Detection of QTL influencing mammary morphology in dairy sheep

Vrcan, Marko

UNIVERSITY OF ZAGREB
FACULTY OF AGRONOMY

DETECTION OF QTL INFLUENCING MAMMARY MORPHOLOGY IN DAIRY SHEEP

MASTER'S THESIS

Marko Vrcan

Zagreb, september, 2022.

UNIVERSITY OF ZAGREB
FACULTY OF AGRONOMY

Master study:

Animal Genetics and Breeding

DETECTION OF QTL INFLUENCING MAMMARY MORPHOLOGY IN DAIRY SHEEP

MASTER'S THESIS

Marko Vrcan

Mentors:

prof. Alen Džidić, PhD

asst. prof. Aroa Suárez-Vega, PhD

Zagreb, september, 2022.

SVEUČILIŠTE U ZAGREBU
AGRONOMSKI FAKULTET

IZJAVA STUDENTA

O AKADEMSKOJ ČESTITOSTI

Ja, **Marko Vrcan**, JMBAG 0178114754, rođen/a 4.12.1997. u Splitu, izjavljujem da sam samostalno izradio diplomski rad pod naslovom:

DETECTION OF QTL INFLUENCING MAMMARY MORPHOLOGY IN DAIRY SHEEP

Svojim potpisom jamčim:

- da sam jedina autorica/jedini autor ovoga diplomskog rada;
- da su svi korišteni izvori literature, kako objavljeni tako i neobjavljeni, adekvatno citirani ili parafrazirani, te popisani u literaturi na kraju rada;
- da ovaj diplomski rad ne sadrži dijelove radova predanih na Agronomskom fakultetu ili drugim ustanovama visokog obrazovanja radi završetka sveučilišnog ili stručnog studija;
- da je elektronička verzija ovoga diplomskog rada identična tiskanoj koju je odobrio mentor;
- da sam upoznata/upoznat s odredbama Etičkog kodeksa Sveučilišta u Zagrebu (Čl. 19).

U Zagrebu, dana _____ _____

Potpis studenta / studentice

SVEUČILIŠTE U ZAGREBU
AGRONOMSKI FAKULTET

IZVJEŠĆE

O OCJENI I OBRANI DIPLOMSKOG RADA

Diplomski rad studenta/ice **Marko Vrcan**, JMBAG 0178114754, naslova

DETECTION OF QTL INFLUENCING MAMMARY MORPHOLOGY IN DAIRY SHEEP

obranjen je i ocijenjen ocjenom _____ , dana _____ .

Povjerenstvo:			potpisi:
1.	prof. Alen Džidić, PhD	mentor	_____
2.	asst. prof. Aroa Suárez-Vega, PhD	co-mentor	_____
3.	assoc. prof. Dragica Šalamon, PhD	member	_____
4.	asst. prof. Ante Kasap, PhD	member	_____

Summary

Of the master's thesis – student **Marko Vrcan,** entitled

DETECTION OF QTL INFLUENCING MAMMARY MORPHOLOGY IN DAIRY SHEEP

Mammary morphology related traits are the most important functional traits in dairy sheep due to their relationship with machine milking aptitude, their influence on udder pathology, mainly mastitis, and on the animal's welfare. This aim of this study was the identification of quantitative trait loci (QTL) associated with mammary morphology in Churra dairy sheep based on data generated with the Illumina OvineSNP50 BeadChip (50K-chip). Two different genome scan QTL mapping analyses were performed based on traditional linkage analysis (LA) and a genome-wide association study (GWAS). The studied population belonged to the Churra dairy Selection Nucleus (ANCHE) and it included a total number of 1,680 ewes distributed in 16 half-sib families. For the four traits considered (Udder depth, Udder attachment, Teat placement and Teat size), we detected a total of 10 significant QTL by LA and 20 significant regions by GWAS (one and two of these significant regions, respectively, reached the 5% genome-wise significance level). For the confidence intervals defined for each of the significant QTL identified, we performed gene and QTL annotation and QTL enrichment analyses. The comparison of our with QTLs previously reported in sheep and cattle for mammary morphology and mastitis resistance related traits showed interesting coincidences that support the genuine nature of the QTL here reported. Based on a reference studies, a list of eight potential functional candidate genes has been identified. The possible direct association of these genes on udder morphology traits of interest, which should be confirmed by future studies, could help to identify genetic markers to increase efficiency of the Churra sheep breeding programe to improve the udder morphology traits considered in the present study.

Keywords: mammary morphology, 50K-chip, QTL, Genome Wide Association study (GWAS), Linkage Analysis (LA)

Sažetak

Diplomskog rada studenta/ice **Marko Vrcan**, naslova

DETEKCIJA QTL KOJI IMAJU UTJECAJ NA SVOJSTVA MLIJEČNE ŽLIJEZDE

Svojstva povezana s morfologijom mliječne žlijezde najvažnija su funkcionalna svojstva kod uzgoja ovaca za proizvodnju mlijeka zbog njihove povezanosti sa strojnom mužnjom, utjecajem na patologiju vimena, uglavnom mastitis, i dobrobiti životinje. Cilj ovog istraživanja bio je identificirati lokus kvantitativnih svojstava (eng. *Quantitative trait loci*, QTL) povezanih sa morfologijom mliječne žlijezde Churra pasmine na temelju podataka generiranih s Illumina OvineSNP50 BeadChip (50K-čip). Provedene su dvije različite QTL analize skeniranja genoma temeljene na tradicionalnoj analizi povezanosti (eng. *Linkage analysis*, LA) i povezanosti preko cijelog genoma (eng. *Genome wise association study*, GWAS). Proučavana populacija Churra pasmine pripadala je selekcijskom nukleusu ANCHE (eng. National Association of Churra Breeders) i uključivala ukupan broj od 1680 ovaca raspoređenih u 16 polubrat/polusestra obitelji. Za četiri proučavana svojstva (dubina vimena, vanjski oblik vimena, položaj sise i veličina sise), identificirali smo ukupno 10 značajnih regija koristeći LA i 20 značajnih regija koristeći GWAS (jedna i dvije navedenih značajnih regija dosegnule su 5% genomsku razinu značajnosti). Za intervale pouzdanosti definirane za svaki od identificiranih značajnih QTL-ova, izvršili smo anotaciju gena i QTL-a kao i analizu obogaćivanja (eng. *enrichment analysis*). Usporedba QTL-ova identificiranih u ovom istraživanju s QTL-ovima prethodno prijavljenim kod ovaca i goveda za morfologiju dojke i osobine povezane s otpornošću na mastitis pokazala je zanimljive podudarnosti. Na temelju referentnih studija identificirano je osam potencijalnih funkcionalnih kandidat gena. Moguća izravna povezanost ovih gena sa zanimljivim morfološkim svojstvima vimena, koja bi trebala biti potvrđena budućim studijama, mogla bi pomoći u identificiranju genetskih markera u svrhu povećanja učinkovitosti uzgojnog programa Churra pasmine za poboljšanje morfoloških svojstava vimena razmatranih u ovoj studiji.

Keywords: svojstva mliječne žlijezde, 50K-chip, QTL, Genome Wide Association study (GWAS), Linkage Analysis (LA)

Contents

1. Introduction

The Human Genome Project, or the determination of the nucleotide sequence of the human DNA, is probably one of the most outstanding scientific projects in the history of civilization (Venter et al., 2001). The work on that project led to the establishment of a new science branch, genomics. Genomics is a scientific area built on the basis of the accomplishments in molecular biology made by Sanger et. al. (1977), who sequenced the first complete genome, called phiX174 virus, which opened the doorway for the development of techniques that can determine nucleotide sequence in nucleic acid molecules. In order to understand the structure and function of these sequences and of the ensuing biological products, genomics is the study of all or a portion of an organism's genetic or epigenetic sequence information (Fadiel et al., 2005). In animal breeding, where the study of genomics is crucial to maximize animal production across a variety of environments, new technologies have been developed as a result of genome-based analyses that deepen our understanding of the genetic basis of production traits (Rexroad et al., 2019). These technologies may be used to optimize animal breeding strategies and high-quality management decisions.

Some authors have divided genomics into two branches: structural genomics, for which the main goal is to visualize the three-dimensional structure of every protein using either experimental, computational, or a mix of both methods (Skolnick et al., 2000), and functional genomics, which instead of focusing on a single gene in the genome, examines the structure, function, and regulation of all genes as well as the dynamic processes including gene transcription, translation, and protein-protein interactions (Bunnik and le Roch, 2013).

Protein analysis has slowly turned into a particular scientific discipline, analogously to genomics, called proteomics (Blackstock and Weir, 1999). Another specific scientific field, comparative genomics has been developed based on genomics results. By discovering the general principles relating to the group of genomes, comparative genomics plays a role in determining the diversity of the living world (Hardison, 2003; Miller et al., 2004).

Together with the development of genomics, the scientific discipline of bioinformatics has been developed. Bayat (2002) defined bioinformatics as the application of computation tools and analysis to the capture and interpretation of biological data. Bioinformatics is an interdisciplinary field, which operates and integrates knowledge from different sciences such as computer science, mathematics, physics, and biology. Also, it is essential for management of data in modern biology and medicine (Bayat, 2002).

Based on the advances of the sequencing technologies in the last years, billions of genome base pairs are being deciphered into hundreds of species. Such extensive work requires the formation of specific databases. These databases aim to make easily accessible the results of scientists' work on deciphering the genomes of numerous organisms available to all interested researchers

worldwide via the Internet. One of the most famous databases is available on the US National Center for Biotechnology Information (NCBI) website. The address of this site is http://www.ncbi.nlm.nih.gov. Another reference repository database for genomic data is Ensembl (Hubbard et al., 2002), which is a project that offers a bioinformatics framework to organize biology around the sequences of big genomes. With verified gene predictions that have been integrated with other data sources, it provides a complete source of stable automatic annotation of the human and animal genome sequences that is accessible as either a flat file or an interactive website. It is a portable system capable of handling extremely big genomes and the related requirements, from sequence analysis to data storage and visualization, are also being developed as an opensource software engineering project. In addition, Europe's primary nucleotide sequence data resource is the European Bioinformatics institute (EBI), available at http://www.ebu.ac.uk. The SWISS-PROT Protein Sequence database is maintained and made available by the EBI in partnership with Amos Bairoch of the University of Geneva. These are more than fifty additional specialized molecular biology databases, along with software and literature of interest to molecular biologists. In addition to the USA supported NCBI repository, one national DNA database to highlight is the Japanese one, available at http://www.nig.oc.jp/home.html9. In recent years, the data included in all the above mentioned databases have been combined into a single database known as GOLD (from the English name Genomes OnLine Database) available at http://www.genomesonline.org26. The future of genomics is applying its achievements in many natural sciences such as agronomy (agrogenomics), pharmacology (Pharmacogenomics), and especially medicine.

The main goal of genetic research is to understand the molecular mechanisms through which genetic variants influence or control the quality of interest. In livestock species, most of the traits of economic interest are quantitative traits in nature. Quantitative traits are qualities that, at the genetic level, do not behave according to the simple Mendelian laws as the genetic segregation of one gene cannot explain their heritage, and they are the result of the action of many genes with small effect (Hayes et al., 2001). Quantitative traits are normally distributed, which means that they present a continuous range of variability and are influenced by environmental and genetic factors as well as the interaction between genes and environment. This is why these traits are also known as complex traits. Compared to Mendelian traits (qualitative), quantitative traits are much more common and affected by many genes with minor effects (Hayes et al., 2001). They require complex statistical analyses because of their complexity. Examples of complex characteristics in domestic animals include milk production, milk component percentages, udder morphology, body morphology, etc.

The genetic study of complex traits requires the use of complex statistical methods based on the polygenetic hypothesis. Hence, quantitative genetics attempts to discover a holistic system of all genes associated with complex and quantitative traits. As we have already mentioned, this strategy could not break down the effect of individual genes, following the Mendelian laws. However, the development of molecular methods and quantitative techniques developed during '90s resulted in the possibility of monitoring quantitative properties of the polygenetic base with a detailed description of the effects of genes on phenotypic variation (Hill, 2010). Later, with the appearance of genomics, significant progress has been made in reaching a better

understanding of quantitative genetics (Hill, 2012). One of the most effective interests of quantitative genetics is the systematic study of the genetic architecture, number, distribution, and interaction of loci and genetic variants that influence the phenotypic variation of economically important traits.

In livestock species, similar to human and mouse, the identification of genetic variants associated with complex traits of interest was initially undertaken through the candidate gene strategy. This strategy is based on the study of genes that, due to their physiological function, can influence biological pathways associated with the expression of specific phenotypes (Moioli et al., 2007). The identification of genetic variants in these genes and the subsequent performance of association analyses was the classical workflow performed with the aim of identifying significant associations of complex genes under study. However, this strategy has as the main limitation the lack of full knowledge of the physiological pathways involved in each phenotype (Gutiérrez-Gil, 2004). As an alternative to the candidate gene studies, and taking advantage of the growing description of genetic markers in the 1980's, non-coding markers, such as microsatellites, were used to scan the genome in search of regions associated with the phenotypic variation observed in complex traits such as milk production, carcass weight, etc. Thus, in the last decade of the 20th century, several mapping projects aiming the identification of the genes of small effect that influence complex traits in livestock species, and which are known as Quantitaitve Trait Loci or QTL, were initiated initially in pigs (Andersson et al., 1994) and cattle (Georges et al., 1995). These studies, based on Linkage Analysis (LA) and family-structured populations, showed in general a limited mapping accuracy although in same cases led to the identification of causal mutations influencing the studied traits, which in this context are known as Quantitaitve Trait Nucleotide (QTN) (Grisart et al., 2004). Genome scans in other species such as sheep were reported in the first decade of the XXI century (Gutiérrez-Gil et al., 2009a;2009b; Raadsma et al., 2009), based, in general, on the analysis of about 150-200 microsatellite genetic markers.

The large progress made on sequencing technologies at the end of the Human Genome Project, led to the development of the next-generation sequencing (NGS) technologies and made possible the completion of sequencing projects in non-model livestock species during the first decade of the present century. These projects resulted in the identification of thousands of single nucleotide polymorphisms (SNPs) across the studied genomes and the development of SNP-chip arrays. Most previous genotyping approaches for SNP genotyping were too slow and expensive to be used in large-scale studies involving hundreds or thousands of SNPs for a large number of DNA samples. Because of this, high-throughput genotyping technologies based on SNP-arrays appeared as an efficient genomic tool for a faster, more efficient, and less expensive procedure to analyze the genome of livestock species, while also triggered the improvement of the existing analysis methods (Tsuchihashi & Dracopoli, 2002). Hence, the availability of SNP-arrays determined the publication of a large number of gene-mapping studies based on the application of the Genome-wide Association Study (GWAS) methodology, which avoids some of the limitations of classical LA and remarkably improves mapping accuracy. In addition, the availability of SNP-chips has allowed the implementation of Genomic Selection (GS), which is a MAS variant based on animal genotypic data (Meuwissen et al., 2001). Additionally, the

democratization of sequencing technologies has allowed the routine study of whole genomes and transcriptomes (Forrest and Carninci, 2009). The enormous amount of information produced in this way means a large challenge for data analysis, but at the same time brings hope of achieving a deeper understanding on the genetic and functional mechanisms that control complex traits of interest in livestock species.

The present Master's thesis has been carried out through the analysis of genomic data provided by the Research Group of Animal Breeding (MEGA) at the University of León (ULE), in Spain. This group has a large expertise in the genetic study of economically important traits in dairy sheep. Several studies conducted by the MEGA-ULE group were initially focused on the study of the environmental factors that influence these traits and the estimation of their genetic parameters (De la Fuente et al., 1996; Othmane et al., 2002; Gutiérrez-Gil et al., 2010). These studies were developed mainly in commercial populations of the Spanish Churra sheep breed to estimate the potential of classical genetic selection to improve the traits of economic interest in the considered population. In relation to molecular studies to identify genes associated with traits of interest in dairy sheep, initial studies of this group were based on the candidate gene strategy (Gutiérrez-Gil et al., 2001; Garcia-Fernández et al., 2011; Pérez, 2016). This group was also responsible of the first LA microsatellite-based genome scans described in dairy sheep for a wide range of traits, including milk production and milk composition traits, mammary and body morphology traits, mastitis and nematode resistance (Gutiérrez-Gil et al., 2007, 2008, 2009a, 2009b, 2011).

After 2009, the availability of the Illumina Ovine SNP50 BeadChip® (Chip-50K), developed by the International Sheep Genomic Consortium (ISGC), provided an efficient tool to undertake research on the ovine genomic architecture of complex traits (International Sheep Genomics Consortium, 2008). This medium-density high-throughput genotyping array, which will be referred from now own as the 50K-Chip, has allowed sheep genomic researchers to conduct GWAS analyses with a large number of markers equally scattered throughout the sheep genome that do not require the family-structured designs as traditional QTL mapping (Becker et al., 2010). As an additional advantage, the information provided by the 50K-Chip can be also evaluated using the concepts of classical Linkage Analysis (LA) or by combining linkage analysis with Linkage Disequilibrium analysis (LD, Linkage Disequilibrium), through the combined LDLA strategy which can be appropriate for commercial populations that have an intrinsic family structure. Using the genotypic information generated with the 50K-Chip in a commercial population of Spanish Churra sheep, the MEGA-ULE group has applied these three types of analysis strategies (LA, LDLA and GWAS) to the study of the genetics underlying milk production traits and gastrointestinal nematode resistance indicator traits (Garcia-Gámez et al., 2012c; 2013; Atlija et al., 2016).

Udder morphological traits are one of the most important functional traits in dairy sheep because they determine the animal's machine milking efficiency and have a significant impact on its functional longevity (Labussière, 1988). Currently, udder morphological qualities are, after milk production traits, considered as selection targets in several of the sheep breeding programs currently undertaken in Spain. In Churra sheep these traits were included as selection

objective in 2002. Visual examination and score assignment on a linear scale described by De la Fuente et al. (1996) is used in the routine measuring of these traits. Despite the moderate to high heritability estimates of these traits (Fernandez et al., 1997), which ensure a good response to selection, a greater understanding of the genetic architecture underpinning mammary morphology traits could improve the effectiveness of classical selection. Apart of the microsatellite-based genome scan reported in Churra sheep by Gutierrez-Gil et al., (2008), according to the SheepQTLdatabase (https://www.animalgenome.org/cgi-bin/QTLdb/OA/index), only two GWAS analyses have been published in relation to morphology of the mammary gland in the Chinese Hu sheep breed. Both studies considered traits related to the number of teats or supernumerary teats (Peng et al., 2017; Zhao et al., 2022). Nowadays, despite the potential of Genomic Selection to improve any trait of economic interest without completing understanding the genetic underlying mechanisms, the interest of identifying QTL and QTN associated to udder morphology is still justified by the fact that the efficiency of Genomic Selection can be improved by adding information of specific mutations that have a direct effect on the phenotype of interest (Teng et al., 2020).

Based on all the above-mentioned points, the global aim of this Master's Thesis memory is the identification and annotation of QTL for mammary morphology traits in a commercial population of Spanish Churra sheep by using the available medium-density 50K-Chip genotypes to identify potential candidate genes that could be used to improve the efficiency of classical or genomic selection of these traits. To ensure the accomplishment of this global objective, the following specific objectives have been defined:

1. Using the medium-density ovine SNP-array to perform QTL detection analysis for four mammary morphology traits in Churra sheep by applying two analysis methods: Linkage Analysis (LA) and genome-wide association study (GWAS) (in order to discover regions influencing mammary morphological traits).
2. Gene and QTL annotation of the QTL regions identified, based on the latest sheep reference genome, Oar_ram_v2.0 and comparative analysis with QTL previous identified in cattle for udder morphology traits.
3. Identification of the most promising candidate genes that could be further studied through complementary –omic analyses including whole genome and transcriptomic sequencing with the aim of identifying genetic markers to increase the efficiency of classical and genomic selection for udder traits in Churra sheep.

2. Literature review

2.1. Socio-economic importance of dairy sheep in Castilla y León, Spain

Sheep (*Ovis aries*) and goats (*Capra hircus*) were among the first domesticated animals. This domestication process is thought to have begun in West Asia 9000 to 12,000 years ago (Stiner et al., 2022). Domesticated sheep and goats offered fiber, meat, and milk to early humans. Small ruminants were still the principal food supply in dry and remote parts of the earth due to their modest growth and adaptability. Currently, the ovine species appears in a variety of breeds and sub-breeds as a result of numerous generations of domestication in the Eurasian and African regions, and later processes of adaptation to different regions and specialization through the implementation of genetic selection programs (Chessa et al., 2009).

Dairy sheep production is a strategic farming alternative in the Mediterranean basin (FAOSTAT, 2018), where the climate limits many agricultural activities and where there is a long tradition of consuming sheep dairy products (e.g., cheese and yoghurt). Ministry of Agriculture, Food and Environment of Spain, Spain accounted in 2020 for 27.3% of the sheep milk delivered to the industry in the European Union (MAPA, 2022). With over 15 million dairy sheep (13 % of the total sheep population) and 556.250 tons of ewe milk in 2020 (MAPA, 2022) Spain is rated the fifth country in the world for ewe milk production (after China, Turkey, Greece, Syria, and Romania) and the second for sheep milk cheese production (after Greece). The Community of Castilla y León, which has 2.4 million dairy ewes and provides 54% of Spain's total ewe milk production, in 2022, accounts for nearly 39.7% of the Spanish dairy industry (MAPA, 2022).

In the Spanish northest region of Castilla y León, the most important dairy sheep breeds are the autochthonous Spanish Churra sheep, and the non-local Assaf breed. Churra is an important Iberian dairy sheep breed used for milk production and, to a lesser extent, for suckling lamb meat production. It is a hardy breed with black patches in the periphery, around the eyes, and the distal legs, and well adapted to the continental climate of the Castilla y León region characterized by long, harsh winters, short springs, and scorching, dry summers. This breed is traditionally reared in semi-extensive farming to take advantage of the stubble. On average, rams' weigh ranges from 65 to 75 kg, and ewes' weigh ranges from 45 to 55 kg, although rams may reach 100 kg and ewes 70 kg (De la Fuente, 1996). In the last years, the number of heads of Churra sheep has suffered an important decreased and the current census is of 110,380 heads (MAPA, 2021). This is mainly due to a sociological problem, as the breeders of Churra sheep are traditional farmers that are being retired and they are not being substituted by young breeders. In addition, the limited number of young sheep breeders of the region are turning to the non-local Assaf breed, which is reared intensively and shows a higher milk production potential. The Assaf breed, which was originally obtained in Israel as a synthetic crossbreed consisting of 5/8 awassi and 3/8 east Friesian, was introduced in the León province in the 1970s (Legaz et al., 2008). The good adaptation of this breed to the climate conditions of Castilla and León led to the establishment of the Spanish Assaf breed through absorption crosses with

animals of some of the autochthonous breeds, such as Churra, Manchega and Castellana dairy breeds (Legaz et al., 2008). Currently, the Spanish Assaf breed has a population of 165,023 heads (ARCA, 2022), and were mostly raised in intensive or semi-intensive production systems in the autonomous community of Castilla y León.

The National Association of Breeders of Selected Churra Sheep (ANCHE) initiated the planning of the breeding program for the Spanish Churra sheep breed in 1984. This breeding program was implemented in 1986, with the technical assistance from the University of Leon's Department of Animal Production (De la Fuente et al., 1996). The main goal of this breeding effort was to increase milk yield while preserving the animal's ability to adapt to adverse environmental circumstances. Milk composition (milk proteins and fat percentage) was also a selection target in the breeding program in order to maximize cheese yield. In 2006, morphological features such as udder morphology, body conformation, and the number of born lambs were introduced also as selection objectives. In 1990s, a meat production Churra breeding program was developed trying to optimize the economic interests of the Churra farms exclusively dedicated to suckling lamb meat production.

2.2. Mammary morphology of dairy sheep and genetic improvement

Implementation of breeding programs for pure-bred dairy sheep was on the rise during the last quarter of the XX century (Barillet et al., 2001), especially in Mediterranean areas. Sheep production is characterized by the fact that each region has its own local breed well-adapted to the corresponding specific environment and to the seasonal availability of pastures. Moreover, the regular increase of recorded ewes about the 1980s was directly associated with the development of selection programs. Hence, the efficiency of selection on milk properties has allowed to consider increasingly complex objectives to address consumers' interests (e.g. milk quality traits such as healthy fatty acid profile, sanitarian conditions) and today encourages to consider as new objectives the ability of the animals to get adapted to the adverse circumstances affecting the production systems (e.g. increasing prices of high-quality protein ingredients of animals' diets, and growing sustainability challenges for the food production system). In 1985, ANCHE started a genetic improvement program of the Spanish Churra sheep for milk production by increasing productivity, using available resources and current knowledge (De la Fuente et al., 1996). Initially, the main objective of the program was milk quantity. Still, considering the negative correlation between milk quantity and protein percentage, in 1998, protein content was included as a selection goal. In addition, it is well known that in dairy species, particular attention should be paid to the mammary gland, as it is essential for two functional characteristics of interest: milkability and resistance to mastitis. The relevance of these functional traits was higher when the percentage of dairy sheep undergoing milking increased, and the European governments tightened the different aspects that help improve the milk's sanitarian and hygienical quality. Because of that, it was suggested that adaptation to machine milking and an increased genetic resistance to mastitis could be considered as breeding objectives in Churra dairy sheep population. The methods used to evaluate mammary gland morphology in the past could not be implemented in the Churra breeding program. The technique described by Labussiere et al. (1981) shows certain deficiencies, such as low speed,

over-invested effort, and costs, making it impossible to apply these methods within the breeding program implemented in commercial farms. Classification by types, as other authors suggested (Sagi and Morag, 1974; Sagi, 1978) was also difficult to apply in practice because the BLUP (from Best Linear Unbiased Prediction) method used for prediction of genetic merit is less suitable for assessing genetic value for non-continuous characteristics or variables included in this evaluation analysis. This problem was solved by setting traits on a linear scale, as previously done in dairy cattle (Thompson et al., 1983; Lucas et al., 1984). The proposed evaluation method to be implemented in Churra sheep was based on the assessment of four essential characteristics of the mammary gland and a fifth one concerning the global definition of mammary gland morphology, all of them evaluated according to a 9 point-based linear scale. The five considered traits are related to milkability and are briefly described here below (Fernandez et al., 1995):

- **Udder depth** is the distance between the junction of the udder and the floor. The hook is used as a reference; an udder with too much depth usually results in deficiencies in the suspensive ligament.
- **Udder attachment** refers to the perimeter of insertion of the udder into the abdominal wall of the ewe. A maximum of 9 points of insertion is considered to be optimal.
- **Teat placement** is defined as the angle of the teat; in this case the optimum is the absolute vertical position of the teats (9 points) directed towards the floor and matches the minimum height of the tanker.
- **Teat size** is determined by the length of the teat. Also, extreme lengths make it impossible to adapt to the standard teat cups, manual milking and suckling. The Churra breed has an average teat length of 3.83 cm (Fernandez et al., 1995). This teat size was associated to a score of 5 on the linear scale, which is considered as the optimum for this trait.
- **Udder shape** determines optimal mammary morphology for milking ability (9 points) and corresponds to the "Udder machine" described by Mikus (1978).

Mammary morphology traits are highly variable, and the heritability estimated for these traits in different sheep populations is moderate to high (Charon, 1990; Gootwine et al., 1980). Churra breed show a medium heritability, varying between 0.16, for the udder depth, up to 0.24, for udder conformation and teat position (Table 1) (Fernández et al., 1997). Because of the high level of heritability, the classical selection is still effective, but a better understanding of the genetic architecture underlying mammary morphology could help to increase the efficiency of selection and, also, the identification of causal variants can increase the efficiency of genomic selection-based methods in relation to these relevant functional traits. Traits related to udder size (depth, width, length) are significantly influenced by the lactation month, flock, and milk yield. In contrast, cistern morphology related traits (cistern height, teat position, teat angle) are significantly affected by flock and parity (Fernandez et al., 1995). At the time of considering the need of including these traits in the Churra sheep breeding program, because the initial focus of the selection scheme had been on productive traits, an increase in the udder depth and a reduction of the teat verticality was observed (Fernandez et al., 1997). Hence, it was evident that it was necessary to introduce mammary morphology traits in the Churra breeding program. In addition, improved milking adaptation was expected to positively affect udder health by

reducing subclinical and clinical mastitis of the ewes as had been observed in other dairy sheep populations (Fernandez et al., 1997; Marie-Etancelin et al., 2001; Bergonier et al., 2003; Legarra and Ugarte, 2005).

The genetic correlations between mammary morphology traits in the Churra sheep are provided in Table 2 (Fernández et al., 1997). Based on these estimations, the importance of the global udder shape trait is highlighted, as it is a compendium of all the other characteristics and presents a high genetic correlation with teat placement (0.96) and udder attachment (0.55). In this way, a selection towards the teat position and udder attachment would imply an improvement of the global udder conformation.

Table 2.2.1. Heritability estimates reported for linear mammary morphology traits in Churra sheep (Fernandez et al., 1997).

Trait	Heritability
Udder depth	0,16 ± 0,04
Udder attachment	0,17 ± 0,05
Teat placement	0,24 ± 0,06
Teat size	0,18 ± 0,05
Udder shape	0,24 ± 0,06

Table 2.2.2. For the five udder morphology traits evaluated in Churra sheep, genetic correlations with milk production (MY) and phenotypic correlation with the logarithmic transformation of the number of somatic cells in milk (logSCS) are presented (Fernández et al., 1997).

Trait	r_g MY	r_p logSCS
Udder depth	0,82	0,13
Udder attachment	-0,02	0,01
Teat placement	-0,34	0,02
Teat size	-0,16	0,18
Udder shape	-0,26	-0,02

Considering the possible influence of mammary morphology traits on udder health, the phenotypic correlations described by Fernández et al. (1997) between mammary morphology traits and the somatic cell score (SCS) indicate that higher teats lengths would be associated with a higher somatic cell count (Table 2). This has been explained because longer teats could be prone to trauma from standard-sized liners or are closer to the ground and possibly have larger holes. Thus, due to the positive correlation between the global udder conformation and the teat size (0.35), the selection based solely on the global udder conformation trait would lead to an excessive increase in the teat, which could be associated with a greater risk of mammary infections. The slightly positive correlation between udder depth and SCS (0.13) at the phenotypic level could be explained because a deeper udder may be more susceptible to traumatic processes (Table 2). Although these traits show a moderate to high heritability, a

better understanding of the genetic architecture underlying mammary morphology could help to increase the efficiency of their genetic selection.

2.3. The evolution of genomics

Genomics is an interdisciplinary field of science dealing with the structure, function, evolution, mapping and editing of the genome. The genome is a complete DNA set of the individuals' of a given species, and includes all the genes of that species as well as their hierarchical three-dimensional structural configuration (Cremer et al., 2006). Considering genetics as the study of individual genes and their role in inheritance, genomics attempts to qualify and quantify all genes of individuals, as well as their interaction and impact on the organism.

Thomas Morgan and his student Alfred Sturtevant established the principle of genetic mapping in 1911. They located a sex-linked gene on the sixth chromosome of a *Drosophila melanogaster*. The focus of gene mapping and linkage analysis is still used in the same way and only the methodology shows today a higher level of development. However, the traits of productive or economic interest in domestic animals show, in general, a continuous phenotypic variation, which means that they can be described in terms of the normal distribution. The classical genetic improvement of these traits was based on the infinitesimal model proposed by Fisher in 1918. This model advocated that an infinite number of genes influence the phenotypic variation, i.e., suggesting an endless number of unrelated loci with an individual small additive effect (Fisher, 1923). However, years later, it was learnt that the amount of genetic material is limited, which means that phenotypic variation is affected by a finite number of genes, depending on the species of the organism. These findings were confirmed by the Human Genome Project, which proved the existence of approximately 25,000 genes or loci (Pennisi, 2003). This knowledge dismissed Fisher's infinitesimal model and confirmed the theory of the finite number of loci influencing phenotypic variation. Furthermore, thanks to the knowledge developed so far, there has been a new interdisciplinary field of science known as genomics. It is a field of study that is divided into two branches: structural genomics, whose main objective is to visualize the three-dimensional structure of every protein using either experimental, computational, or a combination of both methods (Skolnick et al., 2000), and functional genomics, which instead of concentrating on one gene in the genome, examines the structure, function, and regulation of all genes as well as the dynamic processes such as gene transcription, transfection, and translation (Bunnik and le Roch, 2013). Similar to genomics, the study of proteins has gradually evolved into the field of science known as proteomics (Blackstock & Weir, 1999). Based on the findings of genomes, comparative genomics is another specialized branch of science. Comparative genomics contributes to understanding the diversity of the living world by identifying the overarching principles governing the group of genomes (Hardison, 2003; Miller et al., 2004). Quantitative trait loci (QTL) are genomic regions harboring polymorphisms associated to changes or variability in quantitative attributes. The methods and the experimental design to perform studies of QTL mapping are determined by the availability of molecular technologies. In livestock species, microsatellite markers were the mainstay of QTL mapping until the availability of SNP-chip arrays (2005 in chicken, 2008 in cattle, 2009 in sheep, 2012 in pigs) (Muir et al., 2008; Matukumalli et al., 2009; Liu et al., 2013;

Gronnen et al., 2012). Because of the low density of the genetic maps including these markers, the correlation between markers and QTL could only be determined through linkage analysis (LA). In relation to the experimental design, family designs were required, such as line crosses (e.g. Holstein X Charolais) or half-sib families. After the identification of a region as a potential region harboring a putative QTL, more markers had to be genotyped in the considered population to obtain a higher density map of the target region. The aforementioned allows fine-mapping of the QTL in order to develop haplotypes of markers in Linkage Disequilibrium (LD) across the different families under study. For marker-assisted selection (MAS) purposes, and for restricting the potential position of the QTL to detect the causative genetic polymorphism, tightly linked markers were required. However, a vast number of studies undertaken in the different livestock populations showed that traditional QTL mapping based on LA and microsatellites had a limited capacity to identify causal mutations (e.g. Gutierrez-Gil et al., 2009a; 2009b, Raadsma et al., 2009).

The advancement of next-generation sequencing (NGS) after the completed of the Human Genome Project, and the progress of other genome sequencing project for various domestic animals of economic importance during the first decade of the present century, resulted in the identification, across the corresponding genome sequences, of thousands of SNPs, which are the most prevalent genetic markers (Heaton et al., 2002). Most previous genotyping approaches for SNPs were too slow and expensive to be used in large-scale studies involving hundreds or thousands of SNPs in a large number of DNA samples. However, the technological advancements related to high-throughput genotyping technologies resulted in considerable improvements and in faster, more efficient, and less expensive procedures, while also improved the existing analysis methods (Tsuchihashi, 2002). This allowed the simultaneous genotyping for a few to several hundred or thousands of markers in hundreds to thousands of individuals, which are known as SNP-genotyping arrays, or SNP-chips. In addition, the availability of SNP chips has enabled the implementation of genome selection (GS), which is a MAS variant based on animals' data obtained from high-throughput genotyping. Advances in sequencing technologies taken place in the first decade of the XXI century, has allowed the routine use of NGS in sheep genomic projects. It is essential to mention that these high-throughput sequencing technologies have become critical in new branches of science such as genomics, epigenomics, and transcriptomics, which try to understand the mechanisms and functions of the genome as a whole entity. NGS sequencing technology can be used for a variety of purposes, which are briefly commented here: (i) Whole-genome sequencing (WGS) is a comprehensive method for analyzing the entire genome, and it generates unbiased high-complex libraries that deliver highly uniform coverage across the entire genome with very low duplication rates; WGS is used for the identification of causal risk variants or the study of the metagenome of a given sample; (ii) tissue RNA or transcriptome sequencing (RNA-seq) allows to study all the transcripts present in a cell at a given moment and at the metagenome level. Also, if we wish to link certain information in a genome to functional protein production, we need to comprehend the transcriptome; (iii) global methods like chromatin immunoprecipitation sequencing (ChIPseq) or whole genome bisulfite sequencing (WGB-seq) can be used to explore heritable changes in gene expression and other genomic activities without changing the DNA sequence.

2.4. Application of genomics in dairy sheep breeding

2.4.1. The candidate gene approach

In dairy sheep, the initial studies aiming to decipher the genetic basis of the traits of economic interest, published in the 90s, were based on the candidate gene approach. This methodology assesses the potential association of functional candidate genes with the traits under study. At that time, the dairy sheep industry was an important but not-developed part of the small ruminant industry. Different breeders' associations realized that in a sheep breeding program, in addition to wool and meat use, the ewes' milk could also be processed into high quality dairy products such as cheese and milk powder and bring high economic interest to businesses at the local or international level (Li et al., 2022). The nutritional value of sheep's milk is greater than that of goats or cows, due to its higher fat yield, percentage proteins, and minerals contents, which explained the potential of the sheep dairy industry for further development.

The initial research studies using the candidate gene strategy in sheep were focused on the analysis of coding genes of milk proteins such as alphaS1-casein (Piredda et al., 1993) or beta-lactoglobulin (Barillet et al., 2005) because of the interest in cheese yield related traits. Further examples of this strategy are the studies, performed by the MEGA-ULE group assessing the effect of several SNPs in genes coding for enzymes involved in fatty acid synthesis on the profile of these milk components (García-Fernández et al., 2010a; 2010b). In addition, the candidate gene strategy can also be used based on knowledge generated by QTL mapping studies obtained in the same or in different species. As an example of this approach, the MEGA-ULE group studied in Churra sheep the potential influence of genes carrying causal mutations in dairy cattle on milk production traits (García-Fernández et al., 2011). These genes were *GHR*, *DGAT1* and *ABCG2*, which encode, respectively, for the growth hormone, the enzyme acyl-CoA-diacylglycerol acetyltransferase I and the breast cancer resistance protein. The results of this study only revealed significant associations for the *ABCG2* gene, suggesting that the genetic architecture of dairy traits may differ between sheep and cattle, and also between breeds of the same species, as later suggested by Marina et al. (2021). Despite this, the candidate gene approach has several disadvantages, mostly due to a lack of complete knowledge of the physiological processes that interfere with the expression of the genes that influence on complex phenotypes (Gutiérrez-Gil, 2004).

2.4.2. QTL mapping

In sheep, Gutiérrez-Gil et al. (2009a) used microsatellite markers to undertake a genome scan based on LA to find QTLs underlying milk production traits. The LA mapping analyses performed detected QTLs that showed low significance levels and broad confidence intervals. The majority of QTL mapping analyses reported in sheep have been based on interval mapping, which is a methodology where the information of all the genotyped markers within a chromosome is used at the same time, and applies statistical analyses based on the maximum likelihood approach or the regression analysis (Arranz & Gutiérrez-Gil, 2012). The first gene

mapping studies of economic interest in domestic species were carried out in pigs (Andersson et al., 1994) and in dairy cattle (Georges et al., 1995). According to the AnimalQTLdb database (Hu et al., 2013) there are more than 762 published QTL articles for pigs, and 1103 QTL articles for cows, respectively. In sheep, the number of QTL detection studies is much lower (221 publications annotated in the AnimalQTLdb), mainly due to the lower economic relevance of this species and the limited size of the exploited populations.

The first genome scan studies based on microsatellites markers described in dairy sheep were carried out by the MEGA-ULE group, where this Master's thesis has been developed. The analysis of 182 microsatellite markers, distributed throughout the ovine autosomal genome at a regular distance of 20 cM, for a total of 1,400 Churra sheep distributed through 11 half-sib families, allowed the identification of QTL for SCS, milk production, and mammary and body morphology traits (Gutiérrez-Gil et al., 2007, 2008, 2009a, 2011), as well as for gastrointestinal nematodes resistance traits (Gutiérrez-Gil et al., 2009b). It is important to note that in dairy sheep, most QTL studies have been focused on milk production traits and there is a limited number of studies focused on functional traits. To our knowledge, a classical microsatellite-based genome scan for mammary morphology traits has only been reported in Churra sheep (Gutiérrez-Gil et al., 2008). More recently, based on a GWAS study on supernumerary teats performed in the Chinese Wadi sheep breed, four candidate pathways involved in the development of the supernumerary teat phenotype have been detected (Peng et al., 2017).

Because the QTLs found through classical microsatellite-based genome scans showed such broad confidence intervals (CI), they frequently comprise a large number of genes. Because of that further fine-mapping studies were needed to reduce the CI of the potential QTL and make it easier to examine positional candidate genes in the region as likely causal genes. If after the fine-mapping, any of the positional candidate genes functionally corresponds to the trait under study, that gene would be considered both a positional and a functional candidate gene and should be further studied. In Churra sheep, the most significant QTL previously identified by the global genome scan performed for milk production traits was located on chromosome 3 (OAR3) and affected the milk protein percentage (Gutiérrez-Gil et al., 2009a). Using a different commercial population of Churra sheep, García-Gámez et al. (2012b) replicated the same QTL and carried out a fine-mapping study by increasing the density of microsatellite markers in the target region. In this way, the CI interval of the QTL could be replicated and redefined, leading to the identification of a promising functional candidate from the list of positional candidates, the *LALBA* gene, which encodes alpha-lactalbumin, a whey protein (García-Gámez et al. 2012b).

2.4.3. QTL mapping and new genomic tools

The first draft of a human genome sequence was published in 2001 (Venter et al., 2001), marking the start of a genomic era in which sequencing technology advanced at an exponential rate. It reduced the cost and time required to sequence genomes, and it served as a jumping-off point for sequencing the genomes of the most important domestic animals. The chicken (Lucinda et al., 2004) was the first domestic species to be sequenced, followed by the cow

(Bovine Genome Sequencing and Analysis Consortium, 2009), and the pig (Gronnen et al., 2012). The sequencing of the sheep genome was delayed due to the low economic interest of this species. Using the human and cow genomes as a starting point, the ISGC created the first virtual draft of the sheep genome (Dalrymple et al., 2007). The sequence of the bovine genome was rearranged into the sheep genomic framework using synteny blocks described by sheep Bacterial Artificial Chromosomes (BASs) (CSIRO, 2010). Following that, the ISGC made a great effort by releasing the first genuine draft of the sheep genome, which was obtained by sequencing a Texel breed individual. Thousands of SNPs were found in the initial draft (Oar v1.0) of the sheep reference genome. The ISGC used this information to produce the Illumina OvineSNP50 BeadChip (Chip-50K), which was made publicly available in 2009. The ISGC launched the Sheep HapMap project in 2008 as a prologue to the public release of the Chip-50K, which assessed 2,819 animals from 74 sheep breeds scattered over a wide variety of geographical regions throughout the world. A large sheep population with a high diversity of SNPs and a highly effective population size was discovered thanks to this effort. This indicates that sheep were domesticated from a far larger genetic background than cows or pigs (Kijas et al., 2012). Two years later, the ISGC's efforts resulted in the development of a second reference draft of the reference sheep genome (Oar_v2.0), and the data obtained from a male and a female from the Texel breed, which was used to fill gaps in version 1.0 and served to develop a new version of the sheep genome (Oar_v3.1) released in October 2012, and which was the Ensemble available version for many years. Although the next available version, Oar_v4.0, was released in 2015, the next version available in Ensemble for this species was the Oar_rambouillet_v1.0, which was released in 2017. In this version a global analysis of transcription starts sites (TSS) and TSS-Enhancer clusters using Cap Analysis Gene Expression (CAGE) sequencing was carried out including 56 tissue samples obtained from the reference ewe Benz2616. Although the Oar_rambouillet_v1.0 reference version was made available through Ensemble in 2020, it has been already improved as ARS-UIRamb_v2.0 (used in this study). This new assembly has 142 scaffolds, down from 2,640 in the Oar rambouillet_v1.0. Also, this latest version is 15 times more continuous than the first Rambouillet breed assembly, having a contig N50 of 43 Mb. It is important to mention that this version cannot be found on Ensembl, but is available through NCBI, at https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_016772045.1/.

In sheep, QTL detection studies based on the Chip-50K were performed, with the methodology of association analysis at the genome level, or GWAS, or, in fewer cases, the principles of linkage analysis (LA) or linkage analysis combined with linkage disequilibrium (LDLA). The Chip-50K was used in a population of 1,681 Churra sheep belonging to 16 half-sib families of Spanish Churra sheep using a daughter design by the MEGA-ULE group of the University of León. First, a GWAS was conducted in order to identify regions that influenced milk production traits (milk yield, milk protein, fat yield protein percentage, and milk fat percentage) (Garcia-Gamez et al., 2012a). In this study, in addition to identifying new regions associated with such traits, which had not been detected by previous analyzes based on microsatellite markers, the detection of the most significant QTL previously identified on OAR3 for protein percentage was replicated. Increasing marker density allowed the identification of several candidate genes within the new estimated confidence interval (13 cM). Since the top SNP in that region was

located within an intron of the *LALBA* gene, previously identified as a promising candidate, the sequence of this gene was analyzed by Sanger sequencing. Only one of the 13 mutations identified was responsible of an amino acid change. That mutation was located in Exon 1 of the studied gene (*LALBA g.242T>C*) and caused an amino acid change in residue 27 of the encoded protein, Val27Ala. This mutation was reported as the first QTN in dairy sheep (García-Gámez et al., 2012b). The genotypes generated in the same population were then subjected to linkage analysis (LA) and LDLA using the QTLMap software (Filangi et al., 2010). This allowed the comparison of results from QTL analyses performed with the three distinct analytic approaches (García-Gámez et al., 2013). Thus, while the QTL for protein percentage in OAR3 was not detected by LA, it was identified by LDLA and GWAS. Also, a region on OAR2 was identified as being significantly associated with milk production traits by all the considered approaches. Another example of results obtained with these methods was used by Atlija et al. (2016), who performed genome scans using LA, LDLA, and GWAS on a subset of 518 sheep from the population previously studied by García-Gámez et al. (2012a), using data for fecal egg count (FEC) and serum levels of immunoglobulin A (IgA) in natural infection by gastrointestinal nematodes. The results confirmed a QTL previously discovered through a genome scan based on microsatellite markers for the FEC trait on OAR6 (Gutiérrez-Gil et al., 2009b), providing information that could be used in subsequent studies to identify the allelic variations that control this QTL.

This series of studies have shown that, considering the same population, each of the QTL mapping method identifies different types of signals, whereas few QTLs were commonly detected by all of them. Globally, considering all the results generated in this Churra sheep population, for the different traits analyzed, it was suggested that the LDLA analysis is the one that offers the greatest advantage by simultaneously exploiting information from the pedigree information through LA, and by exploiting information obtained at the population level related to LD. Comparison of the three analysis methods was identified as the best approach to reveal all the QTLs that segregate at both the family and the population levels (García-Gámez et al., 2013). For the last lists of QTLs identified for both milk production, SCS, and parasite resistance traits, only the *LALBA* polymorphism was identified as a QTN, which reflects the limited number to identify causal variants of complex genetic effects.

In 2013, as a further step on the production of genomic tools of interest to study the sheep genome, the ISGC produced an ovine high-density chip (Ovine Infinium® HD SNP BeadChip) with 606,006 SNPs (HD-Chip). This chip provided a genomic tool that reduced the genomic gaps to cover the sheep genome to 5 Kb markers (Kijas et al., 2012). The expensive cost of genotyping is the main limitation to apply this new HD-Chip to real sheep breeding schemes. A possible solution to this problem is the imputation of genotypes in the populations previously genotyped with the 50K-Chip, and later partially genotyped with the HD-Chip. The imputation process is based on the genotyping of a subset of animals with the two chips, which avoids the needs to genotype the entire population with the HD-Chip (Chitneedi et al., 2017). Numerous optimized programs for inferring genotypes in livestock populations have been developed as a result of the interest on this approach. We refer to Beagle v4.0 and Flmpute v2.2, which were utilized by Chitneedi et al., (2017) to impute HD-Chip genotypes in the same Churra population

previously studied by García-Gámez et al. (2012a). Based on the imputed genotypes for that populations, this work of the MEGA-ULE group presented updated and more accurate estimations of the LD extent in that ovine population and the identification of runs of homozygosity (Chitneedi et al., 2017).

2.4.4. Whole genome sequencing (WGSeq)

Whole-genome sequencing (WGSeq) data can be used to identify the mutation responsible or causal of a specific QTL effect. The results of a GWAS analysis performed on 1,009 sheep genotyped with the Chip-50K and divided into 33 half-sib families were published by Rupp et al. (2015). A total of 22 QTL were identified using haplotype linkage and association studies. A QTL on OAR3, in particular, provided comparable data in both analyses, making it possible to perform further fine-mapping analyses using WGSeq in a trio that segregated for alternate alleles of the QTL and that showed the greatest divergence for mastitis resistance, based on the indicator trait provided by the SCS phenotype. The studied trio included a heterozygous male for the QTL (*Qq*) and two homozygous daughters with divergent value for the phenotype under study (a *QQ* daughter with very low SCS value, and *qq* daughter with very high SCS value). This approach allowed researchers to identify a mutation in the *SOCS2* gene that causes a change in protein sequence (*p. Arg96Cys*) as being directly responsible for the SCS QTL effect previously found in OAR3. Another example of the implementation of WGSeq in the region previously detected as QTL by LA analysis was described by Gutiérrez-Gil et al. (2018) in relation to the region of OAR20 associated with SSC. In this case, a list of potential causative mutations in positional and functional candidate genes was identified, although the different location of the QTL peak in the different analyzed families hampered the identified of the genuine causal mutation.

2.4.5. High-throughput transcriptome sequencing (RNA-seq)

As a revolutionary tool for transcriptomics high-throughput transcriptome sequencing (RNA-Seq) provides a precise measurement of the levels of transcripts and their isoforms (Wang et al., 2009). As a consequence of its sensitivity and ability to characterize and quantify messenger RNAs in different tissues with great repeatability and low false-positive rate, this type of massive analysis has been used to elucidate the transcriptomics behind economically important traits in livestock species (Wickramasingheet al., 2014). In order to improve milk production, knowing the biological mechanisms that control mammary gland morphology development, as well as, lactation is of great interest. There are no transcriptomic studies focused specifically in mammary gland morphology, however, the gene expression profile of the mammary gland through lactation has been evaluated in dairy cows (Bionaz et al., 2012; Wickramasinghe et al., 2012; Cui et al., 2014; Dai et al., 2018),dairy goats (Shi et al., 2015; Crisà et al., 2016), and dairy sheep (Suárez-Vega et al., 2015a, 2015b, 2016, 2017). Specifically, Suárez-Vega et al. (2015), provides the first integrated overview of the sheep milk mammary gland transcriptome through lactation. In this work, RNA-seq was performed on total RNA extracted from MSCs collected at four lactation stages of the mammary gland after lambing (day 10, 50, 120 and 150 of lactation). Two sheep breeds were included in this study, Spanish Churra and Spanish Assaf,

which have different milk production characteristics. The results of the comparative analyses performed showed that for the two breeds the highest expressed genes across all the lactation stages were caseins and milk whey proteins. In the time series analyses performed individually for each breed, a total of 573 differentially expressed genes (DEGs) were found across all the lactation stages. Among them, 256 DEGs were detected as DEGs in common between both breeds, and give an idea of the biological processes happening in the mammary gland across lactation, with a higher expression of genes related to extracellular matrix remodeling at late lactation stages. Some of the DEGs showing a higher expression in Churra were enriched for gene ontology (GO) terms related to the group of endopeptidase and channel activity. This result is of interest because of the relation of the mentioned terms and the higher cheese yield of Churra sheep compared with Assaf sheep. First identification of genetic variants across all the expressed genes in the lactating mammary gland using a transcriptome approach was conducted by Suárez-Vega et al., (2017). In this study, a total of 216,637 variants were detected in the MSCs transcriptome of eight ewes. From them, 21.44% novel variants, not previously annotated in the SNPdb of the NCBI (https://www.ncbi.nlm.nih.gov/snp/) were detected in QTL regions for milk yield, protein percentage and fat percentage. The most significant term identified in the *KEGG pathway* functional enrichment analysis of the genes positioned in selected QTL regions was one related to protein processing in the endoplasmic reticulum. Additionally, among a list of candidate genes encoding principal milk protein and molecules involved in the lipid metabolism, a total of 504 and 1,063 variants were found in this study. Of them, 20 of these variants were found to have putative relevant effects on the encoded proteins. As a practical result of this work, the SNPs found in this work were suggested as potential useful markers to be included in genotyping platforms, or custom SNP arrays for association analysis in commercial populations, as well as, for genomic selection in the dairy industry. Indeed, a list of 3,194 SNPs identified in the RNA-seq-based study were included in a *custom chip array* and used later to genotype Churra and Assaf ewes and perform a GWAS for milk coagulability and cheese-yield related traits (Marina et al., 2019).

3. Materials and Methods

3.1. Resource Population, Genotypes, and Marker Map

The commercial population of Spanish Churra sheep analyzed in this work has previously been described in detail by García-Gámez et al. (2012a). This population included a total of 1,696 animals: 1,680 ewes divided into 16 half-sib families, with an average of 105 daughters per male, and the 16 corresponding parents, which were insemination rams from the ANCHE Selection Nucleus. DNA was extracted from the daughter's blood and the semen of the 16 purebred rams.

The genotypic dataset analyzed in this work was that used by García-Gámez et al. (2012a), which included the genotyping of the 1,696 animals of the study population with the OvineSNP50 BeadChip (Chip-50K) at AROS Applied Biotechnology AS (Aarhus, Denmark) and the Laboratoire d'Analyses Génétiques pour les Espèces Animales (LABOGENA; Jouy-en-Josas, France). Hence, considering this initial raw dataset of genotypes that included the 1,696 individuals and 54,241 SNPs, the first analysis performed for the present memory was a quality control (QC), performed with the PLINK software (Purcell et al., 2007) following the criteria previously applied by García-Gámez et al. (2012). Firstly, a QC per animal was performed (call rate >90%) followed by a QC per SNP (call rate >95%; minor allele frequency >0.05; correspondence with Hardy-Weinberg equilibrium: $P > 0.00001$). After that, a total of 1,678 individuals, including the 16 rams and 40,291 autosomal SNPs were retained for further QTL mapping analyses. Because the initial marker positions provided by the MEGA-ULE group for the SNPs included in the Chip-50K were based on the sheep reference genome version Oar_v3.1, additional processing steps were performed using PLINK to update the marker positions according to the sheep reference genome Oar_Ramb_v2.0. These positions were considered to create the map files necessary for both the LA and GWAS analyses.

3.2. Phenotypes

Udder depth, udder attachment, teat position, and teat size were the phenotypes here analyzed. The phenotypic raw data, based on the linear scale described by De la Fuente et al., (1996), came from the routine assessments made by the Association of Churra Breeders (ANCHE) among the herds of the Selection Nucleus. The dependent variables used in the QTL mapping analyses were the yield deviations (YD) of the four traits under study. The YDs were calculated as averages of the ewe's raw phenotypic records adjusted for the fixed environmental effects and the common environmental effect (Vanraden and Wiggans, 1991). Multivariate repeatability animal models were used to calculate yield deviations including the following fixed effects in the model: the herd-test-day, the weeks in lactation and the birth order (including as nested covariates, the age of the ewe at parturition, the milk production in the previous lactation and the number of born lambs).

3.3. QTL mapping analysis

The QTL analysis was performed using the QTLMap software (Filangi et al., 2010). This software was developed at the National Institute for Agricultural Research in France (INRA), and it performs interval mapping based on maximum likelihood calculations (LRT, Likelihood Ratio Test). Hence, this package is dedicated to the detection of QTLs in experimental designs of commercial populations based on linkage analysis (LA) and the combination of linkage analysis and linkage disequilibrium (LDLA), following the model of Legarra and Fernando (2009). Using the software analysis option - - calcul 4 for a single trait, the QTL search was carried out every 0.1 cM. The linkage map used for the analysis was based on the physical map provided by the sheep reference genome Oar_Ramb_v2.0 by converting physical distances into genetic distance using the equivalent of 1cM ~1Mb, following the procedure used by García-Gámez et al. (2012a). The sires' polygenic and QTL effects were evaluated during the linkage analysis. The QTLMap software was also used to estimate the chromosomal-wise significance threshold through a total of 1000 permutations (at 0.1 cM steps). We also considered the genome-wise significance threshold, which was based on the chromosome-wise significance thresholds by adjusting for the total number of chromosomes under analysis, as described by previous publications of the MEGA-ULE group (García-Gámez et al., 2013; Atlija et al., 2016). For each QTL identified by the across-family LA scan, linkage-based within-family analyses were performed to identify the corresponding segregating families. To estimate the CIs of the QTL identified by the across-family and within-family LA analyses, the likelihood ratio test (LRT) provided by the analyses was converted into a logarithm odds ratio (LOD) score value, and later the 1-LOD drop-off method, described by Lander and Botstein (1989), was used to define the corresponding CI.

In addition to the LA analyses, for each of the mammary morphology traits considered in this study, individual GWAS analyses were performed using the GCTA software (Yang et al., 2011), by applying the following formulae:

$$Y = \beta_0 + Wu + e,$$

where, $\boldsymbol{Y}$ is the vector of the corrected phenotypes, $\boldsymbol{\beta_0}$ represents the intercept of the regression model, $\mathbf{W}$ is the SNP markers' incidence matrix, $\boldsymbol{u}$ is the vector of random SNP effects, and $\boldsymbol{e}$ is the vector of residual effects. We assume that $\boldsymbol{u}$ and $\boldsymbol{e}$ are normally distributed with a mean of zero and variance of $\boldsymbol{\sigma_u^2}\mathbf{I}$ and $\boldsymbol{\sigma_e^2}\mathbf{I}$, respectively, being $\mathbf{I}$ the identity matrix. The GCTA software corrects for the population substructure during the estimation of the SNP effects.
GWAS significance thresholds are usually estimated using a Bonferroni multiple-test correction. To properly apply the Bonferroni approach with this purpose, by correcting the nominal p-value considering the number of independent SNPs analyzed, we applied the methodology proposed by Gao et al. (2008) to estimate the number of intendent markers analyzed in each chromosome, following previously reported studies in this field (García-Gámez et al., 2012a; Atlija et al., 2016; Marina et al., 2021). Thus, we first calculated the effective number of independent tests based on the LD individually estimated for each

chromosome (M_{eff_C}), and the *indep-pairwise* command parameters were set to SNP window size 50, SNPs per step 5, and r^2 threshold to 0.2. After applying this approach, we considered a total of 16,520 independent markers being analyzed across the genome. To calculate the adjusted significance level for each chromosome, we applied the Bonferroni correction formula for each chromosome as follows:

$$\alpha_C = \alpha_e / M_{eff\_C},$$

where M_{eff_C} presents the number of independently analyzed markers per chromosome, α_e is the chromosome-wise type I error rate (0.05), and α_C the adjusted chromosome-wise significance level. Correcting the total number of independent markers analyzed across the genome, we estimated the genome-wise significance threshold based on the chromosome-wise significance thresholds. Only SNPs with a significant association P-value lower than the chromosome-wise threshold (α_C) were considered to reach the chromosome-wise significance threshold. Only the regions identified by LA or GWAS as chromosome-wise significant QTL were subjected to the subsequent QTL and gene annotation analyses.

3.4. QTL regions and gene annotation

For the significant QTL identified by LA, the CI flanking interval positions in cM were translated to physical positions in the sheep reference genome (pb) and considered as target genomic intervals (TGIs) to perform QTL gene annotation with the GALLO R package, which provides a tool to perform a systematic annoation of positional candidate genes (Fonseca et al., 2020). Similarly, for the significant SNPs detected through the GWAS analyses performed, the corresponding TGIs were defined by considering 2.50 Mb intervals that included a 1,25 Mb region on both sides from the significant SNP. These TGIs were also considered for gene annotation with the previously mentioned package, using the corresponding annotation file for the Oar_Ramb_v2.0 ovine reference genome, which is available at https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_016772045.1/ .

For both, the TGIs identified by LA and GWAS, we also performed a systematic annotation of QTLs reported in previous sheep studies for udder morphology traits and also for the indicator traits of mastitis (e.g. SCC, SCS, etc.), due to the direct link between udder morphology and mastitis problems. For that, the available information, based on the Oar_Ramb_v2.0 reference genome, was downloaded from the SheepQTLdb database (http://www.animalgenome.org/cgi-bin/QTLdb/OA/search; Hu et al., 2013). In addition, the GALLO software was also used to perform a QTL enrichment analysis of the regions of interest with the aim of providing a more clear understanding of the genomic context of the candidate regions by investigating their QTL representativeness and diversity. Briefly, the QTL enrichment analysis function implemented by the GALLO package is based on a hypergeometric test approach, where the number of QTLs annotated within the candidate regions for each QTL type or trait is compared with the observed number of QTLs in the reference database (Fonseca et al., 2020).

4. Results

4.1. QTL regions detected

The LA genome scan identified ten 5% chromosome-wise significant QTLs for three of mammary morphology traits under study, udder attachment, teat placement and teat size and one genome-wide significant QTL for teat placement, whereas no significant QTLs were identified for the Udder depth trait (see the characterization of the QTL regions identified in (Table 3). In contrast, the GWAS analysis performed identified 20 QTLs that are associated with all the studied traits (Table 4). The significant results obtained from each analysis are described in more detail below.

4.1.2. Results of the LA analysis performed for udder traits

The across-family regression analysis was performed for udder depth, udder attachment, teat placement and teat size across the 26 ovine autosomes and identified 10 chromosome-wise significant QTLs. Three of these QTLs were located on OAR1 (*Ovis aries* chromosome) (peak at 101.16 cM), OAR2 (peak at 159.12 cM), and OAR13 (peak at 75.59 cM), and had an effect on Udder attachment. Also, three QTLs were identified on OAR13 (peak at 67.39 cM), OAR14 (peak at 52.78 cM), and OAR22 (peak at 75.59 cM) for Teat placement, whereas the other QTLs located on OAR1 (peak at 71.86 cM), OAR2 (peak at 152.12 cM), OAR3 (peak at 113.15 cM) and OAR14 (peak at 8.97 cM) had an effect on teat size. In addition, no QTL was identified for udder depth. A characterization of the significant QTL identified by the across-family LA analysis, including the corresponding CIs estimated by the 1-LOD drop-off method, and the results of the within-family analyses, are provided in Table 3. The genome-wise significant QTL are indicated in the table in bold font. In addition, the graphic representations of the LRT statistical values obtained in the across-family LA analysis performed for each trait, throughout the ovine autosome, are represented as Manhattan plots in Figure 1.

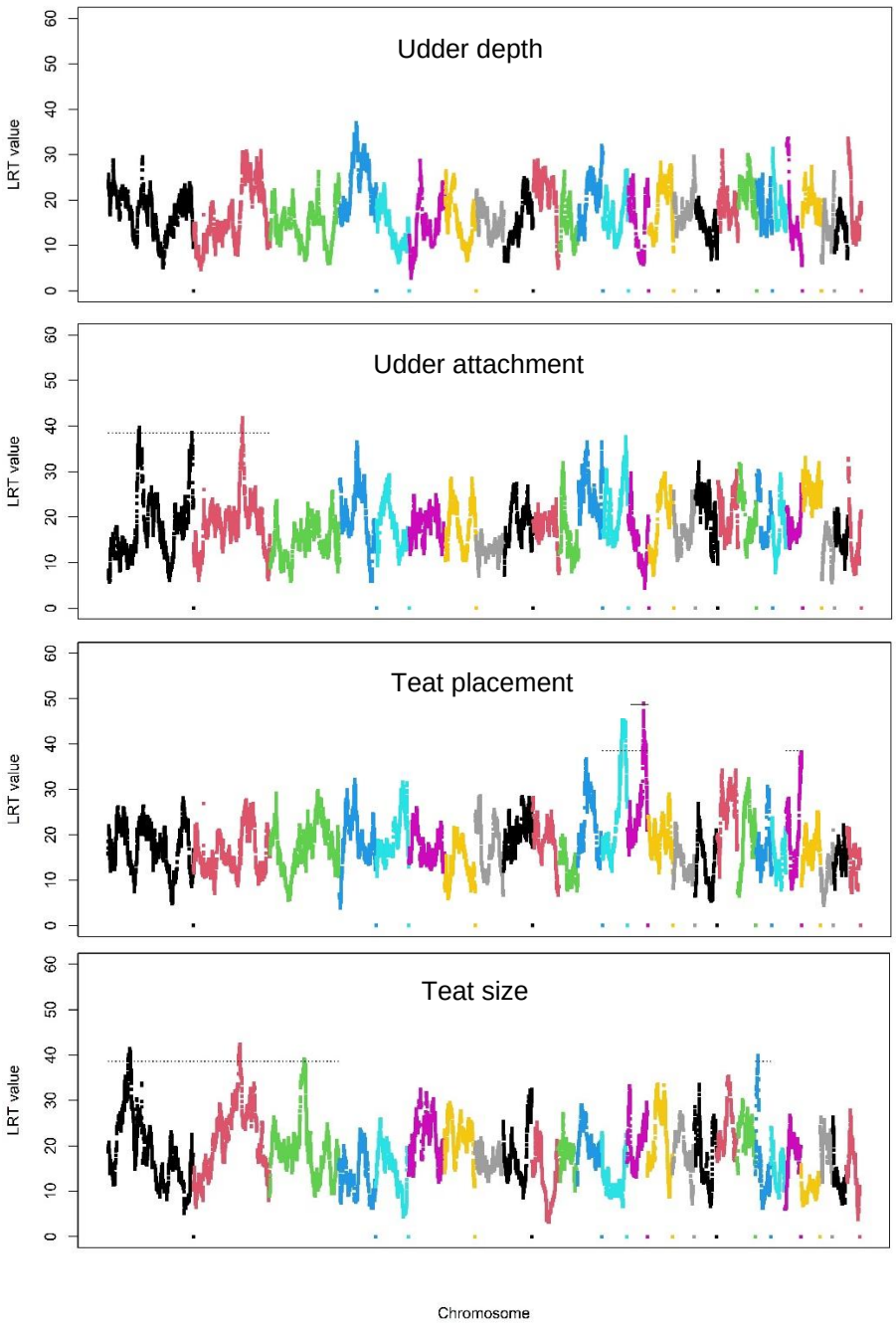

Figure 4.1.2.1. Results of the linkage analysis performed at the across-family level for the four mammary morphology traits under analysis. For the chromosomes harboring a significant QTL, the chromosome-wise (dashed line) and genome-wide significance thresholds are indicated as horizontal lines.

Table 4.1.2.1. Significant chromosome-wise QTLs detected by linkage analysis (LA) for the udder morphology traits analyzed in the present work.

Trait[a]	OAR[b]	Across-family analysis			Within-family analysis	
		Pos of max LRT(cM)[c]	P-value[d]	CI (cM)[e] TGI (Mb)[f]	Segregating family identifier (Pos of max LRT, cM)[h]	Size effect trait units (SD units)[i]
Udder attachment	1	101.16	< 0.05	98.0-103.2	Fam. 8 (278.16)	-0.108 ±0.023 (0.28)
					Fam. 14 (87.86)	-0.174± 0.019 (0.45)
					Fam. 16 (110.96)	0.425± 0.047 (0.04)
	2	159.12	< 0.05	157.7-159.7	Fam. 14 (187.22)	0.091± 0.024 (0.24)
	13	75.59	< 0.05	74.0-76.0	Fam. 5 (83.29)	-0.062± 0.011 (0.16)
					Fam. 6 (69.39)	-0.086± 0.004 (0.38)
					Fam. 10 (5.59)	-0.124± 0.019 (0.59)
Teat placement	13	67.39	< 0.05	65.0- 69.7	Fam. 1 (73.19)	0.081± 0.008 (0.23)
					Fam. 13 (57.39)	0.097± 0.009 (0.32)
	14	**52.78**	**< 0.001 (< 0.05)**	**52.6-52.9**	**Fam. 6 (51.48)**	**-0.070± 0.004 (0.23)**
					Fam. 9 (25.78)	**-0.005± 0.024 (0.02)**
					Fam. 13 (51.68)	**-0.106± 0.008 (0.34)**
					Fam. 16 (52.78)	**0.0348± 0.012 (0.11)**
	22	50.38	< 0.05	48.8-51.5	Fam. 2 (49.48)	0.104± 0.004 (0.34)
					Fam. 6 (8.58)	0.011± 0.004 (0.04)
					Fam. 16 (1.28)	-0.018± 0.098 (0.06)
Teat size	1	71.86	< 0.05	66.2-74.5	Fam. 5 (73.86)	0.088± 0.012 (0.36)
					Fam. 11 (79.06)	-0.053± 0.005 (0.22)
	2	152.12	< 0.05	147.8-153.9	Fam. 5 (132.32)	-0.063± 0.013 (0.26)
					Fam. 8 (152.82)	0.134± 0.018 (0.54)
	3	113.15	< 0.05	109.2-118.0	Fam. 15 (116.75)	0.130± 0.018 (0.05)
					Fam. 14 (116.75)	-0.044± 0.008 (0.18)
	20	8.97	< 0.05	7.5- 10.5	Fam. 7 (14.87)	0.077± 0.007 (0.32)
					Fam. 16 (4.87)	0.127± 0.007 (0.52)

[a] Analyzed traits.

[b] OAR ovine chromosome.

[c] Position of the chromosome (in centiMorgans) at wich the maximum LRT of the LA is reached in the analysis involving 16 half-sib families included in this work (across-family analysis) or the individual analysis of the segregating families (those showing a Pc-value <0.05 in the within-family analysis), respectively.

[d] P_c-value chromosome-wise significant P-value established through 1,000 permutation analysis. (Pg-value genome-wise significant P-value, only if significant) The genome-wise signficant QTL are highlithed in bold font.

[e,i] CI confidence interval (in cM) estimated from the position of the max LRT for the across-family analysis, respectively, following the 1-LOD-drop-off method (García-Gámez et al., 2012)

[f] TGI Target genome interval (Mb) defined as the corresponding genomic regions, according to the reference sheep assembly Ramb_v2.0, to the CI estimated for LA significant QTL.

[g] Estimated size effect of the QTL identified in within-family analysis expressed in trait units (Yield Deviation) and in phenotypic SD of the trait (in brackets).

For most of the QTL detected, the number of segregating families were two or three, except for the OAR2 QTL for udder attachment, with shows a single segregating family, and the OAR14 QTL for teat placement, for which four segregating families were identified (Table 3; Figure 2). Interestingly, the only genome-wise significant QTL was that for which four segregating families were detected. For the significant QTL, the position of the maximum LRT and the CIs estimated in the across-family analysis, were, in general, in correspondence with the positions of the maximum LRT value identified by the within-family analysis, although a certain lag with the across-family peak was observed for some of the segregating families (e.g. Fam. 8 in the QTL of OAR1 for udder attachment, Fam. 16 in the QTL of Oar 22 for teat placement). Again, in relation to the genome-wise significant QTL on OAR14 for teat placement, three out of the four segregating families showed a coincident peak in the same chromosomal region and in correspondence for the across-family peak, which provides a large support for the accuracy of the gene mapping results of this QTL, also considering that it was the most significant one.

Comparing our results with the QTLs previously reported in sheep for mammary morphology and mastitis resistance related traits, we found an interesting coincidence with one QTL reported for Udder depth also on OAR14 through the microsatellite-based genomic scan performed also the Churra sheep by Gutiérrez-Gil et al., (2008). That QTL appears to be close (CI: 53.5-62.6 Mb in Oar_Ramb_v2.0, according the SheepQTLdb) to the region where we have found here the genome-wise significant QTL teat placement on OAR14 (CI: 52.6-52.9 Mb). In addition, when comparing with the QTL described in sheep for milk traits and mastitis resistance traits, we found another interest result for the genome-wise significant OAR14 QTL for teat position as it matched a QTL effects for milk lactose content and SCS, described in a Awassi x Merino backcross family population (Raadsma et al., 2009). Other coincidences were found for the OAR2 QTL reported here for Teat size teat, which matched an association reported for milk protein content in Churra sheep (Gutiérrez-Gil et al., 2009a), and for the QTL on OAR20 for teat size, which was coincident with a QTL for milk production described in a backcross East Friesian x Dorset population (Mateescu and Thonney, 2010).

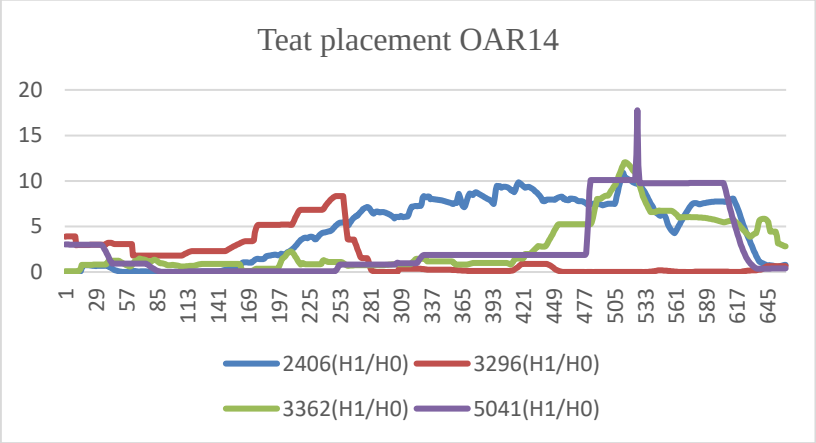

Figure 4.1.2.2. Representation of the LRT profile obtained for the four segregating families identified for the genome-wise significant QTL detected by Linkage Analysis for teat placement on chromosome 14 (OAR14).

4.1.3. Results of the GWAS analysis for udder traits

The GWAS analysis performed identified 18 SNPs that exceeded the 5% chromosome-wise significance level, and which influenced all the four traits under this study. In addition, two SNPs associated with teat placement reached the 5% genome-wide significance threshold (Figure 3, Table 4). Regarding chromosome distribution and traits affected by the significant associations identified, six chromosome-wise significant SNPs for udder depth were distributed on five chromosomes (OAR2, OAR6, OAR10, OAR12 and OAR13); three chromosome-wise significant associations were identified for udder attachment (OAR2 and OAR7); for teat placement, there were three out of the five significant SNPS that reached chromosome-wise significance, distributed on four chromosomes (OAR2, OAR3, OAR16 and OAR 22), whereas the other two where those SNPs reaching 5% genome-wide significance (OAR3 and OAR16); finally, six chromosome-wise significant SNPs were identified and distributed on three chromosomes (OAR1, OAR2 and OAR 16) for teat size (Table 4, Figure 3). The allelic substitution effects of all the significant SNPs identified ranged from 0.11 (for udder depth on OAR13), to 0.33 SD unit (for teat placement on OAR3 (Table 4). As mentioned in the Methods section, for each significant SNP, a TGI was defined considering the LD extent previously reported in Churra sheep (see details in Table 4).

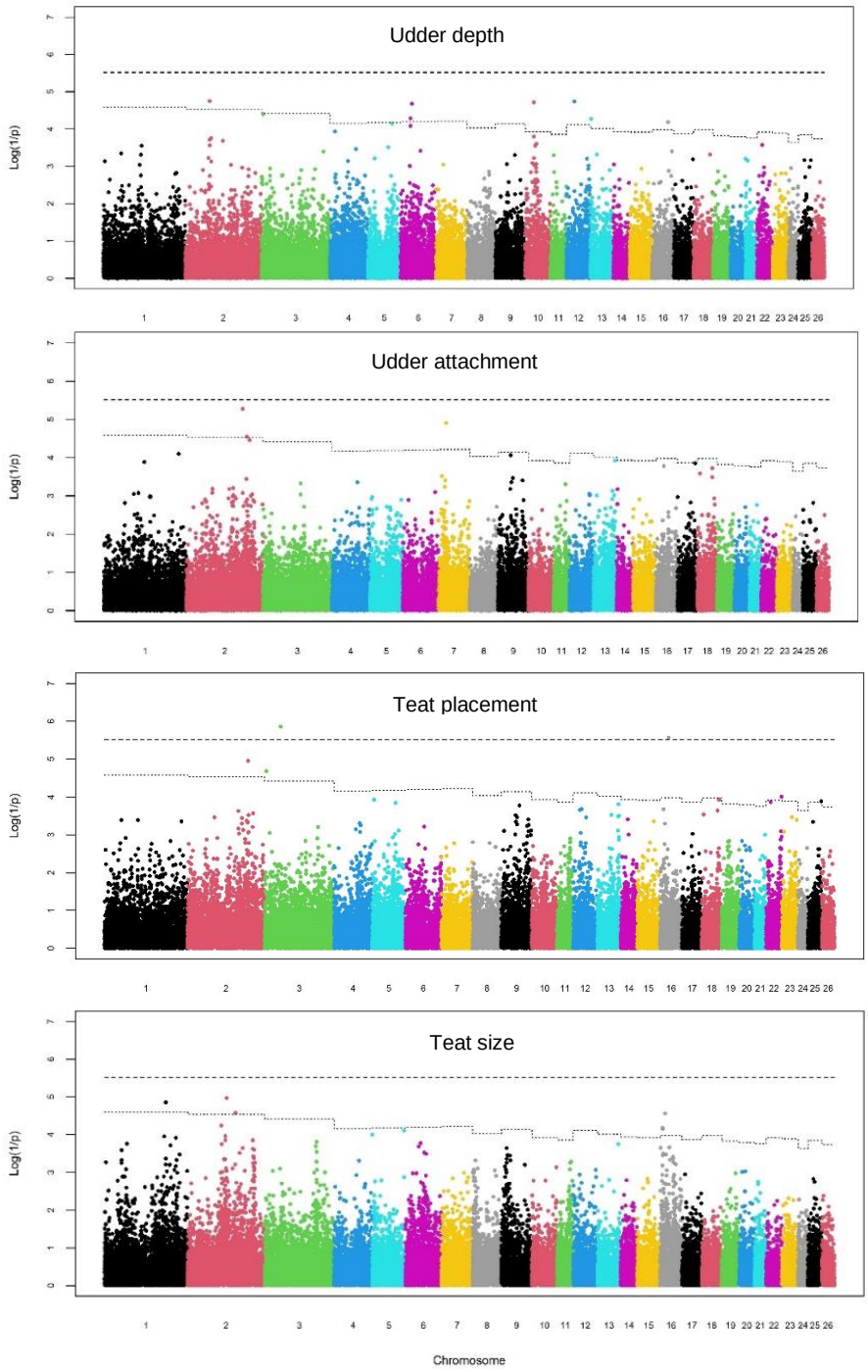

Figure 4.1.3.1. Graphical representation of the statistical profile obtained by the GWAS analysis performed for udder morphology traits in the present work. Horizontal thicker dashed line and horizontal thinner dashed line represent the genome-wise threshold [−log10(P-values) > 5.52], and the chromosome-wise threshold, respectively.

Table 4.1.3.1. Chromosome-wise SNPs significantly associated with mammary morphology traits identified by GWAS

Trait	OAR[a]	Top SNP name	Top SNP position	Allele substitution effect trait units (SD units)[c,d]	Nominal P-value	P_c-value (P_g-value)	TGI (Mb)[f]
Udder depth	2	OAR2_84032911.1	79462166	-0.061 ± 0.014 (0.16)	1.78E-05	0.074	79.34-79.59
	6	OAR6_37861888.1	34566514	0.048 ± 0.012 (0.13)	5.16E-05	0.1	34.44-34.69
	6	OAR6_43064935.1	39382514	-0.051 ± 0.012 (0.13)	2.10E-05	0.041	39.26-39.51
	10	OAR10_30241199.1	30237750	-0.055 ± 0.013 (0.14)	1.92E-05	0.026	30.11-30.36
	12	s37870.1	28872791	-0.077 ± 0.018 (0.20)	1.82E-05	0.024	28.75-29.00
	13	s05752.1	3750945	-0.043 ± 0.011 (0.11)	5.36E-05	0.067	3.63-3.88
Udder attachment	2	OAR2_191269884.1	181423260	-0.08 ± 0.018 (0.25)	5.32E-06	0.22	181.30-181.55
	2	OAR2_208149716.1	197698323	-0.1 ± 0.023 (0.32)	2.82E-05	0.12	197.57-197.82
	7	OAR7_29133786.1	26580982	-0.064 ± 0.015 (0.21)	1.24E-05	0.002	26.46-26.71
Teat placement	2	OAR2_206274047.1	195819615	0.058 ± 0.013 (0.19)	1.11E-05	0.046	195.69-195.94
	3	s75042.1	6176914	-0.054 ± 0.013 (0.18)	2.09E-05	0.078	6.05-6.30
	3	**OAR3_56745860.1**	**53734920**	**0.103 ± 0.021 (0.33)**	**1.39E-06**	**0.005 (0.05)**	**53.61-53.86**
	16	**OAR16_30364510.1**	**28128874**	**0.066 ± 0.014 (0.21)**	**2.72E-06**	**0.003 (0.05)**	**28.00-28.25**
	22	s34954.1	50326312	0.05 ± 0.013 (0.16)	9.88E-05	0.084	50.20-50.45
Teat size	1	OAR1_217573124.1	204192588	-0.042 ± 0.01 (0.17)	1.40E-05	0.062	204.07-204.32
	2	OAR2_134482532.1	127367921	0.046 ± 0.01 (0.19)	1.07E-05	0.044	127.24-127.49
	2	OAR2_163536981.1	155431903	-0.046 ± 0.011 (0.19)	2.69E-05	0.111	155.31-155.56
	16	OAR16_9401082.1	8638778	-0.05 ± 0.013 (0.20)	7.08E-05	0.083	8.51-8.76
	16	OAR16_9278782.1	8790354	-0.038 ± 0.01 (0.16)	6.69E-05	0.078	8.67-8.92
	16	s59518.1	16925238	0.044 ± 0.011 (0.18)	2.75E-05	0.032	16.80-17.05

[a] OAR ovine chromosome

[b] Position of the significant SNPs identified by the GWAS analysis based on Ramb_v2.0 version of the Ovine Genome Assembly (https://www.ncbi.nlm.nih.gov/)

[c, d] Magnitude of the allele substitution effect, and standard error, in trait units (Yield Deviations) and in phenotypic standard deviations (SD) units (in brackets)

[e] Corrected P-values at the 5% chromosome-wise level (and 5% genome wise level) obtained after applying a Gao correction considering the number of independent markers analyzed for each chromosome and for the whole genome, respectively.

[f] TGI Target genomic interval defined for the GWAS significant associations as 250 Kb long intervals centered on the significant SNP. The genes within that interval were extracted a positional candidate genes.

Among the 18 significant GWAS associations reported here, the one located on OAR22 and with effects on teat placement (TGI: 50.20-50.45 Mb) showed a clear overlapping with the chromosome-wise significant QTL detected for the same trait by the LA reported in this work

(TGI: 48.8-51.5 Mb). No other clear overlapping or correspondence was detected between the QTL detected through the two different gene mapping methodologies applied in this work.

4.2. QTL regions and gene annotation

For the significant QTL identified by LA, the CI flanking interval positions in cM were translated to physical positions in the sheep reference genome (pb) and considered as target genomic intervals (TGIs) to perform gene annotation with the GALLO R package, which provides a tool to perform a systematic annotation of positional candidate genes (Fonesca et al., 2020). Similarly, for the significant SNPs detected through the GWAS analyses performed, the corresponding TGIs were defined by considering 2,50 Mb intervals that included a 1,25 Mb region on both sides from the significant SNP. These TGIs were also considered for gene annotation with the previously mentioned package, using the corresponding annotation file for the Oar_Ramb_v2.0 ovine reference genome, which is available at https://www.ncbi.nlm.nih.gov/assembly/GCF_016772045.1 .

For both, the TGIs identified by LA and GWAS, we also performed a systematic annotation of QTLs reported in previous sheep studies for udder morphology traits and also for the indicator traits of mastitis (e.g. SCC, SCS, etc), due to the direct link between udder morphology and mastitis problems. For that, the available information, based on the Oar_Ramb_v2.0 reference genome, was downloaded from the SheepQTLdb database (http://www.animalgenome.org/cgi-bin/QTLdb/OA/search; Hu et al., 2013). In addition, the GALLO software was also used to perform a QTL enrichment analysis of the regions of interest with the aim of providing a more clear understanding of the genomic context of the candidate regions by investigating their QTL representativeness and diversity. Briefly, the QTL enrichment analysis function implemented by the GALLO package is based on a hypergeometric test approach, where the number of QTLs annotated within the candidate regions for each QTL type or trait is compared with the observed number of QTLs in the reference database (Fonseca et al., 2020).

4.3. QTL and gene annotation

The GALLO R package was used to perform a systematic extraction of positional candidate genes, we extracted a total of 1,588 annotated positional candidate genes from the TGIs defined according to the significant results of the LA and GWAS analyses. For the three genome-wise QTL reported in this work, all of them associated to Teat placement, the number of positional candidate genes identified were 79 for the OAR14 identified by LA, and 1 and 2, respectively for the significant associations detected on OAR3 region and OAR16 by GWAS. For these regions, and the rest of significant QTL detected, we present an assessment of potential functional candidate genes, based on comparison with previously reported studies, in the Discussion section of this memory.

In order to better understanding the genetic architecture of the udder morphology traits considered in this work, the integration of multiple sources of biological information is a crucial step. With that aim, we performed QTL annotation and enrichment analyses, based on the

SheepQTLdb information, for all the significant regions detected by both analyses performed, LA and GWAS. The proportion of the different trait type (milk, meat, carcass, health, production, reproduction, and exterior) and the different exterior traits of the annotated SheepQTLdb QTL and Associations that overlapped with the significant regions detected in the LA analysis here described are presented in Figure 4 as a pie chart (A) and as as bar plot (B), respectively. The same plots in relation to the significant TGIs defined based on the significant results of the GWAS analysis here reported are provided in Figure 5. The two most frequent QTL types for the LA QTLs were Meat and Carcass (38.04%) and Milk (22.7%) (Figure 4A), while for the GWAS QTLs the most frequent QTL types were Meat and Carcass (54.41%) and Production (13.24%) (Figure 5A). Note that the Production category includes traits such as digestive system, feed efficiency, feed intake and growth. Also, we performed an in-depth analysis for the Exterior QTL type in order to observe the frequency of each trait associated with specific QTL for the different traits included in this trait category, which are traits related to exterior features of the animals, from udder morphology, coat color, horns, conformation, behavioral, breech traits. The most frequent traits related to Exterior QTLs for the LA QTLs were teat number (1.8%) and horn type (1.25%) (Figure 4B), while for the GWAS QTL, the most frequent exterior trait represented was Horn type (2.5%) (Figure 5B).

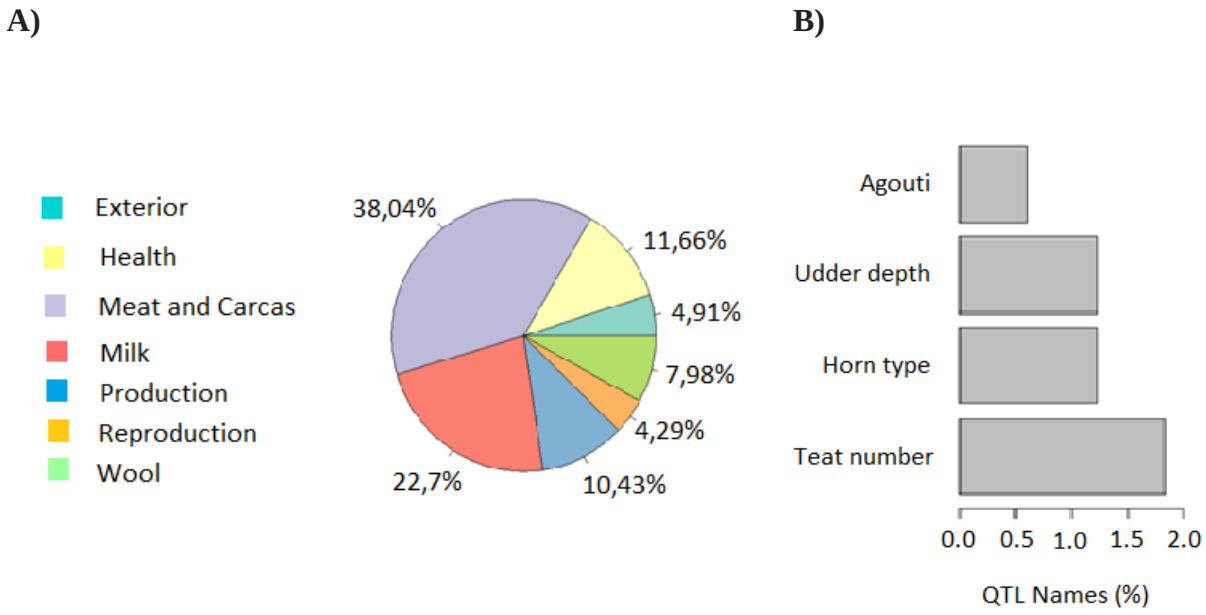

Figure 4.3.1. Percentage of QTL type (pie chart) and traits related to exterior QTLs (bar plots) overlapping with the significant QTL identified in the present work by LA, based on the QTL annotation analyses performed with the GALLO software.

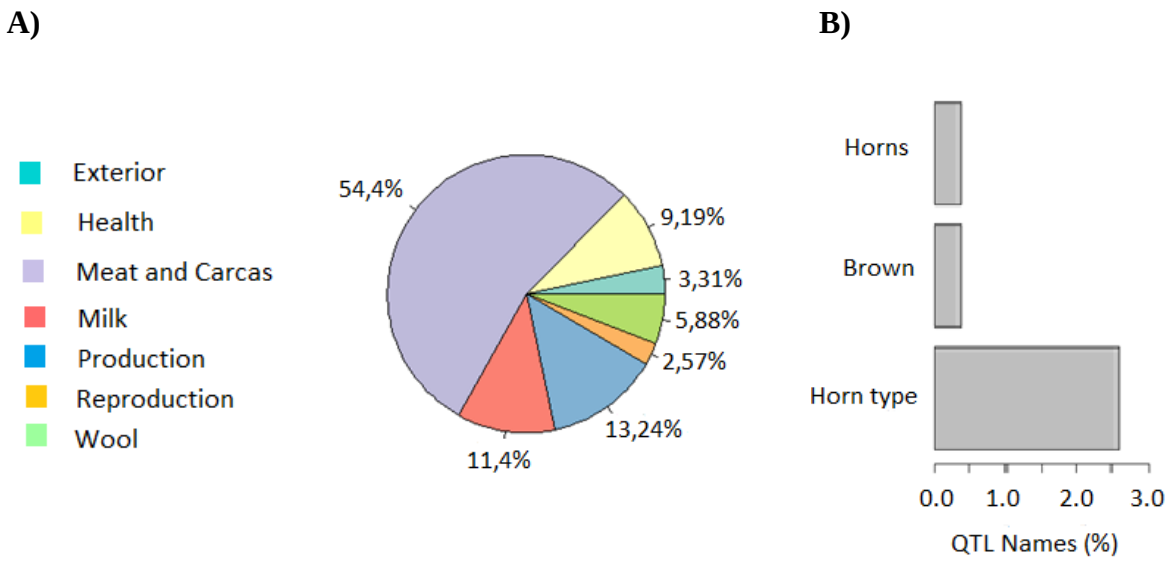

Figure 4.3.2. Percentage of QTL type (pie chart) and traits related to exterior QTLs (bar plots) overlapping with the significant QTL identified in the present work by GWAS, based on the QTL annotation analyses performed with the GALLO software.

To check the importance of the published type QTL representativeness within the regions of interest detected by LA and GWAS, we also performed a QTL enrichment analysis using the appropriate tool of the GALLO R package. The results of this analysis for the candidate regions detected by LA and GWAS are graphically represented in Figure 6 (A and B, respectively), where the x-axis shows a richness factor obtained by the ratio of the number of QTL annotated in the candidate regions and the total number of each QTL in the reference considered database. The most significant enriched terms from QTL annotation (darkest red shade in the circle), as well as the traits with the largest numbers of QTLs (biggest area of the circle) overlapping with our candidate regions were testes weight, in the case of the LA QTLs, and body weight, in the case of the GWAS significant regions. Note that for the results obtained in the case of the LA QTLs, Udder depth was one of the traits showing a largest enrichment.

A) **B)**

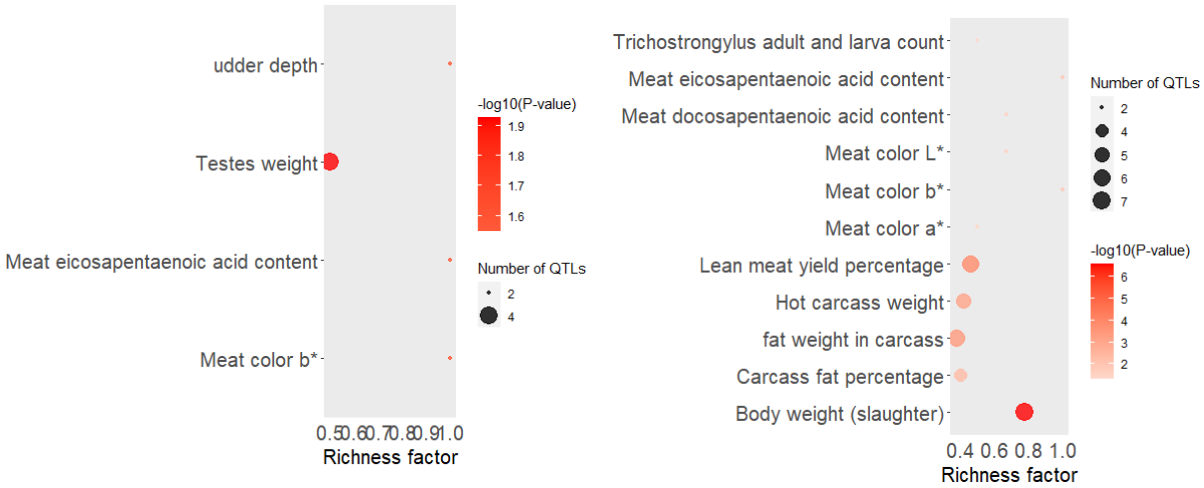

Figure 4.3.3. Bubble plots displaying the most significant results of the type QTL enrichment analysis results performed with the QTLs annotated in SheepQTLdb within the candidate regions identified in the presetn work by LA (**A**) and GWAS (**B**) analyses.

5. Discussion

Over the past 20 years, there has been an increase interest in studying udder morphology in sheep with regard to the suitability of the udder to the suitability for machine milking, particularly in undeveloped countries where dairy sheep have a major economic and social impact. Although, mammary morphology had been already described as an important factor in the machine milkability of dairy ewes (Labussière et al., 1981; Fernández et al., 1995). The overall goal of modern sheep husbandry is to maximize output while simultaneously ensuring good milk quality and better working conditions. The key technology to accomplish these goals is machine milking and udder morphology, which affects milking techniques. Milking techniques are crucial to try to adapt dairy ewes as best as possible to the milking environment (Dzidic et al., 2004). Mammary traits were not initially prioritized as selection criteria for dairy sheep, although, these traits influence a variety of machine milking and manageability factors as well as resistance to mastitis. As a result of long-term and biased selection to improve milk yield, herds of high-yielding sheep have a distorted udder morphology that is brought on by an increased pressure from the udder's weight on its suspensory system (Vrdoljak et. al., 2020). Mastitis is a prominent factor in the premature culling of dairy sheep; thus, the importance of udder health should be considered from both the perspective of the productive longevity of animals, as well as the hygienic quality of the milk produced. Only a few sheep breeds have been researched in relation to somatic cell count (SCC) as an indicator trait of udder health and milk hygiene quality, despite the fact that mastitis, particularly its subclinical form, is one of the primary health issues in dairy sheep flocks and leads to the culling of the animals (Vrdoljak et. al., 2020). However, it is considered that indirect selection for specific udder morphological traits may increase mastitis resistance (Legarra and Ugarte, 2005; Crump et al., 2019). For instance, Barillet (2007) discovered a favorable link between the quantity of somatic cells in milk and various morphological characteristics of the sheep udder, such as the position of the teat, the depth of the udder, and the prominence of the udder suspensory system. Additionally, based on the morphological characteristics of udders assessed during their first lactation, Casu et al. (2006) calculated the chance of udder inflammation during the production life of Mediterranean dairy ewes. Sheep with well-attached udders and less cranially orientated teats were less likely to experience udder inflammation than ewes with deep udders and forward-facing teats (Casu et al., 2010). This suggests that early selection for enhanced udder health based on udder morphology may be a good approach to be implemented in sheep breeding programs.

In addition to classical selection, today Genomic Selection can be used to improve the traits of interest in dairy sheep. Although this approach is still not routinely implemented in Spanish sheep breeding programs, the scenario may change in the next few years and SNP-chip data may be used to improve not only milk production traits but also other functional traits such as udder and body morphology. In this context, the use of genetic markers that have been proven to be associated to the traits to be improved by Genomic Selection could improve the accuracy of the estimated genomic breeding values (Teng et al., 2020).

In the last years, GWAS analyses have developed into a potent method for locating regions affecting phenotypic variation in cattle species and, in some cases, the specific causal mutations. According to our knowledge, this study is the first high-throughput SNP array GWAS focused on linear mammary morphology traits in dairy sheep, and in addition, it presents a genome scan based on LA.

Although strongly associated regions were found to be common to both analytic methods, LA and GWAS, the two strategies used in this study, yielded different results. The classical technique for QTL mapping was for many years was LA, which makes use of the family information provided by the populations considered in this type of studies. Since the availability of SNP-chips, which provide much denser genetic maps, this approach has largely been replaced by the exploitation of LD information across the population through GWAS analyses. However, it has been shown that the methodology that merges these two approaches, through the combined LDLA analysis, can provide a greatest amount of information in specific populations that have a family-based structure as it exploits both, linkage information from the within-family analyses, and LD information obtained at the population level (García-Gámez et al., 2013).

The present study, which was based on the analysis of a half-sib population of dairy ewes, has identified a total of 10 significant QTL regions identified by LA (one genome-wise and nine chromosome-wise significant regions), while the GWAS method identified 22 significant regions (two of them reached genome-wise significance and 20 were chromosome-wise significant). The in-depth comparison of these results, suggests two things: first, that the GWAS method represents an important improvement in the QTL detection compared with traditional LA, and second, that classical LA can only detect QTL in our family-based design if numerous sires are heterozygous (*Qq*) at the same QTL (García-Gámez et al., 2013). Hence, many marker-trait associations that do not fit this assumption but have a true association at the population level will not be detected by LA, but may be detected by GWAS. We compared the location of the significant QTLs found using the LA approach with that of the significant QTL found using the GWAS method, and we identified two QTL effects detected by both methods, one influencing teat placement on OAR22, and one affecting teat size OAR2.

Comparing our results with QTLs previously reported in sheep for mammary morphology and mastitis resistance related traits, we found an interesting coincidence with one QTL previously reported for Udder depth on OAR14, through the microsatellite-based genomic scan performed also in the Churra sheep breed, by Gutiérrez-Gil et al., (2008). That QTL appears to be close (CI: 53.5-62.6 Mb in Oar_Ramb_v2.0, according the SheepQTLdb) to the region where we have found here the genome-wise significant QTL for Teat placement on OAR14 (CI: 52.6-52.9 Mb). Although the two traits are not the same, the coincidence between these results, and the known genetic correlation between the two involved traits (-0,32; Fernández et al., 1997), could suggest that this region on OAR14 harbors a gene or genes that have a direct effect on different aspects of udder morphology, at least in Churra sheep. In addition, when comparing also our results with the QTLs described in sheep for milk production and mastitis resistance traits, we found an interesting coincidence for the genome-wise significant OAR14 QTL

reported here for teat position, as it matched previously reported QTL effects for milk lactose content and SCS described in an Awassi × Merino backcross population (Raadsma et al., 2009). This correspondence may be a reflection of the link between udder morphology and udder health traits, supporting that relevant functional traits, such as udder morphology, should be included as selection objectives in dairy sheep breeding programs. Other coincidences were found for the OAR2 QTL reported here for teat size teat, which matched an association reported for milk protein content in Churra sheep (Gutiérrez-Gil et al., 2009a), and for the QTL on OAR20 for teat size, which was coincident with a QTL for milk production described in a backcross East Friesian x Dorset population (Mateescu and Thonney, 2010).

In the Results section we described the overlapping of the QTL detected in the present work with previously reported QTL in dairy sheep. Despite the limited number of studies reported in sheep for udder morphology, some remarkable coincidences were found. The most interesting overlapping was that found between the genome-wise significant OAR14 QTL for Teat placement and the QTL for udder depth reported by the microsatellite genome scan described by the MEGA-ULE group in a different population of Churra sheep (Gutierrez-Gil et al., 2008). Although the two traits involved in this coincidence are not the same, the estimated genetic correlation between them (-0,32; Fernández et al., 1997) could suggest that this region on OAR14 harbors a gene or genes that have a direct effect on different aspects of udder morphology, at least in Churra sheep. Hence, considering also that this was the LA QTL identified here with the highest statistical support and that that chromosomal region also showed correspondence with a SCC QTL described by Raadsma et al. (2009), we would suggest this region to be further analyzed through a pleiotropy-GWAS analysis with the aim of assessing whether it has a real influence on both udder morphology and udder's health traits. If confirmed, this would be in line with previous authors' suggestions about including udder morphology in dairy sheep breeding programs to obtain an indirect genetic response for udder's health traits (Legarra and Ugarte, 2005).

Apart of the comparative with the SheepQTLdb and taking into account the much larger number of QTL studies reported in dairy cattle than in sheep, we considered of interest, for the interpretation of the QTL results reported here, to do a comparison with the QTL annotated in the CattleQTLdb, in reference to udder type traits, and also to milk production and mastitis resistance traits. As a previous step to that comparison, we had to identify the correspondent orthologous coordinates in the bovine genome to the QTL reported here. For that, the software Remap of the NCBI (https://www.ncbi.nlm.nih.gov/genome/tools/remap) was used to obtain the correspondence between the cattle QTL/Associations genomic coordinates based on the ARS-UCD1.3 reference cattle genome) and the Oar_Ramb_v2.0 sheep reference genome. In general, a direct conversion was obtained, although for some intervals two conversion steps were required. Through this systematic comparison with cattle QTL results, a large number of correspondences were found, with a total of 80 cattle QTL or Associations mapping within the orthologous regions of the QTL reported in the present work (e.g. Schnabel et al. 2005; Brand et al., 2009; Collis et al., 2012; Durán et al., 2017). Considering the analysis method here implemented, a total of 52 cattle QTL/associations were found within the TGI defined by LA in the present study, whereas for the TGI defined on the basis of the significant GWAS QTLs,

28 correspondences where found (results not shown). For the QTL detected by the two methods, correspondence was found for both trait types of QTLs, udder related traits and mastitis resistance indicator traits. Hence, our LA significant QTL regions overlapped with about 29 QTLs for udder traits and 23 for SCC-related traits, whereas for the GWAS detected QTL regions, the identified correspondences with bovine QTLs involved about 16 QTL for udder traits, and 12 for mastitis resistance indicator traits. Focusing on the three genome-wise significant QTL detected in the present study, all of them influencing teat placement, Table 5 provides details about the correspondences found with cattle QTL/Associations for the functional traits considered. As can be seen, the LA genome-wise significant QTL detected on OAR14 overlaps with cattle QTL for udder cleft and teat length in the corresponding orthologous region of bovine chromosome (BTA) 18 (Brand et al., 2009; Schnabel et al., 2005). Whereas, from the two genome-wise significant effects detected by GWAS, only that on OAR16 showed correspondence with a significant association reported for SSC in cattle chromosome 20 (BTA20).

Table 5.1. Correspondence between the genome-wise significant sheep QTLs identified in this work for udder morphology traits and QTLs previously described in cattle in relation to functional traits of interest (mammary morphology and somatic cell count) based on information obtained from the CattleQTLdb database.

Sheep reference character	Analysis	Ovine genomic region Chr:TGI (pb) (Oar_Ramb_v2)	Cattle orthologous interval (pb) (UMD_3.1)	Type of analysis	QTL ID (CattleQTLdb)	Reference
Teat placement	LA	OAR14: 52600000-52900000	BTA18: 53213205-53539982	QTL	Udder cleft (11596)	Brand *et al*. (2009)
				QTL	Teat length (1703)	Schnabel *et al*. (2005)
	GWAS	OAR 3: 53610000-53860000	BTA11: 52960655-53171843	-	-	-
		OAR 16: 28000000-28250000	BTA20: 27859656-28087192	Association	*IGFI* (20740) SCS (122120)	Collis *et al.* (2012) Durán et al. (2017)

Again, the correspondences found for the LA OAR14 QTL for teat placement suggest that this region should be further analyzed in order to identify genetic markers that could be directly used to improve udder morphology traits in dairy sheep. Interestingly, one of these correspondences involves also a teat related trait (teat length), whereas the potential relationship with the udder cleft trait and teat placement could be related to the fact that the udder cleft trait is related to the udder's suspensory ligament's strength. Hence, it could appear logical to consider that variations in teat placement my result from the suspensory ligament's laxity. As a consequence of this, the nipples may be more medially oriented due to an excessively strong suspensory ligament, and more laterally oriented due to an excessively lax suspensory ligament (Prpić et al., 2013). In any case, further study should be performed to confirm if the genome-wise QTL detected on OAR14 for Teat placement may have effects on other mammary gland

morphology traits. Correspondence identified for the geno-wise significant association detected in the present work by GWAS on OAR16 for teat placement (Table 5), and the BTA20 orthologous region harboring the *IGF1* gene (Collis et al., 2012). In this study author showed that the insulin-like growth factor, IGF-I, has endocrine as well as autocrine-paracrine actions on tissue growth. IGF1 ligand is expressed within developing mammary tissue throughout postnatal stages with specific sites of expression in the epithelial and stromal compartments. Based on this, *IGF1* could be suggested as a potential functional candidate gene for the OAR16 QTL detected here for teat placement. In addition, for the same OAR16 QTL region, we also found correspondence with a QTL for SCS in cattle that were previously indicated (Collis et al., 2016). In that study, fine mapping of a previously identified QTL for SCS is presented, and several candidate genes are analyzed. A polymorphism in the *PVRL2* gene was associated with the SCS trait, and the authors noted its potential interest as a gene involved in mammary gland development. One-pass type I membrane glycoprotein with two Ig-like C2-type domains and an Ig-like V-type domain are the products of this gene. One of the plasma membrane elements of adheres junctions is this protein. It also acts as a portal for some mutant strains of the herpes simplex virus and the pseudorabies virus, which are involved in the transmission of these viruses from cell to cell. Hence, that study suggest that the target SCC QTL may actually be the result of a genetic effect that is indeed influencing primarily udder conformation, and only secondly SCS. This theory is supported by the numerous QTL for udder morphology traits that has been found in that region of bovine chromosome BTA18 (Collis et al., 2012), as can be seen in CattleQTLdb.

A later stage of the interpretation of the results reported here was the assessment of the total of 1,588 annotated genes that were extracted as positional candidate genes from the genomic confidence regions defined based on the LA (1,266) and GWAS (322) analyses. For this evaluation, we searched in the literature for studies that could provide a list of reference genes related to udder morphology and udder structure in sheep and/or cattle. Hence, on one hand we considered the transcriptomic study reported by Suarez-Vega et al., (2015a) about the identification of DEGs across the different stages of lactation and, on the other hand, we considered the results of the system-based analyses for udder conformation and health phenotypes presented in French dairy cattle breeds by Marete et al. (2018). Based on these studies, a final list of 764 reference functional candidate genes was considered [575 unique genes identified as DEGs across different lactation time points by Suarez-Vega et al. (2015a), and 189 genes highlighted by the GWAS and other analyses described in Marete et al. (2018)]. When comparing our positional candidate genes, with the reference gene list, we identified eight coincident genes which were considered as potential functional candidate genes for the corresponding QTL regions here identified. Six of these potential functional candidate genes were related to LA QTL regions (*GJA5*, *SLC2A10*, *APOE*, *IP6K3*, *MAPK13*, *SYNGAP1*), and two of them were included within the TGIs of the GWAS-detected QTLs (*ASS1*, *PDE1A*).

From this list of functional candidate genes identified, four had been identified as DEGs in the MSC transcriptome study reported by Suárez-Vega et al. (2015) (*SLC2A10*, *APOE*, *IP6K3*, *MAPK13*), and the other four were considered based on the analyzed presented by Marete et al. (2018) (*GJA5*, *ASS1*, *PDE1A* and *SYNGAP*). The correspondence between these genes and the

QTL here identified is provided in Table 6, where the genome-wise QTL are highlighted in bold font.

Table 5.2. Potential functional candidate genes here identified for the different traits here reported. and their correspondence with the QTL here reported.

Method	Trait	Chr.	Corrected P_c-value (P_q-value)	TGI (bp)	Functional candidate gene identified	Study of reference
LA	Udder attachment	1.	<0.05	98.0-103.2	*GJA5*	Marete et al., (2018)
		13	< 0.05	74.0-76.0	*SLC2A10*	Suarez-Vega et al., (2015a)
	Teat placement	**14**	**< 0.001 (< 0.05)**	**52.6-52.9**	***APOE***	Suarez-Vega et al., (2015a)
	Teat size	20	< 0.05	7.5- 10.5	*IP6K3, MAPK13, SYNGAP1*	Suarez-Vega et al., (2015a) Marete et al., (2018)
GWAS	Teat placement	3	0.078	6.05-6.30	*ASS1*	Marete et al., (2018)
	Teat size	2	0.044	127.24-127.49	*PDE1A*	Marete et al., (2018)

Trying to find a possible link between the biological function of the potential candidate genes identified, we present below a short description of the information obtained through the discussion of results presented by the reference studies considered for this identification of the most promising positional candidate genes extracted from the QTL regions reported in this work.

Due to the relevance suggested by the different results here described for the LA QTL for Teat placement on OAR14, we will first comment on the functional candidate gene identified for this region. The ***APOE*** gene encodes a major apoprotein of the chylomicron. The coding protein of this gene binds to a specific liver and peripheral cell receptor and is essential for the normal catabolism of triglyceride-rich lipoprotein constituents (Miida et al., 2009). According to Suárez-Vega et al. (2015a), the *APOE* gene was identified as DEG in two of the lactation points considered (D10vsD120 and D10vsD150), and through the corresponding Gene Ontology (GO) enrichment analyses this gene appeared as related to the term extracellular region matrix (cellular component category), and to other terms such as development process, multicellular organismal development, response to stimulus, response to stress and defense response (biological processes category). On this regard, we should take into account that the mammary gland is a dynamic tissue that achieves full maturity in the adult. Mammary ducts have luminal cells, associated with myoepithelial cells, surrounded by the basement membrane that separates the epithelium from the stroma. One of the components of the stroma is the extra-cellular matrix (ECM) (laminin, fibronectin, collagen, proteoglycans, etc.), which influences mammary development. The ECM has been reported to modulate mammary epithelial growth and

differentiation in embryonic development, postnatal ductal growth, branching morphogenesis, and carcinogenesis (Fata et al., 2003). Of the myriad of ECM-mammary epithelial cell interactions, integrin signaling will be discussed in more detail. The biochemical and biophysical cues from the extracellular stroma that guide mammary epithelial morphogenesis, homeostasis and malignant transformation will also be described. ECM modulates mammary epithelial growth and differentiation in embryonic development, postnatal ductal growth, branching morphogenesis (Kass et al., 2007). In addition to this gene, Suarez-Vega et al., (2015a) identified several genes related to ECM in relation to the involution process that the mammary gland suffers at the end of lactation, and which is in concordance with the identification of this gene as a DEG in the two across-lactation time comparisons mentioned (D10vsD120 and D10vsD150). This discussion highlights the interest of assessing the role of this gene in relation to the genome-wise QTL here reported on OAR14, and, due to the coincidences described for this region with other udder trait QTL, in both sheep and cattle, not only for Teat placement, but in general, for udder morphology traits, and even, udder's health traits.

For the other genome-wise significant QTL detected in this study, through GWAS, we did not identify any potential functional candidate genes. For the rest of reported QTL in this work, we will comment here briefly about the most promising candidate genes identified. Also, based on the contrasting RNA-Seq analysis presented by Suarez-Vega et al. (2015a) for the D10vsD120 and D10vsD150 comparisons, we identified two potential functional candidates within the TGI defined by our LA genome scan for the OAR20 chromosome-wise QTL with effects on Teat size. The first of these genes is *IP6K3*. A deletion in this gene has revealed its involvement in controlling glucose, insulin, fat mass, body weight and lactate levels in mice (Moritoh et al., 2016). Then, the *MAPK13* gene, which controls a range of cellular functions including proliferation, differentiation, transcription regulation, and development, and serves as an integration point for many biochemical signals (Liu et al., 2020). According to Suárez-Vega et al. (2015a), *MAPK13* is associated to GO biological process terms such as multicellular organismal process, single-multicellular organismal process, single organism process and response to wounding. Finally, the *SLC2A10*, which is located within the TGI defined for the OAR13 QTL identified by LA for TRAIT. This gene was also detected in the previously mentioned time point comparisons and supported the enrichment of GO terms in the biological process category such as cell periphery, plasma membrane, and intrinsic to membrane. This gene encodes for the GLUT10 protein, which influences the regulation of the transforming growth factor-beta (TGF-) signaling pathway (Gordon and Blobe, 2008). This pathway participates in cell growth and division (proliferation), as well as the maturation process that prepares cells to perform certain roles (differentiation) and also participates in the ECM structure, which has already been mentioned in relation to the importance of mammary gland structure and dynamic changes such as involution (Suarez-Vega et al., 2015a). Interestingly, the *SLC4A10* gene has been found as positional candidate genes in a GWAS-based study for mammary structure reported in Canadian Angus cows (Devani et al., 2020).

In relation to the functional candidate genes highlighted in our study due to their correspondence with the results reported by Marete et al. (2018), the *GJA5* gene, which is

located within the TIG defined for the OAR1 QTL influencing Udder attachment and belongs to the connexon gene family. Gap junctions, which are made up of clusters of intercellular channels and serve as a pathway for the transfer of low molecular weight molecules from cell to cell, are made up in part of the encoded protein by this gene. Atrial fibrillation may be linked to mutations in this gene. Although this does not indicates a clear relationship with udder morphology, this gene is also known as Connexin 40, and has been related with genetic diseases of junctions. In relation to this, the permeability of mammary tight junctions, which are semipermeable extracellular structures that are located in proximity to the apical domain of the cell, are related with milk secretion in dairy ewes (Castillo et al., 2008). This highlights that when studying, or improving, at the genetic level mammary morphology in dairy species, we may not only identify, or influence, genetic mechanisms influencing the classical morphological type traits under study, but also functional traits of the mammary gland such as milking speed, or the milking flow.

The other genes identified based on Marete et al. (2018) study included the ***ASS1*** gene, which would be a functional candidate for the association found on OAR3 for Teat placement, encodes the arginosuccinate synthase 1 enzyme. The mentioned enzyme takes part in the urea cycle, a chain of chemical processes that occurs in liver cells. Also, the *PDE1 gene,* contain 3' cyclic phosphate links, which are hydrolyzed by phosphodiesterase's (PDEs), a family of phosphohydrolase's. PDEs control the degradation of second messengers to regulate them. According to Marete et al. (2018) this gene is related to morphogenesis of a branching epithelium; this is the process by which branch anatomical structures are created and arranged. A branch is an offshoot or division of the primary stem. Blood arteries, nerves, lymphatics, and other endothelial or epithelial tubes, such as those present in the mammary gland, are a few examples found in animals. Hence, this gene appears as a promising functional gene to be considered in relation to the OAR2 QTL influencing Teat size, in which TIG was found when contrasting our positional candidate genes with the udder morphology reference gene list considered. Finally, the *SYNGAP* gene, located within the QTL region influencing teat size on OAR20, encodes for a protein termed SynGAP, which is crucial for the function of brain nerve cells. Cell-to-cell communication occurs at the junctions (synapses) between nerve cells where SynGAP is present. The results of Marete et al., (2018) identified as central enriched pathways, based on the genes highlighted by the different analyses presented in relation to udder morphology traits, "Neuropathic pain-signaling in dorsal horn neurons pathway," and, "CREB signaling pathway" and they suggested that this functions could be related to the milk ejection process which may result from the strong integration of sensory input from mammary glands afferents that terminate in the dorsal horn. Hence, the identification of this nerve cell related gene could be of interest in relation to functional aspects of mammary gland in relation to milking ability. The large number of QTLs found in this analysis lends support to the hypothesis that complex multi-gene interactions rather than single genes act together in determining mammary morphology. Our findings represent the initial stages in the discovery of allelic variations that directly regulate the phenotypic diversity seen in adult Churra sheep mammary morphology. The adoption of information derived from these findings into the breeding plans for the Churra sheep breed population would require of the discovery of causal variants or SNPs in substantial LD with the causal variants directly influencing udder morphology trait. Such

variants would be expected to improve the prediction ability of genomic breeding values for udder type traits and, indirectly, for udder milking ability and udder's health status. The results presented here may be considered as a starting point of a future research line of the MEGA-ULE group that combines functional genomic data with genomic variation analysis that could shed light on the biology mechanisms controlling sheep mammary morphology.

6. Conclusion

This study has performed genome scan in Spanish Churra sheep for udder morphology traits, based on a medium density SNP array, Through two different analysis methods, Linkage Analysis and Genome-wide associations analysis, a total of 30 QTL have been identified, three of them, related to the teat placement traits, reaching genome-wise significance. Secondly, the systematic comparisons of the ovine genomic regions harboring the QTL here identified, with previously QTLs reported in both, sheep, and cattle, for mammary morphology traits, milk production and mastitis resistance indicator traits, support the genuine nature of the association of the genomic regions highlighted in this work with udder morphology traits. Finally, from a large list of 1,588 positional candidate genes extracted from the confidence genomic regions identified in this study, and by considering a udder morphology and structure reference list of candidate genes, the present study provides a list of potential functional candidate genes for the QTL detected here. Future studies should be designed in commercial populations to confirm the potential associations suggested here, with the aim of providing useful genomic information to improve the prediction ability of genomic selection programs running in the future in Spanish Churra sheep.

7. Refrences

1. Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., ... & Lundström, K. (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science*, *263*(5154), 1771-1774.
2. Arranz, J. J., & Gutiérrez-Gil, B. (2012). Detection of QTL underlying milk traits in sheep: An update. *Milk production—advanced genetic traits—cellular mechanism—animal management health*, 97-126.
3. Atlija, M., Arranz, J. J., Martinez-Valladares, M., & Gutiérrez-Gil, B. (2016). Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genetics Selection Evolution*, *48*(1), 1-16.
4. Barillet, F. (2007). Genetic improvement for dairy production in sheep and goats. *Small Ruminant Research*, *70*(1), 60-75. Barillet, F. (2007). Genetic improvement for dairy production in sheep and goats. *Small Ruminant Research*, *70*(1), 60-75.
5. Barillet, F., Arranz, J. J., & Carta, A. (2005). Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. *Genetics selection evolution*, *37*(Suppl. 1), S109-S123.
6. Barillet, F., Marie, C., Jacquin, M., Lagriffoul, G., & Astruc, J. M. (2001). The French Lacaune dairy sheep breed: use in France and abroad in the last 40 years. *Livestock Production Science*, *71*(1), 17-29.
7. Bayat, A. (2002). Bioinformatics.(Science, Medicine, and the Future). *British Medical Journal*, *324*(7344), 1018-1023.
8. Becker, D., Tetens, J., Brunner, A., Bürstel, D., Ganter, M., Kijas, J., ... & Drögemüller, C. (2010). Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. *PLoS One*, *5*(1), e8689.
9. Bergonier, D., De Crémoux, R., Rupp, R., Lagriffoul, G., & Berthelot, X. (2003). Mastitis of dairy small ruminants. *Veterinary research*, *34*(5), 689-716.
10. Bionaz, M., Periasamy, K., Rodriguez-Zas, S. L., Hurley, W. L., & Loor, J. J. (2012). A novel dynamic impact approach (DIA) for functional analysis of time-course omics studies: validation using the bovine mammary transcriptome. *PloS one*, *7*(3), e32455.
11. Blackstock, W. P., & Weir, M. P. (1999). Proteomics: quantitative and physical mapping of cellular proteins. *Trends in biotechnology*, *17*(3), 121-127.
12. Bovine Genome Sequencing and Analysis Consortium, Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A., Muzny, D. M., ... & Hitchens, M. E. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, *324*(5926), 522-528.
13. Brand, B., Baes, C., Mayer, M., Reinsch, N., Seidenspinner, T., Thaller, G., & Kühn, C. (2010). Quantitative trait loci mapping of calving and conformation traits on Bos taurus autosome 18 in the German Holstein population. *Journal of dairy science*, *93*(3), 1205-1215.
14. Bunnik, E. M., & Le Roch, K. G. (2013). An introduction to functional genomics and systems biology. *Advances in wound care*, *2*(9), 490-498.

15. Castillo, V., Such, X., Caja, G., Casals, R., Albanell, E., & Salama, A. A. K. (2008). Effect of milking interval on milk secretion and mammary tight junction permeability in dairy ewes. *Journal of dairy science*, *91*(7), 2610-2619.
16. Casu, S., Pernazza, I., & Carta, A. (2006). Feasibility of a linear scoring method of udder morphology for the selection scheme of Sardinian sheep. *Journal of dairy science*, *89*(6), 2200-2209.
17. Casu, S., Sechi, S., Salaris, S. L., & Carta, A. (2010). Phenotypic and genetic relationships between udder morphology and udder health in dairy ewes. *Small Ruminant Research*, *88*(2-3), 77-83.
18. Charon, K. M. (1990). Genetic parameters of the morphological traits of sheep udder. *World Review of animal production*, *25*(1), 73-76.
19. Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., ... & Palmarini, M. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science*, *324*(5926), 532-536.
20. Chitneedi, P. K., Arranz, J. J., Suarez-Vega, A., García-Gámez, E., & Gutiérrez-Gil, B. (2017). Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. *Animal genetics*, *48*(4), 436-446.
21. Collis, E., Fortes, M. R. S., Zhang, Y., Tier, B., Schutt, K., Barendse, W., & Hawken, R. (2012). Genetic variants affecting meat and milk production traits appear to have effects on reproduction traits in cattle. *Animal genetics*, *43*(4), 442-446.
22. Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., & Fakan, S. (2006). Chromosome territories–a functional nuclear landscape. *Current opinion in cell biology*, *18*(3), 307-316.
23. Crisà, A., Ferrè, F., Chillemi, G., & Moioli, B. (2016). RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. *BMC veterinary research*, *12*(1), 1-21.
24. Crump, R. E., Cooper, S., Smith, E. M., Grant, C., & Green, L. E. (2019). Heritability of phenotypic udder traits to improve resilience to mastitis in Texel ewes. *Animal*, *13*(8), 1570-1575.
25. Cui, J., Wang, H., Liu, S., Qiu, X., Jiang, Z., & Wang, X. (2014). Transcriptome analysis of the gill of Takifugu rubripes using Illumina sequencing for discovery of SNPs. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *10*, 44-51.
26. Dai, W. T., Zou, Y. X., White, R. R., Liu, J. X., & Liu, H. Y. (2018). Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. *Functional & integrative genomics*, *18*(2), 125-140.
27. Dalrymple, B. P., Kirkness, E. F., Nefedov, M., McWilliam, S., Ratnakumar, A., Barris, W., ... & Cockett, N. E. (2007). Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome biology*, *8*(7), 1-20.
28. de España, G. (2003). Ministerio de agricultura, pesca y alimentación. *Raza Bovina RUBIA GALLEGA. Available online: https://www. mapa. gob. es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-razas/bovino/rubia-gallega/datos_morfologicos.*

29. De la Fuente, L. F., Fernandez, G., & San Primitivo, F. (1996). A linear evaluation system for udder traits of dairy ewes. *Livestock Production Science*, *45*(2-3), 171-178.
30. Devani, K., Plastow, G., Orsel, K., & Valente, T. S. (2020). Genome-wide association study for mammary structure in Canadian Angus cows. *Plos one*, *15*(8), e0237818.
31. Durán Aguilar, M., Román Ponce, S. I., Ruiz López, F. J., González Padilla, E., Vásquez Peláez, C. G., Bagnato, A., & Strillacci, M. G. (2017). Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *Journal of Animal Breeding and Genetics*, *134*(1), 49-59.
32. Dzidic, A., Kaps, M., & Bruckmaier, R. M. (2004). Machine milking of Istrian dairy crossbreed ewes: udder morphology and milking characteristics. *Small Ruminant Research*, *55*(1-3), 183-189.
33. Fadiel, A., Anidi, I., & Eichenbaum, K. D. (2005). Farm animal genomics and informatics: an update. *Nucleic acids research*, *33*(19), 6308-6318.
34. FAO. (1991). Food and agriculture organization of the united nations FAOSTAT: the statistical database of FAO.
35. Fata, J. E., Werb, Z., & Bissell, M. J. (2003). Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. *Breast cancer research*, *6*(1), 1-11.
36. Fernandez, G., Alvarez, P., San Primitivo, F., & De la Fuente, L. F. (1995). Factors affecting variation of udder traits of dairy ewes. *Journal of Dairy Science*, *78*(4), 842-849.
37. Fernández, G., Baro, J. A., De la Fuente, L. F., & San Primitivo, F. (1997). Genetic parameters for linear udder traits of dairy ewes. *Journal of dairy science*, *80*(3), 601-605.
38. Filangi, O., Moreno, C., Gilbert, H., Legarra, A., Le Roy, P., & Elsen, J. M. (2010, August). QTLMap, a software for QTL detection in outbred populations. In *Proceedings of the 9th world congress on genetics applied to livestock production* (pp. 1-6).
39. Fisher, R. A. (1923). XXI.—On the dominance ratio. *Proceedings of the royal society of Edinburgh*, *42*, 321-341.
40. Fonseca, P. A., Suárez-Vega, A., Marras, G., & Cánovas, Á. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. *GigaScience*, *9*(12), giaa149.
41. Forrest, A. R., & Carninci, P. (2009). Whole genome transcriptome analysis. *RNA biology*, *6*(2), 107-112.
42. Gao, X., Starmer, J., & Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, *32*(4), 361-369.
43. García-Fernández, M., Gutiérrez-Gil, B., García-Gámez, E., Sánchez, J. P., & Arranz, J. J. (2010a). Detection of quantitative trait loci affecting the milk fatty acid profile on sheep chromosome 22: role of the stearoyl-CoA desaturase gene in Spanish Churra sheep. *Journal of dairy science*, *93*(1), 348-357.

44. García-Fernández, M., Gutiérrez-Gil, B., García-Gámez, E., Sánchez, J. P., & Arranz, J. J. (2010b). The identification of QTL that affect the fatty acid composition of milk on sheep chromosome 11. *Animal Genetics*, *41*(3), 324-328.
45. García-Fernández, M., Gutiérrez-Gil, B., Sánchez, J. P., Morán, J. A., García-Gámez, E., Alvarez, L., & Arranz, J. J. (2011). The role of bovine causal genes underlying dairy traits in Spanish Churra sheep. *Animal genetics*, *42*(4), 415-420.
46. Garcia-Gamez, E., Gutierrez-Gil, B., Sahana, G., Sánchez, J. P., Bayón, Y., & Arranz, J. J. (2012a). GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene.
47. García-Gámez, E., Gutiérrez-Gil, B., Sánchez, J. P., & Arranz, J. J. (2012b). Replication and refinement of a quantitative trait locus influencing milk protein percentage on ovine chromosome 3. *Animal Genetics*, *43*(5), 636-641.
48. García-Gámez, E., Gutiérrez-Gil, B., Suarez-Vega, A., De la Fuente, L. F., & Arranz, J. J. (2013). Identification of quantitative trait loci underlying milk traits in Spanish dairy sheep using linkage plus combined linkage disequilibrium and linkage analysis approaches. *Journal of dairy science*, *96*(9), 6059-6069.
49. García-Gámez, E., Sahana, G., Gutiérrez-Gil, B., & Arranz, J. J. (2012c). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC genetics*, *13*(1), 1-11.
50. Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., ... & Zhao, X. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, *139*(2), 907-920.
51. Gootwine, E., Alef, B., & Gadeesh, S. (1980). Udder conformation and its heritability in the Assaf (Awassi× East Friesian) cross of dairy sheep in Israel. In *Annales de génétique et de sélection animale* (Vol. 12, No. 1, pp. 9-13). EDP Sciences.
52. Gordon, K. J., & Blobe, G. C. (2008). Role of transforming growth factor-β superfamily signaling pathways in human disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1782*(4), 197-228.
53. Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J. J., Kvasz, A., ... & Georges, M. (2004). Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences*, *101*(8), 2398-2403.
54. Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., ... & Schook, L. B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, *491*(7424), 393-398.
55. Gutiérrez-Gil, B. (2004) *Detección de regiones genómicas con influencia sobre caracteres de producción láctea y morfología mamaria en el ganado ovino de raza Churra.* Tesis Doctoral. Universidad de León.
56. Gutiérrez-Gil, B., Alvarez, L., De la Fuente, L. F., Sanchez, J. P., San Primitivo, F., & Arranz, J. J. (2011). A genome scan for quantitative trait loci affecting body conformation traits in Spanish Churra dairy sheep. *Journal of dairy science*, *94*(8), 4119-4128.

57. Gutiérrez-Gil, B., Arranz, J. J., Othmane, M. H., De la Fuente, L. F., & San Primitivo, F. (2001). Influencia del genotipo de la b-lactoglobulina ovina sobre caracteres cualitativos y rendimiento quesero individual en la raza Churra. *ITEA*, *22*, 15-17.
58. Gutiérrez-Gil, B., El-Zarei, M. F., Alvarez, L., Bayón, Y., De La Fuente, L. F., San Primitivo, F., & Arranz, J. J. (2008). Quantitative trait loci underlying udder morphology traits in dairy sheep. *Journal of dairy science*, *91*(9), 3672-3681.
59. Gutiérrez-Gil, B., El-Zarei, M. F., Alvarez, L., Bayón, Y., de la Fuente, L. F., San Primitivo, F. y Arranz, J.J. (2009a) "Quantitative trait loci underlying milk production traits in sheep", *Animal Genetics*. Blackwell Publishing Ltd, 40(4), pp. 423–434.
60. Gutiérrez-Gil, B., El-Zarei, M. F., Bayón, Y., Alvarez, L., De la Fuente, L. F., San Primitivo, F., & Arranz, J. J. (2007). Detection of quantitative trait loci influencing somatic cell score in Spanish Churra sheep. *Journal of dairy science*, *90*(1), 422-426.
61. Gutierrez-Gil, B., Esteban-Blanco, C., & Arranz, J. J. (2015). QTLs influencing somatic cell count in sheep: segregation analysis using Whole Genome Sequencing trio analysis. *XVI Jornadas sobre Producción Animal, 19 y 20 de mayo de 2015, Zaragoza, España. Tomo I & II*, 474-476.
62. Gutiérrez-Gil, B., Esteban-Blanco, C., Suarez-Vega, A., & Arranz, J. J. (2018). Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50K SNP chip and whole-genome sequencing. *Journal of dairy science*, *101*(10), 9072-9088.
63. Gutiérrez-Gil, B., Pérez, J., Álvarez, L., Martínez-Valladares, M., de la Fuente, L. F., Bayón, Y., Meana, A., Primitivo, F. S., Rojo-Vázquez, F. A. y Arranz, J. J. (2009b) "Quantitative trait loci for resistance to trichostrongylid infection in Spanish Churra sheep", *Genetics Selection Evolution*, 41(1), p. 46.
64. Gutiérrez-Gil, B., Pérez, J., De la Fuente, L. F., Meana, A., Martínez-Valladares, M., San Primitivo, F., ... & Arranz, J. J. (2010). Genetic parameters for resistance to trichostrongylid infection in dairy sheep. *Animal*, *4*(4), 505-512.
65. Hardison, R. C. (2003). Comparative genomics. *PLoS biology*, *1*(2), e58.
66. Hayes, B. E. N., & Goddard, M. E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*, *33*(3), 1-21.
67. Heaton, M. P., Harhay, G. P., Bennett, G. L., Stone, R. T., Grosse, W. M., Casas, E., ... & Laegreid, W. W. (2002). Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. *Mammalian genome*, *13*(5), 272-281.
68. Hill, WG. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 73-85.
69. Hill, WG. (2012). Quantitative genetics in the genomics era. *Current genomics*, *13*(3), 196-206.
70. Hu, Z. L., Park, C. A., Wu, X. L., & Reecy, J. M. (2013). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic acids research*, *41*(D1), D871-D879.
71. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., ... & Clamp, M. (2002). The Ensembl genome database project. *Nucleic acids research*, *30*(1), 38-41.

72. Kass, L., Erler, J. T., Dembo, M., & Weaver, V. M. (2007). Mammary epithelial cell: influence of extracellular matrix composition and organization during development and tumorigenesis. *The international journal of biochemistry & cell biology*, *39*(11), 1987-1994.
73. Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., ... & International Sheep Genomics Consortium. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS biology*, *10*(2), e1001258.
74. Labussière, J. (1988) "Review of physiological and anatomical factors influencing the milking ability of ewes and the organization of milking", *Livestock Production Science*, 18(3–4), pp. 253–274.
75. Labussière, J., Dotchewski, D., & Combaud, J. F. (1981). Caractéristiques morphologiques de la mamelle des brebis Lacaune. Méthodologie pour l'obtention des données Relations avec l'aptitude à la traite. In *Annales de Zootechnie* (Vol. 30, No. 2, pp. 115-136). EDP Sciences.
76. Lander, E. S., & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, *121*(1), 185-199.
77. Legarra, A., & Fernando, R. L. (2009). Linear models for joint association and linkage QTL mapping. *Genetics Selection Evolution*, *41*(1), 1-17.
78. Legarra, A., & Ugarte, E. (2005). Genetic parameters of udder traits, somatic cell score, and milk yield in Latxa sheep. *Journal of Dairy Science*, *88*(6), 2238-2245.
79. Legaz, E., Álvarez, I., Royo, L. J., Fernández, I., Gutiérrez, J. P., & Goyache, F. (2008). Genetic relationships between Spanish Assaf (Assaf. E) and Spanish native dairy sheep breeds. *Small Ruminant Research*, *80*(1-3), 39-44.
80. Li, R., Ma, Y., & Jiang, L. (2022). Research Progress of Dairy Sheep Milk Genes. *Agriculture*, *12*(2), 169.
81. Lin, J., Bao, Z. K., Zhang, Q., Hu, W. W., Yu, Q. H., & Yang, Q. (2015). Transcriptome analysis of the mammary gland from GH transgenic goats during involution. *Gene*, *565*(2), 228-234.
82. Liu, J., Zhang, L., Xu, L., Ren, H., Lu, J., Zhang, X., ... & Du, L. (2013). Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC genomics*, *14*(1), 1-11.
83. Liu, Y., Chang, Y., & Cai, Y. (2020). circTNFRSF21, a newly identified circular RNA promotes endometrial carcinoma pathogenesis through regulating miR-1227-MAPK13/ATF2 axis. *Aging (Albany NY)*, *12*(8), 6774.
84. Lucas, J. L., Pearson, R. E., Vinson, W. E., & Johnson, L. P. (1984). Experimental linear descriptive type classification. *Journal of Dairy Science*, *67*(8), 1767-1775.
85. Lucinda A., Mardis Elaine R., Wilson Richard K. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695-716.
86. Marete, A., Lund, M. S., Boichard, D., & Ramayo-Caldas, Y. (2018). A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. *PLoS One*, *13*(7), e0199931.

87. Marie-Etancelin, C., Rupp, R., Casu, S., Carta, A., & Barillet, F. (2001, August). New objectives of selection related to udder health, morphology and milkability in dairy sheep. In *52th Annual Meeting of the European Association for Animal Production, Budapest, Hungary* (pp. 26-8).
88. Marina, H., Gutiérrez-Gil, B., Sánchez-Mayor, M., Esteban-Blanco, C., Suárez-Vega, A., Garzón, A., & Arranz, J. J. (2019). Identification of variants related to the coagulation characters of cheese yield of sheep milk using a chip of SNPs designed from RNAseq data. *XVIII Jornadas sobre Producción Animal, Zaragoza, España, 7 y 8 de mayo de 2019*, 501-503.
89. Marina, H., Pelayo, R., Suárez-Vega, A., Gutiérrez-Gil, B., Esteban-Blanco, C., & Arranz, J. J. (2021). Genome-wide association studies (GWAS) and post-GWAS analyses for technological traits in Assaf and Churra dairy breeds. *Journal of Dairy Science*, *104*(11), 11850-11866.
90. Mateescu, R. G., & Thonney, M. L. (2010). Genetic mapping of quantitative trait loci for milk production in sheep. *Animal genetics*, *41*(5), 460-466.
91. Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., ... & Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PloS one*, *4*(4), e5350.
92. Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *genetics*, *157*(4), 1819-1829.
93. Miida, T., & Hirayama, S. (2009). Lipoproteins and their receptors in the central nervous system. *Rinsho byori. The Japanese Journal of Clinical Pathology*, *57*(1), 48-53.
94. Mikus, M., (1978) "Study of the mutual relationship between dimensions of the udder with regard to improvements of sheep for machine milking". En: *2nd International Symposium on mechanical milking of small ruminant*. INRAITOVIC, Alghero, Italy, pp. 102-112
95. Miller, W., Makova, K. D., Nekrutenko, A., & Hardison, R. C. (2004). Comparative genomics. *Annual review of genomics and human genetics*, *5*(1), 15-56.
96. Moioli, B., D'Andrea, M., & Pilla, F. J. S. R. R. (2007). Candidate genes affecting sheep and goat milk quality. *Small Ruminant Research*, *68*(1-2), 179-192.
97. Moritoh, Y., Oka, M., Yasuhara, Y., Hozumi, H., Iwachidow, K., Fuse, H., & Tozawa, R. (2016). Inositol hexakisphosphate kinase 3 regulates metabolism and lifespan in mice. *Scientific reports*, *6*(1), 1-13.
98. Muir, W. M., Wong, G. K., Zhang, Y., Wang, J., Groenen, M. A. M., Crooijmans, R. P. M. A., ... & Cheng, H. H. (2008). Review of the initial validation and characterization of a 3K chicken SNP array. *World's Poultry Science Journal*, *64*(2), 219-226.
99. Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J. C., Lee, J., ... & Reddy, T. B. K. (2021). Genomes OnLine Database (GOLD) v. 8: overview and updates. *Nucleic acids research*, *49*(D1), D723-D733.
100. Othmane, M. H., De La Fuente, L. F., Carriedo, J. A., & San Primitivo, F. (2002). Heritability and genetic correlations of test day milk yield and composition, individual laboratory cheese yield, and somatic cell count for dairy ewes. *Journal of dairy science*, *85*(10), 2692-2698.

101. Peng, W. F., Xu, S. S., Ren, X., Lv, F. H., Xie, X. L., Zhao, Y. X., ... & Li, M. H. (2017). A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (Ovis aries). *Animal genetics*, *48*(5), 570-579.
102. Pennisi, E. (2003). Tracing life's circuitry.
103. Pérez, J. A. M. (2016). *Estudio de la variabilidad genética de genes candidatos implicados en la producción láctea en el ganado ovino* (Doctoral dissertation, Universidad de León).
104. Piredda, G., Papoff, C. M., Sanna, S. R., & Campus, R. L. (1993). Influenza del genotipo della αs1-caseina ovina sulle caratteristiche chimico-fisiche e lattodinamografiche del latte. *Sci. Tecn. latt.-cas*, *44*(3), 135-143.
105. Prpić, Z., Mioč, B., Vnučec, I., Držaič, V., & Pavič, V. (2013). Non-genetic factors of udder morphology traits in Istrian ewes. *Mljekarstvo/Dairy*, *63*(2).
106. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, *81*(3), 559-575.
107. Raadsma, H. W., Jonas, E., McGill, D., Hobbs, M., Lam, M. K., & Thomson, P. C. (2009). Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. *Genetics Selection Evolution*, *41*(1), 1-15.
108. Rexroad, C., Vallet, J., Matukumalli, L. K., Reecy, J., Bickhart, D., Blackburn, H., ... & Wells, K. (2019). Genome to phenome: improving animal health, production, and well-being–a new USDA blueprint for animal genome research 2018–2027. *Frontiers in genetics*, *10*, 327.
109. Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., ... & Tosser-Klopp, G. (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 genetics*, *11*(12), e1005629.
110. Sagi, R. (1978). Udder support as a means for improving milk fractionation in dairy ewes. In *Annales de zootechnie* (Vol. 27, No. 3, pp. 347-353).
111. Sagi, R., & Morag, M. (1974). Udder conformation, milk yield and milk fractionation in the dairy ewe. In *Annales de zootechnie* (Vol. 23, No. 2, pp. 185-192).
112. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.
113. Schnabel, R. D., Sonstegard, T. S., Taylor, J. F., & Ashwell, M. S. (2005). Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Animal Genetics*, *36*(5), 408-416.
114. Scientific, C. (1990). Industrial Research Organization–CSIRO. *Feeding standards for Australian livestock-ruminants. Victoria: Australia Agricultural Council*.

115. Shi, H., Zhu, J., Luo, J., Cao, W., Shi, H., Yao, D., ... & Loor, J. J. (2015). Genes regulating lipid and protein metabolism are highly expressed in mammary gland of lactating dairy goats. *Functional & integrative genomics*, *15*(3), 309-321.
116. Skolnick, J., Fetrow, J. S., & Kolinski, A. (2000). Structural genomics and its importance for gene function analysis. *Nature biotechnology*, *18*(3), 283-287.
117. Stiner, M. C., Munro, N. D., Buitenhuis, H., Duru, G., & Özbaşaran, M. (2022). An endemic pathway to sheep and goat domestication at Aşıklı Höyük (Central Anatolia, Turkey). *Proceedings of the National Academy of Sciences*, *119*(4), e2110930119.
118. Suárez-Vega, A., Gutiérrez-Gil, B., Benavides, J., Perez, V., Tosser-Klopp, G., Klopp, C., ... & Arranz, J. J. (2015a). Combining GWAS and RNA-Seq approaches for detection of the causal mutation for hereditary junctional epidermolysis bullosa in sheep. *PLoS One*, *10*(5), e0126416.
119. Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Robert-Granie, C., Tosser-Klopp, G., & Arranz, J. J. (2015b). Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing. *Scientific reports*, *5*(1), 1-11.
120. Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Tosser-Klopp, G., & Arranz, J. J. (2016). Comprehensive RNA-Seq profiling to evaluate lactating sheep mammary gland transcriptome. *Scientific data*, *3*(1), 1-11.
121. Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Tosser-Klopp, G., & Arranz, J. J. (2017). Variant discovery in the sheep milk transcriptome using RNA sequencing. *BMC genomics*, *18*(1), 1-13.
122. Supratim Mukherjee, Dimitri Stamatis, Jon Bertsch, Galina Ovchinnikova, Jagadish Chandrabose Sundaramurthi, Janey Lee, Mahathi Kandimalla, I-Min A Chen, Nikos C Kyrpides and T B K Reddy. Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucl. Acids Res. (2020) doi: doi.org/10.1093/nar/gkaa983.*
123. Teng, J., Huang, S., Chen, Z., Gao, N., Ye, S., Diao, S., ... & Zhang, Z. (2020). Optimizing genomic prediction model given causal genes in a dairy cattle population. *Journal of dairy science*, *103*(11), 10299-10310.
124. Thompson, J. R., Lee, K. L., Freeman, A. E., & Johnson, L. P. (1983). Evaluation of a linearized type appraisal system for Holstein cattle. *Journal of Dairy Science*, *66*(2), 325-331.
125. Tsuchihashi, Z., & Dracopoli, N. C. (2002). Progress in high throughput SNP genotyping methods. *The pharmacogenomics journal*, *2*(2), 103-110.
126. VanRaden, P. M., & Wiggans, G. R. (1991). Derivation, calculation, and use of national animal model information. *Journal of Dairy Science*, *74*(8), 2737-2746.
127. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Kalush, F. (2001). The sequence of the human genome. *science*, *291*(5507), 1304-1351.
128. Vrdoljak, J., Prpić, Z., Samaržija, D., Vnučec, I., Konjačić, M., & Kelava Ugarković, N. (2020). Udder morphology, milk production and udder health in small ruminants. *Mljekarstvo: časopis za unaprjeđenje proizvodnje i prerade mlijeka*, *70*(2), 75-84.

129. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, *10*(1), 57-63.
130. Wickramasinghe, S., Cánovas, A., Rincón, G., & Medrano, J. F. (2014). RNA-sequencing: a tool to explore new frontiers in animal genetics. *Livestock Science*, *166*, 206-216.
131. Wickramasinghe, S., Rincon, G., Islas-Trejo, A., & Medrano, J. F. (2012). Transcriptional profiling of bovine milk using RNA sequencing. *BMC genomics*, *13*(1), 1-14.
132. Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76-82.
133. Zhao, Y., Pu, Y., Liang, B., Bai, T., Liu, Y., Jiang, L., & Ma, Y. (2022). A study using single-locus and multi-locus genome-wide association study to identify genes associated with teat number in Hu sheep. *Animal Genetics*, *53*(2), 203-211.

Appendix

List of web pages sources:

EMBL's European Bioinformatics Institute.
https://www.ebi.ac.uk/ – pristup 5.05.2022

The Genomes OnLine Database**.**
https://gold.jgi.doe.gov/ – pristup 15.05.2022

Research Organisation of Information and System. National Institute of genetics.
https://www.nig.ac.jp/nig/ – pristup 10.06.2022

Commonwealth Scientific and Industrial Research Organisation.
https://www.csiro.au/ – pristup 13.06.2022

Food and Agriculture Organisation of the united nations
https://www.fao.org/faostat/en/ – pristup 20.06.2022

Ministerio de Agricultura, Pesca y Alimentacion
https://www.mapa.gob.es/en/– pristup 11.04.2022

World Health Organisation.
https://www.who.int/– pristup 10.06.2022

Biography

Marko Vrcan, a student, was born on December 4, 1997, in Split. attended the Franciscan classical gymnasium with public rights in Sinj from 2013 to 2017. Upon completion of undergraduate studies in Animal Science, which lasted from 2017 to 2020, he received Bachelor's degree. In order to improve and gain knowledge about the practical part of the study, he applied for the Erasmus + program, and spent five months at the Faculty of Veterinary in Leon, Spain, at the department of animal production, under the leadership of the MEGA-ULE research group, as part of which this Master's thesis developed. He is currently finishing his graduate studies in Animal Genetics and Breeding at the Faculty of Agriculture in Zagreb. He knows English (understanding C1, speaking B2 and writing C1) and German (understanding B1, speaking A2 and writing A2). As for digital skills, he has outstanding knowledge of using Microsoft Office as well as software intended for population and quantitative genetic analyses (Grain, Magellan, Endog, CFC, Mega7, Arlequin, Pedig, etc.). Furthermore, he is very proficient in the integrated SAS software package for advanced analytics, business intelligence, data management, and predictive analytics, as well as in R studio programming. Also, during his stay in Spain, he gained knowledge about the use of the Linux operating system. The ability to think analytically, in combination with a passion for scientific work, he presents as the main weapon that society needs for quality progress. By nature, he is very enthusiastic and ambitious and is ready for all the adversities and challenges that await him in his scientific career. In his free time, he likes to listen to quality music, play various sports, and have constructive conversations with friends.