – 51.93%), and it was also abundant in soft-paste semi-ripened cheese C (relative abundance of 26.23 – 32.73%). The

Lactococcus genus was also highly prevalent (~40%) in goat Gouda cheese (Salazar et al. 2018). *Lactococcus* are sensitive to low pH (< 7) and temperatures over 42 ºC (Blaya et al., 2018), therefore, they are particularly abundant during the beginning of the ripening process. In Pélardon goat cheese, for instance, *Lactococcus lactis* abundance peaked at day 2 of ripening and then declined until becoming undetectable, neither in core nor rind, at three months of ripening (Penland et al. 2021).

Additional, but less prevalent, LAB genera were also detected in goat cheeses i.e. *Fructilactobacillus* (21.69 – 44.39% in soft semi-ripened cheese C), *Lacticaseibacillus* (22.70 – 44.61% in soft fresh cheese H and 34.01 – 44.26% in semi-hard aged cheese G) and *Leuconostoc* (18.24 – 19.16% and 3.43 – 18.66% in soft semi-ripened cheeses C and P). *Lacticaseibacillus* can inhibit the growth of undesired psychrotrophic bacteria (e.g. *Pseudomonas*), prolonging the shelf-life of cheese (Bassi et al., 2020). In Pélardon cheese, *Lacticaseibacillus paracasei/casei* happened to be the predominant species in cores at 62 and 90 days of ripening, while *Leuconostoc* was one of the most dominant genera during curdling (Penland et al. 2021). *Leuconostoc* bacteria are particularly relevant in cheesemaking because they can produce CO_2, which forms “eyes” on the cheese surface, and they also synthesize volatile compounds, like diacetyl and acetoin, from lactate and citrate (Beresford and Williams, 2004).

Non-LAB microorganisms frequently used as starters were also identified in the six goat cheeses under study. For instance, the flagellated *Hafnia* genus was present in

soft-paste ripened cheeses with mouldy rind (**Figure 2**), being particularly abundant in the rind of cheeses C (19.87%) and P (21.36%). Bacteria from this genus were not detected neither in Pélardon nor Gouda goat cheeses (Salazar et al. 2018, Penland et al. 2021), and this genus was also unreported in the large-scale meta-analysis performed by Walsh et al (2020) evidencing its uncommonness in goat and cattle cheeses. Microorganisms from the genus *Hafnia* have been associated with food spoilage, nevertheless, certain species, like *Hafnia alvei*, are sometimes used as starters since they facilitate the development of flavour by enhancing proteolysis (Morales et al., 2003) and by generating volatile sulfur compounds (Irlinger et al., 2012). Another non-LAB genus of importance is *Staphylococcus* (Walsh et al., 2020), but its abundance in the six goat cheeses was quite low (0 – 0.64%). In contrast, *Staphylococcus equorum* was present in 28.9% of the cheese samples sequenced by Walsh et al. (2020), probably because this species is a frequent ingredient of starter cultures (Place et al., 2003).

Presence of psychrophilic and/or halotolerant bacteria in cheese samples

As shown in **Figure 2,** a high abundance of bacteria from the genera *Pseudomonas* and *Pseudoalteromonas* was detected in fresh soft goat cheeses H (13.92 – 39.80% *Pseudomonas*, 5.63 – 57.47% *Pseudoalteromonas*) and M (3.86 – 21.87% *Pseudomonas*, 0.44 – 4.32% *Pseudoalteromonas*), while *Pseudomonas* (but not *Pseudoalteromonas*) was abundant in soft semi-ripened cheeses P and C (0.84 – 18.85% *Pseudomonas*). *Pseudomonas* are flagellated bacteria that can easily move along the fungal network of the rind (Zhang et al., 2018), so they are commonplace

in this compartment. In our study, *Pseudomonas* were more abundant in the rind of soft fresh M and soft semi-ripened C cheeses, but such trend was not observed in soft fresh H and soft semi-ripened P cheeses (**Figure 2**). Furthermore, *Shewanella* represented 8.78 – 13.51 % and 0.19 – 0.69 % of the microbiome of fresh M and H cheeses, respectively, while its presence in the ripened cheeses was negligible, if any (**Figure 2**). *Shewanella* is a milk contaminant known for producing a putrid odour due to the synthesis of volatile sulfides, amines, and the fishy-smelling compound trimethylamine (Stepaniak, 2022).

The presence of bacteria from the *Pseudomonas*, *Pseudoalteromonas* and *Shewanella* genera in several goat cheeses (**Figures 2 and 3**) might be due to the fact that these microorganisms are psychrophilic or psychrotolerant, being able to grow at very low temperatures. Before pasteurisation, raw milk is kept refrigerated both the farm and in the dairy plant, providing an opportunity for the growth of psychrotolerant bacteria, mainly from the genus *Pseudomonas* (De Jonghe et al., 2011). The synthesis of thermoresistant extracellular proteases and lipases by these microorganisms can provoke the spoilage of pasteurised milk (De Jonghe et al., 2011). On the other hand, post-pasteurisation contamination of refrigerated milk by *Pseudomonas* and other psychrophilic microorganisms present in the facilities or tools of the dairy plant might also occur (Martin et al., 2018).

Figure 1. Relative abundances of bacterial families identified in the rind, middle and core samples from soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses. Bacterial families with relative abundances below 1% have been catalogued as "Other".

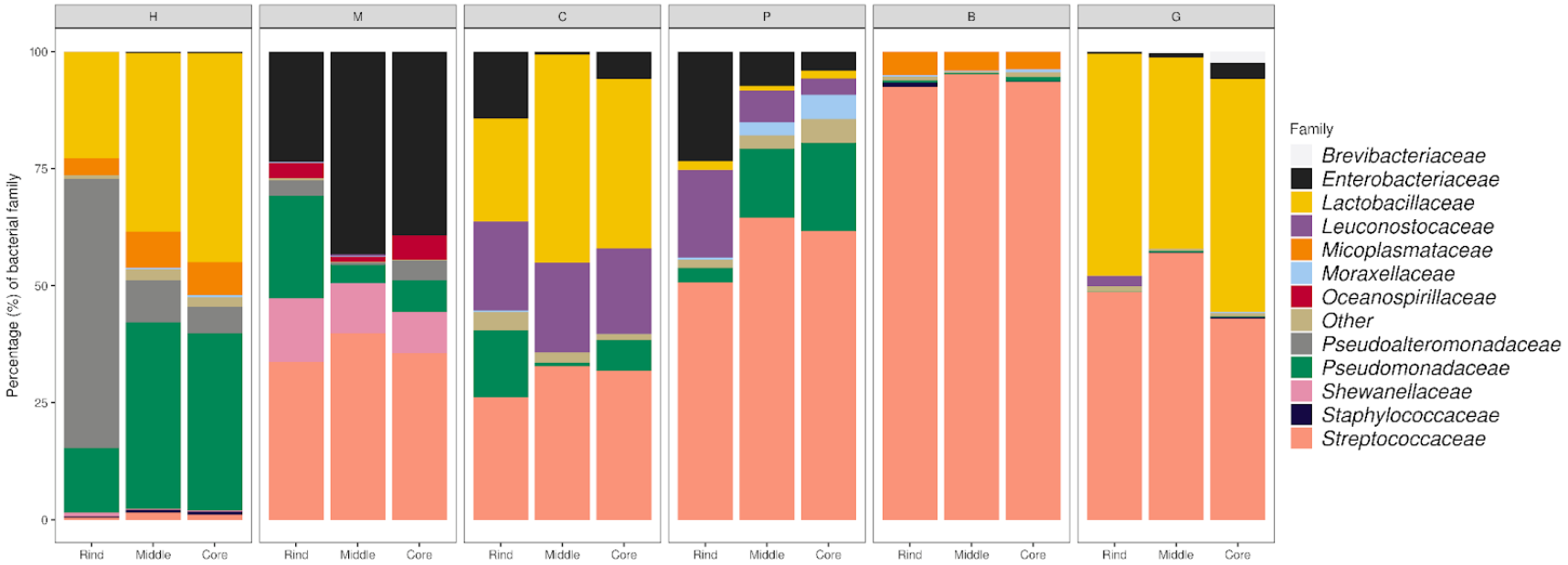


Figure 2. Relative abundances of bacterial genera identified in the rind, middle part and core samples from soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses. Bacterial genera with relative abundances below 1% have been catalogued as "Other".

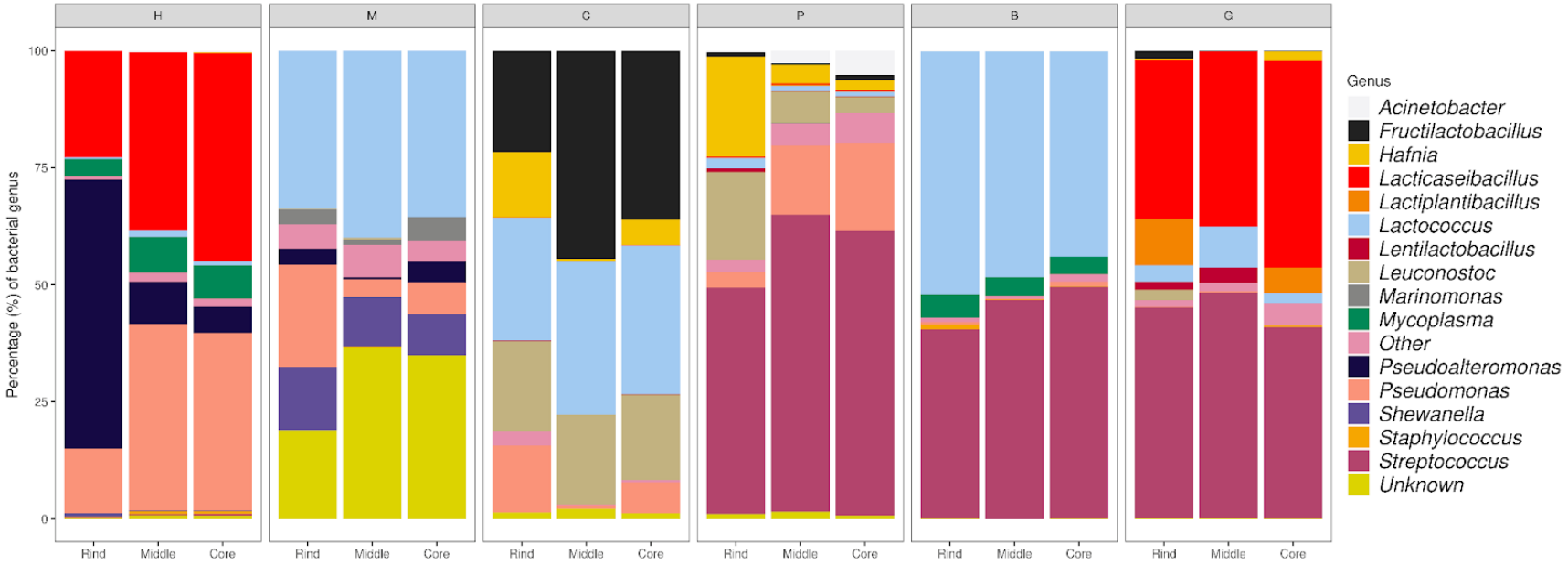
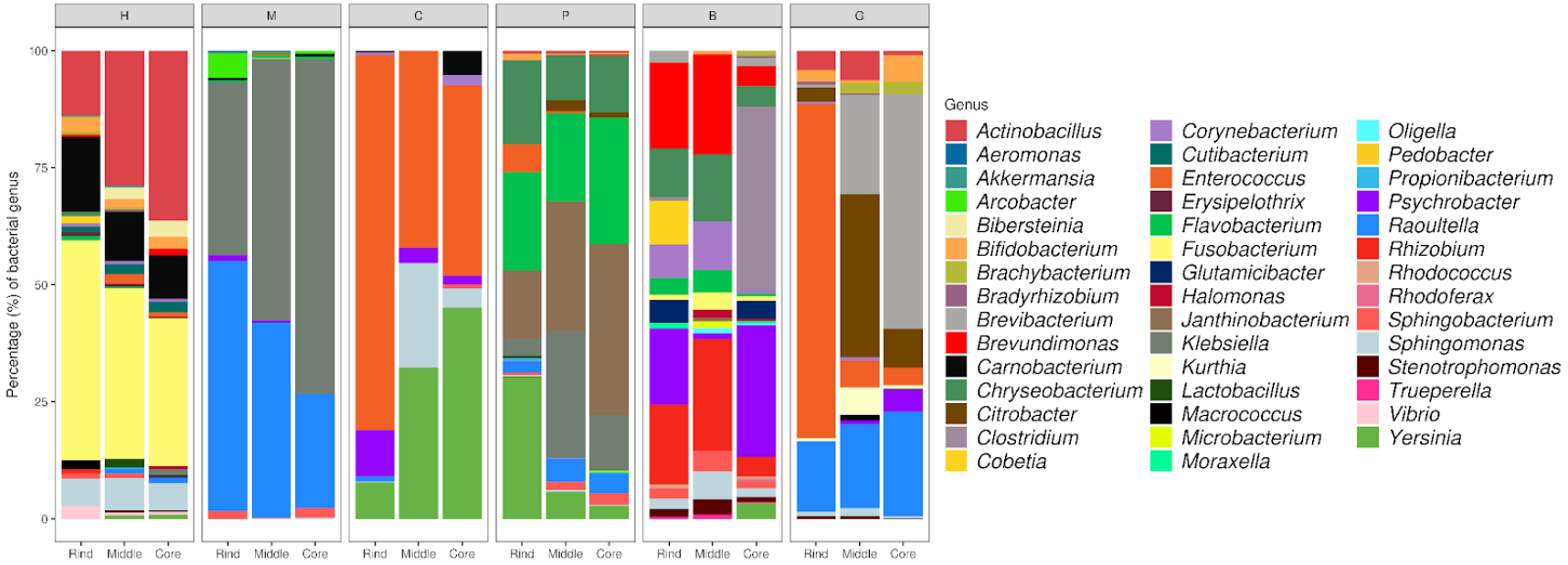


Figure 3. Relative abundances of minority bacterial genera catalogued as "Other" (relative abundances below 1% when considering the whole microbiota) identified in the rind, medial part and core samples from fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses.



Pseudoalteromonas, which is particularly abundant in the two fresh soft goat cheeses, is not only psychrophilic/psychrotolerant but also halotolerant, so we cannot rule out the possibility that microorganisms of this genus contaminated the salt used in cheese brining (Wolfe et al., 2014). *Pseudoalteromona spp.* have been implicated in the appearance of a putrid aroma in cheese due to the production of volatile sulfur compounds (Yeluri Jonnala et al., 2018), but in cheese aged and stored at low temperatures they can also participate in the synthesis of beneficial flavor compounds through lipolysis and proteolysis (Wolfe et al., 2014).

Minority bacterial genera detected in cheese samples

In the six goat cheeses, we have detected multiple low abundant (< 1%) bacterial genera. Among them, we have identified members of *Bifidobacterium* commonly used as probiotics in the cheese industry (Hayaloglu, 2022). The intake of probiotics benefits human health by enhancing the immune system and facilitating food digestion (Kechagia et al., 2013). We have also noticed the presence of bacteria from the genera *Aeromonas, Clostridium, Klebsiella, Moraxella* and *Yersinia* (**Figures 2** and **3**), which have psychrotrophic properties and might affect the quality and safety of cheese (Bintsis and Papademas, 2002; Gómez-Torres et al., 2015; Marino et al., 2017; Sameli et al., 2021). *Moraxella* is a common spoilage microorganism of fresh meat (Gennari et al., 1992), while species from the genera *Clostridium* and *Yersinia* have been involved in food poisoning (Bintsis, 2017).

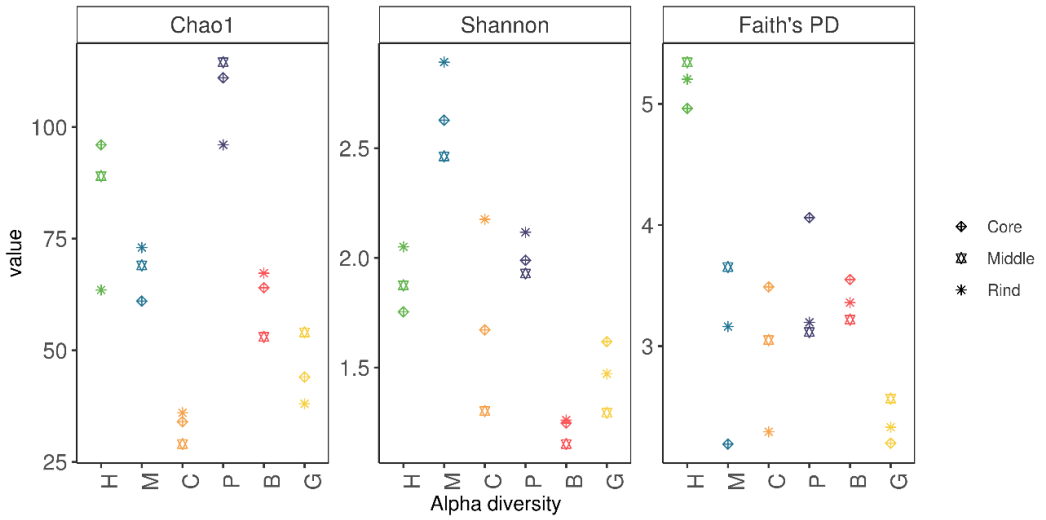
Several minority halophilic genera were present in goat cheeses at very low proportions (0.00001 – 5.1%) e.g. *Acinetobacter*, *Brachybacterium*, *Brevibacterium*, *Vibrio*, *Psychrobacter*, *Flavobacterium* and *Corynebacterium* (**Figure 2** and **3**). Although halophilic bacteria are sometimes associated with food spoilage, several of these genera are of technological interest. For instance, *Acinetobacter* microorganisms have a lipolytic effect during the maturation of Camembert cheese (Addis et al., 2001), and *Brachybacterium*, *Brevibacterium* and *Psychrobacter* produce volatile compounds correlating with a roasty to fruity aroma in cheese (Deetae et al., 2007). Moreover, *Brevibacterium* and *Psychrobacter* improve the flavour and the ripening of cheese and they are commonly found on the rind (Anast et al., 2019; Mayo et al., 2021). These bacterial genera have been also identified in cheeses from ewe and cow (Kothe et al., 2021).

Richness, diversity and clustering of six goat cheeses

The α-diversity analysis based on richness showed values of Chao1 that went from 29 to 114.5 in goat cheeses (**Figure 4** and **Supplementary Table 1**). Although the sample size per cheese in this study is limited, there is an evident trend in the richness of taxa associated with the ageing and type of cheese paste. A longitudinal analysis of the evolution of richness during Gouda cheesemaking supports this observation, as the highest richness was evidenced at the beginning of the process, before the addition of starter LAB, while richness significantly decreased during the manufacture and ageing of cheese (Choi et al., 2020a). As shown in **Figure 4**, The highest Shannon index was estimated for fresh cheese M (Shannon index =

2.46 – 2.89), while hard semi-ripened (B) and semi-hard aged (G) displayed the lowest values for this index (Shannon index of 1.15 – 1.26 and 1.29 – 1.62, respectively). Estimating diversity based on phylogenetic distances of the taxa yielded Faith PD indexes of 2.19–4.06, with the only exception of soft fresh cheese H (Faith's PD = 5.17). Although it is assumed that the core of the cheese is less diverse than the rind because of its higher acidity and anaerobiosis (Choi et al., 2020b), pairwise comparison of α-diversities between the core, middle and rind regions of cheeses sampled in our study did not yield any significant difference (Kruskal-Wallis test, *P*-value ranging from 0.26 to 1).

Figure 4. α-diversity in the rind, middle part and core microbiomes of soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses: (a) richness (Chao1), (b) evenness (Shannon) and (c) phylogenetical diversity (Faith's PD).



β-diversity indices, which reflect differences between microbial communities among samples, are depicted in **Figure 5**. The analysis based on non-phylogenetic Bray-Curtis distances (**Figure 5a**) showed that samples clustered according to the cheese type and not to the portion of the cheese sampled (i.e. core, middle part and rind). The PCoA visualisation of this metric made evident two defined clusters: one formed by soft fresh H and M samples, and the other one comprising soft semi-ripened P and semi-hard aged G cheeses, while soft-semi-ripened C and hard semi-ripened B cheeses were placed in separate locations in the plot (**Figure 5a**). The PCoA plot based on phylogenetic Unifrac distances, both weighted and unweighted, showed a consistent affinity between hard semi-ripened B and soft semi-ripened P cheeses (**Figures 5b** and **5c**). Samples from the soft semi-ripened C cheese made from raw milk did not cluster with the other cheese types in all three analyses.

In accordance with PCoA analyses, the clustering of the cheese samples in a dendrogram based on Aitchison distances and taking into account only the most abundant taxa (relative abundance > 1%) made possible to distinguish two principal groups (**Figure 6a**): one containing soft semi-ripened mouldy rind cheeses (C and P) and semi-hard aged cheese (G), and another group encompassing soft fresh (H and M) and hard semi-ripened (B) cheeses. The inclusion of the less common taxa (< 1%) in the analysis led to a pattern of clustering that was perfectly coherent with the classification of the cheese varieties portrayed in **Table 1** (**Figure 6b**). One of the clusters comprised soft fresh cheeses (H and M), while the other one

encompassed semi-ripened and ripened cheeses (B, G, C and P). The two soft semi-ripened cheeses (C and P) grouped in a secondary node.

Figure 5. Principal coordinate analysis (PCoA) based on the β-diversity of core, middle and rind microbiomes of soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses. β-diversity was estimated on the basis of (a) non-phylogenetic (Bray-Curtis), (b) phylogenetic weighted by abundance (Weighted Unifrac) and (c) phylogenetic based on presence-absence of taxa (Unweigthed Unifrac) dissimilarity indices.

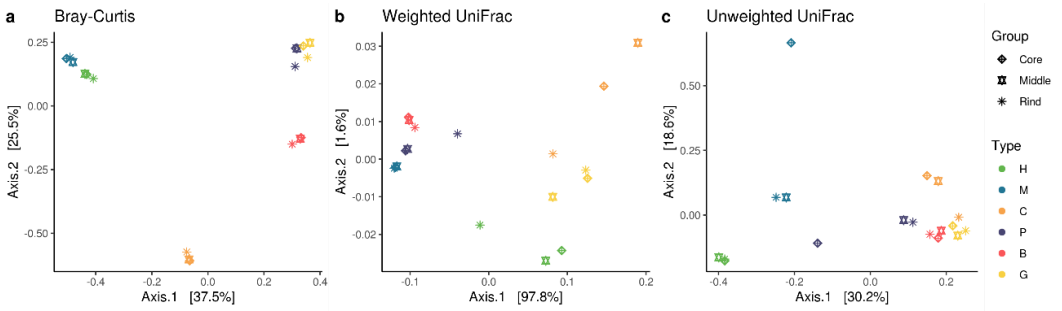
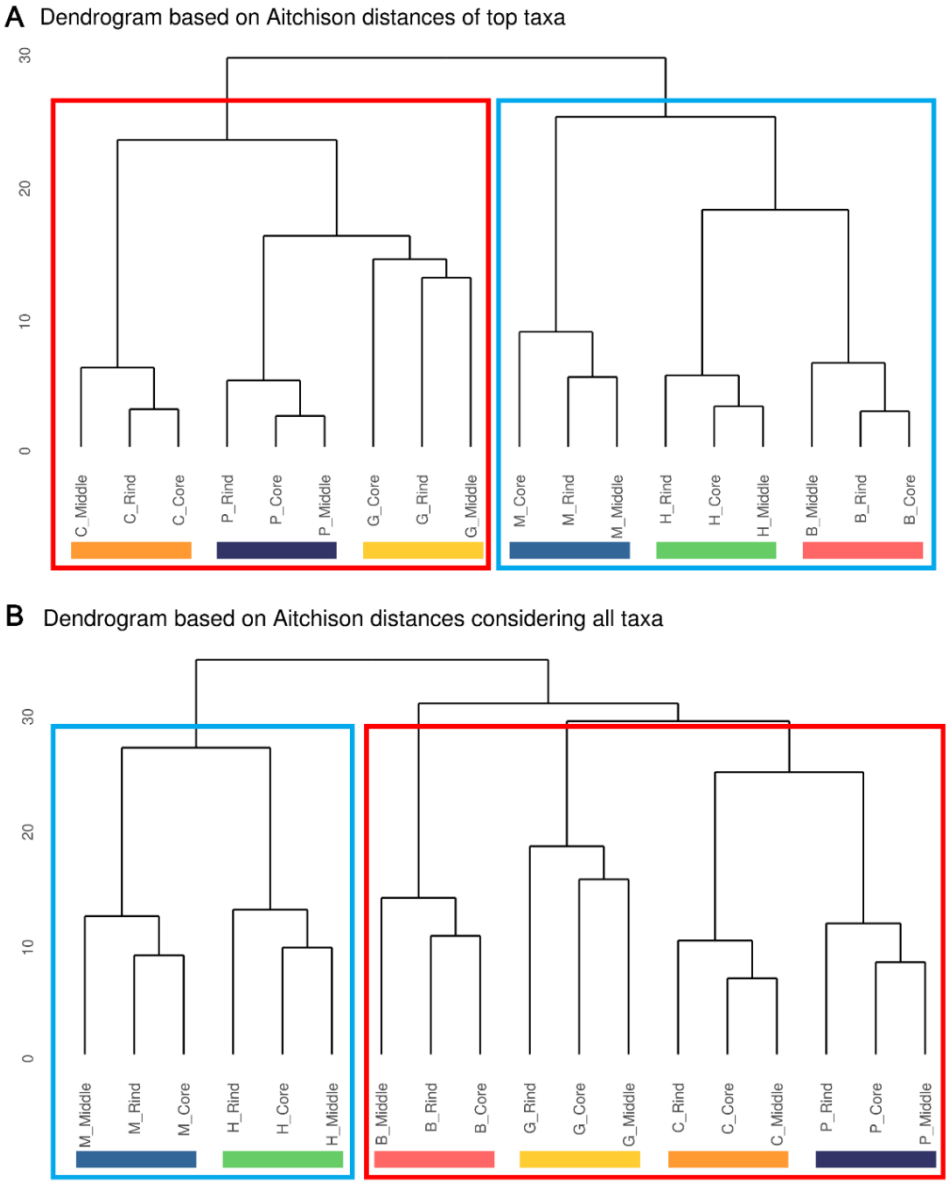


Figure 6. Dendrogram based on Aitchison distances between samples from soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses. Aitchison distances were estimated from the relative abundances of (a) Only taxa with abundances over 1% and (b) All major and minority (< 1%) taxa.



Conclusion

Compared to previous studies focused on the microbiome of single goat cheeses (Gouda and Pélardon), here we provide a more comprehensive view about the composition, abundance and diversity of microbial communities in six caprine cheeses differing in terms of texture, ageing and other technological and organoleptic features. As a whole, we have detected 15 majority and 44 minority (< 1%) bacterial genera in the six goat cheeses. Lactic acid bacteria from the *Lactibacillaceae* (*Frutilactobacillus, Lactococcus* and *Lacticaseibacillus* genera) and *Streptococcaceae* (*Streptococcus* genus) families are clearly predominant in all six goat cheeses. We have also found psychrophilic bacteria from the genera *Pseudomonas* and *Shewanella,* the presence of which is often explained by post-pasteurisation contamination (sometimes associated with food spoilage), as well as of halophilic bacteria such as *Pseudoalteromona,* which can thrive in environments with high salinity. We did not find evidence of an increased microbial α-diversity in the rind, contrary to other reports indicating that the low pH and anaerobiosis of the inner parts of cheese result in a reduction of the diversity of microbial communities. Moreover, in a dendrogram cheese microbiomes clustered according to cheese type (fresh vs ripened) rather than to cheese portion (rind, middle and core), illustrating the strong impact of the ripening process on the composition of the goat cheese microbiome.

Declaration of competing interest

None.

Acknowledgments

This research was funded by grant: PID2019-105805RB-I00 funded by MCIN/ AEI/10.13039/501100011033. We also acknowledge the support of the CERCA programme of the Generalitat de Catalunya and the Center of Excellence Severo Ochoa 2020–2023 (CEX2019–000902-S) grant funded by MCIN/AEI /10.13039/501100011033 and awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). Dailu Guan was funded by a PhD fellowship from the China Scholarship Council (CSC). Maria Luigi-Sierra was funded with a PhD fellowship Formación de Personal Investigador BES-C-2017-079709 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future" Yuliaxis Ramayo is recipient of a Ramon y Cajal post-doctoral fellowship (RYC2019-027244-I) funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future".

References

Addis, E., Fleet, G.H., Cox, J.M., Kolak, D., Leung, T., 2001. The growth, properties and interactions of yeasts and bacteria associated with the maturation of Camembert and blue-veined cheeses. Int. J. Food Microbiol. 69, 25–36. https://doi.org/10.1016/s0168-1605(01)00569-4

Afshari, R., Pillidge, C.J., Dias, D.A., Osborn, A.M., Gill, H., 2020. Microbiota and metabolite profiling combined with integrative analysis for differentiating cheeses of varying ripening ages. Front. Microbiol. 11, 2991. https://doi.org/10.3389/fmicb.2020.592060/bibtex

Anast, J.M., Dzieciol, M., Schultz, D.L., Wagner, M., Mann, E., Schmitz-Esser, S., 2019. Brevibacterium from Austrian hard cheese harbor a putative histamine

catabolism pathway and a plasmid for adaptation to the cheese environment. Sci. Reports. 9, 6164. https://doi.org/10.1038/s41598-019-42525-y

Bassi, D., Gazzola, S., Sattin, E., Dal Bello, F., Simionati, B., Cocconcelli, P.S., 2020. Lactic acid bacteria adjunct cultures exert a mitigation effect against spoilage microbiota in fresh cheese. Microorganisms. 8, 1199. https://doi.org/10.3390/microorganisms8081199

Beresford, T., Williams, A., 2004. The microbiology of cheese ripening, in: Fox, F.P., McSweeney, P.L.H., Cogan, T.M., Guinee, T.P. (Eds.), Cheese: Chemistry, physics and microbiolology (Third edition). Academic Press, pp. 287–317. https://doi.org/10.1016/s1874-558x(04)80071-x

Beresford, T.P., Fitzsimons, N.A., Brennan, N.L., Cogan, T.M., 2001. Recent advances in cheese microbiology. Int. Dairy J. 11, 259–274. https://doi.org/10.1016/S0958-6946(01)00056-5

Bintsis, T., Papademas, P., 2002. Microbiological quality of white-brined cheeses: a review. Int. J. Dairy Technol. 55, 113–120. https://doi.org/10.1046/J.1471-0307.2002.00054.X

Bintsis, T., 2017. Foodborne pathogens. AIMS Microbiol. 3, 529–563. https://doi.org/10.3934/microbiol.2017.3.529

Blaya, J., Barzideh, Z., LaPointe, G., 2018. Symposium review: Interaction of starter cultures and nonstarter lactic acid bacteria in the cheese environment1. J. Dairy Sci. 101, 3611–3629. https://doi.org/10.3168/jds.2017-13345

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K. Bin, Keefe, C.R.,

Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A. V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857. https://doi.org/10.1038/s41587-019-0209-9

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

Camargo, A.C., Costa, E.A., Fusieger, A., Freitas, R. de, Nero, L.A., Carvalho, A.F. de, 2021. Microbial shifts through the ripening of the “Entre Serras” Minas artisanal cheese monitored by high-throughput sequencing. Food Res. Int. 139, 109803. https://doi.org/10.1016/j.foodres.2020.109803

Choi, J., In Lee, S., Rackerby, B., Frojen, R., Goddik, L., Ha, S. Do, Park, S.H., 2020a. Assessment of overall microbial community shift during Cheddar cheese production from raw milk to aging. Appl. Microbiol. Biotechnol. 104, 6249–6260. https://doi.org/10.1007/s00253-020-10651-7

Choi, J., Lee, S.I., Rackerby, B., Goddik, L., Frojen, R., Ha, S. Do, Kim, J.H., Park, S.H., 2020b. Microbial communities of a variety of cheeses and comparison between core and rind region of cheeses. J. Dairy Sci. 103, 4026–4042. https://doi.org/10.3168/jds.2019-17455

Coelho, M.C., Silva, C.C.G., Ribeiro, S.C., Dapkevicius, M.L.N.E., Rosa, H.J.D., 2014. Control of *Listeria monocytogenes* in fresh cheese using protective lactic acid bacteria. Int. J. Food Microbiol. 191, 53–59. https://doi.org/10.1016/j.ijfoodmicro.2014.08.029

De Jonghe, V., Coorevits, A., Van Hoorde, K., Messens, W., Van Landschoot, A., De Vos, P., Heyndrickx, M., 2011. Influence of storage conditions on the growth of *Pseudomonas* species in refrigerated raw milk. Appl. Environ. Microbiol. 77, 460–470. https://doi.org/10.1128/aem.00521-10

Deetae, P., Bonnarme, P., Spinnler, H.E., Helinck, S., 2007. Production of volatile aroma compounds by bacterial strains isolated from different surface-ripened French cheeses. Appl. Microbiol. Biotechnol. 76, 1161–1171. https://doi.org/10.1007/s00253-007-1095-5

Dolci, P., De Filippis, F., La Storia, A., Ercolini, D., Cocolin, L., 2014. rRNA-based monitoring of the microbiota involved in Fontina PDO cheese production in relation to different stages of cow lactation. Int. J. Food Microbiol. 185, 127–135. https://doi.org/10.1016/J.IJFOODMICRO.2014.05.021

Fox, P. F., Singh, T. K., & McSweeney, P. L. H., 1995. Biogenesis of flavour compounds in cheese, in: Malin, E.L., Tunick, M.H. (Eds.), Chemistry of structure-function relationships in cheese. Advances in Experimental Medicine and Biology. Springer, Boston, MA, pp. 59–98. https://doi.org/10.1007/978-1-4615-1913-3_6

Fox, Patrick F., Guinee, T. P., Cogan, T. M., & McSweeney, P. L. H., 2017. Microbiology of cheese ripening, in: Fundamentals of cheese science. Springer, Boston, MA, pp. 333–390. https://doi.org/10.1007/978-1-4899-7681-9_11

Gennari, M., Parini, M., Volpon, D., Serio, M., 1992. Isolation and characterisation by conventional methods and genetic transformation of *Psychrobacter* and *Acinetobacter* from fresh and spoiled meat, milk and cheese. Int. J. Food Microbiol. 15, 61–75. https://doi.org/10.1016/0168-1605(92)90136-q

Gomez, J. M., Gaya, P., Nunez, M., Médina, M., Jesus Gomez, M., Nufiez, M., Medina, M., 1998. *Streptococcus thermophilus* as adjunct culture for a semi-hard cows' milk cheese. Lait. 78, 501–511.

Gómez-Torres, N., Garde, S., Peirotén, Á., Ávila, M., 2015. Impact of *Clostridium spp*. on cheese characteristics: Microbiology, color, formation of volatile compounds and off-flavors. Food Control. 56, 186–194. https://doi.org/10.1016/j.foodcont.2015.03.025

Guidone, A., Zotta, T., Matera, A., Ricciardi, A., De Filippis, F., Ercolini, D., Parente, E., 2016. The microbiota of high-moisture mozzarella cheese produced with different acidification methods. Int. J. Food Microbiol. 216, 9–17. https://doi.org/10.1016/j.ijfoodmicro.2015.09.002

Hayaloglu, A.A., 2022. Microbiology of Cheese, in: McSweeney, P.L.H., McNamara, J.P. (Eds.), Encyclopedia of Dairy Sciences (Third edition). Academic Press, pp. 225–237. https://doi.org/10.1016/B978-0-12-818766-1.00272-5

Irlinger, F., In Yung, S.A.Y., Sarthou, A.S., Delbès-Paus, C., Montel, M.C., Coton, E., Coton, M., Helinck, S., 2012. Ecological and aromatic impact of two Gram-negative bacteria (*Psychrobacter* celer and *Hafnia alvei*) inoculated as part of the whole microbial community of an experimental smear soft cheese. Int. J. Food Microbiol. 153, 332–338. https://doi.org/10.1016/j.ijfoodmicro.2011.11.022

Kamilari, E., Tomazou, M., Antoniades, A., Tsaltas, D., 2019. High throughput sequencing technologies as a new toolbox for deep analysis, characterisation and potentially authentication of protection designation of origin cheeses? Int. J. Food Sci. 5837301. https://doi.org/10.1155/2019/5837301

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kechagia, M., Basoulis, D., Konstantopoulou, S., Dimitriadi, D., Gyftopoulou, K., Skarmoutsou, N., Fakiri, E.M., 2013. Health benefits of probiotics: A review. ISRN Nutr. 2013, 481651. https://doi.org/10.5402/2013/481651

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 26, 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.I., McDonald, D., Melnik, A. V., Morton, J.T., Navas, J., Quinn, R.A., Sanders, J.G., Swafford, A.D., Thompson, L.R., Tripathi, A., Xu, Z.Z., Zaneveld, J.R., Zhu, Q., Caporaso, J.G., Dorrestein, P.C., 2018. Best practices for analysing microbiomes. Nat. Rev. Microbiol. 16, 410–422. https://doi.org/10.1038/s41579-018-0029-9

Kothe, C.I., Bolotin, A., Kraïem, B.F., Dridi, B., Renault, P., 2021. Unraveling the world of halophilic and halotolerant bacteria in cheese by combining cultural, genomic and metagenomic approaches. Int. J. Food Microbiol. 358, 109312. https://doi.org/10.1016/j.ijfoodmicro.2021.109312

Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47, 583–621. https://doi.org/10.1080/01621459.1952.10483441

Lahti, L., Shetty, S., 2019. Microbiome R package. http://microbiome.github.io

Marino, M., Innocente, N., Maifreni, M., Mounier, J., Cobo-Díaz, J.F., Coton, E., Carraro, L., Cardazzo, B., 2017. Diversity within Italian cheesemaking brine-associated bacterial communities evidenced by massive parallel 16S rRNA gene tag sequencing. Front. Microbiol. 8, 2119. https://doi.org/10.3389/fmicb.2017.02119

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

Martin, N.H., Boor, K.J., Wiedmann, M., 2018. Symposium review: Effect of post-pasteurisation contamination on fluid milk quality. J. Dairy Sci. 101, 861–870. https://doi.org/10.3168/jds.2017-13339

Mayo, B., Rodríguez, J., Vázquez, L., Flórez, A.B., 2021. Microbial interactions within the cheese ecosystem and their application to improve quality and safety. Foods. 10, 602. https://doi.org/10.3390/foods10030602

Meola, M., Rifa, E., Shani, N., Delbès, C., Berthoud, H., Chassard, C., 2019. DAIRYdb: A manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products. BMC Genomics. 20, 560. https://doi.org/10.1186/s12864-019-5914-8

Montel, M.C., Buchin, S., Mallet, A., Delbes-Paus, C., Vuitton, D.A., Desmasures, N., Berthier, F., 2014. Traditional cheeses: Rich and diverse microbiota with associated benefits. Int. J. Food Microbiol. 177, 136–154. https://doi.org/10.1016/j.ijfoodmicro.2014.02.019

Morales, P., Fernández-García, E., Nuñez, M., 2003. Caseinolysis in cheese by *Enterobacteriaceae* strains of dairy origin. Lett. Appl. Microbiol. 37, 410–414. https://doi.org/10.1046/j.1472-765x.2003.01422.x

Oksanen, J., Simpson, G.L., Blanchet, F.G, Friendly, M., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbout, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Antoniazi, H.B., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M-H., Ribeiro, E., Smith, T., Stier, A., Cajo, F., Braak, T., Weedon, J., 2022. Vegan: Community ecology package. R package version 2.6-2. https://CRAN.R-project.org/package=vegan

Parente, E., Cogan, T. M., & Powell, I. B., 2017. Starter cultures: General aspects, in: Fox, F.P., McSweeney, P.L.H., Cogan, T.M., Guinee, T.P. (Eds.), Cheese: Chemistry, Physics and Microbiology (Fourth edition). Academic Press, pp. 201–226. https://doi.org/10.1016/b978-0-12-417012-4.00008-9

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A.,

Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2012. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Penland, M., Falentin, H., Parayre, S., Pawtowski, A., Maillard, M.B., Thierry, A., Mounier, J., Coton, M., Deutsch, S.M., 2021. Linking Pélardon artisanal goat cheese microbial communities to aroma compounds during cheese-making and ripening. Int. J. Food Microbiol. 345, 109130. https://doi.org/10.1016/j.ijfoodmicro.2021.109130

Place, R.B., Hiestand, D., Gallmann, H.R., Teuber, M., 2003. *Staphylococcus equorum subsp. linens*, *subsp. nov.*, a starter culture component for surface ripened semi-hard cheeses. Syst. Appl. Microbiol. 26, 30–37. https://doi.org/10.1078/072320203322337281

Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T.P., Ross, R.P., Fitzgerald, G.F., Cotter, P.D., 2013. The complex microbiota of raw milk. FEMS Microbiol. Rev. 37, 664–698. https://doi.org/10.1111/1574-6976.12030

Salazar, J.K., Carstens, C.K., Ramachandran, P., Shazer, A.G., Narula, S.S., Reed, E., Ottesen, A., Schill, K.M., 2018. Metagenomics of pasteurised and unpasteurised gouda cheese using targeted 16S rDNA sequencing. BMC Microbiol. 18, 189. https://doi.org/10.1186/S12866-018-1323-4

Sameli, N., Sioziou, E., Bosnea, L., Kakouri, A., Samelis, J., 2021. Assessment of the Spoilage Microbiota during Refrigerated (4 °C) Vacuum-Packed Storage of Fresh Greek Anthotyros Whey Cheese without or with a Crude Enterocin A-B-P-Containing Extract. Foods. 10, 2946. https://doi.org/10.3390/foods10122946

Smit, G., Smit, B.A., Engels, W.J.M., 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. FEMS Microbiol. Rev. 29, 591–610. https://doi.org/10.1016/j.fmrre.2005.04.002

Stepaniak, L., 2022. Psychrotrophic bacteria: Other psychrotrophs. in: McSweeney, P.L.H., McNamara, J.P. (Eds.), Encyclopedia of Dairy Sciences (Third edition). Academic Press, pp. 369–374. https://doi.org/10.1016/B978-0-12-818766-1.00384-6

Walsh, A.M., Macori, G., Kilcawley, K.N., Cotter, P.D., 2020. Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality. Nat. Food. 1, 500–510. https://doi.org/10.1038/s43016-020-0129-3

Werner, J.J., Koren, O., Hugenholtz, P., Desantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R., Ley, R.E., 2012. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. ISME J. 6, 94–103. https://doi.org/10.1038/ismej.2011.82

Wolfe, B.E., Button, J.E., Santarelli, M., Dutton, R.J., 2014. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. Cell. 158, 422–433. https://doi.org/10.1016/j.cell.2014.05.041

Yeluri Jonnala, B.R., McSweeney, P.L.H., Sheehan, J.J., Cotter, P.D., 2018. Sequencing of the cheese microbiome and its relevance to industry. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01020

Zhang, Y., Kastman, E.K., Guasto, J.S., Wolfe, B.E., 2018. Fungal networks shape dynamics of bacterial dispersal and community assembly in cheese rind microbiomes. Nat. Commun. 9, 336. https://doi.org/10.1038/s41467-017-02522-z

Supplementary files

Supplementary Table 1. α-diversity indices estimated from the rind, middle part and core samples from soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses.

Supplementary Figure 1. Sequencing depth of samples from soft fresh (H and M), soft semi-ripened (C and P), hard semi-ripened (B) and semi-hard aged (G) goat cheeses. The number of amplicon sequence variants (ASVs) is displayed on the *y*-axis, while the *x*-axis shows the sequencing depth expressed as the number of reads.

Journal of Animal Science and Biotechnology BMC

RUNNING HEAD: Inbreeding depression in goats

Genomic patterns of homozygosity and inbreeding depression in Murciano-Granadina goats

María Gracia Luigi-Sierra[1], Almudena Fernández[2], Amparo Martínez[3], Dailu Guan[1], Juan Vicente Delgado[3], Javier Fernández Álvarez[3], Vincenzo Landi[4], Francesc Xavier Such[5], Jordi Jordana[6], María Saura[2]*, Marcel Amills[1, 6]*

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain. [2]Departamento de Mejora Genética Animal, INIA, Carretera de la Coruña km 7,5, Madrid, 28040, Spain. [3]Departamento de Genética, Universidad de Córdoba, Córdoba 14071, Spain. [4]Department of Veterinary Medicine, University of Bari "Aldo Moro", SP. 62 per Casamassima km. 3, 70010 Valenzano (BA), Italy. [5]Group of Research in Ruminants (G2R), Department of Animal and Food Science, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain. [6]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.

Corresponding authors: *

J Anim Sci Biotechnol. 2022 Mar 10;13(1):35. doi: 10.1186/s40104-022-00684-5.

Abstract

Background:

Inbreeding depression can adversely affect traits related to fitness, reproduction and productive performance. Although current research suggests that inbreeding levels are generally low in most goat breeds, the impact of inbreeding depression on phenotypes of economic interest has only been investigated in a few studies based on genealogical data.

Results

We genotyped 1,040 goats with the Goat SNP50 BeadChip. This information was used to estimate different molecular inbreeding coefficients and characterise runs of homozygosity and homozygosity patterns. We detected 38 genomic regions with increased homozygosity as well as 8 ROH hotspots mapping to chromosomes 1, 2, 4, 6, 14, 16 and 17. Eight hundred seventeen goats with available records for dairy traits were analysed to evaluate the potential consequences of inbreeding depression on milk phenotypes. Four regions on chromosomes 8 and 25 were significantly associated with inbreeding depression for the natural logarithm of the somatic cell count. Notably, these regions contain several genes related with immunity, such as *SYK*, *IL27*, *CCL19* and *CCL21*. Moreover, one region on chromosome 2 was significantly associated with inbreeding depression for milk yield.

Conclusions

Although genomic inbreeding levels are low in Murciano-Granadina goats, significant evidence of inbreeding depression for the logarithm of the somatic cell count, a phenotype closely associated with udder health and milk yield, have been

detected in this population. Minimising inbreeding would be expected to augment economic gain by increasing milk yield and reducing the incidence of mastitis, which is one of the main causes of dairy goat culling.

Keywords: Goat, Inbreeding, Milk yield, Murciano-Granadina, Somatic cell score.

Background

Inbreeding is defined as the mating of individuals that are related to each other more closely than the average relationship within the concerned population [1]. In livestock species, the magnitude of inbreeding has been traditionally measured through genealogical information [2]. However, pedigree-based estimates are affected by the depth of the pedigree [2] because founders are assumed to be unrelated and non-inbred [3]. Consequently, inbreeding produced by distant ancestors not included in the pedigree is systematically ignored [4]. Another disadvantage of quantifying inbreeding from pedigree data is that it provides bare expectations about the fraction of the genome which is identical-by-descent (IBD) [3]. With the advent of high-density arrays of single nucleotide polymorphisms (SNPs), it has become possible to estimate genomic inbreeding coefficients which circumvent these limitations [5]. Indeed, important advantages of genomic inbreeding coefficients over their genealogical counterparts are: (i) higher accuracy to differentiate among individuals within the same pedigree, since variation due to Mendelian sampling is captured [4], (ii) higher accuracy to quantify shared ancestry of genetic haplotypes [4], and (iii) the ability to map inbreeding to specific genomic regions [6]. Different types of genomic inbreeding coefficients have been implemented. While inbreeding coefficients based on the proportion of homozygous SNPs (F_{HOM}) just reflect identity-by-state (IBS) allele-sharing proportions [6], coefficients (F_{ROH}) based on measuring the fraction of the genome covered by runs of homozygosity (ROH) estimate IBD allele sharing [4, 7, 8], making possible to disentangle recent from ancient inbreeding [3, 5, 9].

The increase of inbreeding might have adverse consequences on the fitness of livestock populations due to the loss of genetic variability, which can entail a long-term reduction of genetic variance (due to the fixation of alleles) and, consequently, a slowing down of the rate of response to selection in breeding schemes [10, 11, 12]. Moreover, incremented levels of inbreeding might reduce the mean phenotypic performance of livestock populations, a phenomenon known as inbreeding depression (reviewed by Leroy [2]). Although inbreeding depression is particularly intense for fitness and reproduction traits [11], there is broad evidence that it also decreases dairy and growth performances [2, 13, 14 15]. Besides, susceptibility to certain diseases, such as mastitis, is increased in inbred animals [16, 17]. In Holstein cattle, a 1% increase of inbreeding is expected to cause a reduction of $22–24 of lifetime net income per individual [18], while in sheep the average economic loss per ewe amounts to $17 for moderate inbreeding and $36 when inbreeding is close to 50% [18].

Several studies have used genomic methods to determine the levels of inbreeding in goat populations with a broad geographic distribution [19, 20, 21]. A recent investigation carried out by Bertolini et al. [19] revealed that short ROH (< 3 Mb) are particularly abundant in worldwide goat populations. Moreover, five regions on caprine chromosomes (CHI) 11, 12, and 18 contained ROH hotspots that overlapped with signatures of selection [19]. The majority of goat breeds analysed by Bertolini et al. [19] displayed low levels of inbreeding (F_{ROH} < 0.10), with the only exception of certain local breeds with small population sizes (e.g. Dutch Landrace goats) as well as of breeds with insular origins (e.g. Icelandic and

Malagasy goats) which happened to be highly inbred [19, 20]. Despite the fact that inbreeding depression can have adverse effects on the profitability of farmers and animal breeders, very few studies have investigated its consequences on goat production [22, 23]. In this regard, Marete et al. [22] and Deroide et al. [23] estimated, with genealogical methods, the effect of inbreeding depression on the production of Kenya Alpine and Murciano-Granadina goats, respectively, and they found that in both populations such effect was negligible.

The goals of the current work were: (i) to measure the levels of inbreeding in a Murciano-Granadina resource population by using different genomic coefficients, and (ii) to use this information to infer the impact of inbreeding depression on dairy phenotypes recorded in this population.

Methods

Animal material and phenotyping

The animal material comprised 1,040 Murciano-Granadina female goats from 15 farms located in the autonomous region of Andalusia (Spain). Murciano-Ganadina is a local Spanish breed officially created in 1975 by the crossbreeding of Murciano and Granadina goats [24]. Currently, it has a census of 115,105 heads (2020), and its remarkable adaptability to harsh environments as well as its good milking performance (mean of 586 kg/lactation; 5.1% of fat and 3.6% of protein in milk) have made it a very popular breed in Spain and other countries (https://www.mapa.gob.es/es/ganaderia/ temas/zootecnia/razas-ganaderas/razas/catalogo-razas/caprino/murciano-granadina/).

Blood samples were extracted from goats using vacuum tubes coated with EDTA K_3 anticoagulant and stored at – 20 ºC until processing. Phenotypic records for milk yield and composition traits were recorded in the framework of the selection program of the Murciano-Granadina goat breed [24]. Only phenotypes corresponding to the first parity (recorded between the years 2009 and 2017) were taken into consideration. The following phenotypes were recorded in 817 goats: milk yield measured in kilograms at 210 days (MY210), 240 days (MY240) and 305 days (MY305), the natural logarithm of the somatic cell count divided by 1,000 (lnSCC, to convert this value into a somatic cell count please use the formula: $e^{lnSCC} \times 10^3$ cells/mL), fat percentage (FP), protein percentage (PP) and lactose percentage (LP). Milk composition traits were standardised to a lactation of 210 days. Summary statistics of phenotypic records are displayed in **Table 1**.

Table 1. Summary statistics of seven milk production and composition traits recorded in 817 Murciano-Granadina goats.

Traits[a]	Mean	SD
MY210, kg	395.647	131.787
MY240, kg	450.493	142.707
MY305, kg	547.418	179.840
lnSCC	6.278	0.937
FP, %	5.190	0.766
PP, %	3.563	0.351
LP, %	4.865	0.228

[a] **MY210,** milk yield at 210 days of lactation (kg); **MY240,** milk yield at 240 days of lactation (kg); **MY305,** milk yield at 305 days of lactation (kg); **lnSCC,** natural logarithm of the somatic cell count divided by 1,000 (to convert this value into a somatic cell count use the formula: $e^{lnSCC} \times 10^3$ cells/mL); **FP,** fat percentage; **PP,** protein percentage; **LP,** lactose percentage.

Generation of high throughput genotypic data

The isolation of genomic DNA was carried out following a salting-out protocol [25]. Three mL of whole blood were mixed with 4 volumes of Red Cell Lysis Solution (Tris-HCl 10 mM, pH = 6.5; EDTA 2 mM; Tween 20 1%), and this mixture was centrifuged at 2000 g. The supernatant was discarded and the pellet was resuspended in 3 mL of lysis buffer (Tris-HCl 200 mM, pH = 8, EDTA 30 mM, SDS 1%; NaCl 250 mM) plus 100 µl proteinase K (20 mg/mL) and incubated for 3 hours at 55° C. The lysate was chilled, and 1 mL of ammonium acetate 10 M was added to it. After centrifugation at 2000 g for 10 minutes, the supernatant (~4 mL) was transferred to a new tube with 3 mL of isopropanol 96%, and this mixture was centrifuged at 2000 g for 3 minutes. The resulting DNA pellet was washed with 3 mL of ethanol 70% and an additional centrifugation step at 2000 g for 1 minute was performed. The DNA pellet was dried at room temperature, and it was subsequently resuspended in 1 mL of TE buffer (Tris-HCl 10 mM, EDTA 1 mM, pH = 8). Murciano-Granadina goats were genotyped with the Goat SNP50 BeadChip (Illumina Inc., San Diego, CA) by following the instructions of the manufacturer. The goat ARS1 genome [26] was used as reference for inferring the genomic location of the SNPs, and the position and the name of each SNP were updated using the PLINK 1.9 software [27]. Only individuals with at least 95% of SNPs with genotype calls were taken into consideration. Moreover, only SNPs meeting the following requirements were used in the downstream analyses: (i) mapping to autosomes, (ii) displaying a minor allele frequency of 0.05 or higher, (iii) not deviating significantly ($P < 0.00001$) from the Hardy-Weinberg equilibrium, and

(iv) with a genotype call rate over 98%. Data were filtered using PLINK 1.9 [27]. In addition, a principal component analysis (PCA) was carried out with PLINK 1.9 [27] in order to assess population structure. The visualisation of such results was based on the first two components of the PCA.

Detection of runs of homozygosity

The definition of ROH followed six criteria: (i) the minimum length of ROH is 1 Mb, (ii) a ROH must contain at least 15 SNPs, (iii) the density of SNPs per ROH was set to at least 1 SNP every 100 kb, (iv) the maximum distance between consecutive SNPs is 250 Kb, (v) one heterozygous position per ROH is allowed, and (vi) one missing position per ROH is allowed. Said criteria were established based on the density of the genotyping panel, with a mean distance between consecutive SNPs of 51.73 Kb and a mean number of 19.35 SNPs/Mb (**Additional file 1: Table S1**), and several of them are based on the ROH definition established by the AdaptMap Consortium [19, 20]. Runs of homozygosity were identified with the PLINK 1.9 software [27] using a sliding window of 50 SNPs.

Analysis of the genomic patterns of homozygosity

The proportion of homozygosity per site was estimated as the ratio between the number of animals with homozygous genotypes for a particular SNP divided by the number of animals genotyped for that SNP. A sliding window encompassing 30 SNPs was designed to estimate the average of this ratio, and chromosomal patterns of homozygosity were visualised as Manhattan plots using R [28].

Genomic coverage and distribution of ROH were also investigated. The patterns of ROH size and distribution along the genome were analysed and plotted using R [28]. Genomic regions in which ROH are prevalent, the so-called ROH hotspots, were identified by measuring the proportion of animals that harbour a particular SNP occurring within a ROH with regard to the total number of animals genotyped for that SNP. Genomic regions containing the top 1% SNPs most commonly associated with ROH were classified as ROH hotspots [29, 30]. Both highly homozygous regions and ROH hotspots were compared, and overlapping segments were identified. Taking as reference the ARS1 goat genome [26], genes mapping to these overlapping segments were extracted with the Biomart tool of Ensembl [31]. Database for Annotation, Visualization and Integration Discovery (DAVID) Bioinformatics Resources [32] and UniProt [33] were used in Gene ontology analysis to identify over-represented (enriched) gene ontology (GO) terms and KEGG pathways. Amongst other things, DAVID provides biological context for long lists of genes by assigning them to functionally related groups through the use of fuzzy clustering techniques [32]. To assess the significance of gene-enrichment in annotation terms, a Fisher Exact is employed. Only terms with Max.Prob ≤ 0.1, Min.Count ≥ 2 and *P*-value < 0.05 were considered as significantly enriched [32]

Estimation of inbreeding coefficients

Six inbreeding coefficients were estimated at the whole genome level based on genotypic data:

F_{HOM} is defined as the proportion of SNPs with homozygous genotypes [6] and was estimated as:

$$F_{HOM_i} = \frac{H_{o_i}}{S}$$

where H_{Oi} corresponds to the observed number of homozygous genotypes for all the SNPs for each individual *i* and *S* is the number of SNPs for which individual *i* has genotype data. It was computed from the output of the --het command of PLINK 1.9 software [27].

F_{ROH} was estimated as the proportion of the genome covered by ROH by using the following formula:

$$F_{ROH_i} = \frac{L_{ROH_i}}{L_{auto}}$$

where L_{ROHi} corresponds to the sum of the lengths of all ROH present in each individual *i*, and L_{auto} is the total length of the autosomal goat genome covered by SNPs [5]. The same mathematical expression was used to calculate the genomic coverage of ROH with sizes smaller ($F_{ROHShort}$) or larger ($F_{ROHLong}$) than 5 Mb. Such calculations were made to assess the relative importance of distant ($F_{ROHShort}$) versus recent ($F_{ROHLong}$) inbreeding [6], according to the size of the genome and SNP density (**Table 1**). Assuming that the length of the ROH segments follows an exponential distribution with a mean equal to 12g, where *g* corresponds to the number of generations to the closest common ancestor [34], and also assuming that goats have a recombination rate of approximately 1 cM/Mb [35], $F_{ROHShort}$ indicates the inbreeding of an individual from 10 to 50 generations in the past, while $F_{ROHLong}$ estimates the inbreeding from 1 to 9 generations in the past.

F_{IS} coefficient of Wright [36] was calculated with the formula:

$$F_{IS} = \frac{H_{E_i} - H_{O_i}}{H_{E_i}}$$

Where HEi and HOiare the expected and observed heterozygosities of the individual *i*. This coefficient was estimated from the output of the --hardy command of PLINK 1.9 [27].

F_{YANG} coefficient was estimated from the diagonal of the matrix of genomic relationships of Yang based on the correlation between uniting gametes [37, 38]. It was calculated with the following formula:

$$F_{YANG_i} = \frac{1}{S}\sum_{k=1}^{s} \frac{x_k^2 - (1 + 2p_k)x_{k_i} + 2p_k^2}{2p_k(1 - p_k)}$$

where, x_k is the genotype of the individual *i* for the SNP k, and p_k is the frequency of the reference allele in the studied population. The command --ibc of PLINK 1.9 [27] was used to estimate it.

The coefficients F_{HOM} and F_{ROH} were also estimated at the chromosomal level.

Inbreeding depression analyses

The effects of inbreeding depression on dairy traits were investigated using data from 817 goats with available phenotypes (MY210, MY240, MY305, lnSCC, FP, PP and LP). Analyses were performed with the REMLF90 software [39] to implement a restricted maximum likelihood (REML) analysis approach, in which the phenotypic values of each trait in each individual are regressed onto its inbreeding coefficient using a linear mixed model. These analyses were performed

to quantify inbreeding depression at the whole genome scale as well as on a chromosome and regional basis. The model was fitted as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e},$$

where **y** is the vector of observations for each phenotype, β is a vector of fixed effects, including farm (15 levels), year of birth (10 levels), and the linear regression on *F* as a covariate; **a** is the vector of additive genetic effects, **e** is the vector of random residual effects, and **X** and **Z** are incidence matrices relating fixed and random effects to observations.

The significance of the inbreeding effect on the analysed traits was determined by applying a two-tailed hypothesis test. A Z-statistic was estimated with the following general formula:

$$Z = \frac{\bar{x} - \mu_0}{s.e.}$$

Here, $\bar{x}$ corresponds to the regression coefficient representing the effect of the inbreeding over each trait, and μ_0 is the coefficient of inbreeding corresponding to the null hypothesis (in this case is equal to 0), and s.e. is the standard error. The transformation of Z-scores into *P*-values was accomplished with the function *pnorm()* implemented on R [28].

The above analysis was performed by regressing each phenotype onto four genomic inbreeding coefficients (F_{HOM}, F_{ROH}, $F_{ROHLong}$, and $F_{ROHShort}$). In order to detect genomic regions associated with inbreeding depression, analyses at the chromosomal level were performed for traits that in the whole-genome analysis were identified as significantly affected by inbreeding depression (*P*-value < 0.05). Following Saura et al. [6], inbreeding depression was finely mapped by dividing

chromosomes into six segments and performing the analyses reported above in each segment. As previously explained, genes mapping to genomic regions associated with inbreeding depression for a specific trait were retrieved using Biomart [31], and their biological functions were assessed with UniProt [33] and David Bioinformatics Resources version 6.8 [32].

Results

Assessment of homozygosity patterns in Murciano-Granadina goats

A total of 46,689 SNPs and 1,040 animals were selected to investigate the population structure and patterns of homozygosity of Murciano-Granadina goats. The PCA (**Additional file 2: Fig. S1**) indicated that goats clustered, to some extent, in accordance with their farm of origin. We detected 20,312 ROH that were classified as follows: 11,325 had sizes of 0–5 Mb, while 5,470 (5–10 Mb), 2,695 (10–20 Mb), 789 (20–50 Mb) and 33 (> 50 Mb) displayed sizes above 5 Mb. The mean number of ROH per category and per individual varied slightly across farms (**Fig. 1A** and **1B**). The mean ROH number was 19.53 ± 11.89 per individual, with an average length of 6.15 ± 2.05 Mb. As depicted in **Fig. 2**, the majority of the individuals harboured less than 50 ROH, and ROH covered a small proportion of the genome (< 300 Mb, about 10% of the genome). Only 2% of the individuals showed a genomic ROH coverage > 500 Mb, and 1% harboured more than 50 ROH. Larger chromosomes encompassed a greater number of ROH when compared to the smaller ones (**Additional file 3: Fig. S2**), and the correlation coefficient between the number of ROH and chromosome length was 0.92 (P-value < 0.05).

The genome-wide analysis of homozygosity, based on the proportion of homozygous individuals for each genotyped position, made it possible to detect 38 genomic regions with increased homozygosity (harbouring the top 1% of the most homozygous positions) that were scattered on 20 goat chromosomes (CHI), i.e. CHI 1–8, 11, 13–18, 20, 21, 24, 26 and 29 (**Fig. 3A** and **Additional file 4: Table S2**). Eight ROH hotspots mapping to CHI 1, 2, 4, 6, 14, 16, and 17 were identified (**Fig. 3B** and **Additional file 5: Table S3**). One region (i.e. CHI 4:42,552,375–48,378,207 bp) was consistently detected in the genome-wide analysis of homozygosity and ROH. Sixty-six genes mapped to these regions (**Additional file 6: Table S4)** and a functional enrichment analysis revealed 15 GO terms significantly enriched at the nominal level (*P*-value < 0.05). Particularly significant were GO terms related with ferric and copper import into the cell (**Additional file 7: Table S5**).

Estimation of inbreeding coefficients

Genomic inbreeding coefficients reached values of 0.601 ± 0.021 (F_{HOM}); 0.053 ± 0.046 (F_{ROH}); 0.040 ± 0.041 ($F_{ROHLong,}$), 0.014 ± 0.008 ($F_{ROHShort}$), – 0.016 ± 0.035 (F_{IS}) and 0.023 ± 0.047 (F_{YANG}). The magnitude and dispersion of these coefficients are shown in **Fig. 4**. The F_{HOM}, $F_{ROH,}$ $F_{ROHLong}$, F_{IS} and F_{YANG} coefficients were highly correlated, being especially high the correlations between F_{HOM} and F_{ROH}, F_{ROH} and $F_{ROHLong}$ (r = 0.99, *P*-value < 2.2×10^{-16}) and between F_{HOM} and F_{IS} (r = – 1, *P*-value < 2.2×10^{-16}). In contrast, $F_{ROHShort}$ showed the weakest correlations with the remaining inbreeding coefficients ($|r|$= 0.33–0.64),

although their statistical significance (*P*-value < 2.2 ×10^{-16}) was very high (**Table 2**).

Figure 1. (1A) Number of ROH classified according to their length. Purple and yellow bars represent the counts of ROH shorter and longer than 5 Mb, respectively. (1B) Number of ROH classified according to their length category by the farm of origin.

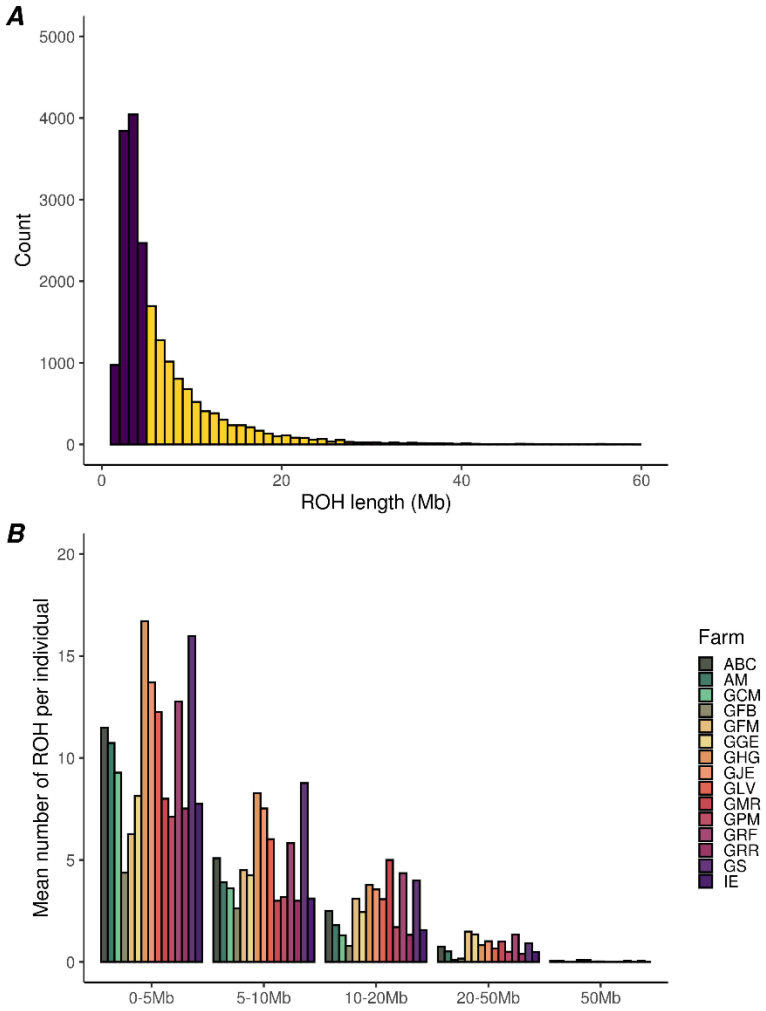


Figure 2. Relationship between ROH number and length in 1,040 Murciano-Granadina goats genotyped with the Goat SNP50 BeadChip.

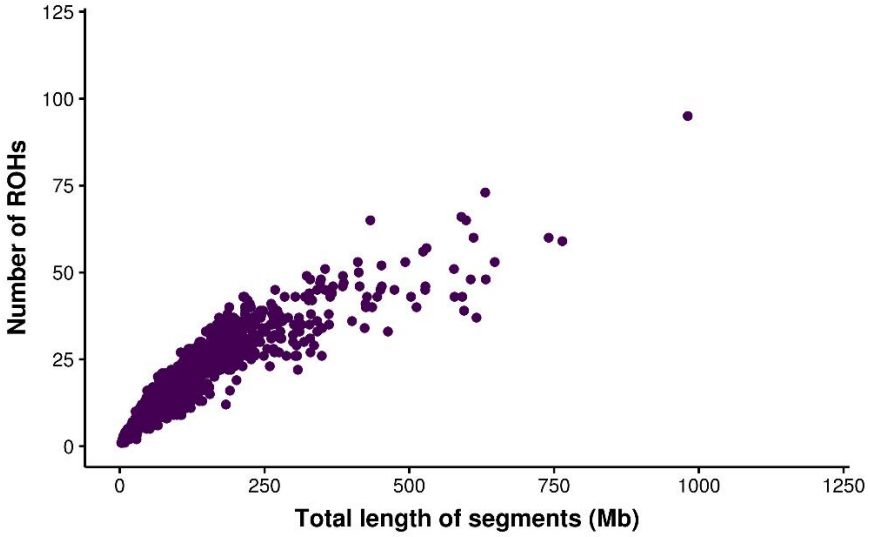


Figure 3. (3A) Proportion of individuals with homozygous genotypes for each SNP marker. The *y*-axis displays the proportion of individuals for which a specific SNP displays a homozygous genotype, while the *x*-axis corresponds to the positional coordinates of SNPs distributed in the 29 caprine autosomes. (3B) ROH hotspots identified in the population of Murciano-Granadina goats under study. The *y*-axis displays the frequency at which a given SNP is found within a ROH in the population; while the *x*-axis corresponds to the positional coordinates of SNPs distributed in the 29 caprine autosomes. Markers above the red line are in the top 1% of each category (homozygosity or frequency of being within a ROH). Markers highlighted in green are located in genomic regions consistently identified as regions of high homozygosity and ROH hotspots.

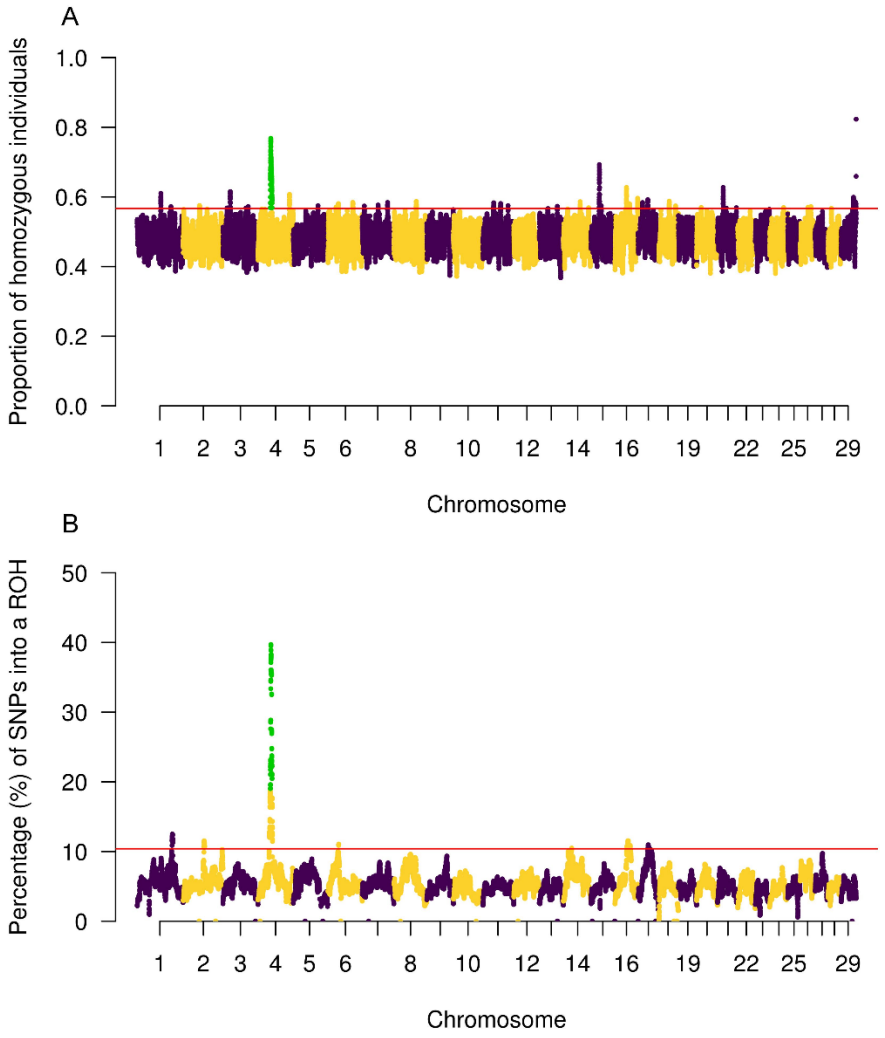
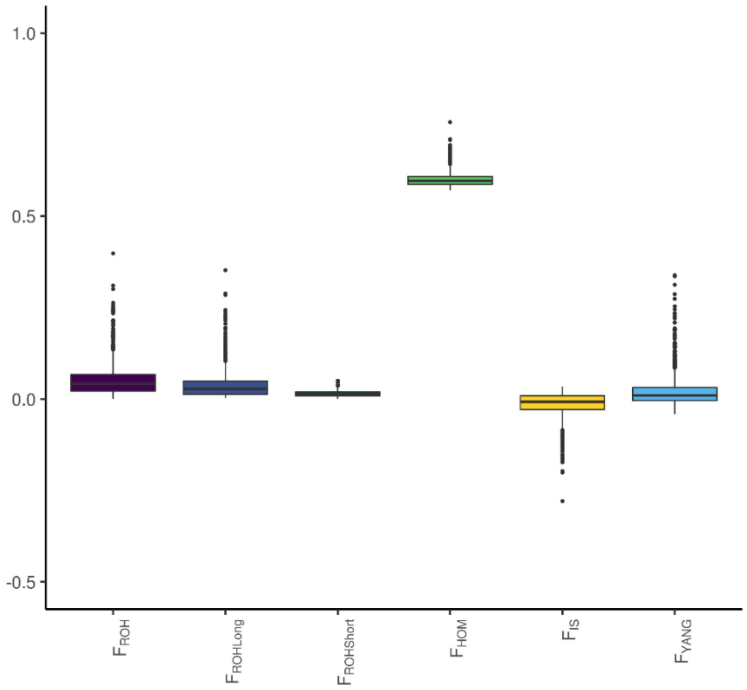


Table 2. Pearson correlations between molecular inbreeding coefficients (*F*) estimated in 1,040 Murciano-Granadina goats.

***F* coefficient**	F_{HOM}	F_{ROH}	$F_{ROHLong}$	$F_{ROHShort}$	F_{IS}	F_{YANG}
F_{HOM}	1					
F_{ROH}	0.99	1				
$F_{ROHLong}$	0.97	0.99	1			
$F_{ROHShort}$	0.64	0.59	0.45	1		
F_{IS}	– 1	– 0.99	– 0.97	– 0.64	1	
F_{YANG}	0.86	0.89	0.92	0.33	– 0.86	1

*All correlation coefficients (lower part of the matrix) were significant (*P*-value < 2.2×10^{-6}).

Figure 4. Boxplots depicting the magnitude and dispersion of molecular inbreeding F_{ROH}, $F_{ROHLong}$, $F_{ROHShort}$, F_{HOM}, F_{IS} and F_{YANG} coefficients estimated in 1,040 female Murciano-Granadina goats. Differences in magnitude between F_{HOM} and the other molecular coefficients are due to the fact that they indicate identity-by-state and identity-by-descent allele-sharing proportions, respectively.



Measurement of inbreeding depression for dairy traits

The natural logarithmic transformation of the somatic cell count divided by 1,000 (lnSCC) as well as milk yield at three different time points (MY210, MY240 and MY305) showed significant evidence of inbreeding depression when analysing F_{HOM} and F_{ROH}. In contrast, no significant effects of inbreeding were identified for coefficients based either on short or long ROH. Increases of 0.1 units of F_{HOM} and F_{ROH} coefficients involved lnSCC increments of 0.29 (*P-value* = 0.037) and 0.127

(*P*-value = 0.038) units, respectively (**Table 3**). At the chromosomal level, significant inbreeding depression for lnSCC was detected on CHI 8 and CHI 25 when regressed on either F_{HOM} or F_{ROH}, while six additional chromosomes (CHI 13, CHI 14, CHI 22, CHI 24, CHI 25 and CHI 27) displayed inbreeding depression for this trait exclusively when it was regressed onto F_{HOM} (**Additional file 8: Table S6**). Four regions containing 666 genes on chromosomes 8 (i.e. CHI 8:37,557,623–56,336,435 bp and CHI 8:75,115,244–93,894,055 bp) and 25 (i.e. CHI 25: 82,419–7,143,084 bp and CHI 25:21,429,255–28,572,340 bp) displayed a significant inbreeding depression for lnSCC based on F_{HOM} (**Additional file 9: Table S7** and **Additional file 10: Table S8)**. After performing functional enrichment analysis, various GO terms and pathways involved in the immune response were significantly enriched only at the nominal level, including integrin domains (i.e. *ITGAM*, *ITGAD*, *ITGAX*, *ITGAL*) and genes with protein kinase activity chemotaxis and cell signalling activities (i.e. *CCL19*, *CCL21*, *CCL24*, *CCL26*, *SYK* and *IL27*). Several of these genes display functions related with innate immunity and inflammatory response e.g. lymphocyte chemotaxis, cellular response to interferon-γ, immunological synapse formation, positive regulation of chemotaxis, cellular response to interleukin-1, monocyte chemotaxis etc. (**Additional file 11: Table S9**).

Table 3. Inbreeding depression estimates for milk yield and composition traits expressed as the change of the phenotypic mean per 0.1 units increase of the corresponding inbreeding coefficient (95% confidence intervals are displayed in brackets).

Trait[b]	F_{HOM}		F_{ROH}		$F_{ROHLong}$		$F_{ROHShort}$	
	Mean ± s.e.[a] (CI 95%)	***P*-value**	**Mean ± s.e.[a] (CI 95%)**	***P*-value**	**Mean ± s.e.[a] (CI 95%)**	***P*-value**	**Mean ± s.e.[a] (CI 95%)**	***P*-value**
MY210	– 48.503 ± 21.124 (– 89.906 to – 7.101)	**0.011**	– 20.492 to ± 9.356 (– 38.830 to – 2.154)	**0.014**	21.231 ± 25.901 (– 29.536 to 71.998)	0.206	19.751 ± 83.308 (– 143.533 to 183.034)	0.406
MY240	– 51.017 ± 24.159 (– 98.369 to – 3.666)	**0.017**	– 21.296 to ± 10.702 (– 42.272 to – 0.320)	**0.023**	34.358 ± 30.841 (– 26.090 to 94.806)	0.133	10.158 ± 93.948 (– 173.980 to 194.296)	0.457
MY305	– 55.719 ± 29.987 (– 114.493 to 3.055)	**0.032**	– 23.196 to ± 13.284 (– 49.233 to 2.841)	**0.040**	48.052 ± 47.583 (– 45.211 to 141.315)	0.156	– 16.834 ± 118.042 (– 248.196 to 214.527)	0.443
lnSCC	0.290 ± 0.162 (– 0.027 to 0.608)	**0.037**	0.127 ± 0.072 (– 0.013 to 0.268)	**0.038**	– 0.268 ± 0.635 (– 1.514 to 0.977)	0.336	– 0.368 ± 0.649 (– 1.640 to 0.905)	0.286
FP	0.119 ± 0.121 (– 0.117 to 0.356)	0.161	0.046 ± 0.053 (– 0.059 to 0.150)	0.196	– 0.016 ± 0.209 (– 0.426 to 0.394)	0.469	– 0.031 ± 0.428 (– 0.871 to 0.809)	0.471
LP	0.029 ± 0.044 (– 0.057 to 0.116)	0.252	0.015 ± 0.019 (– 0.024 to 0.053)	0.226	– 0.090 ± 0.172 (– 0.428 to 0.247)	0.300	– 0.073 ± 0.169 (– 0.404 to 0.257)	0.332
PP	0.071 ± 0.068 (– 0.062 to 0.204)	0.147	0.035 ± 0.030 (– 0.024 to 0.093)	0.122	0.039 ± 0.106 (– 0.169 to 0.246)	0.358	– 0.157 ± 0.257 (– 0.660 to 0.346)	0.270

[a]**s.e.** standard error; CI 95%, 95% confidence interval; [b]**MY210**; milk yield at 210 days of lactation (kg); MY240; milk yield at 240 days of lactation (kg); MY305; milk yield at 305 days of lactation (kg); lnSCC, natural logarithm of the milk somatic cell count divided by 1,000 (to convert this value into a somatic cell count use the formula: $e^{lnSCC} \times 10^3$ cells/mL); FP, milk fat percentage (%); LP, milk lactose percentage (%); PP, milk protein percentage (%). * cells in bold correspond to *P*-values below 0.05.

With regard to the milk yield traits, an increment of 0.1 units of F_{HOM} or F_{ROH} involved a decrease of 48.50 kg (*P*-value = 0.011) and 20.49 kg (*P*-value = 0.014) of milk for MY210; 51.02 kg (*P*-value = 0.017) and 21.30 kg (*P*-value = 0.023) for MY240; and 55.72 kg (*P*-value = 0.032) and 23.20 kg (*P*-value = 0.040) for MY305, respectively (**Table 3**). The analysis at the chromosomal level indicated significant inbreeding depression for MY210 and MY240 on CHI2, CHI3 and CHI11, while for MY305 inbreeding depression was significant only on CHI2 (**Additional file 12: Table S10)**. The region on CHI2, comprised between 22,751,824 and 68,255,473 bp displayed significant inbreeding for all three milk yield traits (**Additional file 13: Table S11)**. A total of 355 genes mapped to this genomic region (**Additional file 14: Table S12)**. The functional enrichment analysis evidenced an overrepresentation of genes involved in mitochondrial and energetic processes, including genes from the PPAR signalling pathway i.e. *CYP27A1*, *ACADL*, *DBI* and *ACSL9* (**Additional file 15: Table S13)**.

Although there is no substantial overlap between regions associated with inbreeding depression and ROH hotspots, the CHI 8:75–93.8 Mb region associated with inbreeding depression for lnSCC and the CHI 2:45.5–68.25 Mb region associated with inbreeding depression for milk yield show positional concordance with two of the 38 regions displaying high homozygosity (i.e. CHI 8:76.25–77.4 Mb and CHI 2:56.13–57.17 Mb).

Discussion

Low inbreeding in Murciano-Granadina goats

For the majority of Murciano-Granadina goats, ROH number and total length were below 50 and 350 Mb, respectively. Moreover, short ROH (< 5 Mb) were more abundant than the medium or long ones (**Fig. 1A**). These patterns are pretty consistent with what has been observed by Bertolini et al. [19] in a worldwide sample of goat breeds. Indeed, Bertolini et al. [19] showed that goat breeds from Southern Europe had, on average, 49 ROH per individual while the genomic coverage per individual was 183.47 Mb. In contrast, breeds from Northern Europe showed higher levels of homozygosity with 98 ROH per individual and genomic coverage of 479.17 Mb.

The inbreeding coefficients F_{ROH}, $F_{ROHShort,}$ and $F_{ROHLong}$ of Murciano-Granadina goats were mainly in the range of 0 to 0.05. In their study, Bertolini et al. [19] reported that about 60% of a worldwide sample of goat breeds displayed low F_{ROH} coefficients (< 0.10), while the remaining ~ 30% and ~ 10% of breeds showed moderate ($0.10 < F_{ROH} < 0.20$) or high (> 0.20) F_{ROH} values. The patterns of low homozygosity that we have observed in Murciano-Granadina goats contrast strongly with what has been reported in certain local breeds, such as Mallorquina, Pyrenean, and Valdostana, which have undergone sharp population bottlenecks [19, 20]. Low inbreeding and homozygosity in the Murciano-Granadina breed are probably explained by its large census size (> 100,000 individuals in the herdbook), the absence of population bottlenecks, and its broad geographic distribution

encompassing more than 4,000 farms across Spain (https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas) and other countries. Noteworthy, the Murciano-Granadina breed was founded by crossing, during the 1970s, two Murciana and Granadina populations with different historical origins [40]. Although genetic differentiation between these two populations was weak [40], this admixture event probably contributed to increase the heterozygosity of the resulting Murciano-Granadina breed. Widespread use of artificial insemination in reproductive management and intensive selection were implemented in the Murciano-Granadina breed a few decades ago [24], so their impact on genetic diversity and inbreeding has probably been quite limited so far.

While correlations between F_{HOM}, F_{ROH}, $F_{ROHLong}$, F_{IS} and F_{YANG} were high, $F_{ROHShort}$ displayed the lowest correlations with the remaining inbreeding coefficients, in line with previous studies focused on other livestock species [6, 13, 41]. Short ROH are mainly derived from ancient inbreeding events [5, 42] and do not reflect the whole autozygosity of the sample. It is also possible that several of these homozygous tracks are identical by state and not by descent, being produced by a low recombination rate or high linkage disequilibrium in unrelated ancestors [30]. Besides, when working with medium density genotype arrays (e.g. 50K SNPs) the detection of short ROH can become quite difficult [43], thus limiting the ability to infer the true proportion of short vs. long ROH in the genome.

Several ROH hotspots are detected in the genomes of Murciano-Granadina goats

We have identified several genomic regions in which ROH are particularly frequent (ROH hotspots). Similar patterns were found when the proportion of homozygous individuals was analysed at a genome-wide level i.e. 35.46% of the most common homozygous regions overlapped with ROH hotspots and these overlapping regions represented 35.42% of the total ROH hotspots. No positional coincidence was detected between ROH hotspots identified by us and those reported by Bertolini et al. [19] in a worldwide sample of goat populations. This finding agrees well with what has been previously observed in sheep [44]. Indeed, ROH hotspots are produced by factors such as positive selection and inversions suppressing recombination, that can differ substantially from population to population [42]. For instance, the characterisation of the products of 5,860 female meioses in *Drosophila melanogaster* by genotyping more than 100 million SNPs made it possible to detect 106,964 recombination events displaying a remarkable degree of intra-specific variation [45]. Factors such as GC content, gene density, distribution of simple repeats and transposable elements, structural variation, and the presence of diverse poorly-characterised sequence motifs might explain the regional variation of the recombination rate across individuals and populations [46].

ROH hotspots contain genes with diverse functions

Regarding the gene content of genomic segments co-localising with both high homozygosity regions and ROH hotspots (CHI 4:42,552,375–48,378,207 bp), the

functional enrichment analysis highlighted several gene ontology terms with nominally significant enrichment (*P*-value < 0.05) (**Additional file 7: Table S5**). From this list of genes, it is worth emphasising STEAP1, STEAP2 and STEAP4 metalloreductases which facilitate the cellular uptake of iron and copper [47]. These proteins modulate the effects of intracellular oxidative stress and inflammation and are involved in multiple biological pathways related with molecular trafficking in the endocytic and exocytic pathways, metabolism, control of cell proliferation and apoptosis and tumour progression [48]. We also found genes with metabolic functions such as insulin-like growth factor binding protein 3 (*IGFBP3*) and insulin-like growth factor binding protein 1 (*IGFBP1*) that participate in the growth and postnatal development of cattle [49, 50]. These results are quite concordant with those reported by Mastrangelo and colleagues [30], who showed that ROH islands identified in Italian bovine breeds contained genes with heterogeneous functions related to milk production, reproduction, immune response, and resistance/susceptibility to infection and diseases.

Effect of inbreeding depression over milk traits

The lnSCC in the Murciano-Granadina population under study averaged 6.25 ± 0.93 units, which is higher than the means reported in primiparous goats from the Alpine (5.09 ± 1.36 units) and Saanen (5.32 ± 1.19 units) breeds [51]. Compared with cows and ewes, goats display higher numbers of somatic cells in milk. Indeed, the apocrine nature of milk secretion in goats increases the proportion of cytoplasmatic particles in milk, a feature, that depending on the measurement method of choice,

could increase the somatic cell count [52]. Besides, somatic cell count is modulated by many factors including the occurrence of bacterial infections, stress, oestrous cycle phase, diet etc. [53]. According to our results (**Table 3**), inbreeding depression increased lnSCC, a feature that is considered adverse because high lnSCC values are often associated with subclinical and clinical mastitis [51, 52]

The magnitude of inbreeding depression for lnSCC estimated from $F_{ROHShort}$ strongly differed from estimates based on the other inbreeding coefficients (**Table 3**), a finding consistent with the moderate ($|r|$ = 0.33–0.64) correlations between $F_{ROHShort}$ and other molecular inbreeding coefficients (**Table 2**). Noteworthy, recent rather than ancestral inbreeding is the main cause of inbreeding depression in mammalian populations [54, 55]. Besides, long stretches of homozygosity usually contain a higher proportion of deleterious mutations than the shorter ones because they are more recent, so deleterious variation has not been yet purged by purifying selection [56].

In Murciano-Granadina goats, a previous study performed by Deroide et al. [23] reported a low percentage of inbreeding (average F = 0.24%). Milk production showed a positive quadratic correlation with inbreeding levels, but such effect was not significant. Deroide et al. [23] also reported that milk fat and dry extract contents experienced a slight increase due to inbreeding. In our study, the dairy trait mostly affected by inbreeding depression was lnSCC. Consistently, Doekes et al. [13] reported that a 1% increase of F_{ROH} involved a 0.86 ± 0.28 unit increase in somatic cell score (days 150 through to 400) recorded in Dutch Holstein-Friesian

dairy cattle. In Iranian cattle, individuals with high inbreeding coefficients tended to have higher somatic cell scores than animals with low inbreeding coefficients [57], and similar results have been reported for Canadian Holstein cattle [58]. Doekes et al. [13] indicated that ancient inbreeding was the main contributor to inbreeding depression for somatic cell score, although such effect was not significant. Somatic cell score is an indicator of the health status of the mammary gland and substantial increases are observed in individuals suffering from mastitis [59]. Inbreeding has been reported to significantly reduce resistance against pathogens in multiple organisms [60, 61, 62], so the significant inbreeding depression observed for lnSCC in Murciano-Granadina goats might be explained, at least in part, by the weakening of the immune defences of the mammary gland. Thus, homozygosity for deleterious mutations might result in the partial or total inactivation of genes related with immunity, and low variability might also compromise the effectiveness of the immune response [63].

Genomic regions associated with inbreeding depression for lnSCC contain several genes related with immunity

When we investigated which enriched clusters are present in the set of genes mapping to chromosomes 8 (37–56 Mb and 75–93 Mb) and 25 (0.082–7 Mb and 21–28 Mb) regions associated with inbreeding depression for lnSCC (**Additional file 10: Table S8**), we found several genes assigned to gene ontologies highly connected with immunity, e.g. integrin-mediated signalling pathway, lymphocyte chemotaxis, monocyte chemotaxis, immunological synapse formation, chemokine

activity, etc. (**Additional file 11: Table S9**). The spleen tyrosine kinase (*SYK*) protein forms part of the integrins cluster and maps to CHI 8: 86,755,291–86,861,895 bp (**Additional file 11: Table S9**). One of the functions of the SYK molecule is to stimulate the phosphorylation of Toll-like receptor 4 (TLR4) [64], which recognises bacterial lipopolysaccharide and induces inflammatory and immune responses [65]. This gene has been described as highly variable in cattle [66], and many *TLR4* polymorphisms and haplotypes have been associated with milk somatic cell count and susceptibility to mastitis [67]. Moreover, the SYK protein induces the recognition of pathogens and cell adhesion and platelet activation [65], and it also affects the proliferation of mammary epithelial cells at several stages of the milking cycle [66]. In the same enriched gene ontologies, we have detected the integrin subunit αM gene (*ITGAM* also known as *CD11b*) which maps to CHI25: 27,221,164–27,264,350 bp and encodes a receptor for lipopolysaccharide [68]. Signalling mediated by TLR4 activates the synthesis of ITGAM/CD11b, which is essential for the migration and adhesion of polymorphonuclear leukocytes to infection sites [68].

As previously indicated, genes related to chemotaxis were significantly enriched at the nominal level (*P*-value < 0.05) (**Additional file 11: Table S9**). This functional category is mainly represented by chemokines, such as chemokine ligand 19 (*CCL19*), 21 (*CCL21*), 24 (*CCL24*), 26 (*CCL26*) and 27 (*CCL27*). Chemokines are essential for the development of the innate immune response since they orchestrate and control the migration of the immune cells (macrophages, monocytes, neutrophils, etc) to sites of infection [69]. During the first stages of mastitis,

chemokines contribute to the stimulation of the cellular immune response against the invading pathogen until acute-phase proteins are expressed [70]. Marsland et al. [71], described how chemokines *CCL19* and *CCL21* participate in the maturation of dendritic cells, allowing them to leverage the T cell response. These chemokines also enhance the migration of leukocytes through lymph and blood circulation and stimulate the production of pathogen-induced proinflammatory cytokines [71]. In the gene set associated with inflammatory response (*P*-value = 0.03), we detected the interleukin 27 (*IL27*) gene which encodes a molecule with both pro and anti-inflammatory effects, thus enhancing the immune response and, at the same time, preventing tissue damage caused by inflammation [72]. Infections caused by Gram-negative bacteria induce IL27 production, and this cytokine interacts with monocytes increasing TLR4 expression and enhancing the LPS-induced inflammatory response [73]. Moreover, IL27 has an autocrine effect on macrophages and monocytes resulting in the amplification of the inflammatory response via cytokine secretion [72].

The genomic region displaying inbreeding depression for milk yield (MY210, MY240 and MY305) was significantly enriched at the nominal level (*P*-value < 0.05) with genes associated with multiple unrelated biological processes (**Additional file 15: Table S13**). The PPAR signalling pathway (*P*-value = 0.032) influences milk production and composition in cattle. Bai and collaborators [74] described an overrepresentation of genes from the PPAR signalling pathway in cows with high daily milk yield in comparison with low yielders. Besides, genes from this pathway are upregulated in cows in the lactation peak when compared to

those in the dry period [75], suggesting a role of these genes not only in the determinism of fat composition [76] but also of milk yield.

Conclusions

Murciano-Granadina goats display low levels of inbreeding (mean F_{ROH} = 0.053 ± 0.046), a finding consistent with the large census size and demographic history of this breed. Four genomic regions on CHI 8:37,557,623–56,336,435, CHI 8:75,115,244–93,894,055 bp, CHI 25: 82,419–7,143,084 and CHI 25:21,429,255–35,715,425 bp were associated with inbreeding depression for lnSCC. Moreover, one region on CHI 2: 22,751,824–68,255,473 was consistently associated with inbreeding depression for three milk yield traits (MY210, MY240 and MY310). Genes encoding integrins, chemokines and pathogen recognition receptors, which play relevant roles in the elicitation of innate immune responses against microbes, mapped to regions associated with inbreeding depression for lnSCC. These results suggest that keeping inbreeding to a minimum, through an adequate reproductive management, might be a useful approach to decrease the incidence of mastitis in Murciano-Granadina goats.

Declarations

Ethics approval and consent to participate

Blood collection is a routine procedure performed by CAPRIGRAN, so it does not require approval by the Ethics Committee on Animal and Human Experimentation of the Universitat Autònoma de Barcelona.

Consent for publication

Not applicable.

Availability of data and materials

Goat SNP50 BeadChip genotypes and milk production phenotypes are accessible at Figshare (https://doi.org/10.6084/m9.figshare.18095825).

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the European Regional Development Fund (FEDER)/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación/Project Reference grant: PID2019-105805RB-I00 and by the CERCA Programme/Generalitat de Catalunya. We also acknowledge the support of the Spanish Ministerio de Ciencia e Innovación for the Center of Excellence Severo Ochoa 2020–2023 (CEX2019-000902-S) grant awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). We also acknowledge the support of the CERCA programme of the Generalitat de

Catalunya. Dailu Guan was funded by a PhD fellowship from the China Scholarship Council (CSC). Maria Luigi-Sierra was funded with a PhD fellowship *Formación de Personal Investigador* (BES-C-2017-079709) awarded by the Spanish Ministry of Economy and Competitivity.

Authors' contributions

MA, JJ, JVD and VL designed the study. JFA, JVD and AM coordinated all tasks involved in phenotype recording. VL did all DNA extractions. MGLS, MS and AF carried out homozygosity and inbreeding depression analyses, with the cooperation of DG. MGLS and MA wrote the first draft of the paper with the cooperation of MS and AF. All authors read and approved the content of the paper.

Acknowledgements

Many thanks to CAPRIGRAN for carrying out phenotype recording and blood sample collection in Murciano-Granadina goats.

References

1. Lush JL. Animal Breeding Plans. Orchard Press; 2013.
2. Leroy G. Inbreeding depression in livestock species: Review and meta-analysis. Anim Genet. 2014;45:618–28.
3. Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. Heredity. 2015;115:63–72.

4. Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 2011;189:237–49.

5. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83:359–72.

6. Saura M, Fernández AIA, Varona L, Fernández AIA, De Cara MÁR, Barragán C, et al. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genet Sel Evol. 2015;47:1.

7. Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain. Am J Hum Genet. 1999;65:1493–500.

8. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet. 2012;91:275–92.

9. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC Genet. 2012;13:70.

10. Kristensen TN, Sørensen AC. Inbreeding - Lessons from animal breeding, evolutionary biology and conservation genetics. Anim Sci. 2005;80:121–33.

11. Charlesworth D, Willis JH. The genetics of inbreeding depression. Nat Rev Genet. 2009;10:783–96.

12. Howard JT, Pryce JE, Baes C, Maltecca C. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J Dairy Sci. 2017;100:6009–24.

13. Doekes HP, Veerkamp RF, Bijma P, De Jong G, Hiemstra SJ, Windig JJ. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. Genet Sel Evol. 2019;52:64.

14. Pryce JE, Haile-Mariam M, Goddard ME, Hayes BJ. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genet Sel Evol. 2014;46:71.

15. Thompson JR, Everett RW, Hammerschmidt NL. Effects of inbreeding on production and survival in Holsteins. J Dairy Sci. 2000;83:2131–8.
16. Sørensen AC, Madsen P, Sørensen MK, Berg P. Udder health shows inbreeding depression in Danish Holsteins. J Dairy Sci. 2006;89:4077–82.
17. Croquet C, Mayeres P, Gillon A, Vanderick S, Gengler N. Inbreeding depression for global and partial economic indexes, production, type, and functional traits. J Dairy Sci. 2006;89:2257–67.
18. Ercanbrack SK, Knight AD. Effects of inbreeding on reproduction and wool production of Rambouillet, Targhee, and Columbia ewes. J Anim Sci. 1991;69:4734–44.
19. Bertolini F, Cardoso TF, Marras G, Nicolazzi EL, Rothschild MF, Amills M. Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. Genet Sel Evol. 2018;50:59.
20. Cardoso TF, Amills M, Bertolini F, Rothschild M, Marras G, Boink G, et al. Patterns of homozygosity in insular and continental goat breeds. Genet Sel Evol. 2018;50:56.
21. Nandolo W, Mészáros G, Banda LJ, Gondwe TN, Lamuno D, Mulindwa HA, et al. Timing and extent of inbreeding in African goats. Front Genet. 2019;10:537.
22. Marete AG, Jung'a JO, Mosi RO. An analysis of inbreeding levels and factors affecting growth and reproductive traits in the Kenya Alpine dairy goat. Livest Res Rural Dev. 2011;23:259.
23. Deroide CAS, Jacopini LA, Delgado J V., Léon JM, Brasil LHA, Ribeiro MN. Inbreeding depression and environmental effect on milk traits of the Murciano-Granadina goat breed. Small Rumin Res. 2016;134:44–8.
24. Delgado JV, Landi V, Barba CJ, Fernández J, Gómez MM, Camacho ME, et al. Murciano-Granadina goat: A Spanish local breed ready for the challenges of the twenty-first century. Sustain Goat Prod Advers Environ. Springer, Cham; 2018. p. 205–19.
25. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988;16:1215.

26. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet. 2017;49:643–50.

27. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

28. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL *http://www.R-project.org/*.

29. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One. 2017;12:e0176780.

30. Mastrangelo S, Tolone M, Sardina MT, Sottile G, Sutera AM, Di Gerlando R, et al. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. Genet Sel Evol. 2017;49:84.

31. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database. 2011;2011:bar030.

32. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.

33. Bateman A. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15.

34. Thompson EA. Identity by descent: Variation in meiosis, across genomes, and in populations. Genetics. 2013;194:301–26.

35. Dumont BL, Payseur BA. Evolution of the genomic rate of recombination in mammals. Evolution. 2008;62:276–94.

36. Wright S. The genetical structure of populations. Ann Eugen. 1951;15:323–54.

37. Villanueva B, Fernández A, Saura M, Caballero A, Fernández J, Morales-González E, et al. The value of genomic relationship matrices to estimate levels of inbreeding. Genet Sel Evol. 2021;53.

38. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010; 42:565–9.

39. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). Commun. No. 28–07. 7th World Congr Genet Appl to Livestock Prod Montpellier, Fr. 2002.

40. Martínez A, Rocha L, Quiroz J, Delgado Bermejo J. Estudio de la diversidad genética intrarracial de la cabra Murciano-Granadina con microsatélites de ADN. Arch Zootec. 2007;56:417–20.

41. Biscarini F, Mastrangelo S, Catillo G, Senczuk G, Ciampolini R. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in Maremmana semi-feral cattle using pedigree and genomic data. Animals. 2020;10:2285.

42. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: Windows into population history and trait architecture. Nat Rev Genet. 2018:19:220–34.

43. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. BMC Genomics. 2018;19:106.

44. Luigi-Sierra MG, Cardoso TF, Martínez A, Pons A, Bermejo LA, Jordana J, et al. Low genome-wide homozygosity in 11 Spanish ovine breeds. Anim Genet. 2019;50:501–11.

45. Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 2012;8:e1002905.

46. Smukowski CS, Noor MAF. Recombination rate variation in closely related species. Heredity. 2011;107:496–508.

47. Ohgami RS, Campagna DR, McDonald A, Fleming MD. The Steap proteins are metalloreductases. Blood. 2006;108;1388–94.

48. Gomes IM, Maia CJ, Santos CR. STEAP proteins: From structure to applications in cancer therapy. Mol Cancer Res. 2012;10:537–87.

49. Simmons RM, Erikson DW, Kim J, Burghardt RC, Bazer FW, Johnson GA, et al. Insulin-like growth factor binding protein-1 in the ruminant uterus: Potential endometrial marker and regulator of conceptus elongation. Endocrinology. 2009;150:4295–305.

50. Hammon HM, Zbinden Y, Sauerwein H, Breier BH, Blum JW, Donkin SS. The response of the hepatic insulin-like growth factor system to growth hormone and dexamethasone in calves. J Endocrinol. 2003;179:427–35.

51. Rupp R, Clément V, Piacere A, Robert-Granié C, Manfredi E. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. J Dairy Sci. 2011;7:3629–34.

52. Paape MJ, Poutrel B, Contreras A, Marco JC, Capuco AV. Milk somatic cells and lactation in small ruminants. J Dairy Sci. 2001;84:E237–44.

53. Lerondelle C, Richard Y, Issartial J. Factors affecting somatic cell counts in goat milk. Small Rumin Res. 1992;8:129–39.

54. Kirin M, McQuillan R, Franklin CS, Campbell H, Mckeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. PLoS One. 2010;5:e13996.

55. Silió L, Rodríguez MC, Fernández A, Barragán C, Benítez R, Óvilo C, et al. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. J Anim Breed Genet. 2013;130:349–60.

56. Pemberton TJ, Szpiech ZA. Relationship between deleterious variation, genomic autozygosity, and disease risk: Insights from the 1000 genomes project. Am J Hum Genet. 2018;102:658–75.

57. Rokouei M, Vaez Torshizi R, Moradi Shahrbabak M, Sargolzaei M, Sørensen AC. Monitoring inbreeding trends and inbreeding depression for economically important traits of Holstein cattle in Iran. J Dairy Sci. 2010;93:3294–302.

58. Miglior F, Burnside EB, Dekkers JCM. Nonadditive genetic effects and inbreeding depression for somatic cell counts of Holstein cattle. J Dairy Sci. 1995;78:1168–73.

59. Menzies PI, Ramanoon SZ. Mastitis of sheep and goats. Vet Clin North Am Food Anim Pract. 2001;17:333–58.
60. Spielman D, Brook BW, Briscoe DA, Frankham R. Does inbreeding and loss of genetic diversity decrease disease resistance? Conserv Genet. 2004;5:439–48.
61. Reid JM, Arcese P, Keller LF, Elliott KH, Sampson L, Hasselquist D. Inbreeding effects on immune response in free-living song sparrows (*Melospiza melodia*). Proc Biol Sci. 2007;274:697–706.
62. Smallbone W, van Oosterhout C, Cable J. The effects of inbreeding on disease susceptibility: *Gyrodactylus turnbulli* infection of guppies, *Poecilia reticulata*. Exp Parasitol. 2016;167:32–7.
63. Siljestam M, Rueffler C. Heterozygote advantage can explain the extraordinary diversity of immune genes. bioRxiv. 2018;347344.
64. Miller YI, Choi S-H, Wiesner P, Bae YS. The SYK side of TLR4: Signalling mechanisms in response to LPS and minimally oxidised LDL. Br J Pharmacol. 2012;167:990–9.
65. Mócsai A, Ruland J, Tybulewicz VLJ. The SYK tyrosine kinase: A crucial player in diverse biological functions. Nat Rev Immunol. 2010;10:387–402.
66. Hou X, Lin L, Xing W, Yang Y, Duan X, Li Q, et al. Spleen tyrosine kinase regulates mammary epithelial cell proliferation in mammary glands of dairy cows. J Dairy Sci. 2016;99:3858–68.
67. Wang X, Xu S, Gao X, Ren H, Chen J. Genetic polymorphism of *TLR4* gene and correlation with mastitis in cattle. J Genet Genomics. 2007;34:406–12.
68. Todd RF, Petty HR. *β*2(CD11/CD18) integrins can serve as signaling partners for other leukocyte receptors. J Lab Clin Med. 1997;129:492–8.
69. Sokol CL, Luster AD. The chemokine system in innate immunity. Cold Spring Harb Perspect Biol. 2015;7:1–20.
70. Brenaut P, Lefèvre L, Rau A, Laloë D, Pisoni G, Moroni P, et al. Contribution of mammary epithelial cells to the immune response during early stages of a bacterial infection to *Staphylococcus aureus*. Vet Res. 2014;45:16.

71. Marsland BJ, Bättig P, Bauer M, Ruedl C, Lässing U, Beerli RR, et al. *CCL19* and *CCL21* induce a potent proinflammatory differentiation program in licensed dendritic cells. Immunity. 2005;22:493–505.

72. Yoshida H, Hunter CA. The immunobiology of interleukin-27. Annu Rev Immunol. 2015;33:417–43.

73. Petes C, Odoardi N, Plater SM, Martin NL, Gee K. IL-27 amplifies cytokine responses to Gram-negative bacterial products and *Salmonella typhimurium* infection. Sci Rep. 2018;8:13704.

74. Bai X, Zheng Z, Liu B, Ji X, Bai Y, Zhang W. Whole blood transcriptional profiling comparison between different milk yield of Chinese Holstein cows using RNA-seq data. BMC Genomics. 2016;17:512.

75. Li Q, Liang R, Li Y, Gao Y, Li Q, Sun D, et al. Identification of candidate genes for milk production traits by RNA sequencing on bovine liver at different lactation stages. BMC Genet. 2020;21:72.

76. Schennink A, Bovenhuis H, Léon-Kloosterziel KM, Van Arendonk JAM, Visker MHPW. Effect of polymorphisms in the *FASN*, *OLR1*, *PPARGC1A*, *PRL* and *STAT5A* genes on bovine milk-fat composition. Anim Genet. 2009;40: 909–16.

Additional files

Additional file 1: Table S1. Number and density of SNPs per chromosome in a population of 1,040 Murciano-Granadina goats genotyped with the Goat SNP50 BeadChip.

Additional file 2: Figure S1. Principal component analysis of 1,040 Murciano-Granadina goats distributed in 15 farms (each farm is indicated with a different colour).

Additional file 3: Figure S2. Number of ROH per chromosome (represented as yellow bars, left axis) and the percentage of each chromosome covered by ROH (represented by a red line, right axis) in 1,040 Murciano-Granadina goats.

Additional file 4: Table S2. Genomic regions associated with the top 1% of homozygosity for each SNP marker in a population of 1,040 Murciano-Granadina goats genotyped with the Goat SNP50 BeadChip.

Additional file 5: Table S3. ROH hotspots identified in the genomes of 1,040 Murciano-Granadina goats.

Additional file 6: Table S4. Genes mapping to regions consistently identified as ROH hotspots and regions with high homozygosity.

Additional file 7: Table S5. Functional enrichment analysis of genes mapping to regions consistently identified as ROH hotspots and regions with high homozygosity.

Additional file 8: Table S6. Inbreeding depression estimates (s.e: standard error) per goat chromosome (CHI) for the natural logarithm of the somatic cell count divided by 1,000 (lnSCC) and their 95% confidence intervals (C.I).

Additional file 9: Table S7. Inbreeding depression estimates (s.e: standard error) for the natural logarithm of the somatic cell count divided by 1,000 (lnSCC) in specific regions of goat chromosomes (CHI) 8 and 25 and their 95% confidence intervals (C.I).

Additional file 10: Table S8. Genes mapping to goat chromosome (CHI) 8 (37-56 Mb and 75-93 Mb) and 25 (0.082-7 Mb and 21-28 Mb) regions associated with

inbreeding depression for the natural logarithm of the somatic cell count divided by 1,000 (lnSCC).

Additional file 11: Table S9. Functional enrichment analysis of genes from the goat chromosome 8 (37-56 Mb and 75-93 Mb) and 25 (0.082-7 Mb and 21-28 Mb) regions associated with inbreeding depression for the natural logarithm of the somatic cell count divided by 1,000 (lnSCC).

Additional file 12: Table S10. Inbreeding depression estimates (s.e: standard error) per goat chromosome (CHI) for milk yield at 210, 240 and 305 days (MY210, MY240 and MY305, measured in kg) and their 95% confidence intervals (C.I).

Additional file 13: Table S11. Inbreeding depression estimates (s.e: standard error) for milk yield at 210, 240 and 305 days (MY210, MY240 and MY305, measured in kg) in specific regions of goat chromosomes (CHI) 2 and their 95% confidence intervals (C.I).

Additional file 14: Table S12. Genes mapping to goat chromosome (CHI) 2: 22-68 Mb region associated with inbreeding depression for milk yield at 210, 240 and 305 days (MY210, MY240 and MY305).

Additional file 15: Table S13. Functional enrichment analysis of genes from the goat chromosome (CHI) 2: 22-68 Mb region associated with inbreeding depression for milk yield at 210, 240 and 305 days (MY210, MY240 and MY305).

bioRxiv

RUNNING HEAD: TRD and GenTrain scores

Impact of SNP calling quality on the detection of transmission ratio distortion in goats

María Gracia Luigi-Sierra[1], Joaquim Casellas[2*], Amparo Martínez[3], Juan Vicente Delgado[3], Javier Fernández Álvarez[3], Francesc Xavier Such[4], Jordi Jordana[2], Marcel Amills[1, 2*]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain. [2]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. [3]Departamento de Genética, Universidad de Córdoba, Córdoba 14071, Spain. [4]Group of Research in Ruminants (G2R), Department of Animal and Food Science, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain.

Corresponding authors: *

doi: https://doi.org/10.1101/2021.06.09.447792

Summary

Transmission ratio distortion (TRD) is the preferential transmission of one specific allele to offspring at the expense of the other one. The existence of TRD is mostly explained by the segregation of genetic variants with deleterious effects on the developmental processes that go from the formation of gametes to fecundation and birth. A few years ago, a statistical methodology was implemented in order to detect TRD signals on a genome-wide scale as a first step to uncover the biological basis of TRD and reproductive success in domestic species. In the current work, we have analyzed the impact of SNP calling quality on the detection of TRD signals in a population of Murciano-Granadina goats. Seventeen bucks and their offspring (N=288) were typed with the Goat SNP50 BeadChip, while the genotypes of the dams were lacking. Performance of a genome-wide scan revealed the existence of 36 SNPs showing significant evidence of TRD. When we calculated GenTrain scores for each one of the SNPs, we observed that 25 SNPs showed scores below 0.8. The allele frequencies of these SNPs in the offspring were not correlated with the allele frequencies estimated in the dams with statistical methods, thus evidencing that flawed SNP calling quality might lead to the detection of spurious TRD signals. We conclude that, when performing TRD scans, the GenTrain scores of markers should be taken into account to discriminate SNPs that are truly under TRD from those yielding spurious signals due to technical problems.

KEYWORDS: Transmission ratio distortion, GenTrain score, genotyping, Mendelian segregation.

Introduction

In diploid organisms, allelic transmission from parents to offspring is expected to followthe Mendelian law of inheritance, implying that both paternal and maternal alleles are transmitted to the progeny following approximately a 1:1 ratio (Huang et al., 2013). A deviation from this ratio, the so-called transmission ratio distortion (TRD), is produced when either the paternal or the maternal allele is preferentially transmitted to the offspring (Fishman and Mcintosh, 2019). A few years ago, Casellas et al. (2014) implemented a new Bayesian methodology to scan TRD for biallelic markers in diploid organisms. This method was later refined by Vázquez-Gómez et al. (2020) to detect TRD even in pedigrees with incomplete trios. This can be achieved by inferring the probability that specific alleles are present in the parent with a missing genotype based on the allele frequencies of the SNP in the overall population. The accuracy of such inference might be substantially affected by the quality of SNP genotypes. In previous studies (Abdalla et al., 2020; Casellas et al., 2017, 2020 Gòdia et al., 2020; Lahoucine etal., 2020; Vázquez-Gómez et al., 2020), the filtering of SNPs used in TRD scans relied fundamentally on two parameters (genotype call rate and minimum allele frequency or MAF). Deviation from HWE has never been used for this purpose because SNPsshowing TRD are expected to display significant departures from HWE. However, it should be noticed that SNPs displaying significant HWE deviations often do so because of genotyping errors (Hosking et al., 2004). In other words, HWE filtering eliminates many unreliable markers which might yield spurious TRD signals just because of technical problems.

The goal of the current work was to make a TRD scan with incomplete trios and then to assess the calling quality of the SNPs that are putatively under TRD by using the GenTrain score implemented in the GenomeStudio software from Illumina (Zhao et al., 2018). The GenTrain score is a metric that fluctuates between 0 (very poor calling quality) and 1 (excellent calling quality) and indicates the reliability of SNP detection based on the distribution of genotypic classes (Pavy et al., 2008; Zhao et al., 2018). By doing so, we aimed to assess the usefulness of the GenTrain score as a complementary metric to be considered in TRD scans as well as to recommend a specific GenTrain score filtering threshold to researchers interested in the detection of TRD.

Materials and methods

Sampling and genotyping of Murciano-Granadina goats

As animal material, we have collected blood samples from 17 bucks and their offspring (N=288) in vacuum tubes with K_3EDTA. These samples have been subsequently stored at -20ºC. Since blood collection is a routine procedure performed by CAPRIGRAN, no approval by the Ethics Committee on Animal and Human Experimentation of theUniversitat Autònoma de Barcelona was required to perform this experiment.Information about the number of offspring per sire is depicted in Table S1. Genomic DNA extractions were performed following the modified salting out procedure described by Guan et al. (2020). Animals were genotyped with the Illumina Goat SNP50 BeadChip (Illumina Inc., San Diego, CA), which contains 54,241 SNP, following the instructions of the manufacturer.

Genotypic data were updated with PLINK 1.9 (Chang et al., 2015) based on the *Capra hircus* genome ARS1 assembly (Bickhart et al., 2017) and the annotation provided by the International Goat Genome Consortium (http://www.goatgenome.org/projects.html#50K_snp_chip). Genotypes were pruned using PLINK 1.9 (Chang et al., 2015). We selected SNPs fulfilling the following criteria: (1) genotype call rate over 95%, (2) minor allele frequency above 0.05 and (3) no missing genotypes in any of the 17 sires. Besides, the percentage ofsires heterozygous for each SNP was estimated from the output obtained with the --hwe command of PLINK 1.9 (Chang et al., 2015) in order to remove SNPs with less than 20% of heterozygosity in the sire population (only SNPs with heterozygous genotypes are informative). The reference allele in this subset of the population was set as the mostcommon allele in all individuals.

Statistical methods

To estimate TRD, we used a frequentist modification (Vázquez-Gómez et al., 2020) of the Bayesian method implemented by Casellas et al. (2014). Assuming two alleles (A1 and A2) and the existence of genotyped heterozygous sires and of ungenotyped dams, this method allows to compute for every marker an α-value which ranges between −0.5 (the A1 allele is not transmitted) and 0.5 (the A2 allele is not transmitted) thus providing an estimate of the magnitude of TRD. Allele frequencies in the ungenotyped dams were inferred by calculating a π-parameter which varies from 0 to 1. The two α and π parameters were estimated by maximizing the likelihood function and the statistical significance of α was assessed

by using a likelihood ratio test (Nelson, 2008). A correction for multiple testing was applied to the P-values obtained from the χ^2 distribution using the false discovery rate approach (FDR) reported by Benjamini & Hochberg (1995) to obtain the corresponding q-values. Markers with α-values above 0.15 or below -0.15 and q-values < 0.05 were considered to show significant TRD. After quality control based on genotype call rate and MAF, the final genotypic data included 42,272 autosomal SNPs.

Results and discussion

The implementation of the TRD test allowed us to identify 2,944 SNPs that were deviating from the Mendelian ratio (α-value > 0.15 or < -0.15). The highest α-value detected in our population was 0.499, while the lowest was -0.428; but for the majority of the genotyped SNPs, α-values were comprised in the [-0.15 to 0.15] interval which indicates the absence of TRD (Fig. 1a). After applying a likelihood ratio test and FDR correction, 36 SNPs were selected as significant (*q-value* < 0.05), from which 15 SNPs had an α-value below -0.15 implying a major transmission of the alternative allele,while the remaining 21 SNPs showed an over transmission (α-value over 0.15) of the reference allele (Fig. 1b).

In order to verify the accuracy of the genotyping of the 36 SNPs displaying significant TRD, we calculated their GenTrain scores with the GenomeStudio software (Illumina Inc., San Diego, CA). Clustering of SNPs with different GenTrain scores and distribution of the GenTrain scores of the genotyped SNPs are depicted in Fig. 1c and Fig. S1. In Table S2, it can be seen that 25 of these SNPs

have GenTrain scores below 0.80, with values ranging from 0.16-0.63 and an average score of 0.51 ± 1.14 (**Group1**). In contrast, eleven SNPs have GenTrain scores above such threshold, with anaverage value of 0.87 ± 0.04 (**Group 2**). For each of these two groups of SNPs, we havecalculated the correlation between allele frequencies of the SNP in the offspring vsallele frequencies inferred for the ungenotyped dams with the methods reported by Vázquez-Gómez et al. (2020). In principle, allele frequencies of parents and their offspring should be significantly and positively correlated. In the **Group 1** of SNPs such correlation was very weak and non-significant (r = -0.007, *P*-value = 0.9733). Even worse, when we retrieved from **Group 1** seventeen SNPs with GenTrain scores between 0.5-0.6, the correlation was -0.0932 (*P*-value 0.7217). In strong contrast, allele frequencies of mothers and offspring were highly correlated in the **Group 2** of SNPs (r = 0.8656, *P*-value = 0.0005). This result implies that the method implemented by Casellas et al. (2014) and subsequently modified by Vázquez-Gómez et al. (2020) works very well in reconstructing allele frequencies in parental individuals without genotypes when SNPs have high GenTrain scores (> 0.80 in our study), which are the vast majority (Fig. 1c).

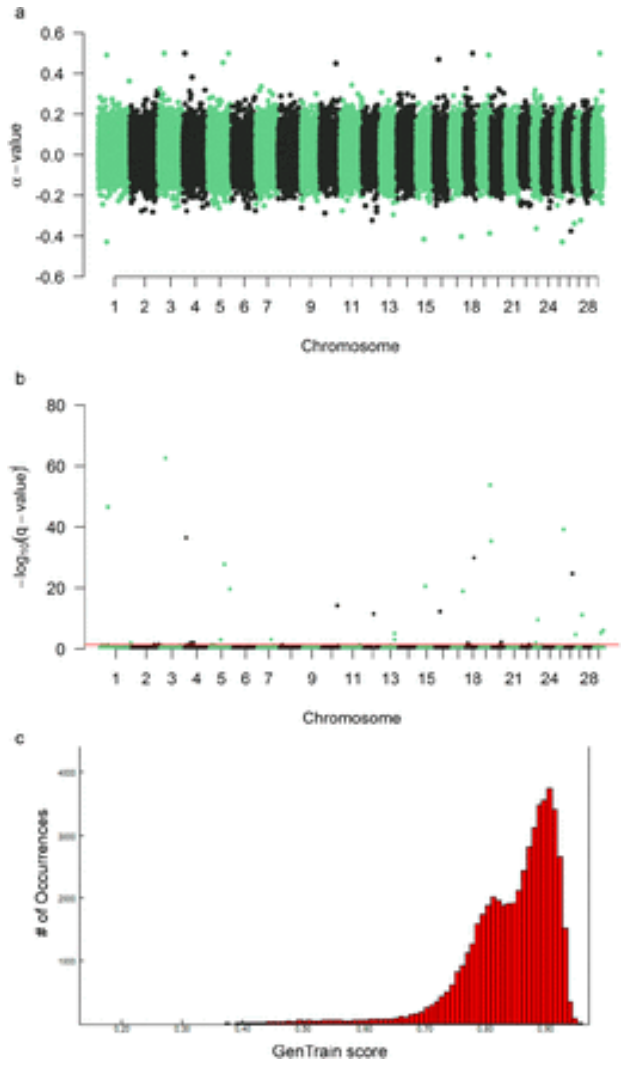
Previous reports have indicated that SNPs with minimum GenTrain scores of 0.25 can be safely used in most analyses (Pavy et al., 2008). Indeed, Pavy et al. (2008) designed highly-multiplexed SNP arrays for the genotyping of black and white spruce and reported that SNPs with GenTrain scores of 0.25 are very reliable and have a low rate ofmissing data. According to Pavy et al. (2008), SNPs with GenTrain scores of 0.25 or more had average call rates above 99%, and for SNPs

with GenTrain scores above 0.4 the rate of missing data became negligible and the average rate of missing data per successful SNP was very low, On the other hand, Guo et al. (2014) indicated that SNPs with GenTrain scores above 0.70 are correctly clustered in most cases, while the clustering of SNPs with GenTrain scores below 0.7 might be more problematic. Based on our results, we conclude that the performance of TRD scans, especially in the case in which full trios are not available, should rely on the establishment of a stringent threshold for SNP calling quality. In our study, family size was small (on average 17-18offspring per sire) and the density of the Goat SNP50 BeadChip (Illumina) is modest. In these conditions, we advise to consider markers with GenTrain scores of 0.80 or higher.

In studies with a more optimal design (particularly regarding family size), such threshold could be determined empirically by analyzing the correlations between the allele frequencies estimated in the parental class without genotypes and those inferredexperimentally in the offspring, which should be significant and positive. This simple approach should facilitate the elimination of spurious TRD signals produced by technical factors in order to concentrate efforts on those that have biological implications.

Fig. 1a. Genome-wide detection of SNP markers that show evidence of transmission ratio distortion in a population comprising 17 sire-families of Murciano-Granadinagoats. The α-value estimated for each SNP is plotted in the *y*-axis while the chromosomal locations of SNPs are indicated in the *x*-axis. **1b.** Manhattan plot indicating the statistical significance (y-axis), expressed as $-\log_{10}$

of the *q*-value, of the α- values calculated for 42,272 SNPs genotyped in a population comprising 17 sire- families of Murciano-Granadina goats. The chromosomal location of each SNP is indicated in the x-axis. The red line corresponds to the threshold of significance which corresponds to a *q*-value = 0.05 expressed in a $-\log_{10}$ scale. **1c.** GenTrain score distribution for 42,272 SNPs genotyped in 305 goats. It can be seen that the vast majority of SNPs have GenTrain scores above 0.70.



Acknowledgments

Many thanks to CAPRIGRAN for carrying out phenotype recording and blood sample collection in Murciano-Granadina goats. This research was funded by the European Regional Development Fund (FEDER)/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación/Project Reference grant: PID2019-105805RB-I00 and by the CERCA Programme/Generalitat de Catalunya. We also acknowledge the support of the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2020-2023 (CEX2019-000902-S) grant awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). We also acknowledge the support of the CERCA programme of the Generalitat de Catalunya. Maria Luigi-Sierra was fundedwith a PhD fellowship "Formación de Personal Investigador" (BES-C-2017-079709) awarded by the Spanish Ministry of Economy and Competitivity.

Availability of data

Genotypes of the 305 Murciano-Granadina sires and offspring and pedigree informationare available in 10.6084/m9.figshare.14686230.

References

Abdalla E.A., Id-Lahoucine S., Cánovas A., Casellas J., Schenkel F.S., Wood B.J. &Baes C.F. (2020). Discovering lethal alleles across the turkey genome using a transmission ratio distortion approach. *Animal Genetics* **51**, 876–89. https://doi.org/10.1111/age.13003

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate: A practicaland powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289–300. https://doi.org/10.2307/2346101

Bickhart D.M., Rosen B.D., Koren S., Sayre B.L., Hastie A.R., Chan S., Lee J., Lam E.T., Liachko I., Sullivan S.T., Burton J.N., Huson H.J., Nystrom J.C., Kelley C.M., Hutchison J.L., Zhou Y., Sun J., Crisà A., Ponce De León F.A., Schwartz J.C., Hammond J.A., Waldbieser G.C., Schroeder S.G., Liu G.E., Dunham M.J., Shendure J., Sonstegard T.S., Phillippy A.M., Van Tassell C.P. & Smith T.P.L. (2017) Single-molecule sequencing and chromatin conformation capture enablede novo reference assembly of the domestic goat genome. *Nature Genetics* **49**, 643–50. https://doi.org/10.1038/ng.3802.

Casellas J., Manunza A., Mercade, A., Quintanilla R. & Amills, M. (2014) A flexible bayesian model for testing for transmission ratio distortion. *Genetics* **198**, 1357-67. https://doi.org/10.1534/genetics.114.169607

Casellas J., Cañas-Álvarez J.J., González-Rodríguez A., Puig-Oliveras A., Fina M., Piedrafita J., Molina A., Díaz C., Baró J.A. & Varona L. (2017) Bayesian analysis of parent-specific transmission ratio distortion in seven Spanish beefcattle breeds. *Animal Genetics* **48**, 93-96 . https://doi.org/10.1111/age.12509

Casellas J., Id-Lahoucine S. & Cánovas A. (2020) Discriminating between allele- and

genotype-specific transmission ratio distortion. *Animal Genetics* **51**, 847-54. https://doi.org/10.1111/age.13007

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M. & Lee J.J. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets.*Gigascience* **4**, 7. https://doi.org/10.1186/s13742-015-0047-8

Fishman L. & Mcintosh M. (2019) Standard deviations: The biological bases of transmission ratio distortion. *Annual Review of Genetics* **53**, 347–72. https://doi.org/10.1146/annurev-genet-112618-043905

Gòdia M., Casellas J., Ruiz-Herrera A., Rodríguez-Gil J.E., Castelló A., Sánchez A. &Clop A. (2020) Whole genome sequencing identifies allelic ratio distortion in

sperm involving genes related to spermatogenesis in a swine model. *DNA Research* **27**, dsaa019. https://doi.org/10.1093/dnares/dsaa019

Guan D., Landi V., Luigi-Sierra M.G., Delgado J.V., Such X., Castelló A., Cabrera B., Mármol-Sánchez E., Fernández-Alvarez J., de la Torre Casañas J.L.R., Martínez A., Jordana J. & Amills M. (2020) Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. *Journal of Animal Science and Biotechnology* **11**, 35. https://doi.org/10.1186/s40104-020-00435-4

Guo Y., He J., Zhao S., Wu H., Zhong X., Sheng Q., Samuels D.C., Shyr Y. & Long J.(2014) Illumina human exome genotyping array clustering and quality control. *Nature Protocols* **9**, 2643-62. https://doi.org/10.1038/nprot.2014.174

Hosking L., Lumsden S., Lewis K., Yeo A., McCarthy L., Bansal A., Riley J., Purvis I.& Xu C.F. (2004) Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *European Journal of Human Genetics* **12**, 395-9. https://doi.org/10.1038/sj.ejhg.5201164

Huang L.O., A. Labbe, & C. Infante-Rivard (2013) Transmission ratio distortion: Review of concept and implications for genetic association studies. *HumanGenetics* **132**, 245–63. https://doi:10.1007/s00439-012-1257-0.

Lahoucine S.I., Casellas J., Lu D., Sargolzaei M., Miller S.P. & Canovas A. (2020) Distortion of Mendelian segregation across Angus cattle genome reveal novellethal haplotype affecting reproduction. *Journal of Animal Science* **98**, 5. https://doi.org/10.1093/jas/skaa278.009

Nelson W.A. (2008) Statistical Methods. In: *Encyclopedia of Ecology* (ed. by S.E.Jørgensen & B.D. Fath), pp. 3350–62. Academic Press, Oxford. https://doi.org/https://doi.org/10.1016/B978-008045405-4.00661-3

Pavy N., Pelgas B., Beauseigle S., Blais S., Gagnon F., Gosselin I., Lamothe M., Isabel N. & Bousquet J. (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: Application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* **9**,21. https://doi.org/10.1186/1471-2164-9-21

Vázquez-Gómez M., de Hijas-Villalba M.M., Varona L., Ibañez-Escriche N., Rosas J.P., Negro S., Noguera J.L. & Casellas, J. (2020) Maternal transmission ratio

distortion in two iberian pig varieties. *Genes* **11**, 1050. https://doi.org/10.3390/genes11091050

Zhao S., Jing W., Samuels D.C., Sheng Q., Shyr Y. & Guo Y. (2018) Strategies for processing and quality control of Illumina genotyping arrays. *Briefings in Bioinformatics* **19**, 765-75. https://doi.org/10.1093/bib/bbx012

Supplementary tables

Table S1. Family size for each one of the 17 Murciano-Granadina sires.

Table S2. Single nucleotide polymorphisms displaying transmission ratio distortion (α-value above 0.15 or below −0.15, q-value < 0.05) in a population composed by 17 families of Murciano-Granadina goats.

Supplementary figures

Fig. S1. GenoPlots of SNPs with different GenTrain scores (GT-scores). It can be seen that when the GT-score is high (0.90), the clustering of each of the three genotypes is quite tight. In contrast, SNPs with GT scores of 0.55 and 0.62 display scattered patterns of clustering for at least one of the genotypes. For the SNP with a GT score of 0.17, the pattern of clustering is not credible.

RUNNING HEAD: Brain expression in goats

A protein-coding gene expression atlas from the brain of pregnant and non-pregnant goats

María Gracia Luigi-Sierra[1], Dailu Guan[1], Manel López-Béjar[2], Encarna Casas[2], Sergi Olvera[2], Jaume Gardela[2], María Jesús Palomo[3], Uchebuchi Ike Osuagwuh[3], Uchechi Linda Ohaneje[3], Emilio Mármol-Sánchez[1], Marcel Amills[1,4*]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra, Spain.
[2]Department of Health and Animal Anatomy, Universitat Autònoma de Barcelona, Bellaterra, Spain.
[3]Department of Animal Medicine and Surgery, Universitat Autònoma de Barcelona, Bellaterra, Spain.
[4]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain.

Corresponding author: *.

In preparation.

Abstract

We have sequenced the transcriptomes of 12 brain tissues in seven female Murciano-Granadina goats, three of which were pregnant, to build an atlas of protein-coding gene expression of the goat brain. Between 14,889 (cerebellar hemisphere) and 15,592 (pineal gland) protein-coding genes were expressed in goat brain tissues, and most of them displayed ubiquitous or broad patterns of expression across tissues. Principal component analysis and hierarchical clustering based on the patterns of mRNA expression revealed that brain tissues tend to group according to their embryonic origin. Differential expression analysis between pregnant and non-pregnant goats evidenced moderate changes of mRNA expression in the frontal cortex, hippocampus, adenohypophysis, pons and pineal gland, and very dramatic changes in the olfactory bulb. Many genes showing differential expression in this organ are related to olfactory function and behavior in humans.

Keywords: Goat, RNA-Seq, gestation, embryonic vesicle, differential gene expression, encephalon.

Introduction

The mammalian brain is an extraordinarily complex organ integrating multiple highly specialized structures involved in the regulation of memory, behavior, learning, sensory function, motor skills and body homeostasis, amongst others. During embryo development, three vesicles emerge from the neural tube which roughly correspond to the forebrain (prosencephalon), midbrain (mesencephalon) and hindbrain (rhombencephalon) [1]. Further subdivisions take place in the forebrain (telencephalon and diencephalon) and hindbrain (metencephalon and myelencephalon), resulting in the formation of five encephalic vesicles [1]. In the bovine embryo, these five encephalic vesicles are visible at 24 days after conception, but they do not become fully developed until 110 days of gestation [2]. Multiple anatomical structures with specialized functions [1] are subsequently derived from the telencephalon (cerebral hemisphere, basal ganglia, hippocampus, olfactory bulb, lateral ventricles, etc.), diencephalon (thalamus, hypothalamus, pineal body, neurohypophysis, infundibulum, third ventricle, etc.), mesencephalon (rostral colliculus, tegmentum, crus cerebri, cerebral aqueduct, etc.), metencephalon (pons, cerebellum, upper part of the fourth ventricle, etc.) and myelencephalon (medulla oblongata, spinal cord lower part of the fourth ventricle, etc.).

Transcriptomic analyses may hold the key to significantly advance our knowledge about the biological functions of brain regions. In mice, patterns of gene expression have been used to establish a molecular atlas of the adult brain [3], and as much as

737 brain structures have been identified with microscopy techniques complemented with other approaches, providing a comprehensive view about the high functional complexity of this organ [4]. An important feature of the brain is its high plasticity, so any atlas of gene expression is necessarily dynamic, not only in space but also in time. For instance, pregnancy in mice is associated with extensive changes in gene expression in the neocortex, cerebellum, hippocampus and hypothalamus, and there is evidence that several of such modifications might be long lasting [5,6].

Very few atlases of brain gene expression have been generated in domestic animals. Recently, Sjöstedt et al. [7] investigated the profiles of gene expression of 10 major mammalian brain regions in humans, mice and pigs. They found that global transcriptomic profiles are, in general, well conserved in these three species, with cerebrum and brainstem regions clustering apart in hierarchical trees and the cerebellum showing a highly divergent profile of gene expression [7]. The mRNA expression of several brain regions has also been reported in sheep [8] and cattle [9], although not in a comprehensive or systematic way. In goats, mRNA expression of two neural tissues, frontal lobe cortex and cerebellum, has also been characterized by RNA-Seq [10]. The main goal of the current work was to establish an atlas of protein-coding gene expression of the caprine brain by sequencing the transcriptomes of 12 encephalic regions in 7 female Murciano-Granadina goats. Given that 3 of these goats were pregnant at the time of slaughter, we have also investigated whether gestation affects the transcriptomic profiles of the 12 brain regions under study.

Results

Sequencing of total RNA from 12 goat tissues

In this work, we have sequenced 84 RNA samples obtained from 12 different brain tissues retrieved from seven Murciano-Granadina goats, i.e. adenohypophysis, cerebellar hemisphere, cerebellar trunk, frontal neocortex, hippocampus, hypothalamus, medulla oblongata, neurohypophysis, olfactory bulb, pineal gland, pons and rostral colliculus (**Fig 1A**). Three of the sampled goats were 1-month pregnant while the other four goats were non-pregnant. The average RNA integrity number (RIN) of the 84 RNA extractions was 7.53 ± 0.62, ranging from 6.2 to 9 (**Table 1**). These values are in agreement with the conservative threshold of RNA degradation (RIN = 6.4-7.9) defined by Gallego Romero et al. [11] for the in-depth analysis of RNA transcripts. As shown in **Table 1**, the mean sequencing depth ranged from 37.79 to 42.12 million of paired-end reads per tissue, and the average alignment rate ranged from 84.44% (adenohypophysis) to 94.35% (pons).

Fig 1. Anatomical location and PCA clustering of 12 goat brain tissues. (**A**) Anatomical location of 12 brain tissues sampled in seven Murciano-Granadina female goats: adenohypophysis, cerebellar hemisphere, cerebellar trunk, frontal neocortex, hippocampus, hypothalamus, medulla oblongata, neurohypophysis, pineal gland, pons, olfactory bulb and rostral colliculus. The embryonic vesicle, i.e. telencephalon and diencephalon in the forebrain, mesencephalon in the midbrain and metencephalon and myelencephalon in the hindbrain, from which each sampled tissue is derived is indicated with different colors.

Adenohypophysis does not originate from any of these five vesicles but from the oral ectoderm. (**B**) Principal component analysis (PCA) of 12 brain tissues from 7 Murciano-Granadina female goats by considering the mRNA expression patterns of all sampled tissues (adenohypophysis, cerebellar hemisphere, cerebellar trunk, frontal neocortex, hippocampus, hypothalamus, medulla oblongata, neurohypophysis, pineal gland, pons, olfactory bulb and rostral colliculus), (**C**) Same PCA excluding hypophysis and cerebellum, and (**D**) Same PCA excluding pineal gland, hypophysis and cerebellum.

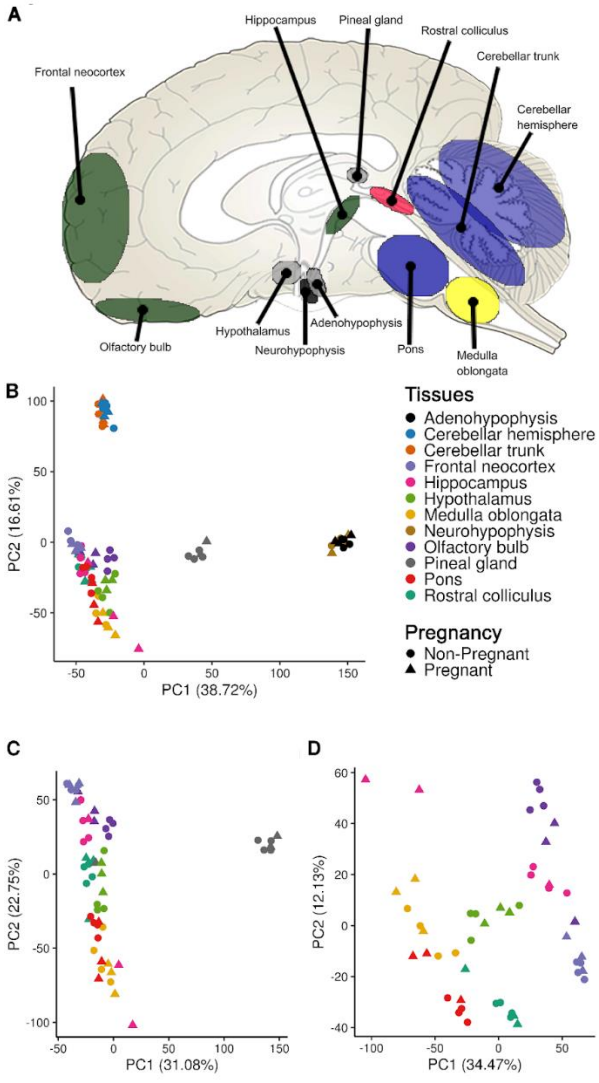


Table 1. Quality of RNA samples and number of genes expressed in 12 brain goat tissues. Average RNA integrity number (RIN), alignment rate and number of expressed genes (CPM > 0.5 in at least 2 samples per tissue) in 12 brain tissues from seven Murciano-Granadina female goats.

Tissue	RIN Mean ± SD	Average alignment rate (%)	Average number of reads (millions) Mean ± SD	Number of expressed protein-coding genes (CPM > 0.5)
Adenohypophysis	7.900 ± 0.428	84.44	37.786 ± 2.348	15,220
Cerebellar hemisphere	8.100 ± 0.486	94.11	41.461 ± 4.571	14,889
Cerebellar trunk	8.143 ± 0.550	94.17	42.117 ± 9.052	14,898
Frontal neocortex	7.400 ± 0.569	94.26	41.361 ± 7.438	15,098
Hippocampus	7.629 ± 0.411	93.98	41.717 ± 7.825	15,437
Hypothalamus	6.943 ± 0.326	93.92	38.618 ± 4.817	15,557
Medulla oblongata	6.843 ± 0.181	94.27	40.141 ± 5.510	15,420
Neurohypophysis	7.771 ± 0.482	87.8	39.443 ± 5.931	15,366
Olfactory bulb	7.214 ± 0.422	94.16	40.139 ± 3.938	15,581
Pineal gland	8.029 ± 0.594	94.23	39.228 ± 7.142	15,592
Pons	7.186 ± 0.654	94.35	39.505 ± 4.233	15,273
Rostral colliculus	7.243 ± 0.412	93.63	38.832 ± 3.179	15,153

Clustering of tissues according to their mRNA expression profile

The total number of protein-coding loci expressed (counts per million, CPM > 0.5 in at least two samples per tissue) in 12 brain tissues from 7 goats ranged

from 14,889 genes for the cerebellar hemisphere to 15,592 genes for the pineal gland. Considering all tissues, 17,054 protein-coding genes passed the quality control (QC) filters and were expressed in at least two samples for each tissue. The principal component analysis (PCA) showed that the majority of samples clustered according to their tissue of origin. We also observed that four tissues, namely adenohypophysis, neurohypophysis, pineal gland and cerebellum (hemisphere and trunk), displayed highly divergent gene expression patterns with regard to the remaining encephalic tissues (**Figs 1B** and **1C**). The removal of these four outlier tissues demonstrated, particularly for the first principal component (PC1, 34.47% of the observed variance), that tissues group in accordance with the embryonic vesicle from which they are derived (**Fig 1D**). Pons and medulla oblongata, which are hindbrain structures, tended to group together and in close proximity with the rostral colliculus, which is derived from the midbrain. Another cluster of tissues was represented by the frontal neocortex, olfactory bulb and hippocampus, which derive from the forebrain telencephalon vesicle. Finally, the hypothalamus, which originates from the forebrain diencephalon vesicle, was located in an intermediate position between these two major clusters (i.e. forebrain vs. midbrain/hindbrain) of brain tissues (**Fig 1D**). Both the hierarchical clustering analysis (**Fig 2**) and the heatmap (**Fig 3**) supported these observations, highlighting that tissue of origin, rather than pregnancy status, was the main factor governing the clustering of samples (**Fig 2**).

Fig 2. Hierarchical clustering of 12 brain tissues from 7 Murciano-Granadina goats based on their mRNA expression profiles. Dendrogram displaying tissue clustering patterns. In most cases, samples group according to their tissue of origin and, moreover, tissues tend to cluster according to the forebrain, midbrain or hindbrain embryonic vesicle they originate from. Exceptions to this general observation are hypothalamus, that groups with rostral colliculus, and two main groups of outliers: cerebellum and glandular tissues (pineal gland and hypophysis). In the right part of the Figure, we indicate the embryonic vesicle from which each tissue is derived. In the embryo, three forebrain (prosencephalon), midbrain (mesencephalon) and hindbrain (rhombencephalon) vesicles are initially formed. Later on, prosencephalon is subdivided into two further vesicles (telencephalon and diencephalon) and so does rhombencephalon (metencephalon and myelencephalon). Adenohypophysis does not originate from any of these five vesicles but from the oral ectoderm.

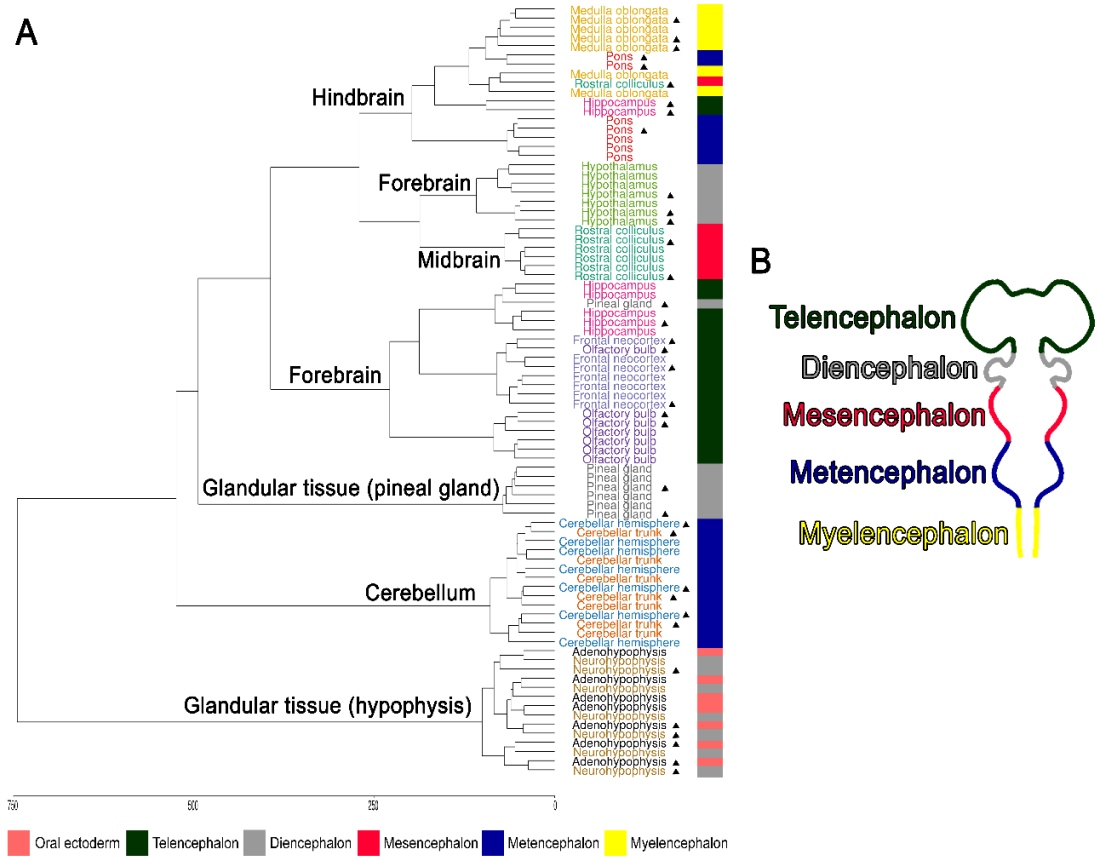
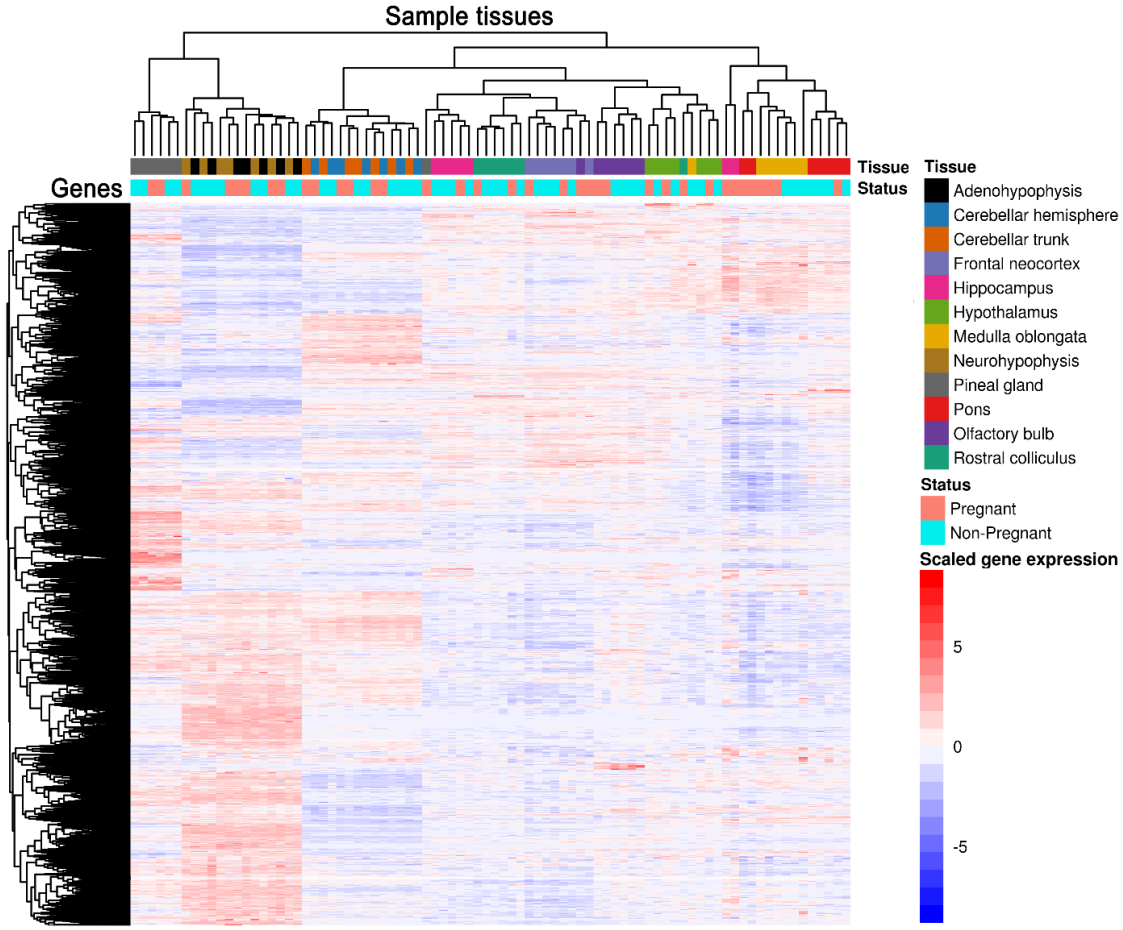

A
Hindbrain
Forebrain
Midbrain
Forebrain
Glandular tissue (pineal gland)
Cerebellum
Glandular tissue (hypophysis)
750
500
250
0
Oral ectoderm
Telencephalon
Diencephalon
Mesencephalon
Metencephalon
Myelencephalon
B
Telencephalon
Diencephalon
Mesencephalon
Metencephalon
Myelencephalon

Fig 3. Heatmap depicting the scaled gene expression of protein-coding loci transcribed in 12 goat brain tissues. The 84 columns correspond to samples from 12 brain tissues retrieved from 3 pregnant and 4 non-pregnant Murciano-Granadina goats, while rows correspond to a set of 17,054 protein-coding genes. It can be seen that tissue of origin, rather than pregnancy, is the main factor determining the degree of similarity between genome-wide mRNA expression patterns across samples.



Tissue-specificity of mRNA expression

Regarding gene expression specificity, the distribution of the tissue-specificity index τ-values was highly skewed to the left (**Fig 4A**). Indeed, 5,826 (34.16%)

protein-coding genes showed τ- values below 0.15, indicative of a highly ubiquitous pattern of expression across all tissues. In contrast, 9,877 (57.92%) protein-coding genes displayed intermediate τ-values (τ = 0.15-0.85) and 1,351 protein-coding genes (7.92%) had a highly tissue-specific profile (τ > 0.85) of mRNA expression (**Figs 4A** and **4B**). More specifically, 469 protein-coding genes showed τ-values of 1, indicating that they are expressed in only one of the sampled tissues (**Figs 4A** and **4C**). From this set of 469 protein-coding genes, 177 were specifically expressed in the pineal gland, evidencing that this tissue is, by far, the one with the highest number of genes with tissue-specific expression (**Fig 4C**). In contrast, the rostral colliculus was the tissue with the lowest number of tissue-specific genes (11 protein-coding genes with τ =1).

KEGG pathway enrichment analysis of highly expressed genes

The results of the Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis for the 1,000 mRNA genes with the highest expression in each tissue (S3A-S3K Tables) revealed several pathways broadly shared across tissues and also reaching strong statistical significance, e.g. synaptic vesicle cycle, endocrine and other factor-regulated calcium reabsorption, endocytosis, phagosome, long-term potentiation, dopaminergic synapse, glutamatergic synapse, circadian entrainment, gap junction, gastric acid secretion and adrenergic signalling in cardiomyocytes (Fig 5). We also detected many pathways related to neurological conditions that attained high statistical significance and were represented in a broad array of encephalic tissues, e.g. Alzheimer disease, Huntington disease, Parkinson

disease, prion disease, amyotrophic lateral sclerosis, and, to a much lesser extent, spinocerebellar ataxia. Pathways related with Vibrio cholerae and Salmonella infection were also commonly found. Protein processing in endoplasmic reticulum, RNA transport and ribosome were amongst the most significant pathways in the adenohypophysis (S3A Table) and neurohypophysis (S3H Table), but not in the other tissues. As said before, in the pineal gland (S3J Table) phototransduction was also a highly significant pathway (q-value = 0.004), while in the remaining tissues it did not reach statistical significance.

Fig 4. Tissue specificity of protein-coding genes expressed in 12 goat brain tissues. (**A**) Histogram representing the tissue specificity of 17,054 protein-coding genes expressed in 12 goat brain tissues (in at least one sample) and surpassing QC filters. The number of genes is indicated in the *y*-axis, while Tau specificity scores (τ, see text for details) are shown in the *x*-axis. (**B**) Number of genes with tissue-specific expression ($\tau > 0.85$) in each of the 12 brain tissues under study. (**C**) Number of genes expressed exclusively in one tissue ($\tau = 1$) in 12 caprine brain regions. Tissues were retrieved from 7 Murciano-Granadina goats.

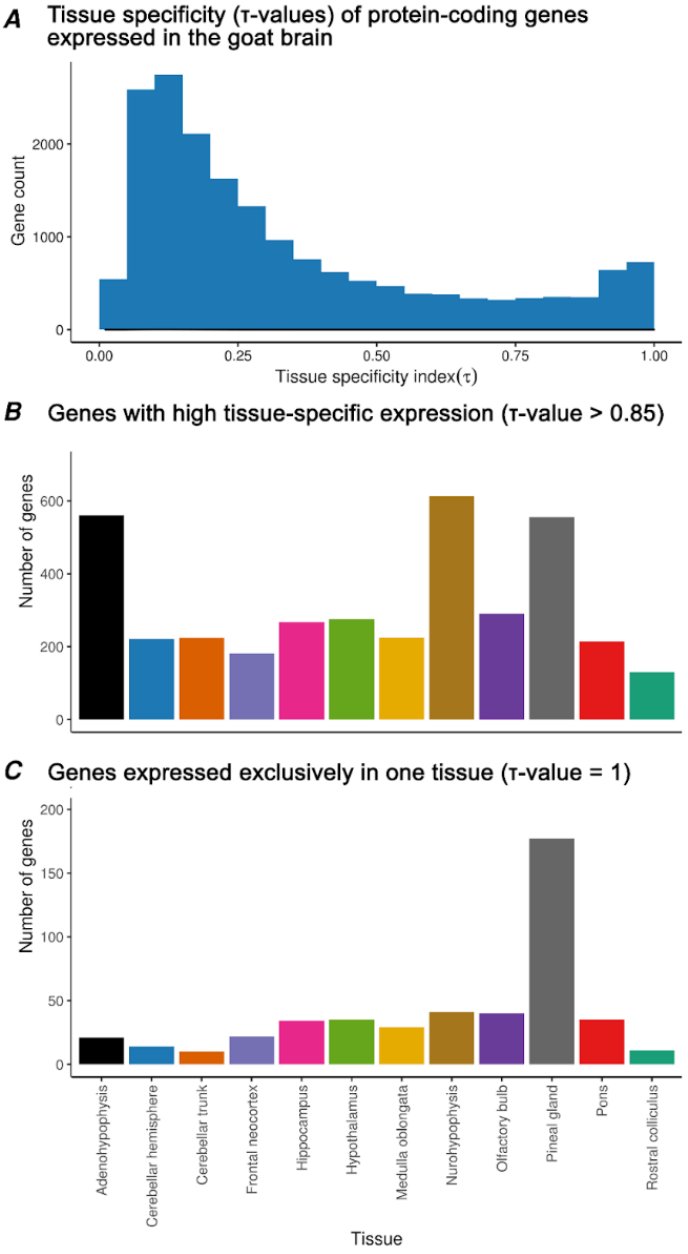
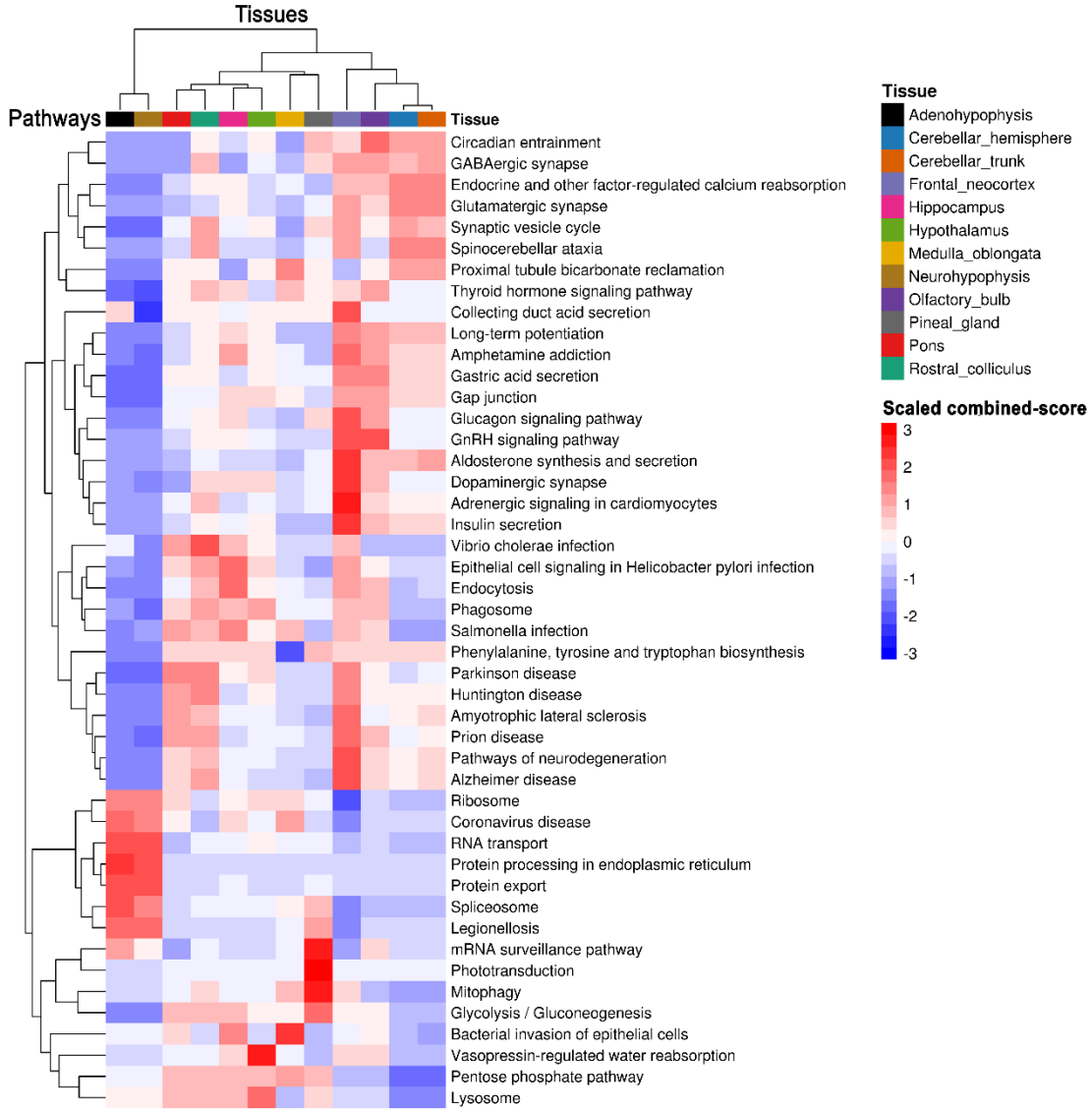

A
Tissue specificity (τ-values) of protein-coding genes expressed in the goat brain
Gene count
2000
1000
0
0.00
0.25
0.50
0.75
1.00
Tissue specificity index(τ)
B
Genes with high tissue-specific expression (τ-value > 0.85)
Number of genes
600
400
200
0
C
Genes expressed exclusively in one tissue (τ-value = 1)
Number of genes
200
150
100
50
0
Adenohypophysis
Cerebellar hemisphere
Cerebellar trunk
Frontal neocortex
Hippocampus
Hypothalamus
Medulla oblongata
Nurohypophysis
Olfactory bulb
Pineal gland
Pons
Rostral colliculus
Tissue

Fig 5. Heatmap depicting the enrichment level of the top 20 enriched pathways in 12 goat brain tissues from seven goats. The top 20 most enriched pathways per tissue were selected. Many pathways were shared across tissues, making a total of 48 unique pathways. The level of expression is displayed as combined scores calculated with the Enrichr software as defined by Chen et al [12].



Changes of brain expression associated with 1-month pregnancy

We were interested to find out which of the brain tissues are more affected by pregnancy. To achieve this goal, we have used the number of differentially expressed genes (DEGs) as an indicator of the strength of such influence. Although low replicate numbers (3 pregnant and 4 non-pregnant goats) decreases the sensitivity to detect DEGs, we consider that such reduction of sensitivity should affect all 12 tissues in a similar manner so it is not expected to bias our results. Indeed, brain tissues from pregnant versus non-pregnant goats showed strong differences in terms of the number of differentially expressed genes (DEGs). Six tissues displayed little changes in their expression levels in response to 1 month-pregnancy: in the cerebellar hemisphere, cerebellar trunk, hypothalamus, medulla oblongata, neurohypophysis and rostral colliculus, only 2, 1, 1, 12, 4 and 1 DEGs were identified, respectively (**Fig 6**). In strong contrast, we observed remarkable changes in the expression profiles of six brain tissues (**Fig 7, S4 Table**): adenohypophysis (201 DEGs, 13 downregulated and 188 upregulated in pregnant goats, **S4A Table**), frontal neocortex (82 DEGs, 37 downregulated and 45 upregulated in pregnant goats, **S4D Table**), hippocampus (70 DEGs, 15 downregulated and 55 upregulated in pregnant goats, **S4E Table**), pineal gland (62 DEGs, 60 downregulated and 2 upregulated in pregnant goats, **S4J Table**), pons (190 DEGs, 57 downregulated and 133 upregulated in pregnant goats, **S4K Table**) and, most remarkably, olfactory bulb (1207 DEGs, 381 downregulated and 826 upregulated in pregnant goats, **S4I Table**).

Pathways significantly associated with DEGs in each tissue are shown in **S5A-S5F Tables** and **S13-S18 Figs**. In the olfactory bulb (**S5D Table**), the majority of biological functions enriched in the set of upregulated DEGs in pregnant goats were associated with axon guidance (*c* = 89.246; *q*-value = 7.29 E-07), dopaminergic synapse (*c* = 79.526; *q*-value = 6.07 E-06), neuroactive ligand-receptor interaction (*c* = 53.217; *q*-value = 5.27 E-06), cholinergic synapse (*c* = 55.712; *q*-value = 1.32 E-04) and phenylalanine, tyrosine and tryptophan biosynthesis (*c* = 53.087; *q*-value = 8.51 E-02). In contrast, downregulated DEGs in the olfactory bulb of pregnant goats were enriched in pathways related to nervous development and control of cellular activities, including migration, differentiation and proliferation (**S5D Table**). Examples of such pathways are calcium signalling (*c* = 22.865; *q*-value = 1.66 E-1), ErbB signalling (*c* = 15.898; *q*-value = 3.32 E-1), ECM-receptor interaction (*c* = 14.828; *q*-value = 3.32 E-1) and tight junction (*c* = 14.492; *q*-value = 2.81 E-1).

By making a literature search for each one of the genes showing differential expression in the olfactory bulb of pregnant vs. non-pregnant goats, we observed that many of them are related to human behavioral traits (**S6 Table).** As shown in **S7 Table**, our literature search also revealed several DEGs associated with the migration and maturation of olfactory bulb interneurons, olfactory bulb morphogenesis and establishment of a functional olfactory neural circuitry.

Fig 6. Volcano plots of differentially expressed genes in six brain tissues little affected by pregnancy. According to our results, when comparing pregnant vs. non-pregnant goats, these six tissues are unaffected or little affected by 1 month-pregnancy: (**A**) cerebellar hemisphere, (**B**) cerebellar trunk, (**C**) hypothalamus, (**D**) medulla oblongata, (**E**) neurohypophysis and (**F**) rostral colliculus. Genes with a fold change below -1.5 and a *q*-value < 0.05 are depicted in blue, while genes with a fold change above 1.5 and a *q*-value < 0.05 are depicted in red. Grey dots represent genes that do not display differential expression (absolute fold change below 1.5 and/or a *q*-value > 0.05).

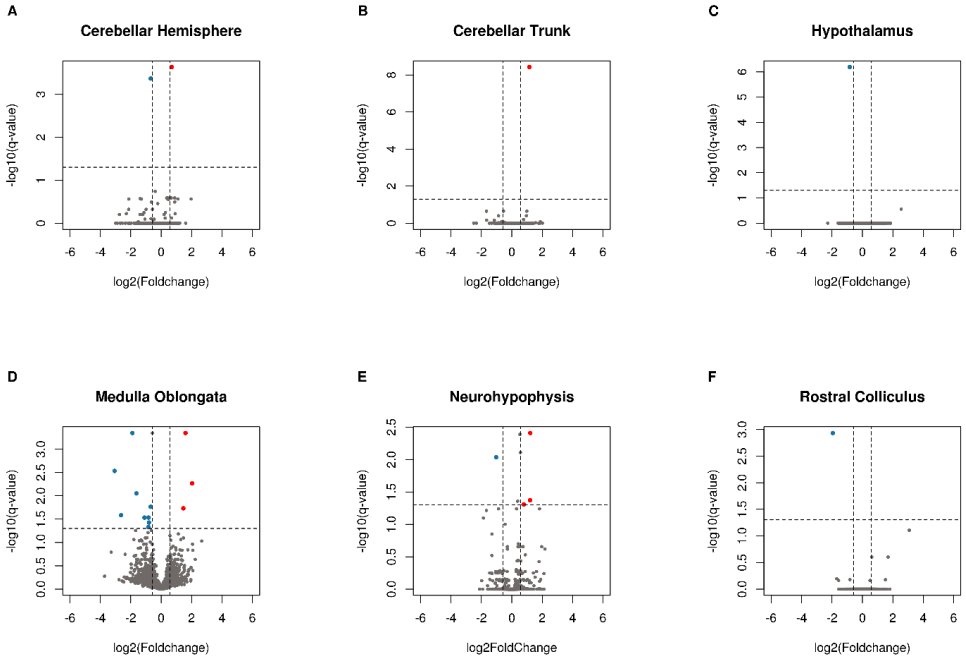
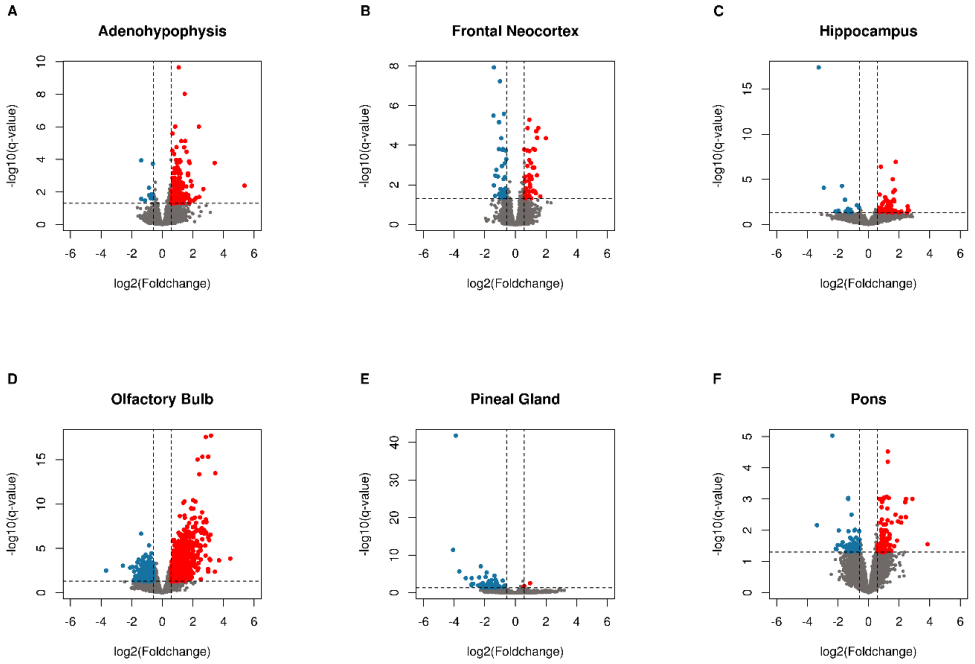


Fig 7. Volcano plots of differentially expressed genes in six brain tissues moderately or strongly affected by pregnancy. According to our results, when comparing pregnant vs. non-pregnant goats, these six tissues are moderately or strongly affected by 1 month-pregnancy: (**A**) adenohypophysis, (**B**) frontal neocortex, (**C**) hippocampus, (**D**) olfactory bulb, (**E**) pineal gland and (**F**) pons. Genes with a fold change below -1.5 and a *q*-value < 0.05 are depicted in blue, while genes with a fold change above 1.5 and a *q*-value < 0.05 are depicted in red. Grey dots represent genes that do not display differential expression (absolute fold change below 1.5 and/or a *q*-value > 0.05).



Discussion

Embryogenesis leaves a durable footprint on the patterns of brain mRNA expression

The majority of the 17,054 protein-coding genes expressed in the goat brain had τ-values below 0.25, implying that they have a ubiquitous pattern of expression. Indeed, only 469 genes (2.75%) were expressed in just one of the 12 sampled tissues (τ = 1). Results obtained in the Genotype-Tissue Expression (GTEx) project were consistent with this finding [13], since only 200 genes showed tissue-specific expression (95% of these genes were exclusively expressed in the testis).

Tissue of origin, rather than pregnancy, was the major factor explaining the clustering of samples. By using microarrays, Ferraz et al. [14] investigated the profiles of expression of 16 porcine tissues and observed that the factor "tissue of origin" accounted for ~11 times more variability than sex or breed. In humans, the GTEx Consortium retrieved 1641 post-mortem samples covering 54 body sites from 175 individuals and reported that tissue type was the primary factor explaining differences in gene expression, a result that is fully consistent with ours [15]. Similar findings have been obtained when building atlases of gene expression in cattle [9], pigs [16] and sheep [8].

We have observed that, in goats, the majority of brain tissues cluster according to their embryonic origin, with the few exceptions of cerebellum, pineal gland and adeno/neurohypophysis which showed highly differentiated patterns of gene expression (**Fig 1B**, **1C** and **1D**). In the comprehensive analysis of porcine tissue expression carried out by Ferraz et al. [14], tissues clustered according to the germ

layer (ectoderm, mesoderm or endoderm) they derive from. Measurement of transcriptome profiles in 24 murine neural tissues also highlighted the existence of a relationship between the cellular position along the anterior-posterior axis of the neural tube and gene expression in different regions from the adult brain [17], as also published by Ortiz et al. [3]. These results and those obtained by us are in agreement with the interpretation that embryogenesis leaves a durable footprint in the profile of mRNA expression of mammalian brain tissues.

Sjöstedt et al. [7] described the patterns of genome-wide expression of protein-coding genes in a number of brain regions from humans, mice and pigs. They found that, in general, transcriptomic patterns are evolutionarily conserved in these three species, with the three forebrain structures (cerebral cortex, hippocampus, and amygdala) grouping together, and the midbrain, thalamus, and pons and medulla forming another cluster of tissues close to the hypothalamus [7]. They also found that the cerebellum behaves as an outlier, a finding fully coherent with ours.

Although transcriptomic profiles of brain regions seem to be well conserved across species, relevant differences also exist. For instance, Sjöstedt et al. [7] indicated that in humans the olfactory bulb clusters with other forebrain structures (cerebral cortex, hippocampus, amygdala, etc.), while in pigs and mice this tissue almost behaves as an outlier. They reasoned that this might be due to the fact that these two latter species have more evolved olfactory systems than humans [7]. In goats, the olfactory bulb is also much more developed than in humans [18], a feature consistent with the fundamental role of the olfactory system in the maternal and social behaviors of ungulates [19]. However, our data indicate that in goats the

pattern of mRNA expression of the olfactory bulb is closely aligned with that of other forebrain tissues (**Fig 1B-1D**). This result agrees much more with the PCA reported by Sjöstedt et al. [7] for human brain tissues than for those corresponding to pigs or mice. A higher resemblance of goats to humans, rather than to pigs, is unexpected because goats and pigs are ungulates [20]. Sjöstedt et al. [7] stated that the functional importance of olfaction might be a key factor explaining the differential patterns of olfactory bulb mRNA expression observed in pigs/mice vs. humans. Such hypothesis is not fully consistent with our results and should be interpreted cautiously until further data are available in other mammalian species.

Cerebellum, pineal gland and hypophysis have highly differentiated patterns of gene expression

Cerebellum displayed a highly differentiated pattern of mRNA expression when compared to other brain structures (**Fig 1B**), a finding that, as mentioned before, is consistent with previous reports [3,7]. Moreover, cerebellum trunk and hemisphere mRNA profiles were quite similar. Cerebellum is strongly specialized in the learning and coordination of motor activities as well as in the triggering of reflex responses, and it has been suggested that it is scarcely connected with cognitive areas of the brain cortex [21]. One of the most distinctive features of the cerebellum is its extraordinarily high cell density: while this organ represents 10% of brain volume, it encompasses 42.0% of all brain cells and 59.8% of all excitatory neurons, mainly due to the conspicuous abundance of tightly packed granular cells [4]. Another relevant feature of the cerebellum is that the same circuit, composed by mossy fibers that excite granule cells that, in turn, excite Purkinje cells, forms a

fundamental unit that is replicated thousands of times [22]. These biological particularities might have contributed to transform the cerebellum into an organ with a very specific transcriptomic profile, not only in goats but also in other mammals.

We also observed highly divergent patterns of gene expression in three glandular tissues: the pineal gland and the adeno/neurohypophysis (**Fig 1B**), a finding consistent with results reported by Harhay et al. [9]. The strong functional specialization of these three anatomical structures in hormonal secretion might explain these findings. In response to light, the pineal gland, which develops from an evagination of neuroepithelium in the dorsal midline of the diencephalon, synthesizes and releases melatonin, which is a key regulator of the circadian sleep-wake cycle and seasonal rhythms [23]. Remarkably, in our study the pineal gland was the second organ showing the highest number of genes with tissue-specific expression ($\tau > 0.85$, **Fig 4B** and **S2A Table**), and presented the highest number of genes exclusively expressed in one tissue (**Fig 4C**). Neurohypophysis and adenohypophysis are also specialized in the secretion of molecules with key physiological roles, and data collected in sheep and cattle indicate that their profiles of mRNA expression are highly differentiated from those of other brain structures [8,9]. The main hormones produced by the adenohypophysis are prolactin, adrenocorticotropic hormone, luteinizing hormone, follicle-stimulating hormone, growth hormone and thyroid-stimulating hormone, which regulate a very diverse set of biological processes including growth, metabolism, lactation, stress and reproduction [24]. One particular feature of the adenohypophysis is that it does not

develop from any of the five neural vesicles but from the oral ectoderm [24]. In contrast, the neurohypophysis has a diencephalic origin and stores vasopressin and oxytocin (both are synthesized in the hypothalamus) regulating diuresis and a broad array of reproduction and behavioral processes, respectively [25]. Despite having completely different embryological origins, histological structure and biological functions, our data indicate that neurohypophysis and adenohypophysis share similar profiles of mRNA expression (**Fig 1B** and **Fig 5**). Our interpretation is that high functional specialization in hormonal secretion might erase, at least partially, the transcriptomic footprint associated with embryogenesis, a hypothesis supported by the distinctive patterns of gene expression observed in both pituitary structures when compared with the remainder encephalic tissues. Further studies will be needed to assess whether such hypothesis is correct.

Heterogeneous effects of pregnancy on the expression profiles of twelve goat brain regions

Amongst the seven individuals sampled in our study, there were three pregnant goats providing the opportunity to investigate the effect of pregnancy on the brain transcriptome. Since the number of replicates for each pregnant (N=3) and non-pregnant (N=4) category is low, sensitivity to detect DEGs is expected to decrease [26]. To circumvent this difficulty, we have used the total number of DEGs as the main criterion to identify which brain regions are mostly affected by 1-month pregnancy. In principle, low number of replicates should affect the number of detected DEGs to a similar extent in all 12 brain tissues, so the number of DEGs

seems to be an appropriate indicator of which brain tissues are more affected by pregnancy.

In comparisons involving cerebellar hemisphere, cerebellar trunk, hypothalamus, medulla oblongata, neurohypophysis and rostral colliculus, the total number of DEGs in pregnant vs non-pregnant goats was very low or inexistent (**Fig 6**). In contrast, in the frontal cortex, hippocampus, adenohypophysis, pons and pineal gland, between 62 and 201 DEGs were detected (**Fig 7**). By far, the organ which displayed the largest number of DEGs was the olfactory bulb (826 upregulated and 381 downregulated genes in pregnant goats). From these data, we conclude that 1-month pregnancy does not have the same effect on all goat brain regions. Ray et al. [5] investigated changes in the mRNA expression of four brain structures (hypothalamus, neocortex, hippocampus and cerebellum) in virgin, pregnant, and postpartum mice, and they found that in the virgin vs. pregnant comparison the number of DEGs was much higher in the hippocampus than in the cerebellum, while hypothalamus and neocortex showed intermediate values. These results support the notion that pregnancy does not affect all brain tissues to the same extent. Brain gene expression during pregnancy is dynamic and it might change depending on the time point under consideration. By using magnetic resonance imaging it has been shown that the brain of primiparous women experiences substantial morphological changes during gestation which mostly affect the right middle temporal gyrus, inferior frontal gyrus and posterior cingulate cortex [27]. A longitudinal morphometric study in mice also provided evidence of transient hypertrophy associated with gestation and/or lactation in the medial preoptic area,

bed nucleus of the stria terminalis, amygdala, caudate nucleus and hippocampus [28]. These findings indicate that certain brain structures undergo significant alterations in their profiles of expression and morphology as a consequence of gestation, and that biological changes are highly dynamic. Thus, the lists of DEGs detected for 12 goat brain tissues at 1-month of gestation might not be representative of the whole gestation but just of this specific time point.

In the goat frontal cortex and the hippocampus, 82 DEGs (45 upregulated and 37 downregulated in pregnant goats) and 70 DEGs (55 upregulated and 15 downregulated in pregnant goats) were detected, respectively (**S4D** and **S4E Tables**). The hippocampus and the frontal neocortex play important roles in modulating memory and learning as well as social behavior [29,30], and there is evidence that hippocampal neurogenesis is affected by pregnancy [31]. Besides, hippocampal lesion in the female rat results in deficient nest construction and reduced pup survival, suggesting that this organ has a key influence on the development of maternal behavior [32].

In the adenohypophysis and pineal gland, 201 DEGs (188 upregulated and 13 downregulated in pregnant goats) and 62 DEGs (2 upregulated and 60 downregulated in pregnant goats) were identified, respectively (**S4A** and **S4J Tables**). Pineal gland and adenohypophysis provide hormonal signals that are fundamental for the maintenance of gestation. For instance, the secretion of maternal pineal melatonin during pregnancy provides photoperiodic information to the fetus, which is essential for achieving a proper neurodevelopment and physical growth [33]. Moreover, the anterior pituitary is strongly enlarged during gestation

in response to hormones produced by the placenta and ovaries, and the secretion of prolactin increases gradually to prepare the female for lactation [34].

Differential mRNA expression in the pons, with 190 DEGs (133 upregulated and 57 downregulated in pregnant goats), is harder to interpret (**S4K Table**). This brain region is mostly involved in the regulation of breathing and sleep, and in relaying information to or from the cerebellum to other brain regions [35]. However, there is evidence that the pons has a basic role in the generation and experience of emotions through the integration of arousal, autonomic function, motor control, and somatosensory signals [35]. Moreover, it has been reported that preoptic area projections to lower brainstem regions affect maternal behavior in postpartum rats [36].

Dramatic changes in the mRNA expression of the olfactory bulb in response to early gestation

The olfactory bulb was the brain region that displayed, by far, the highest number of DEGs (**S4I Table**), implying that this organ is strongly affected by 1 month-pregnancy. The olfactory bulb processes smell information transmitted from olfactory sensory neurons expressing odorant receptors in the olfactory epithelium, and relays it to the olfactory cortex [37]. Importantly, the sense of olfaction is strongly connected to a broad array of behavioral responses related with aversion to food, avoidance of predators, sex arousal, partner preference, and aggression, amongst others [38]. For instance, bilateral bulbectomy of male rats is associated with indifference to the sexual status of females and suppression of mating [39], and it also decreases aggressive behavior between males [40].

Of course, the strong changes in the mRNA expression profile of the olfactory bulb observed in pregnant goats, when compared to their non-pregnant counterparts, are not triggered by odorant signals delivered by the offspring. More likely, this remarkable change of gene expression might be promoted by other brain regions and/or endocrine glands delivering chemical signals to the olfactory bulb. As expected, several DEGs are related with the sense of olfaction (**S7 Table**). More enigmatic are changes in the mRNA expression of a large number of genes known to influence behavioral traits in humans (**S6 Table**). We consider that the disproportionate number of DEGs detected in the olfactory bulb, when compared to other brain tissues, is a robust result demonstrating that this organ is strongly affected by pregnancy. Importantly, this anatomical structure is substantially involved in the development of maternal behavior in mouse, i.e. its complete removal eliminates such capacity and, very often, implies that litters are cannibalized by their mothers soon after parturition [41]. Moreover, intrabulbar infusions of an oxytocin antagonist delays maternal behavior in rats, while oxytocin itself has the opposite effect [42]. In sheep, which are phylogenetically close to goats, disruption of the noradrenergic projections to the olfactory bulb suppresses the establishment of a maternal bond between mother and offspring, probably because such event strongly depends on odor cues [91]. These results support a critical role of the olfactory bulb in the induction of maternal behavior in mammals, although discrepant findings have also been published [91].

It has been proposed that in sheep and goats the increase of estrogen levels in blood during prepartum, combined with the vaginocervical stimulation provoked by the

expulsion of the fetus, trigger modifications in a neural network including the main olfactory system, the medial preoptic area and the paraventricular nucleus of the hypothalamus, in order to induce maternal behavior [44]. Our transcriptomic data demonstrate that extensive areas of the goat brain manifest changes in their expression profiles much before the peripartum period. These alterations might represent an adaptive response towards preparing the pregnant goat for maternity, and we hypothesize that the olfactory bulb might hold a very relevant role in such process.

Conclusions

As a whole, brain tissues in goats display mRNA expression profiles which are strongly coherent with the encephalic vesicle from which they are derived, implying that embryonic development leaves a durable footprint on the transcriptomic landscape of such tissues. Exceptions to this general rule are secretory glands (pineal gland and pituitary) and cerebellum, which show highly divergent patterns of gene expression when compared to other brain tissues, a feature attributable to the highly specialized functions of these organs. We have also demonstrated that 1-month pregnancy does not affect the mRNA expression profile of goat brain tissues to the same extent. These patterns are probably highly dynamic, so the list of altered and unaltered tissues might experience substantial changes throughout pregnancy. At 1 month of gestation, we have observed that the olfactory bulb is the tissue experiencing the more drastic changes in its mRNA expression profile. Many of the genes with altered expression are related with behavioral traits in humans and other

species. This finding is consistent with the key role of this organ in the establishment of an affective bond between mother and offspring, which is the main hallmark of maternity.

Materials and Methods

Sample collection

Seven multiparous non-lactating Murciano-Granadina goats raised in the experimental farm of the Faculty of Veterinary Sciences at the Universitat Autònoma de Barcelona (UAB) were slaughtered because of reasons unrelated with this project. Sampled goats were kept under the same management and environmental conditions had similar ages (6.28 ± 1.38 years). To minimize pain, goats were administered pentobarbital (150 mg/kg) in the jugular vein. Three of these goats were 1-month pregnant at the time of slaughtering. Since animals were killed due to the routine culling process implemented in this experimental farm, no permission from the Ethics Committee on Animal and Human Experimentation at UAB was required.

After slaughtering, goats were transported to the Necropsy Room of the Faculty of Veterinary Sciences at UAB. The cranial vault of each goat was open with a bone saw and twelve brain tissues were carefully dissected and biopsied by an expert anatomist. Sample tissues were drawn from the adenohypophysis, cerebellar hemisphere, cerebellar trunk, frontal neocortex, hippocampus, hypothalamus, medulla oblongata, neurohypophysis, pineal gland, pons, olfactory bulb and rostral

colliculus (**Fig 1A**). Biopsies were immediately submerged in RNAlater (Thermofisher Scientific, Barcelona, Spain) to be stored at -80 ºC until processing.

To purify total RNA, tissue samples were mixed with 1 ml QIAzol (QIAGEN Inc., Barcelona, Spain) and homogenized using the Lysing Matrix D reagent (MP Biomedicals, Santa Ana, CA) in a Precellys 24 tissue homogenizer (Bertin Instruments, Rockville, MD). The extraction of total RNA was performed using the RNeasy lipid tissue mini kit (QIAGEN Inc., Barcelona, Spain) following the protocol described by the manufacturer. The concentration and purity of extracted RNA molecules were analyzed using the Nanodrop ND-1000 spectrophotometer (Thermofisher Scientific, Barcelona, Spain) and RNA integrity was assessed with a Bioanalyzer-2100 equipment (Agilent Technologies, Santa Clara, CA) using the RNA 6000 Nano Kit 4.2 (Agilent Technologies, Santa Clara, CA).

Sequencing of total RNA

Paired-end sequencing (2 x 50 bp) of total RNA was carried out at the Centre Nacional de Anàlisi Genòmica (CNAG). Sequencing methods have been reported by Guan et al. [45]. Briefly, the RNA-Seq library was prepared with the KAPA Stranded mRNA-Seq Illumina Platforms Kit (Roche, Sant Cugat, Spain) by using 500 ng total RNA as template. Oligo-dT magnetic beads were used to enrich the poly-A fraction and subsequently, RNA was fragmented. Strand cDNA synthesis was performed in the presence of dUTP to enforce strand-specificity. The blunt-ended double stranded cDNA was 3'-adenylated and ligated to Illumina adaptors with unique dual indexes and unique molecular identifiers (Integrated DNA

Technologies, Coralville, IA). Enrichment of the ligation product was ensured by performing 15 cycles of polymerase chain reaction amplification. An Agilent 2100 Bioanalyzer equipment was employed to verify the quality of the final library by using the DNA 7500 assay (Agilent Technologies, Inc., Santa Clara, CA). Library sequencing was carried out with a HiSeq 4000 equipment (Illumina, San Diego, CA) in accordance with the protocol for dual indexing advised by the manufacturer. Image analysis, base calling and quality scoring of the sequencing run were checked with the Real-Time Analysis (RTA 2.7.7) tool (Illumina, San Diego, CA) and FASTQ sequence files were subsequently generated.

Quality control, alignment and quantification

The quality of the sequences was assessed with the FastQC software v.0.11.9 [46] and adapters were trimmed using the TrimGalore v.0.6.6 tool [47]. In addition, reads with more than five ambiguous bases (Ns) were removed. Filtered sequences were aligned to the goat ARS1 reference genome [48] with HISAT2 v.2.2.1 aligner [49,50] and gene expression was quantified using StringTie v.2.1.0 [51,52] in accordance with the protocol described in Pertea et al. [53]. The estimated count matrix was obtained from coverage values of each feature (genes in our case) using the dedicated script “prepDE.py” from the StringTie pipeline and based on the following formula [54]:

$$\text{Reads per gene} = \text{coverage} \times \frac{\text{gene length}}{\text{thread length}}$$

Construction of a brain atlas based on the expression of protein-coding genes

Genes expressed in each tissue were catalogued according to their Ensembl biotype [55]. To remove systematic technical effects (e.g. library size), count data from protein-coding genes were normalized with the trimmed mean of M values (TMM) method implemented in the Bioconductor R package MDseq [56]. The corresponding estimated counts per protein-coding gene were normalized to CPM with the edgeR package [57,58]. Finally, data were $\log_2$ transformed adding a pseudo-count of 1. Each tissue was normalized separately. Normalization factors were estimated by jointly considering pregnant and non-pregnant goats because library size features were quite similar in both groups. To perform downstream analyses, only genes with a CPM over 0.5 in at least two samples per tissue were taken into consideration.

A PCA was performed using the stats R package [59,60]. Besides, a heatmap depicting the level of expression of each gene per tissue and sample was plotted. The Pheatmap software [94] was used to visualize the outputs of such analyses. Instead of plotting normalized expression values, data from each row (genes in our case) were Z-scored to a distribution with mean 0 and standard deviation 1. By doing so, we made sure that expression trends are not obscured because of the highly dissimilar expression levels of genes. Additionally, a hierarchical clustering dendrogram was built based on Euclidean distances and applying the Ward D clustering method [61]. The top 1000 genes with the highest expression levels per tissue considering all the individuals were selected to perform pathway enrichment analyses with the Enrichr package [12,62]. The KEGG release 99.0 human database

[63] was used as reference to annotate pathways enriched in the 12 sets of genes (one for each tissue). In order to identify such enriched pathways, a combined score (*c*) was estimated from the output of a traditional Fisher exact test and a Z-score estimated from expected ranks and variances of the set of genes. In accordance with Chen et al. [12] the following formula was applied:

$$c=\log(p) \times z$$

where *c* is the combined score, *p* corresponds to *P*-values estimated with the Fisher exact test, and *z* is the Z-score from the expected rank.

The tissue-specificity expression of protein-coding genes was assessed by calculating the tissue specificity index tau (τ), a quantitative, graded scalar measure of the gene expression profile. A τ-value of 0 would correspond to a housekeeping gene with ubiquitous tissue expression. In contrast, genes with a restricted tissue expression show τ-values above 0.85, while genes with τ=1 can be assumed to be expressed in a single tissue [64]. The formula to estimate the index was defined by Yanai et al. [64]:

$$\tau=i=1n\ (1-xi)\ n-1$$

where n is the number of tissues, and xi corresponds to the expression profile component normalized by the maximal component value for that gene (i.e. the expression of the gene in the tissue where it is most highly expressed), making the estimation sensitive to extreme expression values [64]. This calculation was performed using the tspex v.0.6.1 program [65,66]. Genes with a τ-score below 0.15 were selected and catalogued as ubiquitous, while genes with τ-score over 0.85

were considered tissue-specific. The genes falling into the aforementioned categories were identified and functional enrichment analyses were performed over the two datasets using Enrich R [12,62], as previously explained.

Differential expression and pathway enrichment analyses

The software DEseq2 [67,68] was used to perform a differential expression analysis comparing pregnant vs. non-pregnant goat transcriptomic profiles. Genes with a number of estimated counts below 10 were removed. Correction for multiple testing was applied using the false discovery rate (FDR) method [69]. We considered that a gene is differentially expressed when two conditions are met: absolute logarithm of the fold change ($\log_2$FC) > 1.5 and an associated *q*-value < 0.05. Differentially expressed genes were used to perform pathway enrichment analyses using the Enrichr R package [12]. The annotation of DEGs was performed as explained before, and the enrichment pathway analyses were carried out in independently in the sets of upregulated and downregulated genes. The level of enrichment was estimated by calculating a *c* score [12], as previously defined.

Data availability

The raw data generated in this research are available in the Sequence Read Archive (SRA) under the bioproject accession number PRJNA808876.

Ethics statement

Goats were slaughtered as a result of a routine culling procedure managed by the Service of Farms and Experimental Fields of the Universitat Autònoma de

Barcelona (UAB) completely unrelated with this project, so permission from the Ethics Committee on Animal and Human Experimentation of the UAB was not required.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the European Regional Development Fund (FEDER)/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación/Project Reference grants: AGL2016-76108-R and also by project PID2019-105805RB-I00 funded by MCIN/AEI/10.13039/501100011033. We also acknowledge the support of the CERCA programme of the Generalitat de Catalunya and the Center of Excellence Severo Ochoa 2020–2023 (CEX2019–000902-S) grant funded by MCIN/AEI /10.13039/501100011033 and awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). Dailu Guan was funded by a PhD fellowship from the China Scholarship Council (CSC). Maria Luigi-Sierra was funded with a PhD fellowship Formación de Personal Investigador BES-C-2017-079709 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future". Emilio Mármol-Sánchez was funded with a FPU Ph.D. grant from the Ministerio de Educación (Spain, FPU15/01733).

Authors contributions

Conceptualization—MA, MLB; Formal Analysis—MGL; Funding acquisition—MA; Investigation—MGL, DG; Project administration—MA; Resources—MLB, MA, MGL, DG, EC, SO, JG, MJP, UIO, ULO EMS; Visualization—MGL; Writing—original draft—MA, MGL; Writing—review & editing—EMS, MJP, DG, MLB, MA, MGL, EC, SO, JG, UIO, U.L.OS. All authors read and approved the content of the paper.

Acknowledgments

Thanks to José Luis Ruiz de la Torre Casañas from the Service of Farms and Experimental Fields of the UAB and Dr Xavier Such from the Department of Animal Sciences of the UAB for providing the animal material used in this experiment.

References

1. Kuwar Chhetri P, M Das J. Neuroanatomy, neural tube development and stages. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2022 Jan.

2. Ferreira AO, Vasconcelos BG, Favaron PO, Santos AC, Leandro RM, Pereira FTV, et al. Bovine central nervous system development. Pesqui Vet Bras. 2018; 77:2088–97.

3. Ortiz C, Navarro JF, Jurek A, Märtin A, Lundeberg J, Meletis K. Molecular atlas of the adult mouse brain. Sci Adv. 2020; 26:eabb3446.

4. Erö C, Gewaltig MO, Keller D, Markram H. A cell atlas for the mouse brain. Front Neuroinform. 2018; 12:84.

5. Ray S, Tzeng RY, DiCarlo LM, Bundy JL, Vied C, Tyson G, et al. An examination of dynamic gene expression changes in the mouse brain during pregnancy and the postpartum period. G3 Genes, Genomes, Genet. 2016; 6:221–33.

6. Arbeitman MN. Maternal experience leads to lasting gene expression changes in some regions of the mouse brain. G3 Genes, Genomes, Genet. 2019; 9:2623–28.

7. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science 2020; 367:eaay5947.

8. Clark EL, Bush SJ, McCulloch MEB, Farquhar IL, Young R, Lefevre L, et al. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS Genet. 2017; 13:e1006997.

9. Harhay GP, Smith TPL, Alexander LJ, Haudenschild CD, Keele JW, Matukumalli LK, et al. An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. Genome Biol. 2010; 11:R102.

10. Muriuki C, Bush SJ, Salavati M, McCulloch MEB, Lisowski ZM, Agaba M, et al. A mini-atlas of gene expression for the domestic goat (Capra hircus). Front Genet. 2019; 10:1080.

11. Gallego Romero I, Pai AA, Tung J, Gilad Y. RNA-seq: Impact of RNA degradation on transcript quantification. BMC Biol. 2014; 12:42.

12. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G V., et al. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013; 14:128.
13. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science 2015; 348:660–5.
14. Ferraz ALJ, Ojeda A, López-Béjar M, Fernandes LT, Castelló A, Folch JM, et al.. Transcriptome architecture across tissues in the pig. BMC Genomics 2008; 9:173.
15. Ardlie KG, DeLuca DS, Segrè A V., Sullivan TJ, Young TR, Gelfand ET, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015; 348:648–60.
16. Freeman TC, Ivens A, Baillie JK, Beraldi D, Barnett MW, Dorward D, et al. A gene expression atlas of the domestic pig. BMC Biol. 2012; 10:90.
17. Zapala MA, Hovatta I, Ellison JA, Wodicka L, Del Rio JA, Tennant R, et al. Adult mouse brain gene expression patterns bear an embryologic imprint. Proc Natl Acad Sci USA 2005; 102:10357–62.
18. Kavoi BM, Jameela H. Comparative morphometry of the olfactory bulb, tract and stria in the human, dog and goat. Int J Morphol. 2011; 29:939–46.
19. Keller M, Lévy F. The main but not the accessory olfactory system is involved in the processing of socially relevant chemosignals in ungulates. Front. Neuroanat. 2012; 6:39.
20. Benton MJ, Donoghue PCJ, Asher RJ, Friedman M, Near TJ, Vinther J. Constraints on the timescale of animal evolutionary history. Palaeontol Electron. 2015; 18:1–116.
21. Glickstein M, Doron K. Cerebellum: Connections and functions. Cerebellum. 2008; 7:589–94.
22. Kozareva V, Martin C, Osorno T, Rudolph S, Guo C, Vanderburg C, et al. A transcriptomic atlas of mouse cerebellar cortex coprehensively defines cell types. Nature. 2021; 598:214–219.

23. Patel S, Rahmani B, Gandhi J, Seyam O, Joshi G, Reid I, et al. Revisiting the pineal gland: A review of calcification, masses, precocious puberty, and melatonin functions. Int J Neurosci. 2020; 130:464–75.

24. Le Tissier PR, Hodson DJ, Lafont C, Fontanaud P, Schaeffer M, Mollard P. Anterior pituitary cell networks. Front Neuroendocrinol. 2012; 33:252–66.

25. Santos Fontanez SE, De Jesus O. Neurohypophysis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan.

26. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 2016; 22:839–51.

27. Hoekzema E, Barba-Müller E, Pozzobon C, Picado M, Lucco F, García-García D, et al. Pregnancy leads to long-lasting changes in human brain structure. Nat Neurosci. 2017; 20:287–96.

28. Barrière DA, Ella A, Szeremeta F, Adriaensen H, Même W, Chaillou E, et al. Brain orchestration of pregnancy and maternal behavior in mice: A longitudinal morphometric study. Neuroimage. 2021; 230:117776.

29. Zemla R, Basu J. Hippocampal function in rodents. Curr Opin Neurobiol. 2017; 43:187–97.

30. Reinert S, Hübener M, Bonhoeffer T, Goltstein PM. Mouse prefrontal cortex represents learned rules for categorization. Nature 2021; 593:411–17.

31. Rolls A, Schori H, London A, Schwartz M. Decrease in hippocampal neurogenesis during pregnancy: A link to immunity. Mol Psychiatry 2008;13:468–9.

32. Murphy HM, Wildeman CH. Pup survival and development following hippocampal lesions in the female rat 1. Ohio J sci. 1986; 1986:89–94.

33. Motta-Teixeira LC, Machado-Nils AV, Battagello DS, Diniz GB, Andrade-Silva J, Silva S, et al. The absence of maternal pineal melatonin rhythm during pregnancy and lactation impairs offspring physical growth, neurodevelopment, and behavior. Horm Behav. 2018; 105:146–56.

34. Carlson HE. The pituitary gland in pregnancy and the puerperium. In: Shlomo Melmed editor.The pituitary. Academic press; 2011. p. 397–412.

35. Venkatraman A, Edlow BL, Immordino-Yang MH. The brainstem in emotion: A review. Front Neuroanat. 2017; 11:15.
36. Numan M, Numan MJ. Preoptic-brainstem connections and maternal behavior in rats. Behav Neurosci. 1991; 105:1013–29.
37. Nagayama S, Homma R, Imamura F. Neuronal organization of olfactory bulb circuits. Front Neural Circuits. 2014; 8:98.
38. Sullivan RM, Wilson DA, Ravel N, Mouly AM. Olfactory memory networks: From emotional learning to social behaviors. Front Behav Neurosci. 2015; 9:36.
39. Edwards DA, Griffis KT, Tardivel C. Olfactory bulb removal: Effects on sexual behavior and partner-preference in male rats. Physiol Behav. 1990; 48:447–50.
40. Rowe FA, Edwards DA. Olfactory bulb removal: Influences on the aggressive behaviors of male mice. Physiol Behav. 1971; 7:889–92.
41. Gandelman R, Zarrow MX, Denenberg VH, Myers M. Olfactory bulb removal eliminates maternal behavior in the mouse. Science. 1971; 171:210–1.
42. Yu GZ, Kaba H, Okutani F, Takahashi S, Higuchi T. The olfactory bulb: A critical site of action for oxytocin in the induction of maternal behaviour in the rat. Neuroscience. 1996; 72:1083–8.
43. Pissonnier D, Thiery JC, Fabre-Nys C, Poindron P, Keverne EB. The importance of olfactory bulb noradrenalin for maternal recognition in sheep. Physiol Behav. 1985; 35:361–3.
44. Poindron P, Lévy F, Keller M. Maternal responsiveness and maternal selectivity in domestic sheep and goats: The two facets of maternal attachment. Dev Psychobiol. 2007; 49:54–70.
45. Guan D, Landi V, Luigi-Sierra MG, Delgado JV, Such X, Castelló A, et al. Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. J Anim Sci Biotechnol. 2020; 11:35.
46. FastQC (Version0.11.9). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
47. Trim Galore (Version 0.6.6) https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

48. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet. 2017; 49:643–50.

49. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. Nat Methods. 2015; 12:357–60.

50. HISAT2 (Version 2.2.1). http://daehwankimlab.github.io/hisat2/.

51. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015; 33:290–5.

52. StringTie (Version 2.1.0). https://ccb.jhu.edu/software/stringtie/.

53. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. RNA-seq experiments with HISAT , StringTie and Ballgown. Nat Protoc. 2016; 11:1650–67.

54. StringTie. http://ccb.jhu.edu/software/stringtie/index.shtml?t=manual Accessed 2022 Feb 1.

55. Biotypes. https://m.ensembl.org/info/genome/genebuild/biotypes.html Accessed 2022 Feb 1.

56. MDSeq (Version 1.0.5). https://github.com/zjdaye/MDSeq.

57. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009; 26:139–40.

58. EdgeR (Version 3.36). https://bioconductor.org/packages/release/bioc/html/edgeR.html.

59. R Core team. A language and environment for statistical computing. Vienna, Austria ISBN 3-900051-07-0, URL http//wwwR-project.org/.

60. Stats (Version 4.2.0). https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html.

61. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? J Classif. 2014; 31:274–95.

62. Enrichr (Version 3.0). https://cran.r-project.org/web/packages/enrichR/.

63. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016; 44:D457–62.
64. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005; 212:650–9.
65. Camargo AP, Vasconcelos AA, Fiamenghi MB, Pereira GAG, Carazzole MF. Tspex: A tissue-specificity calculator for gene expression data. Research Square. doi: 10.21203/rs.3.rs-51998/v1.
66. Tspex (Version 0.6.1). https://github.com/apcamargo/tspex.
67. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550.
68. DESeq2 (Version 1.34.0). https://bioconductor.org/packages/release/bioc/html/DESeq2.html.
69. Benjamini, Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995; 57:289–300.

Supplementary information

S1 Table. (A) Catalogue of 5,826 genes ubiquitously expressed in 12 goat brain tissues. (B). Functional enrichment analysis of 5,826 genes ubiquitously expressed in 12 goat brain tissues.

S2 Table. Catalogue and pathway enrichment analyses of genes with high tissue specificity in 12 goat tissues. (**A**) List of 1,351 protein-coding genes with high tissue specificity (tau score > 0.85) in 12 goat brain tissues. (**B**) Pathway enrichment analysis of 560 tissue-specific (tau score > 0.85) protein-coding genes expressed in the adenohypophysis of Murciano-Granadina goats. (**C**) Pathway enrichment analysis of 221 tissue-specific (tau score > 0.85) protein-coding genes expressed in the cerebellar hemisphere of Murciano-Granadina goats. (**D**) Pathway enrichment analysis of 224 tissue-specific (tau score > 0.85) protein-coding genes expressed in the cerebellar trunk of Murciano-Granadina goats. (**E**) Pathway

enrichment analysis of 181 tissue-specific (tau score > 0.85) protein-coding genes expressed in the frontal neocortex of Murciano-Granadina goats. (**F**) Pathway enrichment analysis of 267 tissue-specific (tau score > 0.85) protein-coding genes expressed in the hippocampus of Murciano-Granadina goats. (**G**) Pathway enrichment analysis of 275 tissue-specific (tau score > 0.85) protein-coding genes expressed in the hypothalamus of Murciano-Granadina goats. (**H**) Pathway enrichment analysis of 225 tissue-specific (tau score > 0.85) protein-coding genes expressed in the medulla oblongata of Murciano-Granadina goats. (**I**) Pathway enrichment analysis of 613 tissue-specific (tau score > 0.85) protein-coding genes expressed in the neurohypophysis of Murciano-Granadina goats. (**J**) Pathway enrichment analysis of 290 tissue-specific (tau score > 0.85) protein-coding genes expressed in the olfactory bulb of Murciano-Granadina goats. (**K**) Pathway enrichment analysis of 556 tissue-specific (tau score > 0.85) protein-coding genes expressed in the pineal gland of Murciano-Granadina goats. (**L**) Pathway enrichment analysis of 214 tissue-specific (tau score > 0.85) protein-coding genes expressed in the pons of Murciano-Granadina goats. (**M**) Pathway enrichment analysis of 130 tissue-specific (tau score > 0.85) protein-coding genes expressed in the rostral colliculus of Murciano-Granadina goats.

S3 Table. Pathway enrichment analysis of the top 1000 genes most expressed in the (A) adenohypophysis, (B) cerebellar hemisphere, (C) cerebellar trunk, (D) frontal neocortex, (E) hippocampus, (F) hypothalamus, (G) medulla oblongata, (H) neurohypophysis, (I) olfactory bulb, (J) pineal gland, (K) pons and (L) rostral colliculus of goats.

S4 Table. List of protein-coding genes differentially expressed (absolute $\log_2$ fold-change > 0.58 and *q*-value < 0.05) in the (A) adenohypophysis, (B) cerebellar hemisphere, (C) cerebellar trunk, (D) frontal neocortex, (E) hippocampus, (F) hypothalamus, (G) medulla oblongata, (H) neurohypophysis, (I) olfactory bulb, (J) pineal gland, (K) pons and (L) rostral colliculus from pregnant (N=3) and non-pregnant (N=4) goats.

S5 Table. Pathway enrichment analysis of genes differentially expressed in the (A) adenohypophysis, (B) frontal neocortex, (C) hippocampus, (D) olfactory

bulb, (E) pineal gland and (F) pons of 1 month-pregnant vs non-pregnant goats.

S6 Table. List of articles generated through a bibliographic search describing the behavioral functions of genes that are differentially expressed in the olfactory bulb of pregnant vs. non-pregnant goats.

S7 Table. List of articles generated through a bibliographic search describing the olfactory functions of genes that are differentially expressed in the olfactory bulb of pregnant vs. non-pregnant goats.

S1 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the adenohypophysis of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S2 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the cerebellar hemisphere of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S3 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the cerebellar trunk of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S4 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the frontal neocortex of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S5 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the hippocampus of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S6 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the hypothalamus of Murciano-Granadina goats. Combined scores generated in the

enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S7 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the medulla oblongata of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S8 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the neurohypophysis of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S9 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the olfactory bulb of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S10 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the pineal gland of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S11 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the pons of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S12 Fig. Top 20 pathways enriched in tissue-specific genes expressed in the rostral colliculus of Murciano-Granadina goats. Combined scores generated in the enrichment analysis are shown in the *x*-axis; while pathway denominations are indicated in the *y*-axis.

S13 Fig. Top 20 pathways enriched in genes differentially expressed in the adenohypophysis of pregnant vs. non-pregnant goats. The *x*-axis displays the combined score calculated from the enrichment analysis, while the *y*-axis indicates the denomination of each pathway. Pathways containing upregulated genes in

pregnant goats are displayed in red, while pathways encompassing downregulated genes in pregnant goats are shown in blue.

S14 Fig. Top 20 pathways enriched in genes differentially expressed in the frontal neocortex of pregnant vs. non-pregnant goats. The *x*-axis displays the combined score calculated from the enrichment analysis, while the *y*-axis indicates the denomination of each pathway. Pathways containing upregulated genes in pregnant goats are displayed in red, while pathways encompassing downregulated genes in pregnant goats are shown in blue.

S15 Fig. Top 20 pathways enriched in genes differentially expressed in the hippocampus of pregnant vs. non-pregnant goats. The *x*-axis displays the combined score calculated from the enrichment analysis, while the *y*-axis indicates the denomination of each pathway. Pathways containing upregulated genes in pregnant goats are displayed in red, while pathways encompassing downregulated genes in pregnant goats are shown in blue.

S16 Fig. Top 20 pathways enriched in genes differentially expressed in the olfactory bulb of pregnant vs. non-pregnant goats. The *x*-axis displays the combined score calculated from the enrichment analysis, while the *y*-axis indicates the denomination of each pathway. Pathways containing upregulated genes in pregnant goats are displayed in red, while pathways encompassing downregulated genes in pregnant goats are shown in blue.

S17 Fig. Top 20 pathways enriched in genes differentially expressed in the pineal gland of pregnant vs. non-pregnant goats. The *x*-axis displays the combined score calculated from the enrichment analysis, while the *y*-axis indicates the denomination of each pathway. Pathways containing upregulated genes in pregnant goats are displayed in red, while pathways encompassing downregulated genes in pregnant goats are shown in blue.

S18 Fig. Top 20 pathways enriched in genes differentially expressed in the pons of pregnant goats vs. non-pregnant goats. The *x*-axis displays the combined score calculated from the enrichment analysis, while the *y*-axis indicates the denomination of each pathway. Pathways containing upregulated genes in pregnant

goats are displayed in red, while pathways encompassing downregulated genes in pregnant goats are shown in blue.

4 GENERAL DISCUSSION

4.1 UNDERSTANDING THE GENOMIC ARCHITECTURE OF MILK AND MORPHOLOGICAL TRAITS IN MURCIANO-GRANADINA GOATS.

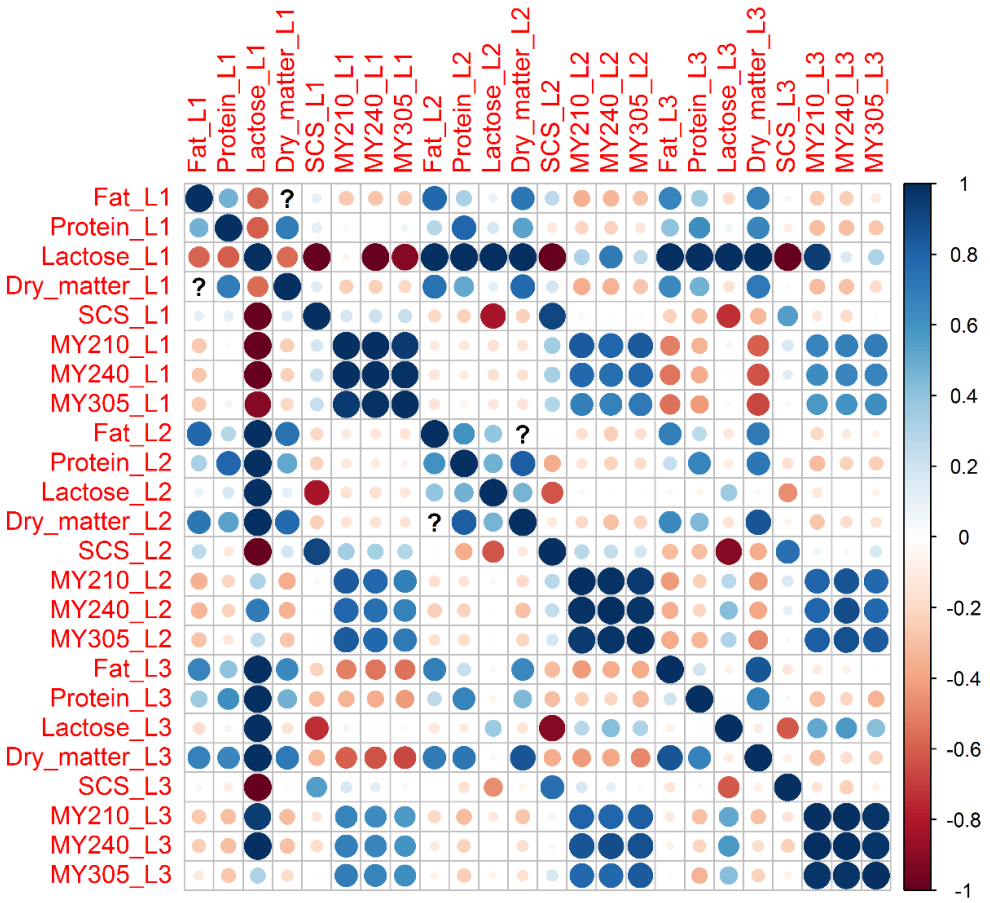
4.1.1 Variation in the casein gene cluster is highly associated with milk protein percentage in Murciano-Granadina goats.

Guan et al. (2020) performed a GWAS for milk production and composition traits recorded during the first lactation of the same Murciano-Granadina goats employed in **study 2** of the current thesis. Guan et al. (2020) reported 24 QTL for milk traits, and three of them reached genome-wide statistical significance, i.e. CHI 6:130.72–130.01 Mb for lactose percentage, CHI 6: 78.90–93.44 Mb for protein percentage and CHI 17: 11.20 Mb for both protein and dry matter percentages. In this thesis, we have also performed a GWAS, but the main difference with the work of Guan et al (2020) is that we have considered phenotypic records from three lactations. Two different approaches have been used: (1) Data from each lactation is analysed independently (so 3 GWAS are performed), (2) A longitudinal analysis jointly considering the three lactations is undertaken. Both approaches consistently identified one QTL on CHI 2 (129.77 – 131.01 Mb), which is associated with lactose percentage, and another QTL on CHI 6 (74.8 – 94.6 Mb) associated with milk protein percentage. The first approach evidenced that QTL detected on lactations 1, 2 and 3 are quite, but not completely, concordant. This could be due to differences in sample size i.e. 917, 805 and 660 goats for lactations 1, 2 and 3,

respectively. To check this, we performed GWAS for each lactation separately considering only the 660 individuals with information for all three lactations, and the results obtained in this way were consistent with those corresponding to the full sample size (data not shown). In the light of this, we conclude that sample size is not the main cause of detecting different QTL across lactations. Alternatively, the genetic determinism of milk traits might not be exactly the same across lactations and certain QTL might show time-dependent effects. Tong et al. (1979) obtained lactation records from 13,544 cows and showed that heritabilities of milk yield in the first and third lactations were 0.25 and 0.17 (32% lower), respectively, and the heritabilities of milk fat percentage were 0.45 and 0.35 (23% lower), respectively. Genetic correlations between milk traits measured in different lactations fluctuated between 0.83 and 0.93 (Tong et al., 1979). We have calculated genetic correlations between milk traits measured in different lactations with GCTA (Yang et al., 2011) and they are shown in **Figure 4.1**. Genetic correlations ranged between 0.72 to 1 (lactations 1 vs. 2); 0.61 to 1 (lactation 1 vs 3) and 0.36 and 0.89 (lactation 2 vs 3). The lowest correlation values were estimated for lactation 3 when compared to lactations 1 and 2, and such coefficients were particularly low for the traits of milk composition (**Figure 4.1**). Overall, these results indicate that milk traits measured in different lactations share a strong common genetic basis, although certain genetic determinants might be lactation-specific. Moreover, the number of lactation significantly affects milk yield and composition of goats, with a lower production during the first lactation and a progressive increase until the third lactation (Ciappesoni et al., 2004). This variability is influenced by morphological and

physiological changes in the goat udder after the first and second parturitions (Ciappesoni et al., 2004).

Figure 4.1. Heatmap depicting the genetic correlations of eight milk production and composition traits recorded during the first three lactations in 660 Murciano-Granadina goats.



The longitudinal analysis of the three lactations allowed the identification of 31 QTLs, and 8 of them reached genome-wide significance. The longitudinal analysis yielded a higher number of QTL (31) than the independent analyses made for each one of the three lactations i.e. 22 (6 with genome wide significance), 29 (3 with

genome wide significance) and 9 (1 with genome wide significance) QTLs for lactations 1, 2 and 3, respectively. All four analyses consistently detected a QTL on CHI 6:74.89–96.5 Mb for protein percentage, and this region was also associated with dry matter percentage in the longitudinal GWAS and in the independent GWAS for lactations 1 and 2. Another association consistently identified in the longitudinal GWAS and the separate GWAS for lactations 1 and 2 was that between the CHI 2:129.80–130.47 region and lactose percentage. Nevertheless, multiple regions with statistical significance for the analysed traits were detected exclusively in the longitudinal analysis, specifically 22 QTLs, three of them with genome-wide significance.

In general, repeated measures are difficult to model using traditional GWAS software (Rönnegård et al., 2016). One possibility is to use averaged phenotypic records, but this complicates the correction for fixed factors and it might result in the reduction of statistical power and the inflation of false positives (Rönnegård et al., 2016). A more powerful approach is to use GWAS software specifically developed to use repeated measurements as input, as it is the case of RepeatABEL (Rönnegård et al., 2016). According to Rönnegård et al. (2016), in the case of unbalanced data or data where there is between year variation, the inclusion of repeated measures in the analysis increases the power of a GWAS to detect causal variants. This might be the primary reason why we detected more QTL in the longitudinal analysis than in the three independent GWAS.

As said, the most relevant result obtained in the longitudinal and independent GWAS is the highly significant QTL on chromosome 6, which is associated to

protein percentage. This result is very meaningful because this QTL shows positional concordance with the cluster of the four casein genes which encode 80% milk proteins. The high polymorphism of the goat *CSN1S1* gene was initially uncovered in electrophoretic studies carried out by Boulanger et al. (1984) and Grosclaude et al. (1987), and later on in-depth analyses at the DNA level were performed (reviewed in Martin et al., 2002 and Moioli et al., 2007). Molecular studies demonstrated the existence of at least 17 variants that can be classified in four groups depending on the CSN1S1 content they are associated to (reviewed in Amills et al., 2012):

- Strong alleles (3.5 g casein/allele): A, B1, B2, B3, B4, C, H, L and M.
- Medium alleles (1.1 g casein/allele): E and I
- Low alleles (0.45 g casein/allele): F, D and G
- Null alleles (0 g casein/allele): 01, 02 and N

Several causal mutations influencing transcript stability and processing have been identified so far in the caprine *CSN1S1* gene, and results obtained in French goat breeds have evidenced that they have a considerable impact on milk protein and fat percentages as well as on micelle size, cheese yield, curd firmness and cheese flavor (Cosenza et al., 2003; Marletta et al., 2007; Ollier et al., 2008; Caravaca et al., 2011; reviewed in Amills et al., 2012). In this regard, the *CSN1S1* locus is a strong candidate gene to explain (at least in part) the strong QTL that we have detected on chromosome 6 for milk protein percentage. Although Caravaca et al. (2008) reported the absence of associations between *CSN1S1* genotypes and milk yield and composition traits in Murciano-Granadina goats, Pizarro Inostroza et al. (2019)

concluded that the inclusion of the *CSN1S1* genotype in a breeding model resulted in a substantial increase of protein production heritability in such breed. Our results are more consistent with those of Pizarro Inostroza et al. (2019), although, we cannot rule out the possibility that the QTL detected by us is explained by the variation of the *CSN2*, *CSN1S2* or *CSN3* genes. Indeed, genetic variants with quantitative effects on milk traits have been discovered in these three genes. For instance, two *CSN1S2* D and 0 variants have been associated to reduced concentrations of CSN1S2 in milk (Ramunno et al., 2001a, 2001b), and two null alleles have been reported for the *CSN2* gene (Persuy et al., 1999; Ramunno et al., 1995), although we do not know if these allelic variants are segregating in the Murciano-Granadina breed. With regard to the *CSN3* gene, Caravaca et al. (2009) provided evidence that the milk from AB and BB goats has significantly higher levels of total casein and protein content than milk from their AA counterparts. Besides, different genotypes in the *CSN1S1* and *CSN3* genes significantly affect the coagulation and curding process in Murciano-Granadina goats, becoming an essential factor for cheesemaking (Caravaca et al., 2011). It would be necessary to genotype selected variants in the four casein genes to ascertain their relative contributions to the protein percentage QTL detected on chromosome 6.

Unlike Martin et al. (2017), we did not find associations between the region harbouring the *DGAT1* gene (CHI 14:81.32–81.33 Mb) and fat content, suggesting that the R251L and R396W variants detected in their study are not segregating in Murciano-Granadina goats. In French breeds, associations between the *CSN1S1* genotype and milk fat content have been reported (Chilliard et al., 2006; Ollier et

al., 2008; Vázquez-Flores et al., 2012), and they have been explained in the light of a coupled and co-regulated secretion of proteins and lipids in milk (Cebo et al., 2012). However, in **study 1** of this thesis, the QTL mapping to the casein cluster was not associated with milk fat percentage suggesting a different mechanism of action of the underlying causal mutations.

Another consistent QTL was located on CHI 2 (129-130 Mb) and had effects on lactose percentage. Interestingly, Costa et al. (2019) detected the same chromosomal region (BTA 2: 5.7-5.8 Mb) as associated with lactose content, and proposed several loci related with transmembrane transport activity (*MFSD6*, *NEMP2*, and *SLC40A1*) as potential candidate genes. However, the relationship between these genes and lactose metabolism is, at best, very tenuous. In our study, we have proposed two candidate genes (*INNP1* and *ORMDL1* loci) but again the link with lactose metabolism or transport is not direct or fully convincing. However, the positional coincidence of this QTL for lactose in goats and cattle suggests the existence of a causal mutation with such effect in both species. An improved functional annotation of this region might be helpful to uncover the genetic basis of this QTL.

4.1.2 Morphological traits have a highly polygenic genetic background.

There is evidence that the morphological conformation of the udder has an impact on milkability (Rovai et al., 2004) as well as on milk yield (Keskin et al., 2005; Montaldo and Martínez-Lozano, 1993). For instance, Capote et al. (2006) showed that deep and well attached udders are significantly and positively correlated with

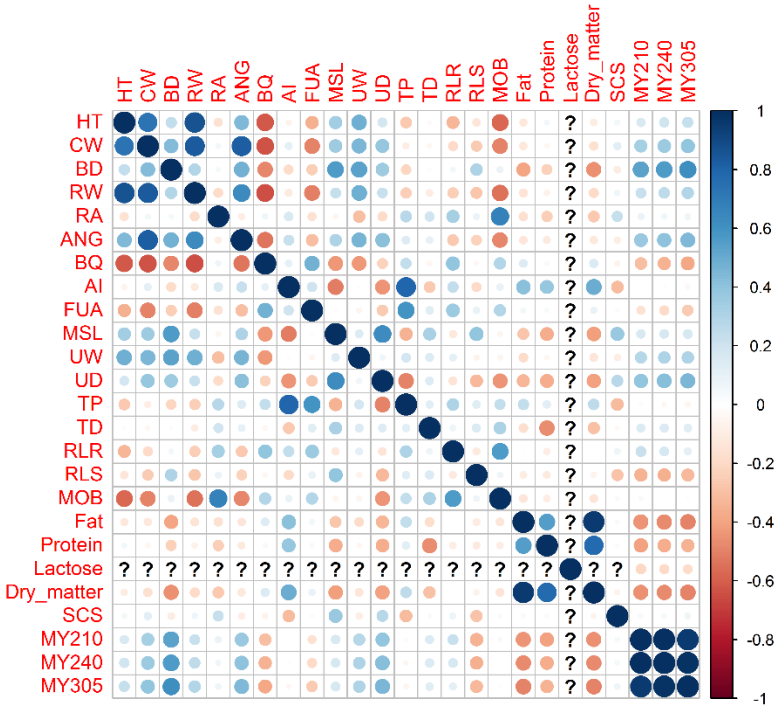
the quantity of produced goat milk. Moreover, teat angle is one of the most important traits that determines the suitability of the udder to machine milking (Peris et al., 1999). In the analyses performed in **study 2**, moderate positive phenotypic correlations were estimated between milk yield at three time points and udder depth (r_p 0.30 – 0.32), and between milk yield and body capacity traits, i.e. chest width (r_p 0.23 – 0.27), body depth (r_p 0.21 –0.25), and also with angularity (r_p 0.21 – 0.23), a trait related with the dairy performance of the goats. We have made a preliminary estimation of genetic correlations between morphologic and milk production traits, that is depicted in **Figure 4.2**. This analysis was not included in **study 2** because it was judged to be too preliminar, but at it would be worth to mention it herewith As previously reported for phenotypic correlations, milk yield at three time points showed significant genetic correlations with body depth (r_g= 0.53 – 0.61) and udder depth (r_g= 0.39 – 0.46). Additionally, the anterior insertion of the udder displayed a genetic correlation of 0.50 with dry matter percentage of milk, and teat diameter was negatively correlated with protein percentage of milk (r_g= -0.46). However, when we analysed the positional concordance of the QTL detected in this thesis for milk production and morphological traits, no substantial match was observed. In other studies, positional coincidences between QTL for dairy and morphology traits has been reported. For instance, a region on CHI 19:26.61 – 26.62 Mb was identified as significantly associated with milk, fat and protein yield, and somatic cell score in Saanen, Alpine, Nubian, Toggenburg and crossbreed goats from New Zealand (Scholtens et al., 2020). In a subsequent analysis in the same New Zealand mixed breed population, Jiang et al. (2022) found

that the very same CHI 19: 26.61–26.62 Mb region was significantly associated with udder conformation traits (i.e. udder depth, front and rear udder attachment). In another study, the CHI 19: 26.61–26.62 Mb interval was associated with milk yield, udder attachment, udder depth and front legs conformation in Saanen × Toggenburg × Alpine goats (Mucha et al., 2018). Indeed, the existence of significant genetic correlations between milk and morphology traits recorded in cattle (Sartori et al., 2018), sheep (Legarra and Ugarte, 2005) and goats (McLaren et al., 2016) points out to the potential existence of pleiotropic genetic determinants with effects on both sets of phenotypes. Possibly, the use of a multivariate GWAS approach would have facilitated the detection of pleiotropic loci in the Murciano-Granadina population, since such method is known to increase the statistical power to identify associations in the case of a shared genetic basis between phenotypes (Ruotsalainen et al., 2021).

Different regions were significantly associated with morphological traits in Alpine and Saanen goats by Martin et al. (2018), suggesting that morphology QTL have a certain degree of population specificity. From the 12 QTLs identified for udder morphology in Murciano-Granadina goats (**study 1**), three were located less than 2 Mb from those reported by Martin et al. (2018) but corresponded to different traits, i.e. two QTLs on CHI 15 and 19 for teat length in Alpine and Saanen goats were close to QTL detected in Murciano-Granadina goats for udder width and teat placement, and one QTL for teat length on CHI 27, for Alpine and Saanen goats, was close to one significant region for udder width in Murciano-Granadina goats. In summary, we have observed a weak positional coincidence between morphology

QTL detected in Alpine and Saanen breeds vs those identified in Murciano-Granadina goats. This result is compatible with the existence of a remarkable degree of genetic heterogeneity amongst breeds with regard to the determinism of morphological phenotypes.

Figure 4.2. Heatmap depicting the genetic correlations of milk production and composition traits and morphological linear scores recorded during the first lactation in 825 Murciano-Granadina goats.



In congruence with other GWAS studies for morphological traits performed in goats (Martin et al., 2018; Mucha et al., 2018), the number and significance of associations between genetic markers and morphological traits are quite low and heterogeneous in Murciano-Granadina goats. In dogs, the genetic analysis of average breed body size and external body dimensions and cranial, dental, and long bone shape and size traits revealed that for most traits a reduced (less than three)

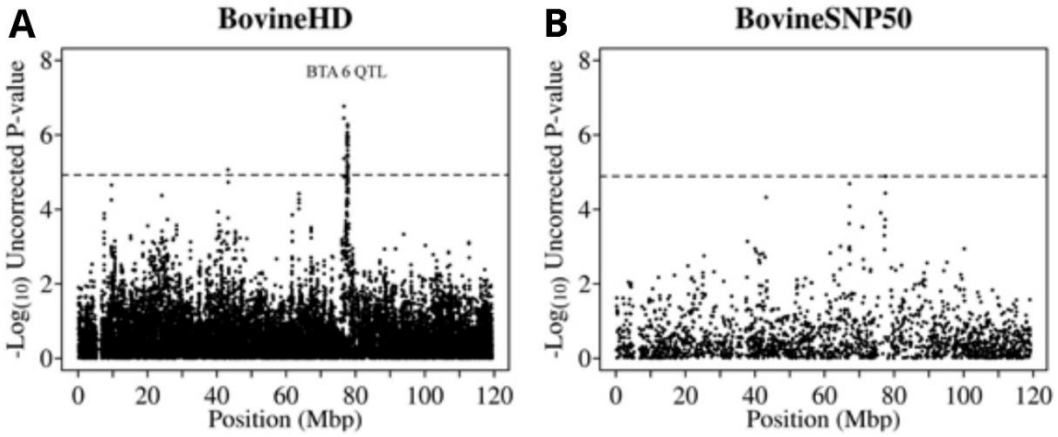
number of QTL explained the majority of phenotypic variation (Boyko et al., 2010). But in dogs, morphological variation is much larger than in goats and it is not measured with linear scores. Although the scarcity of the QTL detected for morphological traits in Murciano-Granadina goats could be due to many technical and biological reasons, being limited sample size one of the most important ones, our results and those reported by other authors (Martin et al., 2018; Mucha et al., 2018) suggest that morphology traits in goats have a highly polygenic architecture represented by a large number of variants with small effects. Indeed, the genetic analysis of stature in cattle (Bouwman et al., 2018) showed that the lead variants in 163 significantly associated genomic regions explained at most 13.8% of the total phenotypic variance, thus suggesting the existence of a large number of mutations with small or very small effects on this trait that remained undetectable despite the large size (58,265 cattle) of the population used in this study. This polygenic nature hinders the identification of genomic regions responsible for the variability of morphological traits due to the small contribution of each variant. In birds, for instance, the heritability of morphological traits is distributed along the genome and positively correlated with the length of the chromosome (Silva et al., 2017), indicating a contribution of genes mapping to multiple locations.

4.1.3 Low and medium-density panels limit the detection of genomic regions associated with traits of economic interest in dairy goats.

One aspect that might have limited the detection of QTL in studies 1 and 2 is the fact that we have used a chip with only 50,000 SNPs. Using an SNP array of

medium density affects the capacity to detect genomic signals in GWAS analysis compared with the usage of high-density chips (hundreds of thousands or millions of SNPs) or whole genome sequences (Tam et al., 2019). As an example, Meredith et al. (2013) performed a GWAS for somatic cell score in 702 Holstein sires by using the BovineHD BeadChip (770,000 SNPs) and compared such results with those obtained with the 50K chip. In the Figure below (**Figure 4.3**), it can be observed that the definition of the chromosome 6 QTL detected for somatic cell score is much preciser and the significance much higher when using the high-density chip that when using the 50K chip.

Figure 4.3. Comparisson of QTL detected for somatic cell score using (A) BovineHD BeadChip (770,000 SNPs) and (B) BovineSNP50 BeadChip (50,000 SNPs). Adapted from Meredith et al. (2013).



Linkage disequilibrium patterns vary across the genome and populations and medium-density SNP arrays might not have enough power to detect significant associations as they overestimate the extent of linkage disequilibrium across markers (Qanbari, 2020). To increase the number of markers, imputation of the SNP

arrays genotype to a high-density panel or whole genome sequences should be performed. In principle, this methodology increases the statistical power of the GWAS and does not require the sequencing of all the individuals included in the assay, but only of a reference population (Howie et al., 2009; Quick et al., 2020). A genotype imputation strategy based on the sequencing of a large number of Murciano-Granadina goats could be very helpful to significantly increase the resolution of the two GWAS (milk and morphological traits) carried out in this thesis, making it possible to increase QTL discovery (although this is also largely dependent on population sample size). The availability of high-density density genotyping SNP panels for goats would also facilitate the implementation of two-step imputation and increase imputation accuracy (Brøndum et al., 2014). Future work, after this thesis, will involve the whole-genome sequence imputation of the genotyped Murciano-Granadina goats to perform QTL discovery and fine mapping, and to integrate such data with molecular information generated through a variety of techniques including ATAC-seq, Methyl-seq, RNA-Seq, and small RNA-Seq in order to identify potential causal mutations.

4.2 THE GENOMIC BASIS OF VIABILITY TRAITS IN MURCIANO-GRANADINA GOATS.

4.2.1 The effect of inbreeding depression on the performance of Murciano-Granadina goats.

Doekes et al. (2021) made a systematic revision of research articles (published from 1990 to 2021) describing the consequences of inbreeding depression on livestock

populations. Out of 154 studies included in this review, only eight were performed in goats, while 65 and 31 focused on cattle and sheep, respectively (Doekes et al., 2021). In **study 3** of this thesis, we aimed to characterise the inbreeding levels of Murciano-Granadina goats as well as to measure the impact of inbreeding depression on milk traits.

The inbreeding coefficient based on ROH (F_{ROH}) estimated in this thesis for Murciano-Grandina goats ranged from 0.001 to 0.402 with an average of 0.054 ± 0.046. This value contrasts strongly with the genealogical inbreeding coefficient reported by Deroide et al. (2016) for Murciano-Granadina goats, which reached a much lower value (F = 0.0024). Although genealogical inbreeding coefficients are often lower than the molecular ones, because ancient inbreeding might not be properly taken into account, this discrepancy is quite remarkable and difficult to explain. If we compare the F_{ROH} coefficient obtained for Murciano-Granadina goats with those obtained for a very diverse set of goat populations characterized in the Adaptmap project (Stella et al., 2018), it is evident that they are quite similar (Bertolini et al., 2018) i.e. ~ 60% of the breeds analyzed in the Adaptmap project displayed low F_{ROH} coefficients (< 0.10), while ~ 30 and ~ 10% of the goat populations showed moderate (F_{ROH} 0.10 – 0.20) or high (> 0.20) F_{ROH} values. More importantly, the lower F_{ROH} values detected in the populations analysed in Adaptmap were 0.02. We conclude that, in terms of inbreeding, the Murciano-Granadina breed is placed at the lower range of what has been reported in a worldwide sample of goats by Bertolini et al. (2018).

Studies in popular dairy cattle breeds report higher inbreeding coefficients than the one calculated by us in Murciano-Granadina breed, as depicted in **Table 4.1** (Dadousis et al., 2022; Zhang et al., 2015a). Mean inbreeding values based on ROH using medium-density panels displayed values up to 0.150 and 0.07 for Holstein and Jersey cattle, respectively (Dadousis et al., 2022; Zhang et al., 2015a). The estimation of F_{ROH} using whole genome sequences yielded higher inbreeding values than the ones estimated in the same populations using medium-density chips, highlighting the effect of marker density on the detection of ROH and, in consequence, on the estimation of inbreeding F_{ROH} coefficients (**Table 4.1**).

This difference in inbreeding levels between dairy cattle and goats can be attributed to causes related with the organization of the breeding schemes and selection intensity. As reviewed by Brito et al. (2021), the rate of inbreeding in dairy cattle has increased over time (**Figure 4.4**) because genomic selection and reproductive biotechnologies, like artificial insemination, embryo transfer and in vitro fertilisation, have increased selection intensity and reduced the number of parents, particularly sires (Weigel, 2001). As an example, popular Holstein sires have over 250,000 daughters and 3,000 tested sons worldwide (Weigel, 2001). This circumstance decreased the genetic variability of dairy cattle populations and augmented the inbreeding rate. Noteworthy, Makanjuola et al. (2020) estimated an effective population size (N_e) ranging from 43 to 66 animals for the Holstein breed and from 64 to 85 animals for Jersey cattle (both populations belonged to the Canadian Dairy Network).

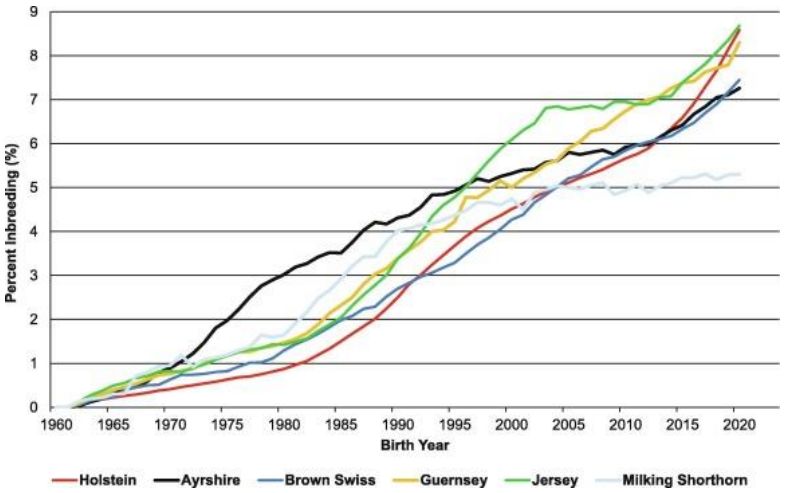
Table 4.1. Inbreeding coefficients based on pedigree records (F_{PED}) and runs of homozygosity (F_{ROH}) estimated on specialised dairy cattle breeds. Information was obtained from Signer-Hasler et al. (2017), Dadousis et al. (2022) and Zhang et al. (2012).

Population	Study	Sample size	Genotyping panel/Sequencing	F_{PED}	F_{ROH}
Brown Swiss	(Signer-Hasler et al., 2017)	281	50K SNPs chip	0.071 ± 0.037	0.091 ± 0.029
Original Branvieh	(Signer-Hasler et al., 2017)	167	50K SNPs chip	0.059 ± 0.023	0.074 ± 0.028
Simmental	(Signer-Hasler et al., 2017)	248	50K SNPs chip	0.092 ± 0.030	0.039 ± 0.023
Holstein	(Signer-Hasler et al., 2017)	2,568	50K SNPs chip	0.057 ± 0.066	0.058 ± 0.025
Italian Holstein	(Dadousis et al., 2022)	F_{PED} N=393,607 F_{ROH} N=95,540	Imputation to 84K SNPs	0.073 ± 0.024	0.150 ± 0.041
Holstein	(Zhang et al., 2015a)	32	50K SNPs chip	0.036	0.066
Jersey	(Zhang et al., 2015a)	27	50K SNPs chip	0.018	0.070
Holstein	(Zhang et al., 2015a)	32	WGS	0.036	0.187
Jersey	(Zhang et al., 2015a)	27	WGS	0.018	0.242

In strong contrast with cattle, dairy goats are usually raised under a semi-extensive management regime. In fact, the average herd size of Murciano-Granadina goats in Spain is 625 heads per farm. According to official reports from the Spanish Ministry of Agriculture, Fisheries and Food, during 2021 only 60 males were used in artificial insemination, while 4,580 sires were used for natural service (https://servicio.mapa.gob.es/arca/flujos.html?_flowId=datosCensalesRaza-flow&tipoOperacion=CONSULTA&id=50154&isMapa=1&formatoPagina=0). Certain anatomical factors and low seminal motility of the sperm of bucks make difficult the implementation of artificial insemination in goats, causing low

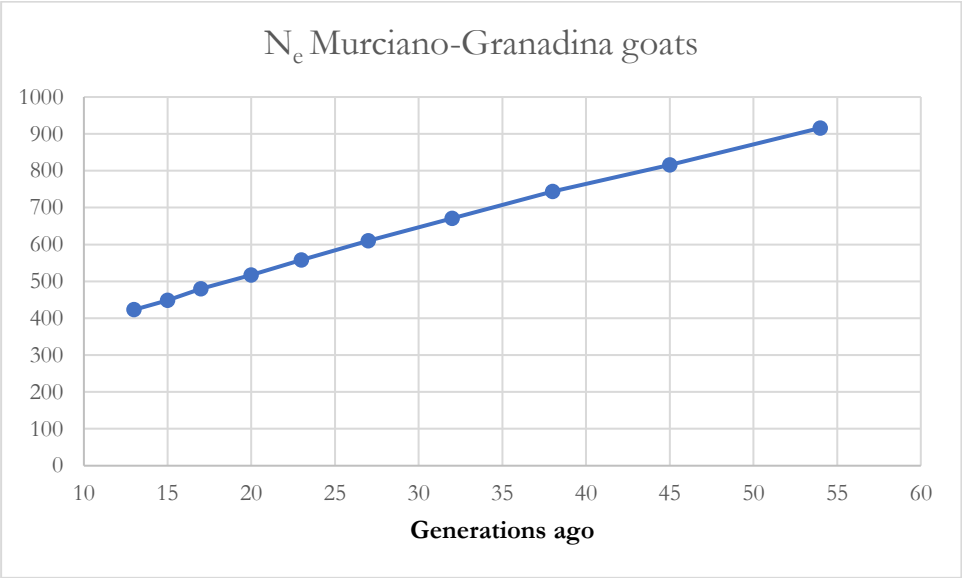
pregnancy rates and economic losses (Mocé et al., 2022). These factors reduce the selection intensity in dairy goats, preventing the usage of a very small number of sires for reproduction purposes and maintaining a higher genetic diversity and lower inbreeding levels than in dairy cattle. According to our results, Murciano-Granadina goats displayed a moderate number of ROH per individual (< 50 ROH). The majority of the ROH were of short length (< 10 Mb), thus suggesting a low amount of recent inbreeding (Thompson, 2013). For this population, we estimated the effective population size with the software SNeP1.1 based on available genotypic data (Barbato et al., 2015). Thirteen generations ago, effective population size was 423 (**Figure 4.4**), a figure considerably larger than those described for dairy cattle.

Figure 4.4. Inbreeding percentage trends of popular specialised dairy cattle breeds in the United States of America. Data published in Brito et al., (2021).



*They obtained it from Council on Dairy Cattle Breeding (https://queries.uscdcb.com), August, 2020.

Figure 4.5. Effective population size (Ne) estimated for 13 to 54 generations ago from the population of 1,040 female Murciano-Granadina goats that was analysed in this thesis.



4.2.1.1 Potential implications of inbreeding depression for milk somatic cell score in Murciano-Granadina goats

Although inbreeding coefficients in the Murciano-Granadina population under study were generally low, we detected a significant effect of inbreeding depression on milk somatic cell count (SCC, computed as the natural logarithm of the somatic cell count divided by 1000, lnSCC). Inbreeding analyses based on molecular data (see **study 3**) evidenced the existence of inbreeding depression for lnSCC and a couple of regions on chromosomes 8 and 25 were associated with it. According to Miglior et al. (1995), a 10% increase of the inbreeding coefficient was associated with a 10.5% increase of the original phenotypic standard deviation of lactation

somatic cell score recorded in Holstein cattle. Although in this case the effect of inbreeding depression on SCC was relatively modest, on average, inbred animals tended to have higher lactation somatic cell scores (Miglior et al. 1995). In Dutch Holstein-Friesian cattle, a 1% increase in F_{ROH} was associated with a 0.86 unit (SE = 0.28) increase in somatic cell score for day 150 through to 400 (Doekes et al., 2019). Such effect was equivalent to 0.05% of the trait averages, respectively. Moreover, recently generated inbreeding was more harmful than the ancestral one, especially for phenotypes related to milk yield (Doekes et al., 2019). Similarly, in Holstein cattle from Iran highly inbred individuals tended to have higher somatic cell scores than those with low inbreeding coefficients (Rokouei et al., 2010).

About the potential implications of inbreeding depression for SCC, several considerations need to be made. Genomic regions displaying inbreeding associated with lnSCC in Murciano-Granadina goats were enriched in genes with immunological functions, e.g. chemokine ligand 19 (*CCL19*), 21 (*CCL21*), 24 (*CCL24*), 26 (*CCL26*) and 27 (*CCL27*); and interleukin 27 (*IL27*). Inbreeding and the consequent decrease of genomic variation, particularly on immune genes, might

have had detrimental effects on natural resistance to pathogens. For instance, cattle with high endogamy levels presented lower levels of γδ T lymphocytes in peripheral blood than calves with lower endogamy (Macedo et al., 2014). Besides, genetic diversity of the major histocompatibility complex (MHC) is essential to the persistence of wild populations and the health of domestic and human populations due to the function of this complex in antigen presentation and the activation of the immune response (Gutiérrez-Reinoso et al., 2022; Kardos et al., 2021; Teixeira and

Huber, 2021). This means that increased homozygosity, the main outcome of inbreeding, lowers the variability of immune-related genes, decreasing the natural resistance of individuals to pathogens, and increases the emergence of recessive hereditary diseases (Lie et al., 2009).

Higher levels of somatic cells are expected in goats, when compared with cattle, due to the apocrine secretion of milk (Paape et al., 2001) and the increment of pro-inflammatory cells (mainly leukocytes) during late lactation in goats (Silanikove et al., 2010). In this species, somatic cell counts above 300,000 cells/mL in the first lactation indicate infection of the udder and leads to a decrease in milk production and alteration of the milk composition (Leitner et al., 2011; Silanikove et al., 2010). As reviewed by Sharma et al. (2011), SCC is an indicator of mammary gland health. An increase in the SCC is often explained by intramammary infections (Sharma et al., 2011). An analysis performed on large dairy goat herds in France points out to udder disorders, including mastitis, as the most frequent cause of culling and mortality, representing over 18% of the causes of dismissal of individuals (Malher et al., 2001). This leads to significant economic losses because of the decrease in milk production, changes in milk composition that affect cheesemaking, and the cost of treatment and culling of individuals. Indeed, modifications of the milk composition alter the clotting properties of milk, negatively affecting the curd firmness, which is of great importance in dairy goat production since cheese is its main commercial product (Leitner et al., 2011). In this way, Silanikove et al. (2010) have proposed different degrees of intramammary infection based on the somatic cell counts and the impact on milk quality, the effect of each category of somatic

cell counts is summarised in **Table 4.2**. Moreover, a significant effect of inbreeding was detected over milk yield, i.e. a decrease ranging from 48.50 to 5.01 and 20.49 to 23.19 Kg of milk at three different timepoints of lactation was reported for the increase in 1% of F_{HOM} and F_{ROH}, respectively. Similar results were reported in cattle by Doekes et al. (2019) that reported a decrease in 36.3 kg of milk at day 305 for every 1% of F_{ROH} increment.

In summary, although inbreeding levels in the Murciano-Granadina goat population under study are low, it is important to implement strategies to keep them at a minimum or, otherwise, unwanted consequences on the incidence of mastitis and milk quality might arise.

Table 4.2. Classification of intramammary infections based on somatic cell count (SCC) and percentage of animals affected, and the estimated losses in terms of milk production and curd formation obtained from Silanikove et al. (2010).

Grade	SCC (cells/mL)	Herd affected (%)	Milk loss (%)	Curd loss (%)
A	≤ 840,000	25	0.8	3.3
B	> 840,000 and < 1,200,000	50	1.5	6.5
C	> 1,200,000 and < 3,500,000	75	2.3	9.8
Not accepted on market*	> 3,500,000			Very bad

*High probability of milk containing pathogens and toxins that might be harmful to human consumption.

4.3 THE MICROBIAL COMPOSITION OF GOAT CHEESES.

Being the main commercial product derived from the milk of dairy goats, the factors that determine the technological and organoleptic properties of cheese have become an important subject of study (Pulina et al., 2018). In **study 1**, we analysed the genomic basis of milk composition that has an important impact on the manufacturing of cheese. Indeed, the solid content of milk (protein, fat, lactose) and the abundance of somatic cells affect curd formation. In **study 4,** the microbial composition of a set of goat cheeses has been investigated by analysing sequences from the ultravariable V3-V4 regions of the 16S rRNA gene. In comparison with cattle, the microbiome of goat cheeses has been poorly characterized. Few studies have reported a description of the microbial communities of goat cheeses and they have focused on very specific kind of cheeses like Pélardon cheese (Penland et al., 2021) and Gouda cheese (Salazar et al., 2018).

In our study, we investigated the microbiomes of 6 types of cheeses, i.e. two fresh and soft-paste cheeses (H and M), two semi-cured, soft-paste cheeses with bloomy rind (C and P), one semi-cured hard-paste cheese (B) and one semi-hard cured cheese (G). As expected, we observed a prevalence of LAB, mostly from the genera *Lactococcus*, *Lactobacillus*, *Leuconostoc* and *Streptococcus*, in all six cheeses. This result is congruent with other studies performed on a wide variety of cheeses (Choi et al., 2020b; Walsh et al., 2020) and with the fact that LAB are often used as starters in cheesemaking to induce curd formation (Quigley et al., 2013).

When comparing the alpha diversity of microbial communities in the three different cheese regions sampled (core, middle part and rind) on a pairwise basis, no significant differences were found. According to a previous report (Choi et al., 2020b), higher alpha diversity is commonly observed in the rind due to its exposure to oxygen and environmental microorganisms, allowing the growth of aerobe taxa that cannot thrive in the inner portions of the cheese because of the lack of oxygen. Other factors that increase the microbial diversity of the rind are related with the manufacturing process and include exposure to environmental factors, like brine or alcohol soaking (Choi et al., 2020b; Irlinger et al., 2015), and also the growth of other microorganisms, like fungi, in the bloomy rind cheeses thus favouring the expansion and dispersion of certain bacterial strains (Zhang et al., 2018). However, in our study the microbial composition and diversity of the rind was not strongly different from that of the middle and core. Although the sample size of the study is very small, we observed an effect of the ageing of the cheese on microbial diversity, with higher alpha diversity in samples from fresh cheeses than cured and semi-cured ones. In Cheddar cheeses, Choi et al., (2020a) reported that microbial diversity was higher in the milk and curd, previous to the addition of LAB starters, than afterwards, and they also observed that diversity tends to decrease along the manufacturing and ageing process.

Besides the organoleptic and technological consequences of microbial composition on cheese, microorganisms can also represent an important source of contamination that puts in risk the quality of the cheese and the health of the consumer. In this way, we detected high relative abundances of psychrophilic bacteria in fresh soft

cheeses H and M, that presented up to 38.9% and 57.47% of relative abundance of *Pseudomonas*. Semi-ripened cheeses P and C showed up to 18.85% of abundance of bacteria from the genus *Pseudomonas*. *Pseudomonas* bacteria have been found in multiple cheese rinds, especially bloomy, as they easily move along the fungal network (Zhang et al., 2018). This genus presents moderate positive Pearson correlations ($r= 0.4$, p-value < 0.05) with the moisture of the cheese (Wolfe et al., 2014), explaining their high proportions in the soft-paste cheeses investigated in our study (H, M, C and P). Controlling cheese contamination by *Pseudomonas* is important to assure the quality and shelf-life of the product. The waste of food induced by microbial spoilage has a negative impact on the food industry. Microbial contamination often causes heavy economic losses by driving undesirable changes in the appearance, texture, flavor, and odor that reduce food quality (Martin et al., 2011). *Pseudomonas* spp. can also have detrimental effects on refrigerated foods because many strains are psychrotolerant. All the steps in the productive chain need to be thoroughly controlled because *Pseudomonas* and other psychrophilic bacteria are common contaminants of dairy foods during the post-pasteurisation processes (De Jonghe et al., 2011).

The main limitation of our microbiome study was reduced sample size. We had just one piece per type of cheese and three different regions of each cheese were sampled (core, middle and rind),. However, microbiota can change in the same type of cheese across pieces or production batches due to the variability of the bacterial content in the facilities, milk, storage, among other factors and technical specificities of the productive chain (Johnson et al., 2021). Given this drawback,

we oriented our study to determine what the microbiomes of the six types of goat cheeses have in common rather than assessing what is different. In the future, the retrieval of a larger cheese sample size and the complementation of bacterial sequencing with other techniques, like metatranscriptomics and volatile compound analysis, should provide a broader perspective not only about the bacterial composition of cheeses, but also about the host-microbiome interactions and their effect on the sensorial properties of cheese. Integrating such data might be helpful to select the appropriate bacterial strains in order to ensure the optimal quality of cheese (Solieri et al., 2013; Yeluri Jonnala et al., 2018).

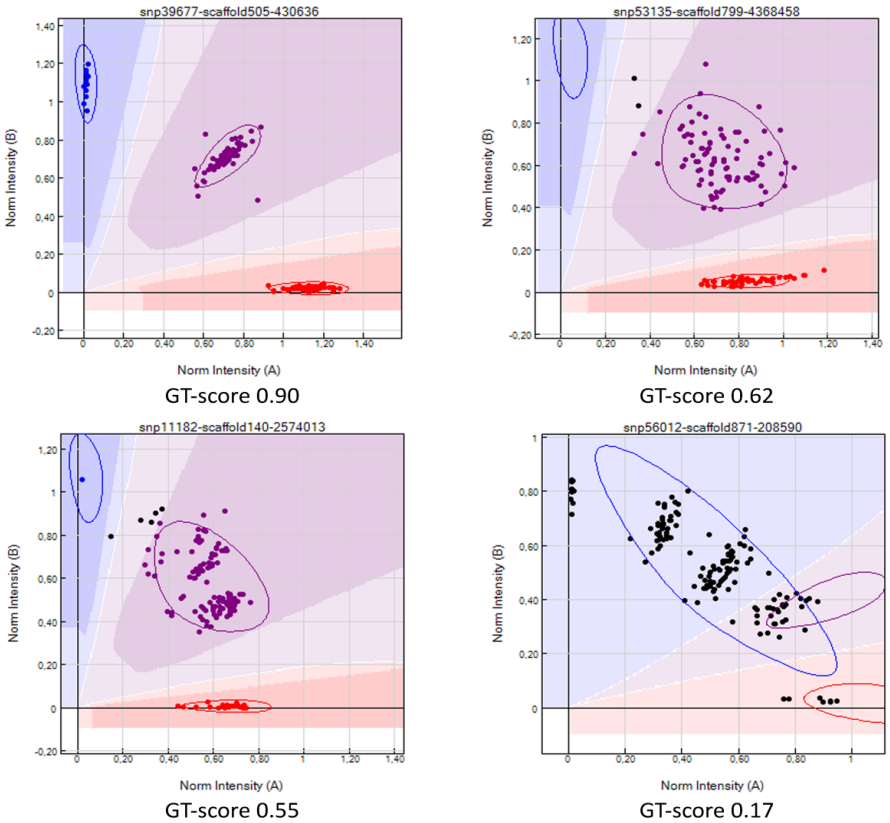
4.4 IDENTIFICATION OF TRANSMISSION RATIO DISTORTION IN MURCIANO-GRANADINA GOATS.

Transmission ratio distortion (TRD) takes place when the two alleles from either parent do not have equal probabilities of being transmitted to the offspring, thus leading to a statistical departure from the Mendelian law of inheritance (Huang et al., 2013). This phenomenon can have multiple causes, including meiotic drive, gametic competition, and embryo lethality (Huang et al., 2013). Transmission ratio distortion is poorly characterised in livestock species (Abdalla et al., 2020; Casellas et al., 2020, 2014; Lahoucine et al., 2020; Vázquez-Gómez et al., 2020) and has not been studied in goats before. In **study 5** of this thesis, we aimed to identify genomic regions containing genetic markers the segregation of which, in heterozygous individuals, deviates from the expected Mendelian 1:1 ratio. The design of our TRD study comprised 17 single-parent families (the genotype of the mother was not

available) formed by one sire and their offspring (from 10 to 30, and 18 per sire on average) All individuals were genotyped with a medium-density GoatSNP50 Illumina Beadchip. Then we used the method reported by Casellas et al. (2012) to detect TRD on a genome-wide basis. This method can be used even when the genotype of one of the parents is missing (the probability that the missing parent has a given genotype for a SNP is inferred on the basis of its frequencies in the general population). Only 36 SNPs showed evidence of being subjected to TRD. It is well known that genotyping errors are a frequent cause of detecting strong Hardy-Weinberg equilibrium departures (Hosking et al., 2004), so we wondered whether our results could be the consequence of genotyping problems rather than of a biological phenomenon leading to the unequal transmission of alleles from parent to offspring. As a quality control measure, we checked the Gentrain scores of these 36 SNPs. The GenTrain score reflects the shape of the genotype clusters and the relative distance between the called clusters and it goes from 0 (very bad quality due to poor cluster differentiation) to 1 (excellent quality). By doing so, we found out that 25 SNPs had low scores (<0.8). In **Figure 4.5** the clustering of the SNPs with different GenTrain scores based on the normalised intensity values of the genotyping assay from Illumina is depicted. The genotypic assignment of variants with GenTrain scores below 0.62 is unclear. In the light of these results, we were afraid that most of the TRD detected in the Murciano-Granadina population of 305 offspring was an artifact produced by errors during the genotyping process, casting doubts about the validity of the whole analysis. We were also aware that a family size of 18 individuals and the lack of genotypic information from dams severely

limited our ability to identify TRD signals with enough confidence. Due to these drawbacks, we just did a technical discussion of the TRD results by emphasizing that the establishment of stringent filters for genotype quality is of paramount importance when carrying out TRD analysis, since genotyping errors can be an important source of spurious TRD signals.

Figure 4.5. Genotype cluster plot based on normalised intensities of genotyping assays generated with the Illumina Goat SNP50 BeadChip (Illumina Inc., San Diego, CA) for SNPs with different GenTrain (GT) scores in a population of Murciano-Granadina goats.



4.5 PREGNANCY INDUCES CHANGES IN THE PROFILE OF MRNA EXPRESSION OF GOAT BRAIN TISSUES.

A fundamental trait that affects offspring survival is maternal behaviour. Maternal care in mammals is provided solely by the female and comprises physiological factors like nutrition during pregnancy, thermoregulation, licking, milking, immunological protection via colostrum, and psychological aspects like learning, protection, and guidance (Dwyer, 2014). A series of neuro-endocrine changes occur in females, during gestation and parturition, triggering maternal behaviour, milk production and other physiological changes (Dwyer, 2014). In **study 6** of this thesis, we have characterized the patterns of mRNA expression of 12 goat brain tissues and we have investigated differential expression in pregnant (N=3) vs non-pregnant goats (N=4).

Between 14,889 (cerebellar hemisphere) and 15,592 (pineal gland) protein-coding genes were expressed in goat brain tissues, and most of them were expressed in multiple tissues. The analysis of the patterns of mRNA expression by principal component analysis and hierarchical clustering demonstrated that caprine brain tissues tend to group according to their embryonic origin. For instance, olfactory bulb, hippocampus, and frontal neocortex, which are derived from the telencephalon vesicle, clustered together, and pons and medulla oblongata (both originate from myelencephalon) grouped in another cluster. Exceptions to this general trend were glandular (hypophysis and pineal gland) and cerebellum tissues, that showed highly divergent mRNA expression profiles when compared to other

tissues (maybe because of their high functional specialization). Zapala et al. (2005) investigated gene expression in 24 neural murine tissues by using microarrays and observed that embryonic cellular position along the anterior–posterior axis influenced the gene expression patterns in adult structures. Moreover, a substantial number of genes related with embryonic development (e.g. homeobox transcription factors) displayed region-specific expression in the adult nervous system (Zapala et al. 2005). These results suggested that the brain of the adult mouse is imprinted to some extent by the initial pattern of gene expression established during embryogenesis, and that such "imprinting" plays a key role in setting up regional specificity and functional links between regions in the adult brain (Zapala et al., 2005). Ortiz et al. (2020) also found evidence that the mRNA expression profiles of a reduced set of genes is enough to capture the spatial complexity of the murine brain, and, more importantly, they demonstrated that such set of loci is enriched in developmental genes, thus suggesting that developmental axes are fundamental to determine adult brain compartmentalization. In another study, Ferraz et al. (2008) showed that tissue of origin is the main factor determining the clustering of samples from different tissues and observed that tissues from the same germ layer (ectoderm, mesoderm or endoderm) tended to cluster together. Globally, these findings provide support to the idea that the patterns of gene expression that are established during the development of goat embryos leave a durable footprint on the mRNA expression profiles of the adult goat brain.

In this work, we had the opportunity to sample pregnant and non-pregnant goats that were going to be culled due to reasons unrelated with our research. We

performed an exploratory differential expression analysis although we were aware that sample size (N=3 – 4) was quite low and limited our ability to detect truly differentially expressed genes. In this regard, it is known that when the number of replicates per group decreases then the number of *genes called significant* declines steadily, with evident changes when this number is equal or lower than eight (Baccarella et al., 2018). Because of the low number of replicates per group used in **study 6**, we decided that it was appropriate to use the number of differentially expressed genes as a signal of the "activation" of the corresponding brain region in response to 1 month-gestation, and at the same time we avoided to put much emphasis on discussing the differential expression of specific genes. Differential expression analyses performed on brain tissues from 1-month pregnant versus non-pregnant goats revealed an "activation" of the adenohypophysis, frontal neocortex, hippocampus, olfactory bulb, pineal gland and pons, which displayed from 62 to 1,207 genes differentially expressed between groups, being the olfactory bulb the anatomical structure with the highest number of differentially expressed genes. The "activation" pattern of these tissues is attributable to pregnancy, as these six tissues come from different neural vesicles and they are not closely located.

Despite the small sample size, we wanted to better understand the dramatic activation of the olfactory bulb in pregnant goats, so we performed an intensive bibliographic revision to retrieve functional information reported for the differentially expressed genes in this brain structure. This analysis was not discussed in the paper corresponding to **study 6** due to reasons stated before (low replicate number), but it would be worth to briefly comment it herewith. The

olfactory bulb plays an important role in processing olfactory information by integrating stimuli from the peripheral and central neural systems (Huart et al., 2013). Interestingly, many of the genes differentially expressed in pregnant vs non-pregnant goats are involved in human behavioural traits, including maternal behavior (e.g. downregulated: *DDC*; upregulated: *DBH, DRD1, NTS* and *HTR2A*), affective behavior, sociability and exploration (e.g. downregulated: *FGFR4, GRIP2, NTRK2* and *RYR3*; upregulated: *HTR2C, FBXO45, GPR3, KCNQ2* and *PLXNA2*), anxiety and depression (e.g. downregulated: *NO2, PAN2* and *H3-3B*; upregulated: *BSCL2, HRH1, SIK2, PDYN, GLRB* and *NRN1*), autism (e.g. downregulated: *MOCOS, CACNA1D, MBD6* and *AUTS2*; upregulated: *RAB39B, BTBD11, KCNQ3, CDH9, CADPS2, DOCK4* and *EXT1*), aggression (e.g. downregulated: *NOS1*; upregulated: *HRH3, PRNP, HNMT* and GRIA3), cognition, memory and learning (e.g. downregulated: *EPHA10, STAT5* and *MMP28*; upregulated: *EPHA6, SYP, SORBS2, ARHGEF4, HCN1, CAMK2N2, MMP17, BTBD9, CLSTN3, STAU2RIMKLA, PAK6, SLC22A4* and *NEURL1*), response to stress (e.g. downregulated: *IFIT1*; upregulated: *PDYN, EPOP, DPYSL2* and *HCN2*), feeding behavior (e.g. upregulated: *NELL2, CXCL14, GPR45, GPR162, NPY* and *ACBD7*) and diverse neuropsychiatric disorders (e.g. downregulated: *SLITRK6, IL1RAPL1, BAHCC1, WDR62* and *SLC6A1*; upregulated: *SCN1A, GABBR2, RTN4R, ACOT7, KIF5A. HECW2, NRG1, SNCA, ATP8A2, WFS1, HTR5A, STX1A* and *JPH3*).

The role of the olfactory bulb in maternal behaviour has been documented in multiple species (Dwyer, 2014; Gandelman et al., 1971; Navarro-Moreno et al.,

2020; Poindron et al., 2007a), including goats (Poindron et al., 2007b). What is particularly interesting from our results is the marked changes of expression that we observe in this brain area at a very early stage of gestation. As said, three of the goats included in this analysis were 1-month pregnant, while goat gestation is approximately 150 days long. Obviously, the change of mRNA expression of the olfactory bulb observed by us is not triggered by parturition or the odour of the pups, two activation stimuli that have been reported previously (Dwyer, 2014; Poindron et al., 2007a). One explanation to such change would be the direct or indirect delivery of hormonal or any other biochemical signals modifying gene expression in the olfactory bulb. During the early gestation of mice and rats, there is a process of neurogenesis induced by prolactin in the forebrain that also affects the olfactory bulb (Larsen and Grattan, 2010; Shingo et al., 2003). The enhancement of neurogenesis in the subventricular zone and olfactory bulb mediated by prolactin influences multiple reproductive behaviors including mating/pregnancy, dominant male pheromone preference in females, and paternal recognition of offspring (Wang et al., 2013).

Of course, the involvement of prolactin, or any other biochemical signal, on the modulation of olfactory bulb gene expression in response to gestation is just a matter of speculation. Further studies are needed to better understand the causes and biological implications of the changes in olfactory bulb mRNA expression elicited by gestation. A confirmatory study with a substantially higher sample size would help to analyse the gene and transcript levels of the different goat brain regions affected by pregnancy in a more reliable way (Ching et al., 2014).

Furthermore, in relation to pregnancy studies, the recording of hormonal concentrations in plasma might help to correlate the expression changes in the brain with hormonal changes generated by pregnancy.

5 CONCLUSIONS

1. The longitudinal GWAS of milk yield and composition traits recorded in Murciano-Granadina goats for three lactations made possible to detect as much as 31 QTL (eight with genome-wide significance). In contrast, 22 QTL (6 with genome-wide significance), 29 QTL (3 with genome-wide significance) and 9 QTL (1 with genome-wide significance) were detected when performing independent GWAS for each one of the three lactations. The most significant and consistent QTL was associated with milk protein percentage and mapped to the casein gene cluster on chromosome 6, demonstrating that the casein genotype is a key determinant of milk quality in Murciano-Granadina goats.
2. The GWAS analysis of morphological traits in Murciano-Granadina goats yielded few significant associations between genetic markers and the studied traits, for instance only two SNPs reached genome-wide significance, i.e. markers rs268273468 (CHI 16:69617700) and rs268249346 (CHI 28:18321523) significantly associated with medial suspensory ligament. There was low positional concordance with morphology QTL reported in the Saanen and Alpine breeds and those detected in Murciano-Granadina goats. These findings are compatible with the existence of genetic heterogeneity and a highly polygenic background regulating morphological traits in goat breeds, although technical factors might be also at play.
3. The analysed population of Murciano-Granadina goats displayed in general low levels of inbreeding (mean F_{ROH} = 0.053 ± 0.04), likely because of its

demographic history and semiextensive managementIn spite of this, inbreeding depression was significant for somatic cell score in milk and milk yield at 210, 240 and 305 days of lactation from Murciano-Granadina goats. Genomic regions CHI 8:37–56 Mb, CHI 8:75–93 Mb, CHI 25:0.08–7 Mb and CHI 25:21–28 Mb were significantly associated with the inbreeding depression for somatic cell score and CHI 2:22.75–68.25 was significantly associated with inbreeding for milh yield at three time points mentioned below. We conclude that inbreding levels needs to be kept at a minimum in Murciano-Granadina goats to avoid its detrimental effects on the sanitary status of the mammary gland.

4. We have characterized the microbiota of six commercial types of goat cheeses and by doing so we have found a predominance of lactic-acid bacteria (i.e. *Lactibacillaceae* and *Streptococcaceae*), which are commonly used as starter cultures during the cheesemaking process to improve curd formation and prevent the growth of pathogens. Psychrophilic bacteria, including *Pseudomonas*, *Shewanella* and *Pseudoalteromona*, associated with post-pasteurization contamination were idenfied in fresh cheeses (>50% of abundance), while *Pseudomonas* were also prevalent in semi-cured mouldy rind cheeses (up to 18% of abundance).
5. Genome-wide detection of genetic markers affected by transmission ratio distortion (TRD) in a population of 17 sires and their offspring (N=305) made possible to identify 36 SNPs deviating from the Mendelian expected 1:1 ratio. The majority of these SNPs had low GenTrain scores, so TRD

signals are probably caused by genotyping problems rather than biological factors. These findings stress the need to set high standards for genotyping quality parameters when performing TRD scans in animal populations.

6. We have sequenced the transcriptomes of 12 brain tissues in seven goats, of which three were pregnant. The clustering of brain tissues based on gene expression was very consistent the embryonic encephalic vesicle of origin, with the only exception of tissues specialized in hormone secretion (i.e. pineal gland and hypophysis) and cerebellum. Embryonic development might leave a durable footprint on gene expression that persists even in adulthood. Moreover, early pregnancy (1 month) in goats induced changes in the transcriptome profile of six encephalic tissues, i.e. adenohypophysis, frontal neocortex, hippocampus, olfactory bulb, pineal gland and pons, being the olfactory bulb the most affected tissue(1,207 differentially expressed genes when comparing pregnnant and non-pregnant goats) This finding is consistent with the fundamental role of the olfactory bulb in the development of maternal care in many mammalian species.

6 REFERENCES

Abdalla, E.A., Id-Lahoucine, S., Cánovas, A., Casellas, J., Schenkel, F.S., Wood, B.J., Baes, C.F., 2020. Discovering lethal alleles across the turkey genome using a transmission ratio distortion approach. Anim. Genet. 51, 876–889. https://doi.org/10.1111/age.13003

Abrams, Z.B., Johnson, T.S., Huang, K., Payne, P.R.O., Coombes, K., 2019. A protocol to evaluate RNA sequencing normalization methods. BMC Bioinformatics 20, 1–7. https://doi.org/10.1186/S12859-019-3247-X/FIGURES/3

Agarwal, G., Kudapa, H., Ramalingam, A., Choudhary, D., Sinha, P., Garg, V., Singh, V.K., Patil, G.B., Pandey, M.K., Nguyen, H.T., Guo, B., Sunkar, R., Niederhuth, C.E., Varshney, R.K., 2020. Epigenetics and epigenomics: Underlying mechanisms, relevance, and implications in crop improvement. Funct. Integr. Genomics 20, 739–761. https://doi.org/10.1007/S10142-020-00756-7/FIGURES/3

Agrawal, A.F., Whitlock, M.C., 2012. Mutation load: The fitness of individuals in populations where deleterious alleles are abundant. Annu. Rev. Ecol. Evol. Syst. 43, 115–135. https://doi.org/10.1146/annurev-ecolsys-110411-160257

Aitken, N., Smith, S., Schwarz, C., Morin, P.A., 2004. Single nucleotide polymorphism (SNP) discovery in mammals: A targeted-gene approach. Mol. Ecol. 13, 1423–1431. https://doi.org/10.1111/J.1365-294X.2004.02159.X

Allali, I., Arnold, J.W., Roach, J., Cadenas, M.B., Butz, N., Hassan, H.M., Koci,

M., Ballou, A., Mendoza, M., Ali, R., Azcarate-Peril, M.A., 2017. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. BMC Microbiol. 17, 1–16. https://doi.org/10.1186/S12866-017-1101-8/FIGURES/7

Amills, M., Jordana, J., Zidi, A., Manuel, J., 2012. Genetic Factors that Regulate Milk Protein and Lipid Composition in Goats, in: Milk Production - Advanced Genetic Traits, Cellular Mechanism, Animal Management and Health. IntechOpen. https://doi.org/10.5772/51716

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Zech Xu, Z., Kightley, E.P., Thompson, L.R., Hyde, E.R., Gonzalez, A., Knight, R., 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2, e00191-16. https://doi.org/10.1128/msystems.00191-16

An, X., Hou, J., Zhao, H., Zhu, C., Yan, Q., Song, Y., Wang, J., Cao, B., 2012. Mutations in Caprine *DGAT1* and *STAT5A* genes were associated with milk production traits. Engineering 4, 30–34. https://doi.org/10.4236/eng.2012.410b008

Analla, M., Jiménez-Gamero, I., Muñoz-Serrano, A., Serradilla, J.M., Falagán, A., 1996. Estimation of genetic parameters for milk yield and fat and protein contents of milk from Murciano-Granadina goats. J. Dairy Sci. 79, 1895–1898. https://doi.org/10.3168/JDS.S0022-0302(96)76558-X

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count

data. Genome Biol. 11, 1–12. https://doi.org/10.1186/GB-2010-11-10-R106/COMMENTS

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S., 2010. FastQC - A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Babraham Bioinforma.

Baccarella, A., Williams, C.R., Parrish, J.Z., Kim, C.C., 2018. Empirical assessment of the impact of sample number and read depth on RNA-Seq analysis workflow performance. BMC Bioinformatics 19, 1–12. https://doi.org/10.1186/S12859-018-2445-2/FIGURES/5

Ballester, M., Sánchez, A., Folch, J.M., 2005. Polymorphisms in the goat beta-lactoglobulin gene. J. Dairy Res. 72, 379–384. https://doi.org/10.1017/S0022029905000981

Ballouz, S., Dobin, A., Gillis, J.A., 2019. Is it time to change the reference genome? Genome Biol. 20, 1–9. https://doi.org/10.1186/S13059-019-1774-4/FIGURES/3

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to

single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/CMB.2012.0021

Barbato, M., Orozco-terWengel, P., Tapio, M., Bruford, M.W., 2015. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front. Genet. 6, 109. https://doi.org/10.3389/FGENE.2015.00109/BIBTEX

Barbieri, M., Manfredi, E., Elsen, J., Ricordeau, G., Bouillon, J., Grosclaude, F., Mahé, M., Bibé, B., 1995. Influence du locus de la caséine αs1 sur les performances laitières et les paramètres génétiques des chèvres de race Alpine. Genet. Sel. Evol. 27, 437. https://doi.org/10.1186/1297-9686-27-5-437

Bateman, A., 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049

Belnoue, L., Malvaut, S., Ladevèze, E., Abrous, D.N., Koehl, M., 2016. Plasticity in the olfactory bulb of the maternal mouse is prevented by gestational stress. Sci. Reports 6, 1–11. https://doi.org/10.1038/srep37615

Benjamini, Yoav ; Hochberg, Y., 1995. Controlling the false discovery rate - A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300. https://doi.org/10.2307/2346101

Beresford, T.P., Fitzsimons, N.A., Brennan, N.L., Cogan, T.M., 2001. Recent advances in cheese microbiology. Int. Dairy J. 11, 259–274. https://doi.org/10.1016/S0958-6946(01)00056-5

Bertolini, F., Cardoso, T.F., Marras, G., Nicolazzi, E.L., Rothschild, M.F., Amills, M., 2018. Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. Genet. Sel. Evol. 50, 59. https://doi.org/10.1186/s12711-018-0424-8

Bickhart, D.M., Rosen, B.D., Koren, S., Sayre, B.L., Hastie, A.R., Chan, S., Lee, J., Lam, E.T., Liachko, I., Sullivan, S.T., Burton, J.N., Huson, H.J., Nystrom, J.C., Kelley, C.M., Hutchison, J.L., Zhou, Y., Sun, J., Crisà, A., Ponce De León, F.A., Schwartz, J.C., Hammond, J.A., Waldbieser, G.C., Schroeder, S.G., Liu, G.E., Dunham, M.J., Shendure, J., Sonstegard, T.S., Phillippy, A.M., Van Tassell, C.P., Smith, T.P.L., 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat. Genet. 49, 643–650. https://doi.org/10.1038/ng.3802

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S.,

Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K. Bin, Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A. V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bordi, A., De Rosa, G., Napolitano, F., Litterio, M., Marino, V., Rubino, R., 1994. Postpartum development of the mother-young relationship in goats. Appl. Anim. Behav. Sci. 42, 145–152. https://doi.org/10.1016/0168-1591(94)90154-6

Boulanger, A., Grosclaude, F., Mahé, M.-F., 1984. Polymorphisme des caséines α(s1) et α(s2) de la chèvre (Capra hircus). Genet. Sel. Evol. 16, 157.

https://doi.org/10.1186/1297-9686-16-2-157

Bouwman, A.C., Daetwyler, H.D., Chamberlain, A.J., Ponce, C.H., Sargolzaei, M., Schenkel, F.S., Sahana, G., Govignon-Gion, A., Boitard, S., Dolezal, M., Pausch, H., Brøndum, R.F., Bowman, P.J., Thomsen, B., Guldbrandtsen, B., Lund, M.S., Servin, B., Garrick, D.J., Reecy, J., Vilkki, J., Bagnato, A., Wang, M., Hoff, J.L., Schnabel, R.D., Taylor, J.F., Vinkhuyzen, A.A.E., Panitz, F., Bendixen, C., Holm, L.E., Gredler, B., Hozé, C., Boussaha, M., Sanchez, M.P., Rocha, D., Capitan, A., Tribout, T., Barbat, A., Croiseau, P., Drögemüller, C., Jagannathan, V., Vander Jagt, C., Crowley, J.J., Bieber, A., Purfield, D.C., Berry, D.P., Emmerling, R., Götz, K.U., Frischknecht, M., Russ, I., Sölkner, J., Van Tassell, C.P., Fries, R., Stothard, P., Veerkamp, R.F., Boichard, D., Goddard, M.E., Hayes, B.J., 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat. Genet. 50, 362–367. https://doi.org/10.1038/s41588-018-0056-5

Boyko, A.R., Quignon, P., Li, L., Schoenebeck, J.J., Degenhardt, J.D., Lohmueller, K.E., Zhao, K., Brisbin, A., Parker, H.G., VonHoldt, B.M., Cargill, M., Auton, A., Reynolds, A., Elkahloun, A.G., Castelhano, M., Mosher, D.S., Sutter, N.B., Johnson, G.S., Novembre, J., Hubisz, M.J., Siepel, A., Wayne, R.K., Bustamante, C.D., Ostrander, E.A., 2010. A simple genetic architecture underlies morphological variation in dogs. PLoS Biol. 8, 49–50. https://doi.org/10.1371/journal.pbio.1000451

Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of

southern Wisconsin. Ecol. Monogr. 27, 325–349. https://doi.org/10.2307/1942268

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. https://doi.org/10.1038/nbt.3519

Brito, L.F., Bedere, N., Douhard, F., Oliveira, H.R., Arnal, M., Peñagaricano, F., Schinckel, A.P., Baes, C.F., Miglior, F., 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. Animal 15, 100292. https://doi.org/10.1016/J.ANIMAL.2021.100292

Brøndum, R.F., Guldbrandtsen, B., Sahana, G., Lund, M.S., Su, G., 2014. Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. BMC Genomics 15, 1–8. https://doi.org/10.1186/1471-2164-15-728/TABLES/3

Buenrostro, J.D., Wu, B., Chang, H.Y., Greenleaf, W.J., 2015. ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 2015, 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Burrows-Wheeler Aligner [WWW Document], n.d. URL http://bio-bwa.sourceforge.net/ (accessed 7.11.22).

Bush, W.S., Moore, J.H., 2012. Chapter 11: Genome-Wide Association Studies. PLoS Comput. Biol. 8, e1002822.

https://doi.org/10.1371/JOURNAL.PCBI.1002822

Caballero, A., Villanueva, B., Druet, T., 2021. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. Evol. Appl. 14, 416–428. https://doi.org/10.1111/eva.13126

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

Cañón, J., García, D., García-Atance, M.A., Obexer-Ruff, G., Lenstra, J.A., Ajmone-Marsan, P., Dunner, S., 2006. Geographical partitioning of goat diversity in Europe and the Middle East. Anim. Genet. 37, 327–334. https://doi.org/10.1111/J.1365-2052.2006.01461.X

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

Capote, J., Argüello, A., Castro, N., López, J.L., Caja, G., 2006. Short communication: Correlations between udder morphology, milk yield, and milking ability with different milking frequencies in dairy goats. J. Dairy Sci. 89, 2076–2079. https://doi.org/10.3168/jds.S0022-0302(06)72276-7

Caravaca, F., Amills, M., Jordana, J., Angiolillo, A., Agüera, P., Aranda, C., Menéndez-Buxadera, A., Sánchez, A., Carrizosa, J., Urrutia, B., Sànchez, A.,

Serradilla, J.M., 2008. Effect of αs1-casein (*CSN1S1*) genotype on milk CSN1S1 content in Malagueña and Murciano-Granadina goats. J. Dairy Res. 75, 481–484. https://doi.org/10.1017/S0022029908003609

Caravaca, F., Ares, J.L., Carrizosa, J., Urrutia, B., Baena, F., Jordana, J., Badaoui, B., Sànchez, A., Angiolillo, A., Amills, M., Serradilla, J.M., 2011. Effects of αs1-casein (*CSN1S1*) and κ-casein (*CSN3*) genotypes on milk coagulation properties in Murciano-Granadina goats. J. Dairy Res. 78, 32–37. https://doi.org/10.1017/S002202991000083X

Caravaca, F., Carrizosa, J., Urrutia, B., Baena, F., Jordana, J., Amills, M., Badaoui, B., Sánchez, A., Angiolillo, A., Serradilla, J.M., 2009. Short communication: Effect of alphaS1-casein (*CSN1S1*) and kappa-casein (*CSN3*) genotypes on milk composition in Murciano-Granadina goats. J. Dairy Sci. 92, 2960–2964. https://doi.org/10.3168/JDS.2008-1510

Cardona, S.J.C., Cadavid, H.C., Corrales, J.D., Munilla, S., Cantet, R.J.C., Rogberg-Muñoz, A., 2016. Longitudinal data analysis of polymorphisms in the κ-casein and β-lactoglobulin genes shows differential effects along the trajectory of the lactation curve in tropical dairy goats. J. Dairy Sci. 99, 7299–7307. https://doi.org/10.3168/JDS.2016-10954

Carillier, C., Larroque, H., Robert-Granié, C., 2014. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genet. Sel. Evol. 46, 1–10. https://doi.org/10.1186/S12711-014-0067-3/TABLES/6

Casellas, J., Cañas-Álvarez, J.J., González-Rodríguez, A., Puig-Oliveras, A., Fina, M., Piedrafita, J., Molina, A., Díaz, C., Baró, J.A., Varona, L., 2017. Bayesian analysis of parent-specific transmission ratio distortion in seven Spanish beef cattle breeds. Anim. Genet. 48, 93–96. https://doi.org/10.1111/age.12509

Casellas, J., Gularte, R.J., Farber, C.R., Varona, L., Mehrabian, M., Schadt, E.E., Lusis, A.J., Attie, A.D., Yandell, B.S., Medrano, J.F., 2012. Genome scans for transmission ratio distortion regions in mice. Genetics 191, 247–259. https://doi.org/10.1534/genetics.111.135988

Casellas, J., Id-Lahoucine, S., Cánovas, A., 2020. Discriminating between allele- and genotype-specific transmission ratio distortion. Anim. Genet. 51, 847–854. https://doi.org/10.1111/age.13007

Casellas, J., Manunza, A., Mercader, A., Quintanilla, R., Amills, M., 2014. A flexible bayesian model for testing for transmission ratio distortion. Genetics 198, 1357–1367. https://doi.org/10.1534/genetics.114.169607

Castañeda-Bustos, V.J., Montaldo, H.H., Torres-Hernández, G., Pérez-Elizalde, S., Valencia-Posadas, M., Hernández-Mendo, O., Shepard, L., 2014. Estimation of genetic parameters for productive life, reproduction, and milk-production traits in US dairy goats. J. Dairy Sci. 97, 2462–2473.

Castañeda-Bustos, V.J., Montaldo, H.H., Valencia-Posadas, M., Shepard, L., Pérez-Elizalde, S., Hernández-Mendo, O., Torres-Hernández, G., 2017. Linear and nonlinear genetic relationships between type traits and productive

life in US dairy goats. J. Dairy Sci. 100, 1232–1245. https://doi.org/10.3168/jds.2016-11313

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., Wilson, J.F., 2018. Runs of homozygosity: Windows into population history and trait architecture. Nat. Rev. Genet. 19, 220–234. https://doi.org/10.1038/nrg.2017.109

Cebo, C., Lopez, C., Henry, C., Beauvallet, C., Ménard, O., Bevilacqua, C., Bouvier, F., Caillat, H., Martin, P., 2012. Goat αs1-casein genotype affects milk fat globule physicochemical properties and the composition of the milk fat globule membrane. J. Dairy Sci. 95, 6215–6229. https://doi.org/10.3168/jds.2011-5233

Chakraborty, D., Sharma, N., Kour, S., Sodhi, S.S., Gupta, M.K., Lee, S.J., Son, Y.O., 2022. Applications of omics technology for livestock selection and improvement. Front. Genet. 13, 774113. https://doi.org/10.3389/fgene.2022.774113

Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods 69, 330. https://doi.org/10.1016/J.MIMET.2007.02.005

Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., Cramer, C.L., Huang, X., 2015. Bridger: A new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol. 16, 30. https://doi.org/10.1186/s13059-015-0596-2

Chao, A., 1984. Nonparametric Estimation of the Number of Classes in a Population Nonparametric Estimation of the Number of Classes in a Population. Source Scand. J. Stat. 11, 265–270.

Charlesworth, D., Willis, J.H., 2009. The genetics of inbreeding depression. Nat. Rev. Genet. 10, 783–796. https://doi.org/10.1038/nrg2664

Charlton, S., Ramsøe, A., Collins, M., Craig, O.E., Fischer, R., Alexander, M., Speller, C.F., 2019. New insights into Neolithic milk consumption through proteomic analysis of dental calculus. Archaeol. Anthropol. Sci. 11, 6183–6196. https://doi.org/10.1007/S12520-019-00911-7/FIGURES/2

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128. https://doi.org/10.1186/1471-2105-14-128

Chilliard, Y., Rouel, J., Leroux, C., 2006. Goat's alpha-s1 casein genotype influences its milk fatty acid composition and delta-9 desaturation ratios. Anim. Feed Sci. Technol. 131, 474–487. 323 https://doi.org/10.1016/J.ANIFEEDSCI.2006.05.025

Ching, T., Huang, S., Garmire, L.X., 2014. Power analysis and sample size estimation for RNA-Seq differential expression. RNA 20, 1684–1696. https://doi.org/10.1261/rna.046011.114

Choi, J., In Lee, S., Rackerby, B., Frojen, R., Goddik, L., Ha, S. Do, Park, S.H., 2020a. Assessment of overall microbial community shift during Cheddar

cheese production from raw milk to aging. Appl. Microbiol. Biotechnol. 104, 6249–6260. https://doi.org/10.1007/S00253-020-10651-7/FIGURES/6

Choi, J., Lee, S.I., Rackerby, B., Goddik, L., Frojen, R., Ha, S. Do, Kim, J.H., Park, S.H., 2020b. Microbial communities of a variety of cheeses and comparison between core and rind region of cheeses. J. Dairy Sci. 103, 4026–4042. https://doi.org/10.3168/jds.2019-17455

Ciappesoni, G., Přibyl, J., Milerski, M., Mareš, V., 2004. Factors affecting goat milk yield and its composition. Czech J. Anim. Sci. 49, 465–473. https://doi.org/10.17221/4333-CJAS

Clark, E.L., Archibald, A.L., Daetwyler, H.D., Groenen, M.A.M., Harrison, P.W., Houston, R.D., Kühn, C., Lien, S., Macqueen, D.J., Reecy, J.M., Robledo, D., Watson, M., Tuggle, C.K., Giuffra, E., 2020. From FAANG to fork: Application of highly annotated genomes to improve farmed animal production. Genome Biol. 21, 285. https://doi.org/10.1186/S13059-020-02197-8

Clark, E.L., Bush, S.J., McCulloch, M.E.B., Farquhar, I.L., Young, R., Lefevre, L., Pridans, C., Tsang, H., Wu, C., Afrasiabi, C., Watson, M., Whitelaw, C.B., Freeman, T.C., Summers, K.M., Archibald, A.L., Hume, D.A., 2017. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). PLoS Genet. 13, e1006997. https://doi.org/10.1371/journal.pgen.1006997

Coelho, M.C., Silva, C.C.G., Ribeiro, S.C., Dapkevicius, M.L.N.E., Rosa, H.J.D., 2014. Control of Listeria monocytogenes in fresh cheese using protective

lactic acid bacteria. Int. J. Food Microbiol. 191, 53–59. https://doi.org/10.1016/j.ijfoodmicro.2014.08.029

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13. https://doi.org/10.1186/s13059-016-0881-8

Cosenza, G., Illario, R., Rando, A., Di Gregorio, P., Masina, P., Ramunno, L., 2003. Molecular characterization of the goat CSN1S1(01) allele. J. Dairy Res. 70, 237–240. https://doi.org/10.1017/S0022029903006101

Costa, A., Schwarzenbacher, H., Mészáros, G., Fuerst-Waltl, B., Fuerst, C., Sölkner, J., Penasa, M., 2019. On the genomic regions associated with milk lactose in Fleckvieh cattle. J. Dairy Sci. 102, 10088–10099. https://doi.org/10.3168/JDS.2019-16663

Coulson, A., Sulston, J., Brenner, S., Karn, J., 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. 83, 7821–7825. https://doi.org/10.1073/pnas.83.20.7821

Cremonesi, P., Capoferri, R., Pisoni, G., Del Corvo, M., Strozzi, F., Rupp, R., Caillat, H., Modesto, P., Moroni, P., Williams, J.L., Castiglioni, B., Stella, A., 2012. Response of the goat mammary gland to infection with *Staphylococcus aureus* revealed by gene expression profiling in milk somatic and white blood cells. BMC Genomics 13, 540. https://doi.org/10.1186/1471-2164-13-540

Dadousis, C., Ablondi, M., Cipolat-Gotet, C., van Kaam, J.T., Marusi, M., Cassandro, M., Sabbioni, A., Summer, A., 2022. Genomic inbreeding coefficients using imputed genotypes: Assessing different estimators in Holstein-Friesian dairy cows. J. Dairy Sci. 105, 5926–5945. https://doi.org/10.3168/JDS.2021-21125

de Asís Ruiz Morales, F., Genís, J.M.C., Guerrero, Y.M., 2019. — Special Issue — Current status, challenges and the way forward for dairy goat production in Europe. Asian-Australasian J. Anim. Sci. 32, 1256–1265. https://doi.org/10.5713/ajas.19.0327

De Jonghe, V., Coorevits, A., Van Hoorde, K., Messens, W., Van Landschoot, A., De Vos, P., Heyndrickx, M., 2011. Influence of storage conditions on the growth of pseudomonas species in refrigerated raw milk. Appl. Environ. Microbiol. 77, 460–470. https://doi.org/10.1128/AEM.00521-10

Delgado, J.V., Landi, V., Barba, C.J., Fernández, J., Gómez, M.M., Camacho, M.E., Martínez, M.A., Navas, F.J., León, J.M., 2018. Murciano-Granadina goat: A Spanish local breed ready for the challenges of the twenty-first century, in: Sustainable Goat Production in Adverse Environments. Springer, Cham, pp. 205–219. https://doi.org/10.1007/978-3-319-71294-9_15

Denholm, R., De Stavola, B.L., Hipwell, J.H., Doran, S.J., Holly, J.M.P., Folkerd, E., Dowsett, M., Leach, M.O., Hawkes, D.J., Dos-Santos-Silva, I., 2018. Circulating growth and sex hormone levels and breast tissue composition in young nulliparous women. Cancer Epidemiol. Biomarkers Prev. 27, 1500–

1508. https://doi.org/10.1158/1055-9965.EPI-18-0036

Denoyelle, L., de Villemereuil, P., Boyer, F., Khelifi, M., Gaffet, C., Alberto, F., Benjelloun, B., Pompanon, F., 2021. Genetic variations and differential DNA methylation to face contrasted climates in small ruminants: An analysis on traditionally-managed sheep and goats. Front. Genet. 12, 1809. https://doi.org/10.3389/fgene.2021.745284

Deroide, C.A.S., Jacopini, L.A., Delgado, J. V., Léon, J.M., Brasil, L.H.A., Ribeiro, M.N., 2016. Inbreeding depression and environmental effect on milk traits of the Murciano-Granadina goat breed. Small Rumin. Res. 134, 44–48. https://doi.org/10.1016/j.smallrumres.2015.12.008

Desire, S., Mucha, S., Coffey, M., Mrode, R., Broadbent, J., Conington, J., 2018. Pseudopregnancy and aseasonal breeding in dairy goats: Genetic basis of fertility and impact on lifetime productivity. Animal 12, 1799–1806. https://doi.org/10.1017/S1751731117003056

Dettori, M.L., Pazzola, M., Paschino, P., Pira, M.G., Vacca, G.M., 2015. Variability of the caprine whey protein genes and their association with milk yield, composition and renneting properties in the Sarda breed. 1. The *LALBA* gene. J. Dairy Res. 82, 434–441. https://doi.org/10.1017/S0022029915000461

Dey, S., Chamero, P., Pru, J.K., Chien, M.S., Ibarra-Soria, X., Spencer, K.R., Logan, D.W., Matsunami, H., Peluso, J.J., Stowers, L., 2015. Cyclic regulation of sensory perception by a female hormone alters behavior. Cell

161, 1334–1344. https://doi.org/10.1016/j.cell.2015.04.052

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Doekes, H.P., Bijma, P., Windig, J.J., 2021. How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. Genes. 12, 926. https://doi.org/10.3390/genes12060926

Doekes, H.P., Veerkamp, R.F., Bijma, P., De Jong, G., Hiemstra, S.J., Windig, J.J., 2019. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. Genet. Sel. Evol. 52, 54. https://doi.org/10.1186/s12711-019-0497-z

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J.J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J.J., Wang, W., 2012. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). Nat. Biotechnol. 31, 135–141. https://doi.org/10.1038/nbt.2478

Dubeuf, J.P., Boyazoglu, J., 2009. An international panorama of goat selection and breeds. Livest. Sci. 120, 225–231.

https://doi.org/10.1016/j.livsci.2008.07.005

Dwyer, C.M., 2014. Maternal behaviour and lamb survival: From neuroendocrinology to practical application, in: Animal. Cambridge University Press, pp. 102–112. https://doi.org/10.1017/S1751731113001614

Eichler, E.E., 2019. Genetic variation, comparative genomics, and the diagnosis of disease. N. Engl. J. Med. 381, 64–74. https://doi.org/10.1056/nejmra1809315

Ercanbrack, S.K., Knight, A.D., 1991. Effects of inbreeding on reproduction and wool production of Rambouillet, Targhee, and Columbia ewes. J. Anim. Sci. 69, 4734–4744. https://doi.org/10.2527/1991.69124734x

Evershed, R.P., Payne, S., Sherratt, A.G., Copley, M.S., Coolidge, J., Urem-Kotsu, D., Kotsakis, K., Özdoğan, M., Özdoğan, A.E., Nieuwenhuyse, O., Akkermans, P.M.M.G., Bailey, D., Andeescu, R.R., Campbell, S., Farid, S., Hodder, I., Yalman, N., Özbaşaran, M., Biçakci, E., Garfinkel, Y., Levy, T., Burton, M.M., 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. Nature 455, 528–531. https://doi.org/10.1038/nature07180

Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Falconer, D.S., Mackay, T.F.C., 1989. Introduction to quantitative genetics, Third. ed, Trends in Genetics. Benjamin-Cummings Pub Co.

Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-

throughput information: Effects of SNP density and genotyping errors. Genet. Sel. Evol. 45, 42. https://doi.org/10.1186/1297-9686-45-42

Ferraz, A.L.J., Ojeda, A., López-Béjar, M., Fernandes, L.T., Castelló, A., Folch, J.M., Pérez-Enciso, M., 2008. Transcriptome architecture across tissues in the pig. BMC Genomics 9, 173. https://doi.org/10.1186/1471-2164-9-173

Finkbeiner, E., Haindl, M., Muller, S., 2011. The SUMO system controls nucleolar partitioning of a novel mammalian ribosome biogenesis complex. EMBO J. 30, 1067–1078. https://doi.org/10.1038/emboj.2011.33

Fishman, L., Mcintosh, M., 2019. Standard deviations: The biological bases of transmission ratio distortion. Annu. Rev. Genet. 53, 347–372. https://doi.org/10.1146/annurev-genet-112618-043905

Foissac, S., Djebali, S., Munyard, K., Vialaneix, N., Rau, A., Muret, K., Esquerré, D., Zytnicki, M., Derrien, T., Bardou, P., Blanc, F., Cabau, C., Crisci, E., Dhorne-Pollet, S., Drouet, F., Faraut, T., Gonzalez, I., Goubil, A., Lacroix-Lamandé, S., Laurent, F., Marthey, S., Marti-Marimon, M., Momal-Leisenring, R., Mompart, F., Quéré, P., Robelin, D., Cristobal, M.S., Tosser-Klopp, G., Vincent-Naulleau, S., Fabre, S., Der Laan, M.H.P. Van, Klopp, C., Tixier-Boichard, M., Acloque, H., Lagarrigue, S., Giuffra, E., 2019. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. BMC Biol. 17, 1–25. https://doi.org/10.1186/s12915-019-0726-5

Fox, P.F., Guinee, T.P., Cogan, T.M., McSweeney, P.L.H., 2017. Microbiology of

cheese ripening, in: Fundamentals of cheese science. Springer, Boston, MA, p.p. 333–390. https://doi.org/10.1007/978-1-4899-7681-9_11

Fox, P.F., Singh, T.K., McSweeney, P.L.H., 1995. Biogenesis of flavour compounds in cheese. Adv. Exp. Med. Biol. 367, 59–98. https://doi.org/10.1007/978-1-4615-1913-3_6

Fu, W., Wang, R., Yu, J., Hu, D., Cai, Y., Shao, J., Jiang, Y., 2021. GGVD: A goat genome variation database for tracking the dynamic evolutionary process of selective signatures and ancient introgressions. J. Genet. Genomics 48, 248–256. https://doi.org/10.1016/J.JGG.2021.03.003

Gandelman, R., Zarrow, M.X., Denenberg, V.H., Myers, M., 1971. Olfactory bulb removal eliminates maternal behavior in the mouse. Science 171, 210–211. https://doi.org/10.1126/science.171.3967.210

Gaspar, J.M., 2018. Improved peak-calling with MACS2. bioRxiv 496521. https://doi.org/10.1101/496521

Giani, A.M., Gallo, G.R., Gianfranceschi, L., Formenti, G., 2020. Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput. Struct. Biotechnol. J. 18, 9–19. https://doi.org/10.1016/j.csbj.2019.11.002

Gibson, J., Morton, N.E., Collins, A., 2006. Extended tracts of homozygosity in outbred human populations. Hum. Mol. Genet. 15, 789–795. https://doi.org/10.1093/HMG/DDI493

Goddard, M., 2009. Fitness traits in animal breeding programs, in: Adaptation and

fitness in animal populations. Springer, Dordrecht, pp. 41–52. https://doi.org/10.1007/978-1-4020-9005-9_3

Gòdia, M., Casellas, J., Ruiz-Herrera, A., Rodríguez-Gil, J.E., Castelló, A., Sánchez, A., Clop, A., 2020. Whole genome sequencing identifies allelic ratio distortion in sperm involving genes related to spermatogenesis in a swine model. DNA Res. 27, dsaa019. https://doi.org/10.1093/dnares/dsaa019

Gómez-Carpio, M, Miranda, J, León, JM, Pleguezulos, J & Delgado, J., 2012. Análisis preliminar de la estimación de parámetros genéticos para caracteres morfológicos lineales en la raza caprina Murciano Granadina, in: Congresso Ibérico Sobre Recursos Genéticos Animais. Évora.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883

Grosclaude, F., Mahé, M.-F., Brignon, G., Di Stasio, L., Jeunet, R., 1987. A Mendelian polymorphism underlying quantitative variations of goat αs1-casein. Genet. Sel. Evol. 19, 399. https://doi.org/10.1186/1297-9686-19-4-399

Guan, D., Landi, V., Luigi-Sierra, M.G., Delgado, J.V., Such, X., Castelló, A., Cabrera, B., Mármol-Sánchez, E., Fernández-Alvarez, J., de la Torre

Casañas, J.L.R., Martínez, A., Jordana, J., Amills, M., 2020. Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. J. Anim. Sci. Biotechnol. 11, 35. https://doi.org/10.1186/s40104-020-00435-4

Gunia, M., Phocas, F., Arquet, R., Alexandre, G., Mandonnet, N., 2011. Genetic parameters for body weight, reproduction, and parasite resistance traits in the Creole goat. J. Anim. Sci. 89, 3443–3451. https://doi.org/10.2527/jas.2011-3872

Guo, J., Jiang, R., Mao, A., Liu, G.E., Zhan, S., Li, L., Zhong, T., Wang, L., Cao, J., Chen, Y., Zhang, G., Zhang, H., 2021. Genome-wide association study reveals 14 new SNPs and confirms two structural variants highly associated with the horned/polled phenotype in goats. BMC Genomics 22, 769. https://doi.org/10.1186/s12864-021-08089-w

Gutiérrez-Reinoso, M.A., Aponte, P.M., García-Herreros, M., 2022. A review of inbreeding depression in dairy cattle: Current status, emerging control strategies, and future prospects. J. Dairy Res. 89, 3–12. 333 https://doi.org/10.1017/S0022029922000188

Harhay, G.P., Smith, T.P.L., Alexander, L.J., Haudenschild, C.D., Keele, J.W., Matukumalli, L.K., Schroeder, S.G., Van Tassell, C.P., Gresham, C.R., Bridges, S.M., Burgess, S.C., Sonstegard, T.S., 2010. An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. Genome Biol. 11, R102.

https://doi.org/10.1186/gb-2010-11-10-r102

Hayes, B.J., Lewin, H.A., Goddard, M.E., 2013. The future of livestock breeding: Genomic selection for efficiency, reduced emissions intensity, and adaptation. Trends Genet. 29, 206–214. https://doi.org/10.1016/j.tig.2012.11.009

Henn, B.M., Botigué, L.R., Bustamante, C.D., Clark, A.G., Gravel, S., 2015. Estimating the mutation load in human genomes. Nat. Rev. Genet. 16, 333–343. https://doi.org/10.1038/nrg3931

Hoekzema, E., Barba-Müller, E., Pozzobon, C., Picado, M., Lucco, F., García-García, D., Soliva, J.C., Tobeña, A., Desco, M., Crone, E.A., Ballesteros, A., Carmona, S., Vilarroya, O., 2017. Pregnancy leads to long-lasting changes in human brain structure. Nat. Neurosci. 20, 287–296. https://doi.org/10.1038/nn.4458

Hölzer, M., Marz, M., 2019. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. Gigascience 8, 1–16. https://doi.org/10.1093/gigascience/giz039

Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., Riley, J., Purvis, I., Xu, C.F., 2004. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur. J. Hum. Genet. 12, 395–399. https://doi.org/10.1038/sj.ejhg.5201164

Howie, B.N., Donnelly, P., Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association

studies. PLoS Genet. 5, e1000529. https://doi.org/10.1371/journal.pgen.1000529

Hu, Z.L., Park, C.A., Reecy, J.M., 2022. Bringing the Animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. Nucleic Acids Res. 50, D956–D961. https://doi.org/10.1093/nar/gkab1116

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, L.O., Labbe, A., Infante-Rivard, C., 2013. Transmission ratio distortion: Review of concept and implications for genetic association studies. Hum. Genet. 132, 245–263. https://doi.org/10.1007/s00439-012-1257-0

Huart, C., Rombaux, P., Hummel, T., 2013. Plasticity of the human olfactory system: The olfactory bulb. Molecules 18, 11586–11600. https://doi.org/10.3390/molecules180911586

Id-Lahoucine, S., Cánovas, A., Jaton, C., Miglior, F., Fonseca, P.A.S., Sargolzaei, M., Miller, S., Schenkel, F.S., Medrano, J.F., Casellas, J., 2019. Implementation of Bayesian methods to identify SNP and haplotype regions with transmission ratio distortion across the whole genome: TRDscan v.1.0. J. Dairy Sci. 102, 3175–3188. https://doi.org/10.3168/jds.2018-15296

Irlinger, F., Layec, S., Hélinck, S., Dugat-Bony, E., 2015. Cheese rind microbial communities: Diversity, composition and origin. FEMS Microbiol. Lett. 362,

1–11. https://doi.org/10.1093/femsle/fnu015

Ithurbide, M., Huau, C., Palhière, I., Fassier, T., Friggens, N.C., Rupp, R., 2022. Selection on functional longevity in a commercial population of dairy goats translates into significant differences in longevity in a common farm environment. J. Dairy Sci. 105, 4289–4300. https://doi.org/10.3168/jds.2021-21222

Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. Nature Genetics. 33, 245–254. https://doi.org/10.1038/ng1089

Jenness, R., 1980. Composition and Characteristics of Goat Milk: Review 1968−1979. J. Dairy Sci. 63, 1605–1630. https://doi.org/10.3168/jds.S0022-0302(80)83125-0

Jiang, A., Ankersmit-Udy, A., Turner, S.A., Scholtens, M., Littlejohn, M.D., Lopez-Villalobos, N., Proser, C.G., Snell, R.G., Lehnert, K., 2022. A *Capra hircus* chromosome 19 locus linked to milk production influences mammary conformation. J. Anim. Sci. Biotechnol. 13, 1–10. https://doi.org/10.1186/s40104-021-00667-y

Johnson, J., Curtin, C., Waite-Cusic, J., 2021. The cheese production facility microbiome exhibits temporal and spatial variability. Front. Microbiol. 12, 381. https://doi.org/10.3389/fmicb.2021.644828

Kandiel, M.M.M., Watanabe, G., Sosa, G.A., Abou El-Roos, M.E.A., Abdel-Ghaffar, A.E., Li, J.Y., Manabe, N., El Azab, A.E.S.I., Taya, K., 2010.

Profiles of circulating steroid hormones, gonadotropins, immunoreactive inhibin and prolactin during pregnancy in goats and immunolocalization of inhibin subunits, steroidogenic enzymes and prolactin in the corpus luteum and placenta. J. Reprod. Dev. 56, 243–250. https://doi.org/10.1262/jrd.09-159S

Kang, B., Wang, J., Zhang, H., Shen, W., El-Mahdy Othman, O., Zhao, Y., Min, L., 2022. Genome-wide profile in DNA methylation in goat ovaries of two different litter size populations. J. Anim. Physiol. Anim. Nutr. 6, 239–249. https://doi.org/10.1111/JPN.13600

Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Hauser, S., Hedrick, P.W., Miller, J.M., Tallmon, D.A., Chris Funk, W., 2021. The crucial role of genome-wide genetic variation in conservation. Proc. Natl. Acad. Sci. U. S. A. 118, e2104642118. https://doi.org/10.1073/pnas.2104642118

Keller, M.C., Visscher, P.M., Goddard, M.E., 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 189, 237–249. https://doi.org/10.1534/genetics.111.130922

Kenny, L.C., McCarthy, F., 2017. ‘Omic’ research, in: Introduction to Research Methodology for Specialists and Trainees. Cambridge University Press, pp. 97–103. https://doi.org/10.1017/9781107585775.014

Keskin, S., Kor, A., Karaca, S., Mirtagio Iu, H., 2005. A study of relationships between milk yield and some udder traits by using of path analysis in Akke?i

goats. J. Anim. Vet. Adv. 4, 547–550.

Khatib, H., 2021. Transgenerational epigenetic inheritance in farm animals: How substantial is the evidence? Livest. Sci. 250, 104557. https://doi.org/10.1016/j.livsci.2021.104557

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, E.S., Cole, J.B., Huson, H., Wiggans, G.R., Van Tassel, C.P., Crooker, B.A., Liu, G., Da, Y., Sonstegard, T.S., 2013. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS One 8, 80813. https://doi.org/10.1371/journal.pone.0080813

Klarin, D., Damrauer, S.M., Cho, K., Sun, Y. V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., Chaffin, M., Small, A.M., Huang, J., Tang, H., Lynch, J.A., Ho, Y.L., Liu, D.J., Emdin, C.A., Li, A.H., Huffman, J.E., Lee, J.S., Natarajan, P., Chowdhury, R., Saleheen, D., Vujkovic, M., Baras, A., Pyarajan, S., Di Angelantonio, E., Neale, B.M., Naheed, A., Khera, A. V., Danesh, J., Chang, K.M., Abecasis, G., Willer, C., Dewey, F.E., Carey, D.J., Concato, J., Gaziano, J.M., O'Donnell, C.J., Tsao, P.S., Kathiresan, S., Rader, D.J., Wilson, P.W.F., Assimes, T.L., 2018. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. 50, 1514–1523. https://doi.org/10.1038/s41588-018-0222-9

Klug, W.S., Cummings, M.R., Palladino, M.A., Spencer, C.A., 2011. Concepts of genetics, 10th ed. Benjamin Cummings, San Francisco, California.

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.I., McDonald, D., Melnik, A. V., Morton, J.T., Navas, J., Quinn, R.A., Sanders, J.G., Swafford, A.D., Thompson, L.R., Tripathi, A., Xu, Z.Z., Zaneveld, J.R., Zhu, Q., Caporaso, J.G., Dorrestein, P.C., 2018. Best practices for analysing microbiomes. Nat. Rev. Microbiol. 16, 410–422. https://doi.org/10.1038/s41579-018-0029-9

Kothe, C.I., Bolotin, A., Kraïem, B.F., Dridi, B., Renault, P., 2021. Unraveling the world of halophilic and halotolerant bacteria in cheese by combining cultural, genomic and metagenomic approaches. Int. J. Food Microbiol. 358, 109312. https://doi.org/10.1016/j.ijfoodmicro.2021.109312

Krueger, F., Andrews, S.R., 2011. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Kumar, S., Gonzalez, E.A., Rameshwar, P., Etchegaray, J.P., 2020. Non-coding RNAs as mediators of epigenetic changes in malignancies. Cancers 12, 1–32. https://doi.org/10.3390/cancers12123657

Labbe, A., Huang, L.O., Infante-Rivard, C., 2013. Transmission ratio distortion: A neglected phenomenon with many consequences in genetic analysis and population genetics, in: Epigenetics and omplex traits. Springer New York, NY, p.p. 265–285. https://doi.org/10.1007/978-1-4614-8078-

5_12/TABLES/6

Lahoucine, S.I., Casellas, J., Lu, D., Sargolzaei, M., Miller, S.P., Canovas, A., 2020. 47 Distortion of Mendelian segregation across Angus cattle genome reveal novel lethal haplotype affecting reproduction. J. Anim. Sci. 98, 5. https://doi.org/10.1093/jas/skaa278.009

Larsen, C.M., Grattan, D.R., 2010. Prolactin-induced mitogenesis in the subventricular zone of the maternal brain during early pregnancy is essential for normal postpartum behavioral responses in the mother. Endocrinology 151, 3805–3814. https://doi.org/10.1210/EN.2009-1385

Legarra, A., Ugarte, E., 2005. Genetic parameters of udder traits, somatic cell score, and milk yield in Latxa sheep. Journal of Dairy Sci. 88, 2238–2245. https://doi.org/10.3168/jds.S0022-0302(05)72899-X

Leitner, G., Merin, U., Silanikove, N., 2011. Effects of glandular bacterial infection and stage of lactation on milk clotting parameters: Comparison among cows, goats and sheep. Int. Dairy J. 21, 279–285. https://doi.org/10.1016/j.idairyj.2010.11.013

Leroy, G., 2014. Inbreeding depression in livestock species: Review and meta-analysis. Anim. Genet. 45, 618–628. https://doi.org/10.1111/age.12178

Lévy, F., Keller, M., 2009. Olfactory mediation of maternal behavior in selected mammalian species. Behav. Brain Res. 200, 336–345. https://doi.org/10.1016/j.bbr.2008.12.017

Li, B., Dewey, C.N., 2011. RSEM: Accurate transcript quantification from RNA-

Seq data with or without a reference genome. BMC Bioinformatics 12, 1–16. https://doi.org/10.1186/1471-2105-12-323/TABLES/6

Li, C., Zhu, J., Shi, Hengbo, Luo, J., Zhao, W., Shi, Huaiping, Xu, H., Wang, H., Loor, J.J., 2020. Comprehensive transcriptome profiling of dairy goat mammary gland identifies genes and networks crucial for lactation and fatty acid metabolism. Front. Genet. 11, 878. https://doi.org/10.3389/fgene.2020.00878

Li, C.C., Horvitz, D.G., 1953. Some methods of estimating the inbreeding coefficient. Am. J. Hum. Genet. 5, 107–117.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J., Tibshirani, R., 2013. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. Stat. Methods Med. Res. 22, 519–536. https://doi.org/10.1177/0962280211428386

Li, R., Fu, W., Su, R., Tian, X., Du, D., Zhao, Y., Zheng, Z., Chen, Q., Gao, S., Cai, Y., Wang, X., Li, J., Jiang, Y., 2019. Towards the complete goat pan-genome by recovering missing genomic segments from the reference genome. Front. Genet. 10, 1169. https://doi.org/10.3389/fgene.2019.01169

Li, R., Yang, P., Dai, X., Asadollahpour Nanaei, H., Fang, W., Yang, Z., Cai, Y., Zheng, Z., Wang, X., Jiang, Y., 2021. A near complete genome for goat genetic and genomic research. Genet. Sel. Evol. 53, 1–17. https://doi.org/10.1186/s12711-021-00668-5

Li, Y., Tollefsbol, T.O., 2011. DNA methylation detection: Bisulfite genomic sequencing analysis. Methods Mol. Biol. 791, 11–21. https://doi.org/10.1007/978-1-61779-316-5_2

Liao, Y., Smyth, G.K., Shi, W., 2014. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lie, H.C., Simmons, L.W., Rhodes, G., 2009. Does genetic diversity predict health in humans? PLoS One 4, 6391. https://doi.org/10.1371/journal.pone.0006391

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lozupone, C., Knight, R., 2005. UniFrac: A new phylogenetic method for comparing microbial communities. Appl. Environ. Microbiol. 71, 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Luigi-Sierra, M.G., Landi, V., Guan, D., Delgado, J.V., Castelló, A., Cabrera, B., Mármol-Sánchez, E., Alvarez, J.F., Gómez-Carpio, M., Martínez, A., Such, X., Jordana, J., Amills, M., 2020. A genome-wide association analysis for

body, udder, and leg conformation traits recorded in Murciano-Granadina goats. J. Dairy Sci. 103, 11605–11617. https://doi.org/10.3168/jds.2020-18461

Macedo, A.A., Bittar, J.F.F., Bassi, P.B., Ronda, J.B., Bittar, E.R., Panetto, J.C.C., Araujo, M.S.S., Santos, R.L., Martins-Filho, O.A., 2014. Influence of endogamy and mitochondrial DNA on immunological parameters in cattle. BMC Vet. Res. 10, 1–10. https://doi.org/10.1186/1746-6148-10-79

Makanjuola, B.O., Miglior, F., Abdalla, E.A., Maltecca, C., Schenkel, F.S., Baes, C.F., 2020. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J. Dairy Sci. 103, 5183–5199. https://doi.org/10.3168/jds.2019-18013

Malher, X., Seegers, H., Beaudeau, F., 2001. Culling and mortality in large dairy goat herds managed under intensive conditions in western France. Livest. Prod. Sci. 71, 75–86. https://doi.org/10.1016/S0301-6226(01)00242-1

Manfredi, E., Piacere, A., Lahaye, P., Ducrocq, V., 2001. Genetic parameters of type appraisal in Saanen and Alpine goats. Livest. Prod. Sci. 70, 183–189. https://doi.org/10.1016/S0301-6226(01)00180-4

Manfredi, E., Piacere, A., Ricordeau, G., Elsen, J.M., 1995. The polymorphism of casein alpha s1 and goat selection, in: Rencontres Autour Des Recherches Sur Les Ruminants". pp. 167–170.

Marletta, D., Criscione, A., Bordonaro, S., Maria Guastella, A., 2007. Casein polymorphism in goat's milk. Le Lait, INRA Ed. 87, 491–504.

https://doi.org/10.1051/lait:2007034

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/EJ.17.1.200

Martin, N.H., Murphy, S.C., Ralyea, R.D., Wiedmann, M., Boor, K.J., 2011. When cheese gets the blues: Pseudomonas fluorescens as the causative agent of cheese spoilage. J. Dairy Sci. 94, 3176–3183. https://doi.org/10.3168/JDS.2011-4312

Martin, P., Palhière, I., Maroteau, C., Bardou, P., Canale-Tabet, K., Sarry, J., Woloszyn, F., Bertrand-Michel, J., Racke, I., Besir, H., Rupp, R., Tosser-Klopp, G., 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. Sci. Reports 2017 71 7, 1–13. https://doi.org/10.1038/s41598-017-02052-0

Martin, P., Palhière, I., Maroteau, C., Clément, V., David, I., Klopp, G.T., Rupp, R., 2018. Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. J. Dairy Sci. 101, 5214–5226. https://doi.org/10.3168/jds.2017-13625

Martin, P., Szymanowska, M., Zwierzchowski, L., Leroux, C., 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. Reprod. Nutr. Dev. 42, 433–459. https://doi.org/10.1051/RND:2002036

McClure, S.B., Magill, C., Podrug, E., Moore, A.M.T., Harper, T.K., Culleton,

B.J., Kennett, D.J., Freeman, K.H., 2018. Fatty acid specific δ13C values reveal earliest Mediterranean cheese production 7,200 years ago. PLoS One 13, e0202807. https://doi.org/10.1371/JOURNAL.PONE.0202807

McLaren, A., Mucha, S., Mrode, R., Coffey, M., Conington, J., 2016. Genetic parameters of linear conformation type traits and their relationship with milk yield throughout lactation in mixed-breed dairy goats. J. Dairy Sci. 99, 5516–5525. https://doi.org/10.3168/jds.2015-10269

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H., Wilson, J.F., 2008. Runs of homozygosity in European populations. Am. J. Hum. Genet. 83, 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

Mellado, M., Valdéz, R., García, J.E., López, R., Rodríguez, A., 2006. Factors affecting the reproductive performance of goats under intensive conditions in a hot arid environment. Small Rumin. Res. 63, 110–118. 345 https://doi.org/10.1016/j.smallrumres.2005.02.016

Meredith, B.K., Berry, D.P., Kearney, F., Finlay, E.K., Fahey, A.G., Bradley, D.G., Lynn, D.J., 2013. A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. Front. Genet. 4, 229. https://doi.org/10.3389/fgene.2013.00229

Merlot, E., Quesnel, H., Prunier, A., 2013. Prenatal stress, immunity and neonatal health in farm animal species. Animal 7, 2016–2025. https://doi.org/10.1017/S175173111300147X

Miglior, F., Burnside, E.B., Dekkers, J.C.M., 1995. Nonadditive genetic effects and inbreeding depression for somatic cell counts of Holstein cattle. J. Dairy Sci. 78, 1168–1173. https://doi.org/10.3168/jds.S0022-0302(95)76734-0

Miller, R.R., Montoya, V., Gardy, J.L., Patrick, D.M., Tang, P., 2013. Metagenomics for pathogen detection in public health. Genome Med. 5, 1–14. https://doi.org/10.1186/GM485/TABLES/4

Miranda, J.C., León, J.M., Pieramati, C., Gómez, M.M., Valdés, J., Barba, C., 2019. Estimation of genetic parameters for peak yield, yield and persistency traits in Murciano-Granadina goats using multi-traits models. Animals 9, 411. https://doi.org/10.3390/ani9070411

Moaeen-ud-Din, M., Danish Muner, R., Khan, M.S., 2022. Genome wide association study identifies novel candidate genes for growth and body conformation traits in goats. Sci. Rep. 12, 9891. https://doi.org/10.1038/s41598-022-14018-y

Mocé, E., Mocé, M.L., Lozano-Palazón, S.A., Bernácer, J., Martínez-Granell, M.M., Esteve, I.C., Bernat, F., Contreras, S.J., Villalba, I., Gómez, E.A., 2022. Fertility prediction in dairy goats from Murciano-Granadina breed: The role of sperm evaluation and female traits. Animal 16, 100525. https://doi.org/10.1016/j.animal.2022.100525

Moioli, B., D'Andrea, M., Pilla, F., 2007. Candidate genes affecting sheep and goat milk quality. Small Rumin. Res. 68, 179–192. https://doi.org/10.1016/j.smallrumres.2006.09.008

Montaldo, H., Martínez-Lozano, F.J., 1993. Phenotypic relationships between udder and milking characteristics, milk production and California mastitis test in goats. Small Rumin. Res. 12, 329–337. https://doi.org/10.1016/0921-4488(93)90068-S

Montel, M.C., Buchin, S., Mallet, A., Delbes-Paus, C., Vuitton, D.A., Desmasures, N., Berthier, F., 2014. Traditional cheeses: Rich and diverse microbiota with associated benefits. Int. J. Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2014.02.019

Mucha, S., Mrode, R., Coffey, M., Kizilaslan, M., Desire, S., Conington, J., 2018. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. J. Dairy Sci. 101, 2213–2225. https://doi.org/10.3168/jds.2017-12919

Mucha, S., Tortereau, F., Doeschl-Wilson, A., Rupp, R., Conington, J., 2022. Animal board invited review: Meta-analysis of genetic parameters for resilience and efficiency traits in goats and sheep. Animal. https://doi.org/10.1016/j.animal.2022.100456

Muriuki, C., Bush, S.J., Salavati, M., McCulloch, M.E.B., Lisowski, Z.M., Agaba, M., Djikeng, A., Hume, D.A., Clark, E.L., 2019. A Mini-Atlas of Gene Expression for the Domestic Goat (*Capra hircus*). Front. Genet. 10, 1080.

https://doi.org/10.3389/fgene.2019.01080

Navarro-Moreno, C., Sanchez-Catalan, M.J., Barneo-Muñoz, M., Goterris-Cerisuelo, R., Belles, M., Lanuza, E., Agustin-Pavon, C., Martinez-Garcia, F., 2020. Pregnancy changes the response of the vomeronasal and olfactory systems to pups in mice. Front. Cell. Neurosci. 14, 429. https://doi.org/10.3389/fncel.2020.593309

Nazari-Ghadikolaei, A., Mehrabani-Yeganeh, H., Miarei-Aashtiani, S.R., Staiger, E.A., Rashidi, A., Huson, H.J., 2018. Genome-wide association studies identify candidate genes for coat color and mohair traits in the Iranian Markhoz Goat. Front. Genet. 9, 105. https://doi.org/10.3389/fgene.2018.00105

Nejati-Javaremi, A., Smith, C., Gibson, J.P., 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Anim. Sci. 75, 1738–1745. https://doi.org/10.2527/1997.7571738X

Noguera-Solano, R., Ruiz-Gutierrez, R., Rodriguez-Caso, J.M., 2013. Genome: Twisting stories with DNA. Endeavour. 37, 213–209. https://doi.org/10.1016/j.endeavour.2013.05.003

Ollier, S., Chauvet, S., Martin, P., Chilliard, Y., Leroux, C., 2008. Goat's αs1-casein polymorphism affects gene expression profile of lactating mammary gland. Animal 2, 566–573. https://doi.org/10.1017/S1751731108001584

Ortiz, C., Navarro, J.F., Jurek, A., Märtin, A., Lundeberg, J., Meletis, K., 2020. Molecular atlas of the adult mouse brain. Sci. Adv. 6, eabb3446.

https://doi.org/10.1126/sciadv.abb3446

Osorio, J.S., Jacometo, C.B., Zhou, Z., Luchini, D., Cardoso, F.C., Loor, J.J., 2016. Hepatic global DNA and peroxisome proliferator-activated receptor alpha promoter methylation are altered in peripartal dairy cows fed rumen-protected methionine. J. Dairy Sci. 99, 234–244. https://doi.org/10.3168/JDS.2015-10157

Paape, M.J., Poutrel, B., Contreras, A., Marco, J.C., Capuco, A.V., 2001. Milk somatic cells and lactation in small ruminants. J. Dairy Sci. 84, E237–E244. https://doi.org/10.3168/jds.s0022-0302(01)70223-8

Paiva, R.D.M., de Sousa, J.E.R., Ferreira, J., Cunha, E.E., de Paiva, M.P.S.L.M., Gouveia, A.M.G., Facó, O., 2020. Population structure and effect of inbreeding on milk yield of Saanen goats in Brazilian production systems. Small Rumin. Res. 192, 106194. https://doi.org/10.1016/J.SMALLRUMRES.2020.106194

Pardo-Manuel de Villena, F., Sapienza, C., 2001. Nonrandom segregation during meiosis: The unfairness of females. Mammalian Genome 12, 331–339. https://doi.org/10.1007/s003350040003

Parente, E., Cogan, T.M., Powell, I.B., 2017. Starter cultures: General aspects, in: Cheese: Chemistry, physics and microbiology: Fourth Edition. Academic Press, pp. 201–226. https://doi.org/10.1016/B978-0-12-417012-4.00008-9

Park, P.J., 2009. ChIP–seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. 10, 669–680. https://doi.org/10.1038/nrg2641

Pena, R.N., Sánchez, A., Folch, J.M., 2000. Characterization of genetic polymorphism in the goat beta-lactoglobulin gene. J. Dairy Res. 67, 217–224. https://doi.org/10.1017/S0022029900004155

Penland, M., Falentin, H., Parayre, S., Pawtowski, A., Maillard, M.B., Thierry, A., Mounier, J., Coton, M., Deutsch, S.M., 2021. Linking Pélardon artisanal goat cheese microbial communities to aroma compounds during cheese-making and ripening. Int. J. Food Microbiol. 345, 109130. https://doi.org/10.1016/j.ijfoodmicro.2021.109130

Pérez-Méndez, J.A., Roibás, D., Wall, A., 2020. Somatic cell counts, reproduction indicators, and technical efficiency in milk production: A stochastic frontier analysis for Spanish dairy farms. J. Dairy Sci. 103, 7141–7154. https://doi.org/10.3168/JDS.2019-17146

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R., Baldi, F., 2017. Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 18, 255–271. https://doi.org/10.1111/age.12526

Peris, S., Caja, G., Such, X., 1999. Relationships between udder and milking traits in Murciano-Granadina dairy goats. Small Rumin. Res. 33, 171–179. https://doi.org/10.1016/S0921-4488(99)00017-6

Persuy, M.A., Printz, C., Medrano, J.F., Mercier, J.C., 1999. A single nucleotide deletion resulting in a premature stop codon is associated with marked reduction of transcripts from a goat beta-casein null allele. Anim. Genet. 30, 444–451. https://doi.org/10.1046/J.1365-2052.1999.00547.X

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295. https://doi.org/10.1038/nbt.3122

Pissonnier, D., Thiery, J.C., Fabre-Nys, C., Poindron, P., Keverne, E.B., 1985. The importance of olfactory bulb noradrenalin for maternal recognition in sheep. Physiol. Behav. 35, 361–363. https://doi.org/10.1016/0031-9384(85)90309-9

Pizarro Inostroza, M.G., Landi, V., González, F.J.N., Jurado, J.M.L., Martínez, A.M., Álvarez, J.F., Bermejo, J.V.D., 2019. Does the acknowledgement of αs1-casein genotype affect the estimation of genetic parameters and prediction of breeding values for milk yield and composition quality-related traits in murciano-Granadina? Animals 9, 679. https://doi.org/10.3390/ani9090679

Poindron, P., Lévy, F., Keller, M., 2007a. Maternal responsiveness and maternal selectivity in domestic sheep and goats: The two facets of maternal attachment. Dev. Psychobiol. 49, 54–70. https://doi.org/10.1002/dev.20192

Poindron, P., Terrazas, A., de la Luz Navarro Montes de Oca, M., Serafín, N., Hernández, H., 2007b. Sensory and physiological determinants of maternal behavior in the goat (*Capra hircus*). Horm. Behav. 52, 99–105. https://doi.org/10.1016/j.yhbeh.2007.03.023

Porter, V., Alderson, L., Hall, S.J.G., Sponenberg, D.P., 2016. Mason's World

Encyclopedia of Livestock Breeds and Breeding. CABI (Ed.), Oxfordshire, pp. 350–352.

Pulina, G., Milán, M.J., Lavín, M.P., Theodoridis, A., Morin, E., Capote, J., Thomas, D.L., Francesconi, A.H.D., Caja, G., 2018. Invited review: Current production trends, farm structures, and economics of the dairy sheep and goat sectors. J. Dairy Sci. 101, 6715–6729. https://doi.org/10.3168/jds.2017-14015

Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13, 70. https://doi.org/10.1186/1471-2156-13-70

Purfield, D.C., McParland, S., Wall, E., Berry, D.P., 2017. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One 12, e0176780. https://doi.org/10.1371/journal.pone.0176780

Qanbari, S., 2020. On the extent of linkage disequilibrium in the genome of farm animals. Front. Genet. 10, 1304. https://doi.org/10.3389/fgene.2019.01304

Quick, C., Anugu, P., Musani, S., Weiss, S.T., Burchard, E.G., White, M.J., Keys, K.L., Cucca, F., Sidore, C., Boehnke, M., Fuchsberger, C., 2020. Sequencing and imputation in GWAS: Cost-effective strategies to increase power and genomic coverage across diverse populations. Genet. Epidemiol. 44, 537–549. https://doi.org/10.1002/gepi.22326

Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T.P., Ross, R.P., Fitzgerald,

G.F., Cotter, P.D., 2013. The complex microbiota of raw milk, in: FEMS Microbiology Reviews. pp. 664–698. https://doi.org/10.1111/1574-6976.12030

Rahmatalla, S.A., Arends, D., Reissmann, M., Wimmers, K., Reyer, H., Brockmann, G.A., 2018. Genome-wide association study of body morphological traits in Sudanese goats. Anim. Genet. 49, 478–482. https://doi.org/10.1111/age.12686

Ramunno, L., Cosenza, G., Pappalardo, M., Longobardi, E., Gallo, D., Pastore, N., Di Gregorio, P., Rando, A., 2001a. Characterization of two new alleles at the goat *CSN1S2* locus. Anim. Genet. 32, 264–268. https://doi.org/10.1046/J.1365-2052.2001.00786.X

Ramunno, L., Longobardi, E., Pappalardo, M., Rando, A., Di Gregorio, P., Cosenza, G., Mariani, P., Pastore, N., Masina, P., 2001b. An allele associated with a non-detectable amount of alpha s2 casein in goat milk. Anim. Genet. 32, 19–26. https://doi.org/10.1046/J.1365-2052.2001.00710.X

Ramunno, L., Mariani, P., Pappalardo, M., Rando, A., Capuano, M., Di Gregorio, P., Cosenza, G., 1995. Un gene ad effetto maggiore sul contenuto di caseina β nel latte di capra, in: XI Convegno ASPA. Grado, Italy.

Ray, S., Tzeng, R.Y., DiCarlo, L.M., Bundy, J.L., Vied, C., Tyson, G., Nowakowski, R., Arbeitman, M.N., 2016. An examination of dynamic gene expression changes in the mouse brain during pregnancy and the postpartum period. G3 Genes, Genomes, Genet. 6, 221–233.

https://doi.org/10.1534/g3.115.020982

Reber, I., Keller, I., Becker, D., Flury, C., Welle, M., Drögemüller, C., 2015. Wattles in goats are associated with the FMN1/GREM1 region on chromosome 10. Anim. Genet. 46, 316–320. https://doi.org/10.1111/age.12279

Reinhold, K., 2002. Maternal effects and the evolution of behavioral and morphological characters: A literature review indicates the importance of extended maternal care. J. Hered. 93, 400–405. https://doi.org/10.1093/jhered/93.6.400

Ribeiro Júnior, J.C., de Oliveira, A.M., Silva, F. de G., Tamanini, R., de Oliveira, A.L.M., Beloti, V., 2018. The main spoilage-related psychrotrophic bacteria in refrigerated raw milk. J. Dairy Sci. 101, 75–83. https://doi.org/10.3168/jds.2017-13069

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rokouei, M., Vaez Torshizi, R., Moradi Shahrbabak, M., Sargolzaei, M., Sørensen, A.C., 2010. Monitoring inbreeding trends and inbreeding depression for economically important traits of Holstein cattle in Iran. J. Dairy Sci. 93, 3294–3302. https://doi.org/10.3168/jds.2009-2748

Romeyer, A., Poindron, P., Orgeur, P., 1994. Olfaction mediates the establishment of selective bonding in goats. Physiol. Behav. 56, 693–700.

https://doi.org/10.1016/0031-9384(94)90229-1

Rönnegård, L., McFarlane, S.E., Husby, A., Kawakami, T., Ellegren, H., Qvarnström, A., 2016. Increasing the power of genome wide association studies in natural populations using repeated measures – evaluation and implementation. Methods Ecol. Evol. 7, 792–799. https://doi.org/10.1111/2041-210X.12535

Rovai, M., Thomas, D.L., Berger, Y.M., Caja, G., 2004. Udder morphology and effects on milk production and ease of milking in dairy sheep. Proc. 10th Gt. Lakes Dairy Sheep Symp. Wisconsin 4–6.

Rupp, R., Clément, V., Piacere, A., Robert-Granié, C., Manfredi, E., 2011. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. J. Dairy Sci. 94, 3629–3634. https://doi.org/10.3168/jds.2010-3694

Ruotsalainen, S.E., Partanen, J.J., Cichonska, A., Lin J., Benner, C., Surakka, I., et al., 2021. An expanded analysis framework for multivariate GWAS connects inflammatory biomarkers to functional variants and disease. Eur J Hum Genet 29, 309–324. https://doi.org/10.1038/s41431-020-00730-8

Salazar, J.K., Carstens, C.K., Ramachandran, P., Shazer, A.G., Narula, S.S., Reed, E., Ottesen, A., Schill, K.M., 2018. Metagenomics of pasteurized and unpasteurized gouda cheese using targeted 16S rDNA sequencing. BMC Microbiol. 18, 189. https://doi.org/10.1186/s12866-018-1323-4

Salque, M., Bogucki, P.I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt,

M., Evershed, R.P., 2013. Earliest evidence for cheese making in the sixth millennium bc in northern Europe. Nature. 493, 522–525. https://doi.org/10.1038/nature11698

Sanger, F., Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94, 441–448. https://doi.org/10.1016/0022-2836(75)90213-2

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sartori, C., Guzzo, N., Mazza, S., Mantovani, R., 2018. Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. Animal, 12, 906-914.

Saura, M., Fernández, A.I.A., Varona, L., Fernández, A.I.A., De Cara, M.Á.R., Barragán, C., Villanueva, B., 2015. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genet. Sel. Evol. 47, 1. https://doi.org/10.1186/s12711-014-0081-5

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470. https://doi.org/10.1126/SCIENCE.270.5235.467

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009.

Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09

Scholtens, M., Jiang, A., Smith, A., Littlejohn, M., Lehnert, K., Snell, R., Lopez-Villalobos, N., Garrick, D., Blair, H., 2020. Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. J. Anim. Sci. Biotechnol. 11, 1–14. https://doi.org/10.1186/s40104-020-00453-2

Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–1092. https://doi.org/10.1093/bioinformatics/bts094

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., Blaxter, M., Barton, G.J., 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 22, 839–851. https://doi.org/10.1261/rna.053959.115

Selvaggi, M., Laudadio, V., Dario, C., Tufarelli, V., 2014. Major proteins in goat milk: An updated overview on genetic variability. Mol. Biol. Rep. 41, 1035–1048. https://doi.org/10.1007/s11033-013-2949-9

Sewalem, A., Miglior, F., Kistemaker, G.J., Van Doormaal, B.J., 2006. Analysis of the relationship between somatic cell score and functional longevity in

Canadian dairy cattle. J. Dairy Sci. 89, 3609–3614. https://doi.org/10.3168/jds.S0022-0302(06)72400-6

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sharma, A., Lee, J.S., Dang, C.G., Sudrajad, P., Kim, H.C., Yeon, S.H., Kang, H.S., Lee, S.H., 2015. Stories and challenges of genome wide association studies in livestock - a review. Asian-Australasian J. Anim. Sci. 28, 1371–1379. https://doi.org/10.5713/ajas.14.0715

Sharma, A., Park, J.-E., Chai, H.-H., Jang, G.-W., Lee, S.-H., Lim, D., 2017. Next generation sequencing in livestock species- A Review. J. Anim. Breed. Genomics 1, 23–30. https://doi.org/10.12972/jabng.20170003

Sharma, N., Singh, N.K., Bhadwal, M.S., 2011. Relationship of somatic cell count and mastitis: An overview. Asian-Australasian J. Anim. Sci. 24, 429–438. https://doi.org/10.5713/AJAS.2011.10233

Shingo, T., Gregg, C., Enwere, E., Fujikawa, H., Hassam, R., Geary, C., Cross, J.C., Weiss, S., 2003. Pregnancy-stimulated neurogenesis in the adult female forebrain mediated by prolactin. Science 299, 117–120. https://doi.org/10.1126/science.1076647

Signer-Hasler, H., Burren, A., Neuditschko, M., Frischknecht, M., Garrick, D., Stricker, C., Gredler, B., Bapst, B., Flury, C., 2017. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet. Sel. Evol. 49, 1–13. https://doi.org/10.1186/s12711-017-0358-6

Silanikove, N., Leitner, G., Merin, U., Prosser, C.G., 2010. Recent advances in exploiting goat's milk: Quality, safety and production aspects. Small Rumin. Res. 89, 110–124. https://doi.org/10.1016/j.smallrumres.2009.12.033

Silva, C.N.S., McFarlane, S.E., Hagen, I.J., Rönnegård, L., Billing, A.M., Kvalnes, T., Kemppainen, P., Rønning, B., Ringsby, T.H., Sæther, B.E., Qvarnström, A., Ellegren, H., Jensen, H., Husby, A., 2017. Insights into the genetic architecture of morphological traits in two passerine bird species. Heredity 119, 197–205. https://doi.org/10.1038/hdy.2017.29

Singh, K., Molenaar, A.J., Swanson, K.M., Gudex, B., Arias, J.A., Erdman, R.A., Stelwagen, K., 2012. Epigenetics: A possible role in acute and transgenerational regulation of dairy cow milk production. Animal 6, 375–381. https://doi.org/10.1017/S1751731111002564

Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., Huang, J., Du, Y., Lin, L., Dong, Z., Yang, L., Liu, X., Jiang, H., Xu, X., Wang, J., Yang, H., Bolund, L., Mardinoglu, A., Zhang, C., Von Feilitzen, K., Lindskog, C., Pontén, F., Luo, Y., Hökfelt, T., Uhlén, M., Mulder, J., 2020. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science 367, eaay5947. https://doi.org/10.1126/science.aay4106

Solieri, L., Dakal, T.C., Giudici, P., 2013. Next-generation sequencing and its potential impact on food microbial genomics. Ann. Microbiol. 63, 21–37. https://doi.org/10.1007/s13213-012-0478-8

Soneson, C., Delorenzi, M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14, 1–18. https://doi.org/10.1186/1471-2105-14-91

Stella, A., Nicolazzi, E.L., Van Tassell, C.P., Rothschild, M., Colli, L., Rosen, B., et al., 2018. AdaptMap: exploring goat diversity and adaptation. Genet Sel Evol 50, 61. https://doi.org/10.1186/s12711-018-0427-5

Talouarn, E., Bardou, P., Palhière, I., Oget, C., Clément, V., Tosser-Klopp, G., Rupp, R., Robert-Granié, C., 2020. Genome wide association analysis on semen volume and milk yield using different strategies of imputation to whole genome sequence in French dairy goats. BMC Genet. 21, 1–13. https://doi.org/10.1186/s12863-020-0826-9

Tebani, A., Afonso, C., Marret, S., Bekri, S., 2016. Omics-based strategies in precision medicine: Toward a paradigm shift in inborn errors of metabolism investigations. Int. J. Mol. Sci. 17, 1555. https://doi.org/10.3390/ijms17091555

Teixeira, J.C., Huber, C.D., 2021. The inflated significance of neutral genetic diversity in conservation genetics. Proc. Natl. Acad. Sci. U. S. A. 118, e2015096118. https://doi.org/10.1073/pnas.2015096118

Théoret-Gosselin, R., Hamel, S., Côté, S.D., 2015. The role of maternal behavior and offspring development in the survival of mountain goat kids. Oecologia 178, 175–186. https://doi.org/10.1007/s00442-014-3198-x

Thompson, E.A., 2013. Identity by descent: Variation in meiosis, across genomes,

and in populations. Genetics 194, 301–326. https://doi.org/10.1534/genetics.112.148825

Tilahun, Y., Gipson, T.A., Alexander, T., McCallum, M.L., Hoyt, P.R., 2020. Genome-wide association study towards genomic predictive power for high production and quality of milk in american alpine goats. Int. J. Genomics 2020, 6035694. https://doi.org/10.1155/2020/6035694

Tong, A.K.W., Kennedy, B.W., Moxley, J.E., 1979. Heritabilities and genetic correlations for the first three lactations from records subject to culling. J. Dairy Sci. 62, 1784–1790. https://doi.org/10.3168/jds.S0022-0302(79)83497-9

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., Donnadieu-Tonon, C., Eggen, A., Heuven, H.C.M., Jamli, S., Jiken, A.J., Klopp, C., Lawley, C.T., McEwan, J., Martin, P., Moreno, C.R., Mulsant, P., Nabihoudine, I., Pailhoux, E., Palhière, I., Rupp, R., Sarry, J., Sayre, B.L., Tircazes, A., Wang, J., Wang, W., Zhang, W., Ajmone, P., Amills, M., Boitard, S., Faraut, T., San Cristobal, M., Servin, B., Cheng, W., Cheng, S., Liu, X., Pan, S., Song, C., Xu, X., Ye, C., Zhang, B., Lv, J., Li, X.X., Ren, L., Shi, P., Yu, J., Faruque, O., Lenstra, H., Poli, M.A., Zhao, J., Rui, S., Zhang, Y., Stella, A., Li, X.X., Valentini, A., Zhao, S., 2014. Design and characterization of a 52K SNP chip for goats. PLoS One 9, e86227. https://doi.org/10.1371/journal.pone.0086227

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren,

M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515. https://doi.org/10.1038/nbt.1621

Trerotola, M., Relli, V., Simeone, P., Alberti, S., 2015. Epigenetic inheritance and the missing heritability. Hum. Genomics 9, 17. https://doi.org/10.1186/s40246-015-0041-3

Trujillo, A.J., Jordana, J., Guamis, B., Serradilla, J.M., Amills, M., 1998. Review: Polymorphism of the caprine α s1-casein gene and its effect on the production, composition and technological properties of milk and on cheese making and ripening. Food Sci. Technol. Int. 4, 217–235. https://doi.org/10.1177/108201329800400401

Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., Posthuma, D., 2021. Genome-wide association studies. Nat. Rev. Methods Prim. 1, 59. https://doi.org/10.1038/s43586-021-00056-9

Vailati-Riboni, M., Palombo, V., Loor, J.J., 2017. What are omics sciences?, in: Periparturient Diseases of Dairy Cows: A Systems Biology Approach. Springer International Publishing, pp. 1–7. https://doi.org/10.1007/978-3-319-43033-1_1

Vaisvila, R., Ponnaluri, V.K.C., Sun, Z., Langhorst, B.W., Saleh, L., Guan, S., Dai, N., Campbell, M.A., Sexton, B.S., Marks, K., Samaranayake, M.,

Samuelson, J.C., Church, H.E., Tamanaha, E., Corrêa, I.R., Pradhan, S., Dimalanta, E.T., Evans, T.C., Williams, L., Davis, T.B., 2021. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res. 31, 1280–1289. https://doi.org/10.1101/gr.266551.120

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980

Vázquez-Flores, F., Montaldo, H.H., Torres-Vázquez, J.A., Alonso-Morales, R.A., Gayosso-Vázquez, A., Valencia-Posadas, M., Castillo-Juárez, H., 2012. Additive and dominance effects of the αs1-casein locus on milk yield and composition traits in dairy goats. J. Dairy Res. 79, 367–374. https://doi.org/10.1017/S0022029912000350

Vázquez-Gómez, M., de Hijas-Villalba, M.M., Varona, L., Ibañez-Escriche, N., Rosas, J.P., Negro, S., Noguera, J.L., Casellas, J., 2020. Maternal transmission ratio distortion in two iberian pig varieties. Genes. 11, 1050. https://doi.org/10.3390/genes11091050

Vernikos, G.S., 2020. A review of pangenome tools and recent studies, in: The pangenome: Diversity, dynamics and evolution of genomes. Springer International Publishing, pp. 89–112. https://doi.org/10.1007/978-3-030-38281-0_4

Walsh, A.M., Macori, G., Kilcawley, K.N., Cotter, P.D., 2020. Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality.

Nat. Food 1, 500–510. https://doi.org/10.1038/s43016-020-0129-3

Wang, S., Li, F., Liu, J., Zhang, Y., Zheng, Y., Ge, W., Qu, L., Wang, X., 2020. Integrative analysis of methylome and transcriptome reveals the regulatory mechanisms of hair follicle morphogenesis in cashmere goat. Cells 9, 969. https://doi.org/10.3390/CELLS9040969

Wang, W., Pan, Y.W., Wietecha, T., Zou, J., Abel, G.M., Kuo, C.T., Xia, Z., 2013. Extracellular signal-regulated kinase 5 (*ERK5*) mediates prolactin-stimulated adult neurogenesis in the subventricular zone and olfactory bulb. J. Biol. Chem. 288, 2623–2631. https://doi.org/10.1074/jbc.M112.401091

Weigel, K.A., 2001. Controlling inbreeding in modern breeding programs. J. Dairy Sci. 84, E177–E184. https://doi.org/10.3168/jds.s0022-0302(01)70213-5

Whittaker, R.H., 1972. Evolution and measurement of species diversity. Taxon 21, 213–251. https://doi.org/10.2307/1218190

Wolfe, B.E., Button, J.E., Santarelli, M., Dutton, R.J., 2014. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. Cell 158, 422–433. https://doi.org/10.1016/j.cell.2014.05.041

Wright, S., 1922. Coefficients of inbreeding and relationship. Am. Nat. 56, 330–338. https://doi.org/10.1086/279872

Xiao, P., Zhong, T., Liu, Z., Ding, Y., Guan, W., He, X., Pu, Y., Jiang, L., Ma, Y., Zhao, Q., 2020. Integrated analysis of methylome and transcriptome changes

reveals the underlying regulatory signatures driving curly wool transformation in chinese zhongwei goats. Front. Genet. 10, 1263. https://doi.org/10.3389/fgene.2019.01263

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.W., Li, Y., Xu, X., Wong, G.K.S., Wang, J., 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30, 1660–1666. https://doi.org/10.1093/BIOINFORMATICS/BTU077

Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yeluri Jonnala, B.R., McSweeney, P.L.H., Sheehan, J.J., Cotter, P.D., 2018. Sequencing of the Cheese Microbiome and Its Relevance to Industry. Front. Microbiol. 9, 1020. https://doi.org/10.3389/FMICB.2018.01020

Zapala, M.A., Hovatta, I., Ellison, J.A., Wodicka, L., Del Rio, J.A., Tennant, R., Tynan, W., Broide, R.S., Helton, R., Stoveken, B.S., Winrow, C., Lockhart, Daniel J., Reilly, J.F., Young, W.G., Bloom, F.E., Lockhart, David J., Barlow, C., 2005. Adult mouse brain gene expression patterns bear an embryologic imprint. Proc. Natl. Acad. Sci. U. S. A. 102, 10357–10362. https://doi.org/10.1073/pnas.0503357102

Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proc. Natl. Acad. Sci. U. S. A. 105,

11597–11604. https://doi.org/10.1073/pnas.0801317105

Zhang, L., Wang, F., Gao, G., Yan, X., Liu, H., Liu, Z., Wang, Zhixin, He, L., Lv, Q., Wang, Zhiying, Wang, R., Zhang, Y., Li, J., Su, R., 2021. Genome-wide association study of body weight traits in Inner Mongolia Cashmere goats. Front. Vet. Sci. 8, 752746. https://doi.org/10.3389/fvets.2021.752746

Zhang, Q., Calus, M.P., Guldbrandtsen, B., Lund, M.S., Sahana, G., 2015a. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet. 16, 88. https://doi.org/10.1186/s12863-015-0227-7

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M.S., Sahana, G., 2015b. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics 16, 542. https://doi.org/10.1186/s12864-015-1715-x

Zhang, W.Y., Yuan, Y., Zhang, H.Y., He, Y.M., Liu, C.L., Xu, L., Yang, B.G., Ren, H.X., Wang, G.F., Guang-Xin, E., 2022. Genetic basis investigation of wattle phenotype in goat using genome-wide sequence data. Anim. Genet. 53, 700–705. https://doi.org/10.1111/age.13235

Zhang, Y., Kastman, E.K., Guasto, J.S., Wolfe, B.E., 2018. Fungal networks shape dynamics of bacterial dispersal and community assembly in cheese rind microbiomes. Nat. Commun. 9, 1–12. https://doi.org/10.1038/s41467-017-02522-z

Zidi, A., Casas, E., Amills, M., Jordana, J., Carrizosa, J., Urrutia, B., Serradilla, J.M., 2014. Genetic variation at the caprine lactalbumin, alpha (*LALBA*) gene

and its association with milk lactose concentration. Anim. Genet. 45, 612–613. https://doi.org/10.1111/AGE.12152

Zonaed Siddiki, A.M.A.M., Miah, G., Islam, M.S., Kumkum, M., Rumi, M.H., Baten, A., Hossain, M.A., 2020. Goat genomic resources: The search for genes associated with its economic traits. Int. J. Genomics 2020, 5940205. https://doi.org/10.1155/2020/5940205

7 ANNEXES

All supplementary materials and related documents are available to download at the following link:

https://figshare.com/articles/thesis/Exploring_the_molecular_basis_of_dairy_production_and_fitness_traits_in_Murciano-Granadina_goats_with_genomic_metagenomic_and_transcriptomic_approaches_-_Supplementary_materials/21155467

DOI: 10.6084/m9.figshare.21155467