UNIVERSITÀ DEGLI STUDI DI SASSARI
UNIVERSITY OF SASSARI

SCUOLA DI DOTTORATO IN
SCIENZE VETERINARIE
PHD SCHOOL IN
VETERINARY SCIENCE

INDIRIZZO: **Riproduzione, Produzione e Benessere Animale (XXVIII CICLO)**
CURRICULUM: ***Animal Reproduction, Production and Welfare (28TH COURSE)***

ANALYZING THE GENETIC VARIATION OF DAIRY SARDA AND SPANISH MEAT OVINE BREEDS

Docenti Guida
Supervisors
DOTT. MICHELE PAZZOLA (Università degli Studi di Sassari)

PROF. MARCEL AMILLS ERAS (Universitat Autònoma de Barcelona)

Direttore
Director
PROF. SERGIO LEDDA

Tesi di dottorato della
PhD thesis by
DOTT.SSA ANTONIA NOCE

ANNO ACCADEMICO *ACADEMIC YEAR* 2014 – 2015

Iniziare un nuovo cammino spaventa,
ma dopo ogni passo ci rendiamo conto
di quanto fosse pericoloso rimanere fermi.

R. Benigni

Contents

Summary. Sheep are an important source of milk, meat and wool. In the last century, artificial selection schemes and production system management led to the subdivision of sheep in specific and well defined breeds. Nowadays, many of these populations have become extinct due to competition with more productive breeds and the progressive abandonment of rural activities.

The objectives of the current thesis was to investigate if variation at regulatory sites of the milk protein genes is associated with traits of economic interest in dairy Sarda sheep, as well as the genetic characterization of Spanish meat ovine breeds, by using high throughput genotyping and sequencing platforms.

Sarda sheep is a breed specialized in the production of milk (5% of the world ovine milk production) which is completely transformed into cheese. Our hypothesis was that variation at regulatory sites of the casein genes, which determine 80% of the total milk protein fraction, may have important effects on the composition and coagulation properties of milk. In this thesis, we have partially resequenced the promoters and the 3'UTR of the four casein genes in 25 Sarda sheep. In doing so, we have found 29 SNPs that, in combination with 3 previously reported SNPs mapping to the α-lactalbumin (*LALBA*) and β-lactoglobulin (*BLG*) genes, were genotyped with a multiplex TaqMan Open Array Real-Time PCR assay in 760 Sarda sheep with records for milk composition and coagulation properties. Our results revealed a high level of polymorphism in the promoter and 3' UTR regions of ovine casein genes in Sarda sheep (1 SNP every 200bp). Moreover, we found significant associations between *CSN1S2* and *CSN3* genotypes and milk protein and casein contents, while *CSN1S1* and *CSN2* genotypes were significantly associated with curd firming time.

Furthermore some SNPs co-localized with *in silico* predicted transcription factor and miRNA binding sites.

Population genetics analyses of eleven Spanish ovine breeds was performed by genotyping, with the Infinium 50K Ovine SNP BeadChip (Illumina), a total of 140 individuals belonging to the Xisqueta, Ripollesa, Segureña, Gallega, Roja Mallorquina and Canaria de Pelo breeds. This dataset was merged with 50K genotypes generated by the International Sheep Genomics Consortium for the other five Spanish breeds (Churra, Ojalada, Castellana, Latxa and Rasa Aragonesa). Multidimensional scaling and Admixture analyses of these 11 ovine Spanish breeds revealed that Canaria de Pelo, Roja Mallorquina, Latxa and Churra are clearly differentiated populations, while the remaining seven breeds are almost indistinguishable from each other. Moreover, we performed a second experiment aiming to estimate the amount of shared vs breed-specific variation in five Spanish breeds (Xisqueta, Ripollesa, Gallega, Roja Mallorquina and Canaria de Pelo) through a RNA-seq approach. In this way, we sequenced pools of 10 individuals per breed. Our results showed a similar amount of shared (~184,000 SNPs) vs pool-specific (~146,000 SNPs) variants. Moreover, the number of pool-specific single nucleotide polymorphisms (SNPs) in Canaria de Pelo (56,116 SNPs) and Xisqueta (45,826 SNPs) was 2-5 fold higher than in Gallega (9,384 SNPs), Ripollesa (15,817 SNPs) and Roja Mallorquina (19,143 SNPs). This finding suggest that a significant part of the variation of Spanish sheep populations might be breed-specific, thus justifying the implementation of conservation plans aimed to preserve these valuable genetic resources.

1. Introduction

1.1 THE CONCEPT OF BREED.

Domestic animals represent for humans an abundant source of food, fiber, and power which is nevertheless represented by a very small number of species. It has been calculated that there are approximately 20 completely and 25 partially domesticated species (Notter et al. 2013). On the other hand, genetic diversity within domesticated species is remarkably high. Livestock species evolved under the controlled breeding of humans for millennia but the invention of the concept of breed, a reproductively closed population with a well established phenotype, is relatively recent (Notter et al. 2013).

According to the Food and Agriculture Organization (FAO) of the United Nations (1999), one of the potential definitions of "breed" would be:

"Either a sub specific group of domestic livestock with definable and identifiable external characteristics that enable it to be separated by visual appraisal from other similarly defined groups within the same species or a group for which geographical and/or cultural separation from phenotypically similar groups has led to acceptance of its separate identity".

The concept of breed was developed in the 19th century in England, when livestock breeders began to control mating and selected the presumed superior individuals to improve local groups of animals. These activities were also supported by the establishment of herd/stud books which contained the pedigrees of animals assumed to be purebred (Notter et al. 2013). Since then, genealogical registries,

together with the establishment of breed standards, have proved to be important tools for selecting and monitoring domestic animals. The word "pedigree" derives from the old French term "*pié de grue*" which means "foot of a crane" (Leroy, 2011), as an upside down foot of a bird symbolizes a genealogical tree with progenitors and their descendants.

Contemporary breeds were created after many generations of selection and reproductive isolation. They represent the division of the genetic diversity within each domestic species, *i.e* they can be considered as distinctive genetic entities, and possess specific combinations of genes and associated phenotypic characteristics. Each breed, as a source of useful genes, is able to provide specific sets of genetically defined characteristics: *e.g.* the prolific Russian Romanov sheep are able to produce three to six lambs at a time. However livestock breeds are not fixed, immutable genetic entities because evolutionary forces as gene flow, drift and selection modify within-breed genetic diversity. Thus, all livestock breeds maintain an evolutionary potential and changes caused by both natural and artificial selection as well as by neutral processes are not rare (Notter et al. 2013).

Selective breeding is the basis of genetic improvement programs and individuals are chosen in order to have offspring with optimal production abilities. In order to achieve an effective selective response, there must be a sufficient level of genetic variation within populations and the selected trait should be heritable (Hill W.G. 2013). In the last decades, utilization of global specialized and highly productive breeds has been a crucial step in the development of modern animal farming. Highly productive cosmopolite breeds were used to improve animal

productions but such choice was not always positive for indigenous breeds, which were progressively replaced by their commercial counterparts and, in consequence, experienced sustained demographic declines (Notter et al. 2013).

As regards to local breeds, a worldwide catalogue can be obtained at the DAD-IS database (2015). DAD-IS is the acronym of the Domestic Animal Diversity Information System and it is hosted by the FAO. DAD-IS constitutes an information system describing the genetic diversity of animals and it is also a tool for the implementation of conservation strategies and the management of animal genetic resources. It provides information about the origin, census and main characteristics of thousands of breeds as well as management tools, a library of references and links and contacts of regional and national coordinators for the management of animal genetic resources. DAD-IS has a key role in the overall strategy of the FAO for the management of animal genetic resources, and serves as the virtual structure for achieving this goal.

Thousands of animal breeds with a worldwide distribution have been listed in the DAD-IS database. Many local breeds in Europe, and mainly in the cattle species, are at the verge of extinction or have an endangered status (Gandini et al. 2004; Zander et al. 2013). This loss of genetic resources has been attributed mainly to changes in production systems, the sustained replacement of local genotypes with improved breeds and the uncontrolled use of crossbreeding.

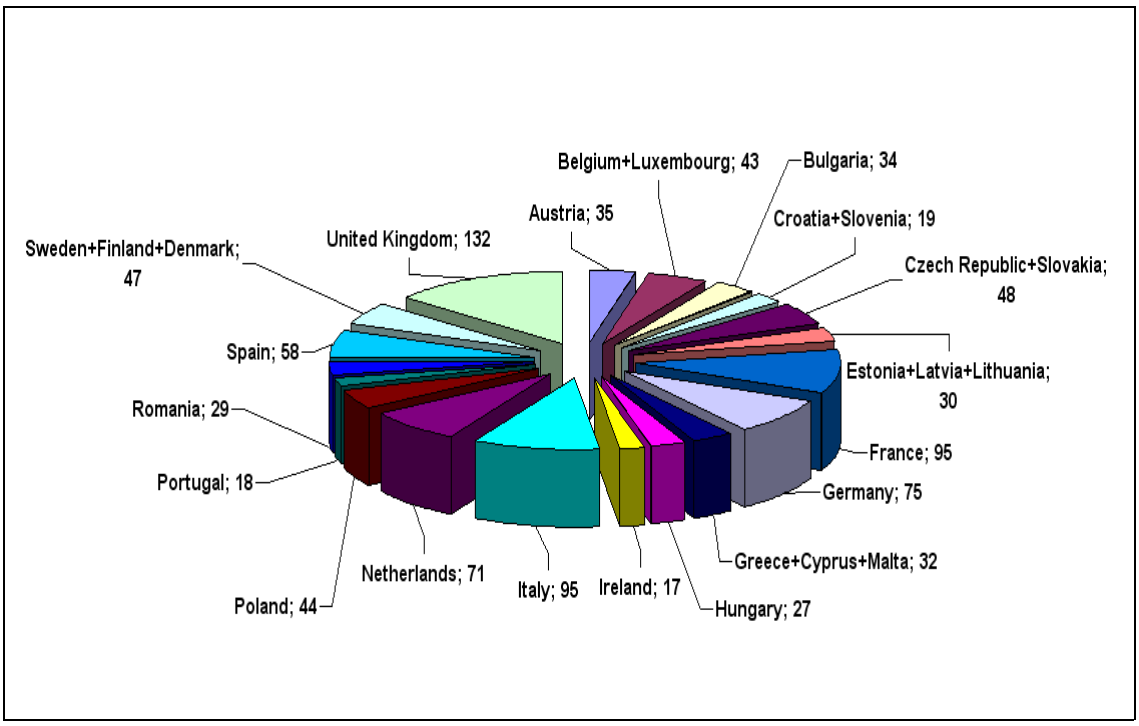


Figure 1. Distribution and number of sheep breeds in the individual or aggregate countries of the European Union (data from DAD-IS, 2015).

Besides, socio-economic factors, as the progressive abandonment of low income rural activities and changes in consumer preferences, have played a role in the demographic decline of local breeds (Rege and Gibson. 2003). Last but not least, farming systems have evolved towards intensive breeding conditions and the establishment of specialized highly productive breeds causing a gradual decrease of the value of local multi-purpose breeds (Zander et al. 2013).

One of the possible strategies for the recovery of local breeds is linked to the identification and relative enhancement of the so-called "typical" products with Protected Designation of Origin (PDO) and Protected Geographical Indication (PGI) labels, whose definition includes the geographic origin of the raw material and its

relationship with the livestock breeds, social and cultural traditions of the production area (Scintu and Piredda. 2007).

1.2 THE DIVERSITY OF SHEEP BREEDS AT A WORLDWIDE SCALE.

Since their domestication 8,000-9,000 years ago in the Fertile Crescent (Ryder M.L. 1964), sheep have become an important economic resource, particularly in developing countries (Meadows et al. 2005). In the last century, artificial selection has been particularly intense in order to produce genetically superior animals through the implementation of evaluation methods based on quantitative genetics and artificial insemination schemes. Management practices led to the subdivision of sheep in specific and well defined breeds (Kijas et al. 2012). Selection was mainly focused on a wide array of morphological and environmental adaptation traits as well as on wool, meat and milk production. This process involved the emergence of a wide range of phenotypic and genetic differences between breeds (Meadows et al. 2005).

Molecular genetics is able to provide important information about the relationships amongst breeds and their levels of genetic variation (Meadows et al. 2005). Several studies focused on mitochondrial diversity revealed the existence of two principal ovine haplotype groups: clade A, which is mainly constituted by breeds from Asia and New Zealand, and clade B made up of breeds from Europe, the Near East, and New Zealand (Wood and Phua. 1996; Hiendleder et al. 1998). Variation within autosomal microsatellites has been also used to make inferences about population history and examine the relationships amongst sheep breeds (Meadows et

al., 2005). Kijas et al. (2012) analyzed a total of 74 sheep breeds with a worldwide distribution using the Illumina Ovine 50K BeadChip array (around 52,000 single nucleotide polymorphism markers) in order to understand the genetic consequences of both domestication and selection. That comprehensive study confirmed the existence of high levels of gene flow and introgression amongst ovine breeds, as reported in previous mitochondrial studies (Wood and Phua. 1996; Hiendleder et al. 1998). In this way, Kijas et al. (2012) evidenced the existence of an extensive haplotype sharing amongst ovine breeds, an observation clearly consistent with abundant gene flow as well as with the short time of divergence of those populations. Kijas et al. (2012) also suggest a low level of conservation in linkage disequilibrium phase between breeds and a weak global population structure.

The aforementioned results, which have a high resolution because they are based on thousands of markers, indicate that, from a genetic point of view, the concept of "breed" is quite relative. Breeds are strongly admixed and, by this reason, they share a substantial amount of genetic variation. Besides, there is not a genetic threshold (*e.g.* an F_{ST} value) that can be used as a reference to decide if a population can be considered, or not, as a separate breed. In the absence of such reference, breeds are often defined on the basis of morphological or productive criteria. A critical question would be to ascertain which fraction of the total variation of sheep breeds is shared amongst two or more breeds and which part is breed-specific (*i.e.* those variants that can only be found in a single breed). This issue is very important because the genetic conservation of breeds has been mainly justified by the

conviction that they harbor alleles that are unique and that could have phenotypic consequences on traits of economic interest.

1.3 THE SARDA SHEEP AND ITS ECONOMIC IMPORTANCE.

According to the FAOSTAT (2013), the leading countries in milk production are Greece, with 705 thousands tonnes of ewe milk produced per year, Romania (630 thousand tonnes), Spain (600 thousand tonnes), Italy (380 thousand tonnes) and France (260 thousand tonnes).

Ewe milk production in Italy is mostly concentrated in Sardinia. Sardinia is an insular region with a total human population of 1.6 million, a total area of 24,000 km^2 and, in consequence, a very low population density: 69 vs. 201 people/km^2 in Italy. It is the second largest island of the Mediterranean Sea, after Sicily, and the climate is characterized by hot, windy and dry summers. Agriculture and livestock farming are relevant components of the Sardinian economy, representing 3.4% of the regional gross domestic product (CRENoS. 2012). Dairy sheep farming is the most representative sector of Sardinian livestock husbandry and plays a crucial role in maintaining the traditions of rural communities (Vagnoni et al. 2015). About 3.2 million sheep are reared in Sardinia (IZS. 2013), being mostly bred in dairy specialized farms. According to Vacca et al. (2008), 5% of the world total amount of ewe milk is produced in Sardinia. The majority of milk is transformed into cheese, being particularly important the three PDO labels Pecorino Romano, Pecorino Sardo and Fiore Sardo (EU 1996; EU 2009). Noteworthy, milk and dairy products are the leading sectors of the Italian food industry with a gross income of about 13.5 billion

euros (ISMEA, 2002). One million tonnes of cheese are produced on a yearly basis, being particularly relevant the two PDO cow milk hard-cheese Parmigiano Reggiano and Grana Padano (275,000 tonnes per year, Di Cagno and Gobetti. 2011). Pecorino Romano is the most famous and appreciated cheese produced from sheep milk with 33,000 tonnes per year (Di Cagno and Gobetti. 2011). It represents about 80% of sheep and goat cheese PDO labels in Italy (Pirisi et al. 2011) and its success is based on a well-established export market in the USA (Macciotta et al. 1999).

In Sardinia, milk is generally produced by the Sarda local breed, which is raised under semi-extensive to semi-intensive conditions (Carta et al. 2009). In Sardinian sheep farming (Carta et al. 2009; Pazzola et al. 2013; Pazzola et al. 2014), lactating ewes are pasture-fed and supplementation is based on commercial concentrates given during the milkings. Milking is performed by automatic manually-operated milking-machines twice a day, in the morning at 6:00 a.m. and in the evening at 16:00 p.m. Reproduction is based on natural mating but artificial insemination is being increasingly used by technicians and farmers. Lambing normally occurs in October-December for multiparous sheep and January-March for primiparous ewes. Ewes enter the dry period in summer (July-August) when pasture availability is scarce because vegetation is approaching desiccation. Lambs are milk fed by their dams, weaned at the age of one month and slaughtered as a PGI label meat product, the suckling lamb “Agnello di Sardegna” (EU. 2001). Almost the totality of sheep in Sardinia belong to the autochthonous Sarda breed (Figure 2). Sarda is the top-ranked sheep breed in Italy, with an estimated population of more than three million heads (Vacca et al. 2013). Moreover, a total of 350,000 animals

distributed in 2,600 farms are officially registered in the official flock book (Asso.Na.Pa., 2013).

Figure 2. Sarda ram and ewe (photographs by M. Pazzola).

Moioli et al. (2006) claimed that Sarda sheep has played, on a smaller scale, the same role as Holstein Friesian cows because it spread out in the Italian peninsula and gradually replaced many other low-yielding breeds. Indeed, the Sarda breed has been imported in many regions of central Italy, mainly in Lazio and Tuscany, and also in several Mediterranean countries, to be reared as a purebred or crossed to local sheep to improve milk production (Dettori et al. 2015).

The first flock book of the Sarda sheep was established at a local level in 1927 (Casu. 1971) but selection schemes were progressively implemented since the 1960s (Carta et al. 2009). The standard has been officially published at a national level by the Italian Ministry of Agriculture and Forestry in 1987. According to this standard (Asso.Na.Pa. 2013), Sarda sheep are defined by their medium size and white fleece; withers height is 71 and 63 cm and live weight 59 kg and 42 kg for rams and ewes, respectively.

Sarda is perfectly adapted to the typical Mediterranean environment and, in relation to its medium size, milk yield can be considered as high. Nevertheless, many attempts have been made in the past to further improve productions. Specialized non-autochthonous breeds, as the East Friesian, have been imported and crossbred with the Sarda but almost the totality of those attempts have been unsuccessful because the novel "synthetic" types were not able to cope with the extensive husbandry and the harsh climate of Sardinia (Sanna et al. 2001).

1.4 SPANISH MEAT OVINE BREEDS: ORIGINS, CONSERVATION STATUS AND PRODUCTIVE ABILITIES.

The evolution of the Spanish sheep census has been practically stable from 1993 to 2000, but it has experienced a significant decline since then (Figure 3). Spanish sheep population in 2013 comprised 16 millions of heads, so Spain can be considered the second leading country of the European Union after the United Kingdom, with about 33 million heads.

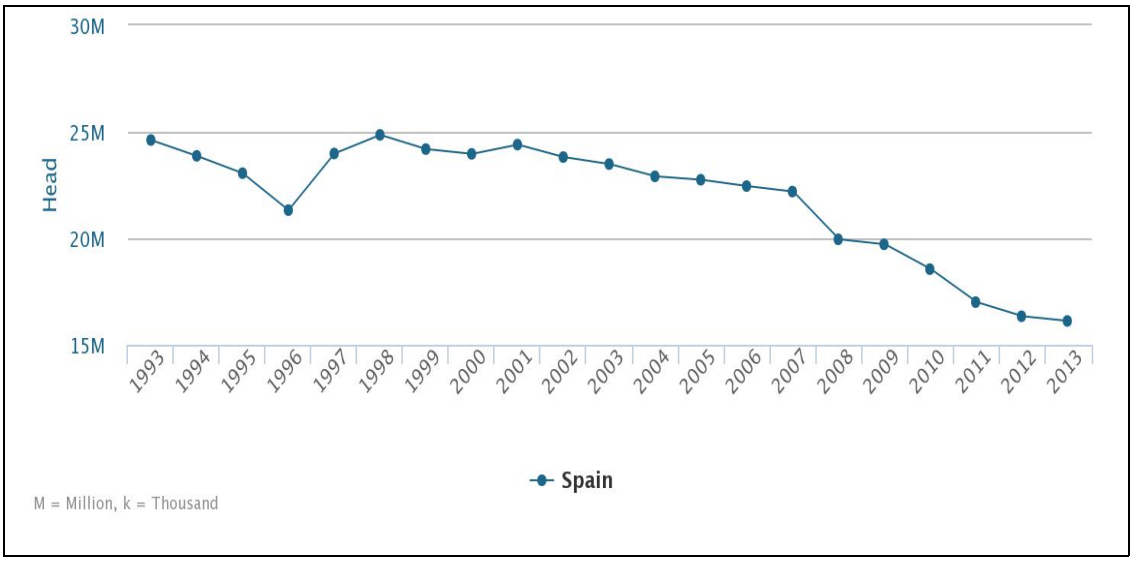


Figure 3. Number of heads of the sheep species in Spain in the period 1993-2013 (FAOSTAT, 2013).

Sheep farming is concentrated in certain regions of Spain, as Castilla y Leon, Castilla La Mancha and Extremadura (Figure 4). Because of competition and replacement with highly productive sheep from foreign countries, among the 43 officially recognized Spanish breeds, 33 of them have an endangered status.

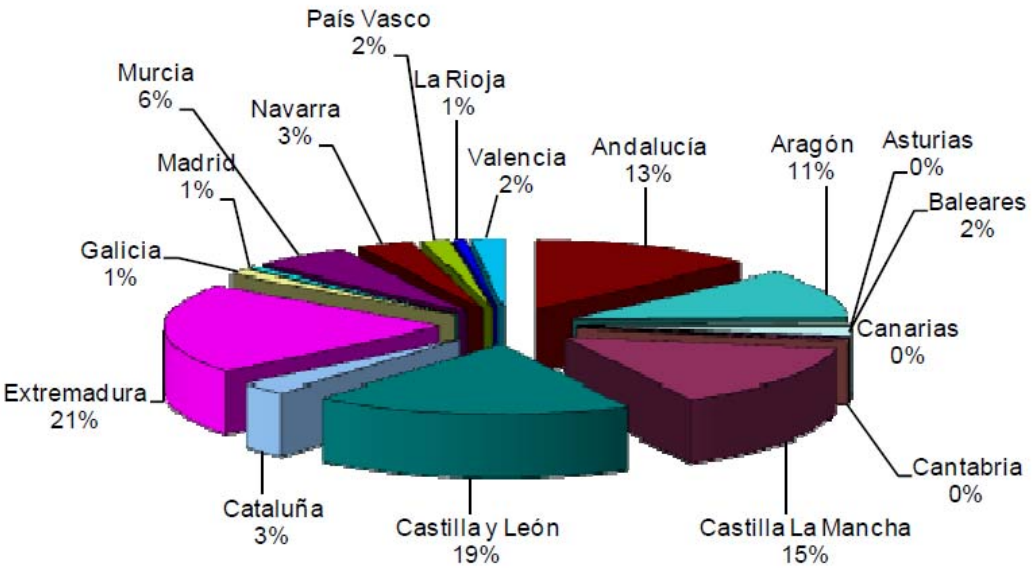


Figure 4. Distribution of the sheep census in Spain (Ministerio de Agricultura Alimentación y Medio Ambiente, Spain, 2013).

A sharp decrease in the production of ovine meat has taken place in the last decade (Figure 5), making evident the difficulties that this important breeding sector is going through nowadays.

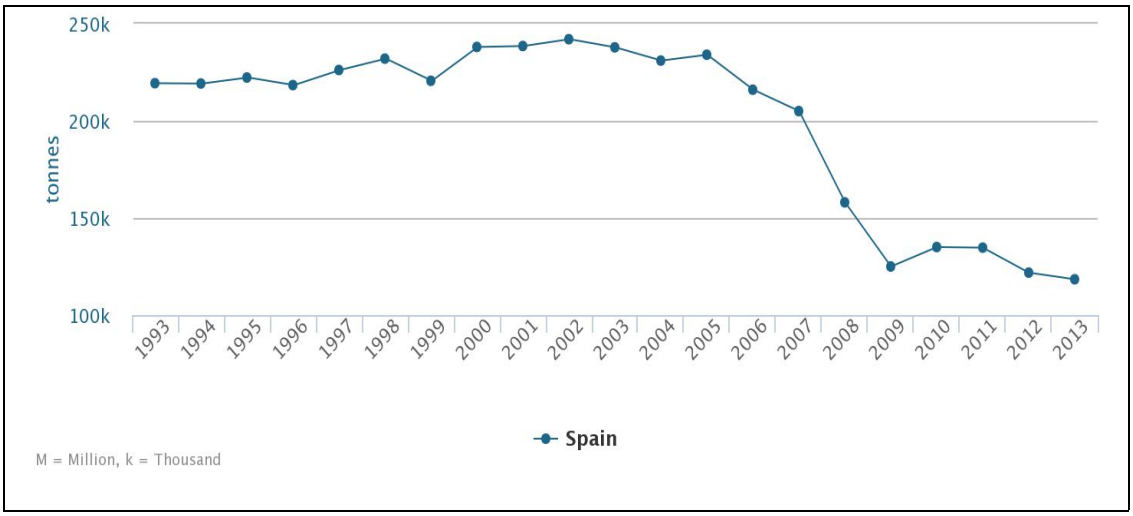


Figure 5. Production of sheep meat in Spain during the period 1993-2013 (FAOSTAT, 2013).

Spanish ovine breeds are characterized by a high genetic diversity and a weak to moderate population structure (Pedrosa et al. 2007; Arranz et al. 2001, Calvo et al. 2011).

In the present thesis five Spanish sheep breeds with an endangered status have been analysed in order to improve the available information about their levels of diversity and genetic relationships *i.e.* Ripollesa, Xisqueta, Canaria de Pelo, Roja Mallorquina and Gallega (Figure 6). A comprehensive description of these breeds can be found at the website of the Spanish Ministry of Agriculture, Food and Environment (Ministerio de Agricultura Alimentación y Medio Ambiente, Spain, 2013).

Ripollesa is a Catalan breed characterized by a convex profile, elongated proportions and varying size, ranging from medium to large. Coat has large spots, brown or white, on the head and limbs. It is a meat breed and was subjected to many

crossings with numerous Spanish and foreign breeds which threatened the maintenance of its genetic purity.

Xisqueta sheep also have a Catalan origin and they are uniformly white but display a black pigmentation at the tip of the ears, around the eyes, lips and distal part of the limbs. The original nucleus is located in the north of the Pallars Jussà region (Northwest Catalonia), specifically in the valleys of Manyanet and Vall Fosca. This breed spread throughout the upper Pyrenees central regions of Lleida and Huesca, but its area of influence has been gradually constrained to mountainous areas.

The Roja Mallorquina is a red coated sheep breed from the Majorca Island that emerged by crossing two distinct ovine lineages from Southern Europe and North Africa. This insular breed is well adapted to grazing, especially in a dry climate.

Canaria de Pelo is an aboriginal hair sheep from the Canarian archipelago (specially in Tenerife), with a large body size and prone to accumulate adipose tissue. The coat can be dark or reddish cherry colored. Canaria de Pelo sheep were exported to the Americas after their discovery by Cristopher Columbus.

Finally, Gallega sheep are located in Galicia and they are medium sized, white-coated and horned sheep with a straight profile.

Ripollesa

Xisqueta

Roja Mallorquina

Canaria de Pelo

Gallega

Figure 6. The five Spanish sheep breeds analysed in the present thesis (pictures from Spanish Ministry of Agriculture, Food and Environment (Ministerio de Agricultura Alimentación y Medio Ambiente, Spain 2015).

1.5 THE GENETICS OF MILK TRAITS IN SHEEP.

Sheep milk production is, in quantitative terms, less relevant than that of cattle (2% of the total milk production in the world) but it has a significant economic impact in some geographic areas, as in the Middle East and Mediterranean countries (Ramos and Juarez. 2011). Ewe milk is mainly devoted to cheese-making and ovine breeds are strongly associated with certain territories and cultural conceptions (Boyazoglu and Morand-Fehr. 2011). In the European Union, sheep milk production is continuously increasing and in 2013 the total yield was 2,8 million tonnes (Figure 7).

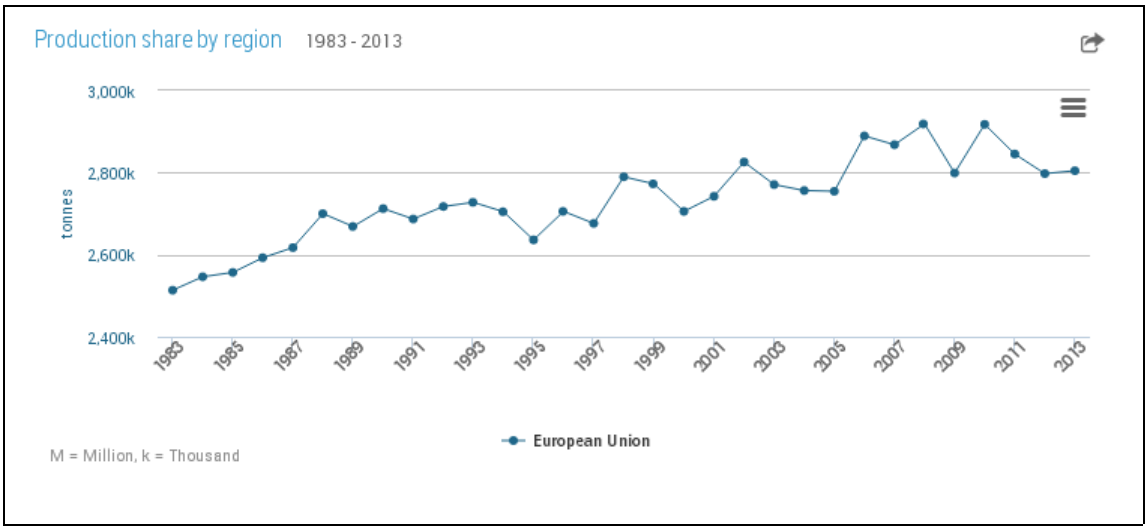


Figure 7. Production of ewe milk in the European Union in the period 1983-2013 (FAOSTAT, 2013).

Regardless of the productive specialization of sheep breeds, milk represents from 65 to 75% of the whole farming income and, as a consequence, it is one of the main targets of breeding strategies aimed to increase profits (Barillet et al. 2005). Selective pressure has been focused on improving not only the "classic" traits as milk yield and composition, but also the "novel" ones as udder morphology and natural

resistance to mastitis (Barillet. 2007). In the past decades, several studies mapped quantitative trait loci (QTL) and enriched our knowledge about the biology and genetic architecture of ovine complex traits (Raadsma et al. 2009). A major objective of QTL studies is to identify loci that can be used in the context of marker assisted selection (Khatkar et al. 2004). In the last five years the development of a high throughput genotyping platform provided the opportunity to explore the genome-wide variability of livestock species and identify regions influencing traits of economic interest. Genome-Wide Association (GWA) studies, which investigate the relationships amongst the genotypes of thousands of single polymorphism nucleotide (SNP) markers and production traits, are currently replacing traditional QTL linkage mapping analyses as an approach to identify causal mutations (Becker et al. 2010; García-Gámez et al. 2011; Zhao et al. 2011).

The first GWA study for milk production traits in sheep was performed by García-Gámez et al. (2012). They genotyped, using the Illumina 50K BeadChip, a total of 1,696 Churra sheep with records for milk yield (MY), protein percentage (PP), fat percentage (FP), protein yield (PY) and fat yield (FY). Significant trait-associated regions were detected at ovine chromosomes (OAR) OAR2, OAR3, OAR6 and OAR20 (Figure 8).

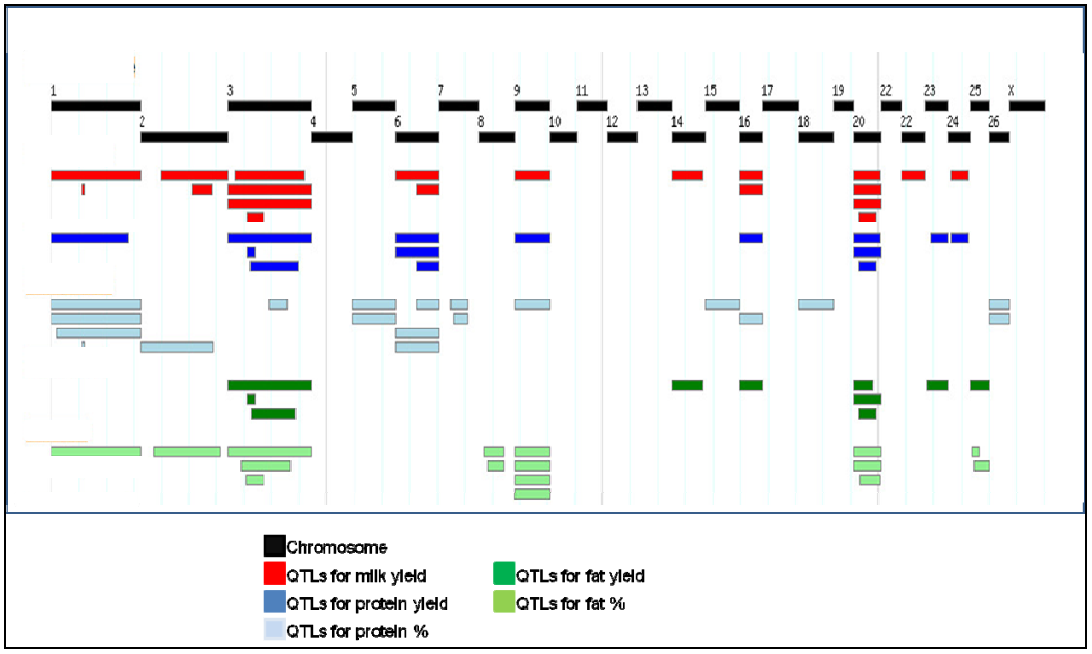


Figure 8. Meta-assembly of ovine QTL showing the genome location of QTLs for the main milk traits (adapted from Raadsma et al., 2009).

In OAR2, three overlapping QTL effects were identified, at approximately 55 cM, for MY, PY, and FY. Those results were confirmed in a subsequent study performed by the same authors using linkage disequilibrium and linkage analysis (LDLA) genome scans methods (García-Gámez et al. 2013)

The most significant associations with dairy traits identified in OAR2 co-localize with the Insulin-like growth factor binding protein-like 1 (IGFBPL1) gene (Gonda et al. 2007). IGFBPL has a domain homologous to the insulin-like growth factor binding proteins (IGFBPs) that can be bound by insulin like growth factors (IGFs). Interestingly, IGFs have been proposed to be associated with milk performance in cattle (Mullen et al. 2011). Another interesting QTL, found in OAR 20 at 23.7–29.6 Mb, includes the major histocompatibility complex class-II genes

(MHC-II) and it is associated with PP, FP and FY, as already reported in other previous studies (Gutiérrez-Gil et al. 2009; Barillet et al. 2005).

Garcia-Gamez et al. (2012) identified the marker OAR3_147028849, located in the third intron of, the α-lactalbumin (*LALBA*) gene which maps to OAR3, as the SNP with the most highly significant association detected for PP and FP. Alpha-lactalbumin is a major whey protein that forms a subunit of the lactose synthase binary complex. For this reason it plays a critical role in the control of milk synthesis and secretion of milk. The putative causal mutation is a Val27Ala substitution that produces a reduction in the activity of the lactose synthase enzyme. There are no other available studies about the influence of the ovine *LALBA* polymorphisms on milk traits but it has been reported that in *LALBA*-deficient mice milk protein and fat concentrations suffer dramatic changes (Stinnakre et al. 1994).

Another candidate gene located in the ovine chromosome 3 and widely studied because its effect on milk composition (Feligini et al. 1998) is β-lactoglobulin (*BLG*). Several studies have shown that *the BLG* gene has three allelic variants, A, B and C. Significant associations were found between the A and B variants, present in many breeds, with milk production (Nudda et al. 2000) as well as with fat and protein contents (Dario et al. 2008). With regard to the C variant, it is a rare allele that only segregates in the Merino breed (Barillet et al. 2005).

García-Gamez et al. (2012) have also reported chromosome-wise significant associations with PP in the OAR6 region that contains the casein cluster (DU430803_572 SNP, 85.44 Mb). Caseins are the most important milk protein fraction and they are classified into four types (Ginger and Grigor. 1999).: α_{s1}-casein

(*CSN1S1*), α_{s2}-casein (*CSN1S2*), β-casein (*CSN2*) and κ-casein (*CSN3*). Milk protein polymorphisms have been considered as a potential source of information for the selection of dairy ruminants (Barillet et al. 2005). For instance, the *CSN1S1* genotype is used as selection criteria in French goats because of its association with milk composition and cheese yield.

Ewe milk is a product with high fat and protein contents, and it is mainly used in the elaboration of commercial and hand-made cheeses and yogurts (Fuertes et al. 1998). The protein content and composition of milk are important to the cheese manufacturers because they are major factors in determining the yield and quality of the final product (Ng-Kwai-Hang et al. 1981; Hurtaud et al. 1993). Furthermore, caseins are the main part of the curd, whereas the serum protein remains in the whey fraction.

In general, heritabilities of sheep dairy traits are moderate indicating that a response to selection can be expected. Heritability values for milk, protein and fat yields are 0.27-0.32, 0.22-0.27 and 0.23-0.29, respectively (Barillet and Boichard. 1987). In contrast, heritability values for milk composition traits are considerably high: 0.49-0.62 and 0.47-0.53 for fat and protein percentages, respectively (Barillet and Boichard, 1987). However, Casu et al. (1975), in a study focused on the Sarda breed, reported that heritability is higher for fat than for protein content. Othmane et al. (2002), using a dataset of more than 7,000 records measured in Spanish Churra sheep, showed that heritability and repeatability for protein and casein contents were similar and relatively high (0.23 and 0.21, respectively) whereas for fat content they were quite low (0.06-0.08). As a whole, those results provide a scientific rationale for

the use of protein and casein contents as selection criteria in dairy sheep. In addition, selection for protein content is recommended over selection for casein content given their high genetic correlation (0.99) and that the former is cheaper to measure.

1.6 GENETICS OF CARCASS AND MEAT QUALITY TRAITS.

Henchion et al. (2014) reported that the total world consumption of ovine meat, calculated as an aggregate of ovine and goat meat, was around 12 millions tonnes in 2014. That amount is much lower than that recorded for cattle (63 millions tonnes), pig (105 millions) and poultry (90 millions) but the overall trend for ovine meat in the period 1990-2009 evidenced an increasing percentage of total (40%), and *per capita* consumption (12%).

Farming systems for producing ovine meat are essentially based on the production of mutton and lambs. Consumer preferences for a specific carcass weight differ amongst countries. The mean weights of carcasses are 8 kg in Portugal (Santos et al. 2007), 9 kg in Italy (Cifuni et al. 2000), 11 kg in Spain (Ripoll et al. 2008) and from 16 to 23 kg in Northern Europe (Beriain et al. 2000). Carcasses are normally paid on the basis of weight but they are also evaluated and scored for quality and composition attributes. Tenderness, cut size, fat cover, marbling, meat and fat color are evaluated to define carcass quality (Harrington and Kempster. 1989) and meet consumer's demands (Stanford et al. 1998).

All factors linked to meat production and carcass quality can be modified by selection as well as by changing the processing environment (Hopkins and Mortimer. 2014). In particular, genetics evaluation methods can predict breeding values in the

context of selection schemes. A summary of genes with effects on sheep meat quality and muscling has been reported by Tellam et al. (2012).

A recent report has summarized the heritabilities of the main traits associated with meat quality by showing that, in general, they fluctuate from high (shear force and intramuscular fat, 0.44-0.48) to low (tenderness and color, 0.14-0.15) values (Hopkins and Mortimer. 2014). Information about QTL associated with sheep growth and meat production traits is relatively scarce. Currently 801 QTLs are annotated in the sheep QTL database (www.animalgenome.org/QTLdb), but only a small fraction affects meat production (Hu et al. 2013). OAR1, OAR2, OAR3 are the chromosomes containing the highest number of trait associated regions (72, 142, 67 respectively). Recently Zhang et al. (2013) performed a GWAS study for 11 growth and meat production traits using a random design in a sheep population. Interestingly, three significant SNPs were located on OAR2 close to RPL7 (Ribosomal Protein L7), SMC2 (Structural Maintenance of Chr2) and MBD5 (Methyl-CpG-Binding Domain) genes, that are associated with muscle development and body weight in different species (Du et al. 2012).

Furthermore, another GWAS study performed in Australian Merino sheep stressed the importance of OAR6 as a major QTL region for body weight (BW). In this way, 13 SNPs mapping to the OAR6 36.15-38.56 Mb region showed a strong association with this trait (Al-Mamun et al. 2015). The NCAPG and LCORL genes are highly conserved among mammalian species and they have been associated with multiple body size traits, suggesting a common evolutionary pathway across diverse mammalian species. A full-genome scan in sheep detected QTL with small to

moderate effects on body composition and body weight on chromosomes 1, 6, 7, 9, 10, 14, 16 and 23 (Cavanagh et al. 2010) and confirmed others already detected in previous studies on chromosomes 2, 11 ,4, 18 and 20 (Raadsma et al. 2009; Karamichou et al. 2005 and 2006; Walling et al. 2001 and 2004). Among the major genes associated with sheep growth and meat production traits, myostatin (*MSTN*) is one of the most characterized ones. Several QTL are located close to *MSTN* on OAR2 (Marcq at al. 2002). Clop at al. (2006) performed a QTL scan in a Romanov × Texel F_2 population to investigate muscle hypertrophy of Texel sheep. They showed that a g+6723G-A transition in the 3' UTR region of the ovine myostatin gene (*GDF8*) creates an illegitimate target site for miRNA1 and miRNA206, that are strongly expressed in the skeletal muscle, causing a down regulation in myostatin levels and contributing to muscular hypertrophy.

Other major genes loci are Callipyge (*CLPG*), which maps to OAR18, and the rib-eye muscling (REM) locus at OAR18. The *CLPG* locus influences muscularity, a positive attribute for meat production, but its effect is often accompanied by a significant decrease in meat tenderness (Fahrenkrug et al. 2000). In contrast, REM is responsible for the hypermuscled phenotype known as Carwell that does not have detrimental effects on meat tenderness and intramuscular fat deposition (Walling et al. 2001; 2002).

1.7 ANALYZING THE DIVERSITY OF SHEEP BY GENOME SEQUENCING AND RNA-SEQ.

One fundamental goal of genetics is to identify the causal mutations that determine phenotypes. The development of next generation sequencing methods (NGS) has provided new opportunities to achieve this ambitious endeavour (Pareek et al. 2011). The basic principle of NGS involves the massively parallel sequencing of DNA molecules attached to a solid surface (Mardis 2008a, 2008b; Metzker, 2010). The recent introduction of instruments capable of producing millions of DNA sequence reads in a single run is rapidly changing the landscape of genetics, providing the ability to answer biological questions in a shorter time (Mardis. 2008, De Wit et al. 2012).

Next-generation DNA sequencing platforms such as Roche/454, Illumina and the Applied Biosystems SOLiD generate reads with different lengths. Each of these methods has unique sequencing chemistries (Figure 9). Roche/454 and SOLiD use beads as library binding surface, while the Illumina platform employs a flow cell, where DNA molecules flanked by adapters are amplified by bridge PCR prior to sequencing (Knief. 2014). Depending on the desired read length, Illumina sequencing provides highly accurate sequencing with a low error rate, even within repetitive sequence regions (Wickramasinghe et al. 2014). Sequencing with Roche/454 is based on the pyrosequencing technology, where the release of one pyrophosphate molecule during nucleotide incorporation triggers the cleavage of oxyluciferin by luciferase and the emmision of light. SOLiD sequencing, in contrast, is based on an oligo ligation and detection technology catalized by a DNA ligase.

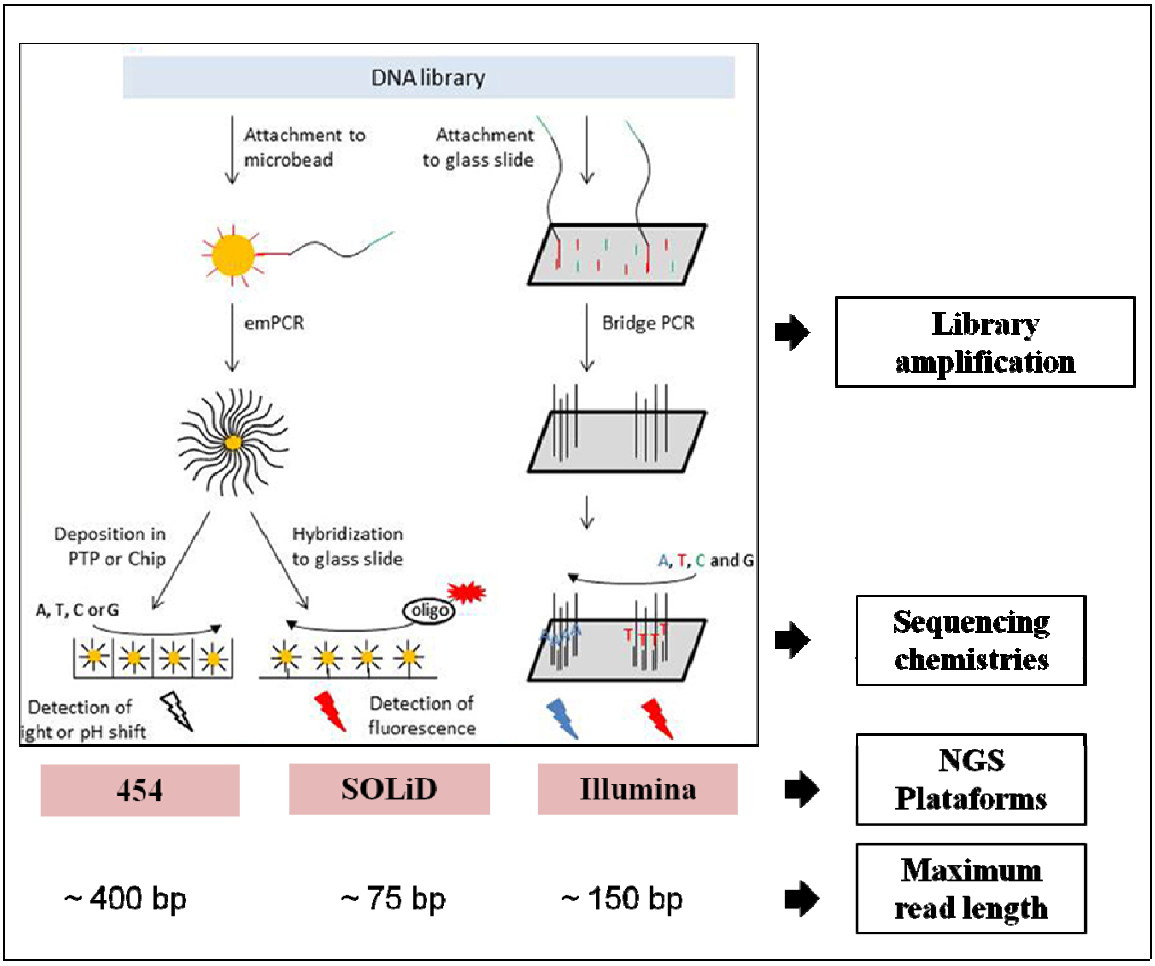


Figure 9. Sequencing process of the most commonly used NGS platforms. (adapted from Knief, 2014).

Exome sequencing is a technique that involves the sequencing of all the protein-coding regions of a genome. A related approach, called RNA-Seq, consists on the sequencing of the complete set of cellular RNA molecules (the transcriptome), thus providing an overall quantitative and qualitative picture of gene expression (Wickramasinghe et al. 2014). Compared with microarrays, RNA-Seq is a very powerful technology because it has a low background noise, high sensitivity and requires less RNA sample. In addition, RNA-Seq could capture almost all of the

expressed transcripts (microarrays can only detect those that are probed). RNA-Seq can be applied to various research purposes, such as quantifying the expression of transcripts (Mortazavi et al., 2008), detecting gene fusions (Maher et al. 2009; Pflueger et al. 2011), inferring alternative splicing (Gan et al. 2010; Sultan et al. 2008) and identifying single nucleotide polymorphisms (SNPs) in transcribed regions (Chepelev et al. 2009). The main consideration when designing RNA-Seq experiments is tissue sampling because gene expression varies across space (body location) and time. As shown in Figure 10, the basic steps of a RNA-seq experiment are the following. 1) The RNA extracted, total or fractionated, is converted into a cDNA library with adaptors ligated to one or both ends of each molecule, and 2) Each fragment, with or without amplification, is sequenced in a high-throughput platform (Roche/454, Illumina, Applied Biosystems SOLiD) to obtain short sequences by single or paired-end sequencing (Wang et al. 2009).

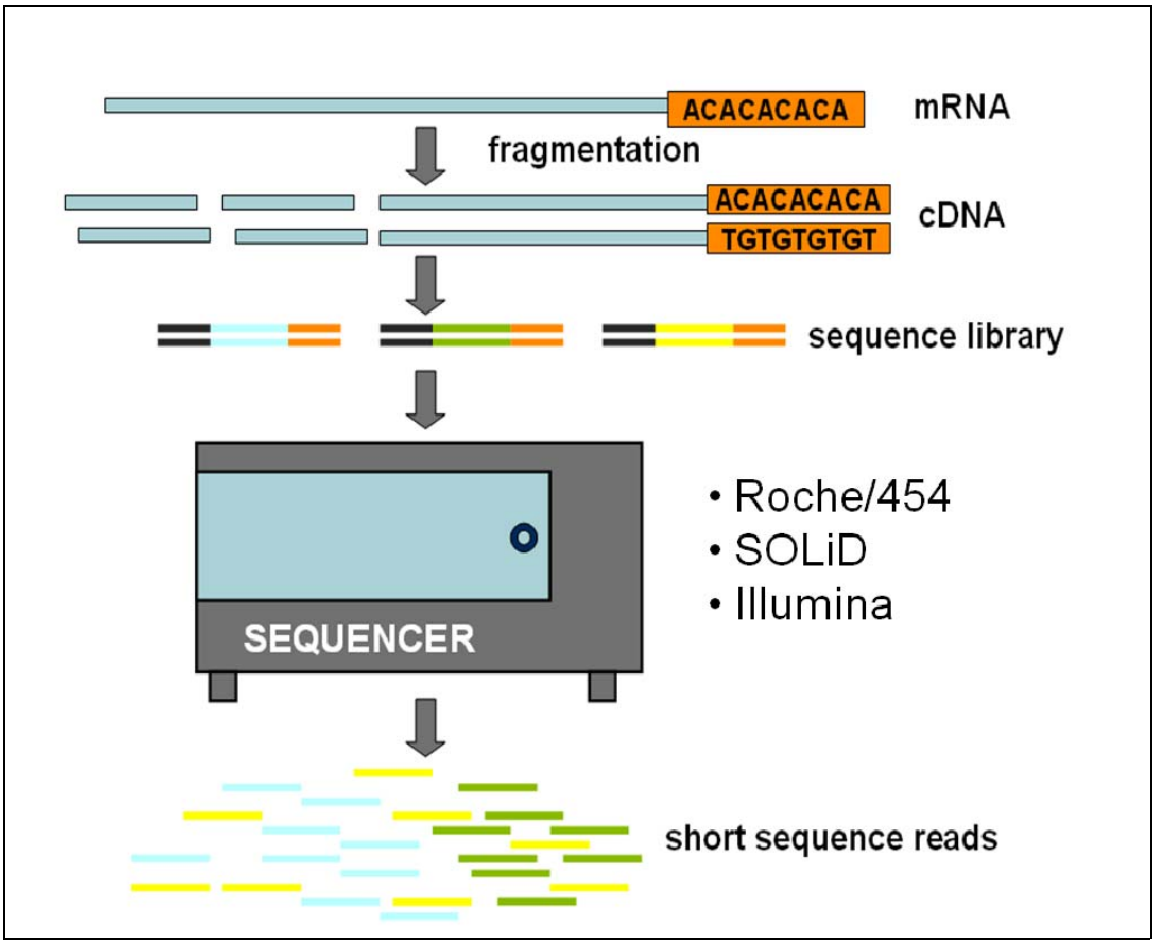


Figure 10. Scheme of a RNA-Seq experiment (adapted from Wang et al. 2009)

The large amount of data generated by RNA-Seq must be analyzed with robust and efficient algorithms implemented in dedicated bioinformatic software tools. They generally carry out read alignment and transcript assembly, and quantification tasks (Trapnell et al. 2012). A typical pipeline of RNA-Seq data analysis is shown in Figure 11. The first step is the quality control of the raw reads obtained by sequencing. With this procedure, sequence quality, sequencing depth, reads duplication rates (clonal reads), alignment quality, nucleotide composition bias, PCR bias, GC bias, coverage uniformity and other relevant parameters are properly

assessed. Tools available for data quality control are FastQC Galaxy, Fastx-Toolkit and Shortread (Trapnell et al. 2012)

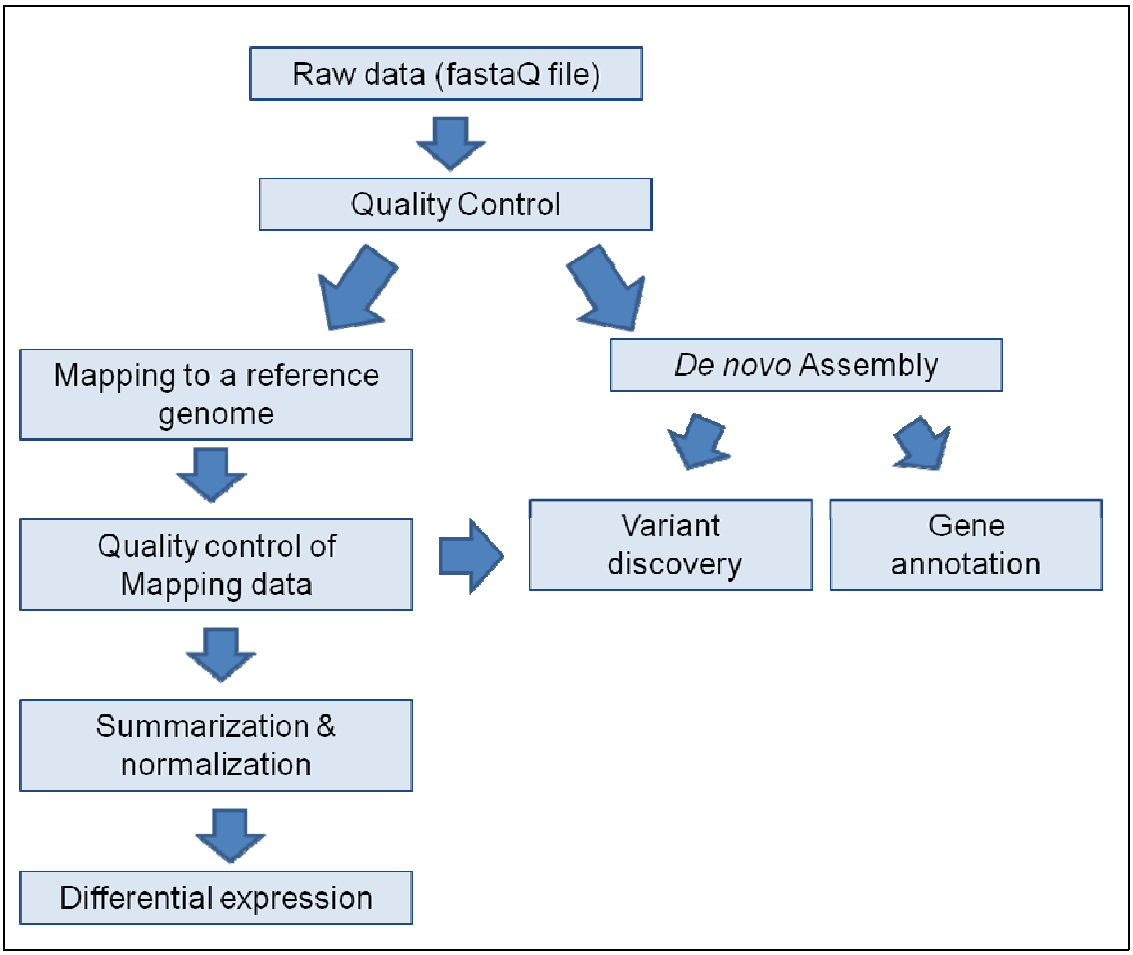


Figure 11. General workflow to perform RNA-seq data analysis.

The second step is the alignment of reads to unique locations on a reference genome. Several softwares are available to perform this task e.g. Bowtie, BWA and Maq and CASAVA (Trapnell and Salzberg. 2009). The complexity of the genome has a direct effect on the efficiency and accuracy of sequence mapping (Wang et al. 2008). The resulting aligned reads are then aggregated and summarized to obtain

quantification values for downstream analyses (Oshlack et al. 2010) using the software Cufflinks or ALEXA-seq (Trapnell and Salzberg. 2009) .

Another important step in the analysis of RNA-Seq data is length normalization. The amount of a given transcript can be expressed as Reads per kilobase transcript per million reads (RPKM), which reflects the molar concentration of a transcript in the starting sample normalized for RNA length and the total number of reads (Mortazavi et al. 2008). Another scaling normalization method relies on the concept of quantile and scaling by trimmed means of M-values normalization (TMM), which uses the raw data to estimate appropriate scaling factors that can be used in downstream statistical analysis procedures, thus accounting for the sampling properties of RNA-seq data (Cloonan et al. 2008). After normalization, RNA-seq data can be utilized for differential expression analysis using programs such as DESeq (Anders and Huber. 2010), EdgeR (Robinson et al. 2010) and NBPSeq (Di et al. 2011) that are based on a negative binomial distribution, or DEGseq and Myrna that are rely on Poisson generalize linear model distribution (Oshlack et al. 2010).

Recently, the use of RNA-seq has been extended to the discovery of SNPs (Canovas et al. 2010, Djari et al. 2013, De Wit et al. 2012). SNP discovery by RNA-seq requires specific filters to prevent false positives. Often, mapped reads do not show a perfect match with the reference genome. Each mismatch needs to be checked in order to evaluate if it is an artifact or a real polymorphism. Besides, when about half of the same reads include a specific mismatch it means that the analyzed genotype may correspond to a heterozygote individual.

Examples of softwares used for SNP discovery are SAMtools (Li et al. 2009), Picard and GATK (McKenna et al. 2010; DePristo et al. 2011). Moreover, commercial software, as CLC Bio, is also available for RNA-seq analysis providing a complete workflow for all kind of analysis (Wickramasinghe et al. 2014). When a reference genome is not available, it needs to be assembled "*de novo*" by searching for overlaps between short reads and thereby building longer and longer "contigs" (De Wit et al. 2012).

Sequencing of the human genome took 15 years until completion, but current NGS methods can do the job in a few days. The reduced costs of sequencing of individual genomes paves the way for carrying out phylogenetic studies, gene annotations and identification of rearrangements or expansions, SNPs discovery, genetic population analysis and dissecting the genetic basis of phenotypic traits (Miller et al. 2015). Nowadays, whole genome sequencing and annotation of different mammalian genomes have been completed and updated animal genome assemblies can be accessed at the Ensembl server (Chakravarti S. 2015). An international consortium completed the first mapping, sequencing, assembly and annotation of the sheep genome after eight years of hard work (Sheephapmap. 2015). This project started with a first sheep virtual genome which was assembled into 1,172 sheep bacterial artificial chromosome comparative genome contigs which covered 91.2% of the human genome (Dalrymple et al. 2007) and a cytogenetic map of 566 loci (Goldammer et al. 2009). The interim assembly version OAR v2.0 for sheep was released in 2010 and it was built using Illumina GA sequence from one female and one male Texel individuals (International Sheep Genomics Consortium et

al. 2010) being subsequently improved (OAR v3.1 release) in the second half of 2011. The final sheep genome assembly, OAR v3.1, has a contig N50 length of ~40 kb and a total assembled length of 2.61 Gb, with ~99% anchored onto the 26 autosomes and the X chromosome. Next generation Sequencing has been also used in sheep for developing SNP arrays. The 1.5K pilot sheep SNP array project used two subsets of the sheep genome (i.e. BAC end sequences and expressed sequence tags) for resequencing purposes. The project resulted in 49,077 usable sequence with around 1 SNP being detected for every ~250 bp of sequence. About 50% of these SNPs had the ability to be multiplexed into a SNP chip. In 2008 the pilot chip was then tested on 413 sheep from different breeds using Illumina technology and it found a core set of 384 SNP which identify population substructure (Kijas et al. 2009). In a subsequent study, Kijas et al. (2012) optimized a panel of around 50,000 SNPs, that are currently included in the Illumina ovine 50K BeadSNP chip, that were genotyped in 2,819 individuals representing 74 breeds. This study evidenced the existence of a substantial gene flow amongst ovine breeds and the absence of a clear relationship between the amount of genetic variation and distance to the domestication center.

2. Objectives

Research presented in the current thesis was performed through a collaboration between the University of Sassari, Italy, and the Center of Research in Agricultural Genomics, Spain, (project RZ2011-00015-C03-01).

The basic goals of this research were:

1. To characterize the variation of regulatory regions of the ovine casein genes in Sarda sheep, one of the most important dairy breeds in Italy. We also aimed to carry out association analyses, for 32 selected polymorphisms, with a wide array of milk traits recorded in Sarda sheep.

2. To characterize the diversity of five endangered Spanish breeds (Roja Mallorquina, Gallega, Xisqueta, Ripollesa and Canaria de Pelo) by using a high throughput genotyping platform and compare it with that of other previously characterized Spanish breeds. We also wanted to investigate the amount of shared vs breed-specific variation in the five populations mentioned above by using a RNA-seq approach.

Two experiments were designed and the related research papers were prepared. These are original and unpublished, and they are going to be submitted to peer-reviewed scientific journals.

3. Research papers

RUNNING HEAD: SNPs at the regulatory regions of ovine casein genes

Variations at regulatory regions of the milk protein genes are associated with milk traits and coagulation properties in the Sarda sheep

A. Noce*†, M. Pazzola*, M. L. Dettori*, M. Amills†, A. Castellò†, A. Cecchinato‡, G. Bittante‡, G. M. Vacca*

*Department of Veterinary Medicine, University of Sassari, Via Vienna 2, 07100 Sassari, Italy.

†Department of Animal Genetics, Center for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB), Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

‡Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE) University of Padova, viale dell'Università 16 - 35020 Legnaro (PD), Italy

Address for correspondence

M. Pazzola, Dipartimento di Medicina Veterinaria, Università degli Studi di Sassari, via Vienna, 2 - 07100 Sassari, Italia.

E-mail: pazzola@uniss.it

SUMMARY

Regulatory variation at the ovine casein genes could have important effects on the composition and coagulation properties of milk. Herewith, we have partially resequenced the promoters and the 3'UTR of the four casein genes in 25 Sarda sheep. Alignment of these sequences allowed us to identify a total of 29 SNPs. These levels of polymorphism (one SNP every 200 bp) are remarkably high if compared with SNP densities estimated in human genic regions (~1 SNPs per kb). The 29 SNPs identified in our resequencing experiment, plus 3 previously reported SNPs mapping to the α-lactalbumin (*LALBA*) and β-lactoglobulin (*BLG*) genes, were genotyped with a multiplex TaqMan Open Array Real-Time PCR assay in 760 Sarda sheep with records for milk composition and coagulation properties. Association analysis revealed the existence of significant associations between *CSN1S2* and *CSN3* genotypes and milk protein and casein contents. Moreover, genotypes at *CSN1S1* were significantly associated with rennet coagulation time, curd firming time and curd firmness, while *CSN2* was associated with curd firming time. These results suggest that SNPs mapping to the promoters and 3'UTR of ovine casein genes may exert regulatory effects on gene expression.

Keywords milk protein genes, single nucleotide polymorphism, Sarda sheep breed, milk traits, milk coagulation property

Introduction

Dairy sheep farming is a well established economic activity in the countries of the Mediterranean Basin and in particular in those of the European Union (Boyazoglu and Morand-Fehr, 2001). A good example of the successful adaptation of rural economy to landscape is represented by the Sardinian semi-extensive sheep farming (Carta et al., 2009). Sardinia is an insular region of Italy which produces about 5% of the whole world's ewe milk (Vacca et al., 2008), and it is also the birthplace of a very successful dairy specialized breed: the Sarda sheep. Because of their great adaptation to the hot and semi-arid environment of Sardinia and their remarkable milk yields, Sarda sheep have been exported to many European and North African countries to be crossed with local varieties (Ordás et al., 1997; Carta et al., 2009; Vacca et al., 2010). Traditionally, milk yield has been the main criterion determining the payment of milk by the cheese making industry and, as a consequence of this, augmenting milk yield has become the main goal of ovine selective programs (Carta et al., 2009). However, the negative genetic correlation between milk yield and composition has led to an undesirable decline in milk fat and protein content. Because such tendency could have a detrimental effect on cheese-making, selection goals have became increasingly focused on increasing milk protein content (Raynal-Ljutovac et al., 2008).

Caseins, the main protein fraction of milk, have been classified into four types: α_{s1}-casein (Cn), encoded by the *CSN1S1* gene, β-Cn (*CSN2*), α_{s2}-Cn (*CSN1S2*) and κ-Cn (*CSN3*). In sheep, the casein genes have been mapped to chromosome 6, where they form a cluster spanning 250 kbp (Echard et al., 1994; Barillet et al. 2005). Whey

proteins include α-lactalbumin (*LALBA*) and β-lactoglobulin (*BLG*) which are encoded by loci mapping to ovine chromosome 3. Polymorphism of the sheep milk protein genes has been investigated both at the protein and genomic levels (Ordás et al., 1997), and its current nomenclature is mainly based on criteria previously established for other ruminant species (Calvo et al., 2013). Although several variants have been discovered and investigated (Corral et al., 2013), the amount of knowledge about the variation of ovine casein genes is much lower than that obtained in cattle and goats. According to Selvaggi et al. (2014), five to eight variants are known to occur at the three calcium-sensitive caseins while *CSN3* has been reported to be monomorphic. To the best of our knowledge only one association analysis between the polymorphism of ovine casein genes and milk traits has been performed so far (Corral et al. 2010).

Variability at regulatory regions, such as gene promoters and 3'UTR, could have broad effects on milk traits by modulating the transcriptional rate, structure and stability of mRNA transcripts. Indeed, Georges et al. (2007) propose that regulatory polymorphisms might be one of the main sources of variation for productive traits in livestock. In the current work, we aimed to explore this hypothesis by characterizing, with a Sanger resequencing approach, the promoter and the 3'UTR diversity of milk protein genes in Sarda sheep. With such information, we planned to carry out an association analysis between milk protein gene polymorphisms and milk traits and coagulation properties recorded in 760 Sarda sheep.

Materials and methods

Phenotype recording and genomic DNA isolation

A total of 760 Sarda ewes descending from 108 different bucks were sampled in 21 farms. Description of farms and animals, which were officially registered in the flock book of the Sarda breed, is reported in Pazzola et al. (2014) and Vacca et al. (2015). Groups of ewes from each farm ranged from 17 to 56, and from each sire ranged from one to 28. Ewes were between two and seven months after parturition and, as regard the different levels of parity, from the first (primiparous) to the seventh. A single sampling day was performed at each farm. Individual milk samples were collected from each ewe in a 200 mL sterile plastic container during the afternoon milking and daily milk yield (morning plus evening milking) was recorded the very same day of milk sampling. Blood samples were taken from the jugular vein of each ewe in K_3EDTA vacuum tubes (BD Vacutainer, Plymouth, UK) and genomic DNA was obtained with the Puregene DNA isolation kit (GENTRA) and quantified using the Nano-Drop ND-1000 spectrophotometer (NanoDrop products, Wilmington, USA).

Milk samples were refrigerated at 4°C and analysed not lather than 24 hours after collection. Fat, protein, casein and lactose content, and pH were obtained by a MilkoScan FT6000 (Foss Electric, Hillerød, Denmark) and the International Dairy Federation (IDF) standard 141C:2000; total bacterial count (TBC) by a BactoScan FC150 (Foss Electric Hillerød, Denmark) and transformed to log-bacterial count to normalize the distribution (LBC = $\log_{10}$ TMC + 3); somatic cell count (SCC) using a Fossomatic 5000 (Foss Electric, Hillerød, Denmark) and the IDF 148-2:2006 method, and later transformed into the logarithm score (SCS = $\log_2$ SCC$\times 10^{-1}$+3).

Milk coagulation properties (MCP) were obtained using the Formagraph instrument (Foss Italia, Padova, Italy) and the procedures reported in Pazzola et al. (2014). The following traits were recorded: RCT (rennet coagulation time in minutes); k20 (curd firming time, min); a30, a45 and a60 (curd firmness, respectively 30, 45 and 60 minutes after rennet addition). RCT was considered missing for milk samples which coagulated 30 min after rennet addition (n = 6); k20 was missing when the width of the graph attained 20 mm later than 5 minutes after rennet addition (n = 4); a30 and a60 were missing when the result was 0 (a30: n = 1; a60: n = 1); no missing result was recorded for a45.

Resequencing of the ovine casein regulatory regions

Individual blood samples were randomly collected from 25 unrelated Sarda ewes, to obtain genomic DNA and to partially sequence casein promoters and 3'UTR.. Two overlapping PCR (Polymerase Chain Reaction) reactions were designed to amplify the promoter region of each of the four casein genes, whereas a single PCR reaction was needed to amplify the 3'UTR. Primers (Supplementary File Table S1) were designed with the Primer Express V2.0 software (Applied Biosystems, Waltham, MA, USA), using as a template the sheep reference genome sequence retrieved from the Ensembl database (http://www.ensembl.org). All PCR reactions were performed in a 15 µL reaction mixture containing 1.5 µL of 10X PCR buffer, 2.5mM $MgCl_2$, 0.3 µM of each primer, 0.25 mM of each dNTP, 50 ng of DNA, 0.75 U Taq Gold DNA polymerase (Applied Biosystem). The reaction mixture was heated at 95 ºC for 10 min, followed by 35 cycles each consisting of denaturation at 95 ºC for 1 min,

annealing (Supplementary File Table S1) for 1 min and extension at 72 ºC for 1 min, and then a final extension at 72 ºC for 10 min.

The PCR products were purified by using the ExoSAP-IT PCR Cleanup kit (Affymetrix, Santa Clara, CA) and sequenced in both directions with the same primers used in the amplification reaction. Sequencing reactions were prepared with the Big Dye Terminator Cycle Sequencing Kit v1.1 (Applied Biosystems) and they were electrophoresed in an ABI 3730 DNA Analyzer (Applied Biosystems). A total amount of 172 DNA target regions were successfully amplified and sequenced. Sequences were aligned and edited with MEGA 5 (http://www.megasoftware.net/). Amplicons corresponding to the promoters and 3'UTRs of the four casein genes had the following sizes: *CSN1S1*, 828 and 421bp; *CSN*2, 985 and 210bp; *CSN1S2*, 1309 and 776bp; *CSN3*, 925 and 622bp. All sequences were submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank) and an accession number was provided (Supplementary File Table S2).

Polymorphism discovery and design of a Taq Man OpenArray Real-Time PCR assay

Analysis of the sequences with the SeqScape v.2.5 software (Applied Biosystems) allowed us to detect 29 single nucleotide polymorphisms scattered in the four casein genes. A TaqMan Open Array Real-Time PCR assay was designed to genotype these 29 markers plus three additional SNPs retrieved from the Ensembl database and mapping to the *BLG* (2 SNPs) and *LALBA* (1 SNP) genes. The full panel of 32 SNPs is reported in Table 1.

The selected SNPs and their flanking sequences (60 nucleotides upstream and downstream), were submitted to the Custom TaqMan Assay Design Tool web site (https://www5.appliedbiosystems.com/tools/cadt/, Life Technologies) to ascertain if they were suitable to be genotyped with TaqMan Open Array multiplex assay. Genotyping was performed at the Servei Veterinari de Genètica Molecular (http://sct.uab.cat/svgm/en) by using a 12K Flex QuantStudio equipment and following the instructions of the manufacturer. The TaqMan Genotyper software v.1.3 (Applied Biosystems) was used to visualize genotypes. Further analysis of individual genotypes was performed using the SVS software (version 7, Golden Helix Inc.). Samples or SNPs with call rates < 0.9 were removed from the dataset. The AliBaba 2.1 Search engine (Grabe et al., 2002, http://www.gene-regulation.com/pub/programs/alibaba2/index.html) was used to identify transcriptional factor binding sites in the analyzed promoter regions and TargetScan v.7.0 (Agarwal et al., 2015) was employed to identify predicted microRNA targets in 3' UTR regions.

Haplotypes, linkage disequilibrium and SNPs blocks

The Haploview software package (Barrett et al., 2005) was used to estimate minor allele frequencies (MAF), observed and expected heterozygosities, and to test if genotype distributions were consistent with the Hardy-Weinberg equilibrium model. The Haploview program was also used to determine and plot pairwise linkage disequilibrium (LD) measures (D′ and r2), to define the LD blocks using the Gabriel

criteria (Gabriel et al., 2002) and to estimate haplotype frequency and haplotype tagging SNPs (htSNPs).

Statistical analysis

The association analysis between single SNPs and milk traits was performed with the MIXED procedure of SAS (SAS Inst. Inc., Cary, NC) and according to the following linear model:

$$y_{ijklmn} = \mu + genotype_i + farm_j + parity_k + DIM_l + sire_m(genotype_i) + e_{ijklmn} \quad (1)$$

where y_{ijklmn} is the observed trait; μ is the general mean; $genotype_i$ is the fixed effect of the *i*th genotype at the polymorphic sites (i = 2 to 3); $farm_j$ is the fixed effect of *j*th farms (j = 1 to 21); $parity_k$ is the fixed effect of the *k*th parity of the ewes (j = 1 to 4 or more); DIM_l is the fixed effect of the *l*th class of days in milking as the measure of the stage of lactation (i = 1 to 5; class 1: ≤ 80 days, class 2: 81 to 120 days, class 3: 121 to 160 days; class 4: 161 to 200 days; class 5: ≥ 201 days); $sire_m(genotype_i)$ is the random effect of *m*th sire (m = 1 to 108) nested within the *i*th genotype and and $e_{ijklmno}$ is the random residual.

The model (1) was also used to achieve the association analysis regarding the haplotypes at each of the four LD blocks. To avoid a bias of results linked to the low number of animals for each genotype level at the LD blocks, only genotypes with a frequency higher than 0.050, which corresponds to a minimum of 38 out of the 760 total sampled ewes, were submitted to the model. Multiple comparison of the least

square means was performed using the Bonferroni method and model effects were declared significant at $P < 0.05$.

RESULTS

Polymorphism discovery in the ovine casein genes

Resequencing of the promoters and 3'UTR of the ovine casein genes in Sarda sheep revealed the existence of 29 SNPs. The majority of these SNPs (28 out of 29) had been previously reported in the Ensembl database (http://www.ensembl.org) confirming that SNPs detected by us are not sequencing artifacts. The average density of SNPs at promoter and 3'UTR was approximately 1 SNP every 200 bp and 1 SNP every 225 bp, respectively. These levels of variation are remarkably high, particularly if we take into account that the sequence panel only contained 25 individuals. The densities of SNPs were also quite similar amongst the *CSN1S1*, *CSN1S2*, *CSN2* and *CSN3* genes (in the range of 1 SNP every 198-239 bp). A number of these SNPs were predicted to map to putative transcription factor and microRNA binding sites. For instance, two SNPs at the 3'UTR of the *CSN2* gene co-localized with predicted target sites for microRNAs miRNA-216a and miRNA-183. Moreover, in the *CSN1S2* gene, two SNPs at the promoter region matched binding sites for the C/EBPα1p and C/EBPβ transcription factors. A list of SNPs mapping to these motifs predicted to be functional is reported in Table S3.

The levels of variation for each SNPs and the results of the Hardy-Weinberg test are shown in Table 1. Three out of the 32 SNPs were classified as undetermined

(rs401368035, rs590028369, rs408578846) because of the low quality in PCR reactions of all the 760 analysed samples. Moreover, five SNPs were monomorphic: c.-2025 A>G, rs408450512, rs421588492, rs403429881, rs429593714. Undetermined and monomorphic SNPs were excluded from the association analyses. Besides, several individual samples were not included because they caused technical problems. Analysis of genotype frequencies showed that only two SNPs, rs410816096 and rs593666385, both at the *CSN3* promoter region, did not match the Hardy-Weinberg equilibrium expectations.

Association analysis between single SNP genotypes and milk traits

Descriptive statistics of milk traits recorded in 760 Sarda ewes are reported in Table S4. Results of the association analysis between the fixed effects included in the statistical model (farm, parity, stage of lactation and each single SNP) and milk traits are reported in Table S5. The fixed factors parity, stage of lactation and specially farm had a very strong influence on the distribution of almost all the milk traits. Regarding the effect of each single SNPs, least square means of milk traits displaying significant associations with casein genotypes are reported in Table 2.

Daily milk yield was not affected by any of the 24 SNPs. Milk protein and casein contents and k20 displayed the most significant associations with genotypes at casein genes. In this regard, three SNPs at the promoter regions of the *CSN1S2* locus, rs398328750, rs424698365 and rs404755583, and *CSN3*, rs593666385, were significantly associated with both milk protein and casein percentages. Moreover, the heterozygous genotypes always corresponded to intermediate protein and casein

contents, when compared with the homozygous genotypes, suggesting an additive mode of action.

As regard k20, the two loci that showed the most consistent and significant associations with this trait were *CSN1S1* and *CSN2*, though a significant association was also observed for *CSN3* genotype.

Several other associations, scattered throughout Table 2, were observed for lactose and fat content, a30 and RCT. Moreover, the *LALBA* genotype was associated with RCT but not with lactose content.

Association analysis between haplotypes blocks and milk traits

To determine the haplotype structure of the ovine casein cluster, some SNPs were excluded from our dataset: *i.e*. the ones mapping to the whey-protein genes *BLG* and *LALBA*, and 5 SNPs within the casein cluster genes which were monomorphic in the analysed Sarda population. The resulting set of 20 polymorphisms covered about 230 kb and the average distance between SNPs was 89 bp, ranging between 1 bp and 1962 bp.

The LD analysis of the casein gene cluster (Figure 1) showed the existence of four LD blocks, numbered from 1 to 4. Blocks 1, 2 and 3 were contiguous and encompassed the calcium-sensitive caseins. Block 1 included three haplotypes determined by the two SNPs of the *CSN1S1* promoter region (rs419330297 and rs400042158). Block 2 encompassed four haplotypes, derived from four SNPs (rs410096855, rs398437598, rs398328750 and rs596875693), and covered 78 kb spanning from the *CSN1S1* 3'UTR region to the promoter regions of *CSN2* and

CSN1S2. Block 3 comprised three haplotypes determined by five SNPs at the *CSN1S2* gene (promoter: rs424698365 and rs404755583, 3'UTR: rs398919262, rs421554515 and rs410928088). Finally, Block 4 covered 9 kb and yielded three haplotypes derived from three SNPs at *CSN3*: two at the promoter region and one at 3'UTR. The rs598701385 SNP was probably a recombination hotspot and, as a consequence, it was excluded from this block. Common haplotypes defined by the polymorphisms at each LD block, haplotype tagging SNPs (htSNPs) and frequencies are reported in Figure 2. Frequencies of haplotypes at each of the four LD blocks are reported in Table S6.

At each block, haplotypes with frequencies higher than 0.05 were included in the association analysis with milk traits. Results of the association analysis between the fixed effects (each single block, farm, parity and stage of lactation) and milk traits are reported in Table S7. Least square means of milk traits with a significant *P*-value, according to each block and haplotype, are reported in Table 3.

Farm, parity and stage of lactation showed significant influences on many traits, but these were fewer in comparison with the association analysis regarding the single SNPs. Haplotypes at Block 3 was significantly associated with protein and casein contents ($P < 0.05$) and the highest means corresponded to individuals harbouring the CGGAC haplotype. Besides, Block 4 was associated with RCT ($P < 0.05$) and the shortest coagulation time corresponded to ewes carrying the TTA haplotype.

DISCUSSION

A high level of variation in the regulatory regions of the ovine casein cluster

One of the main goals of the current work was to define the amount of SNP diversity mapping to the regulatory regions (promoters and 3'UTR) of the ovine casein genes. We did so because there are strong evidences that regulatory mutations with effects on gene expression could be an important source of phenotypic variation in domestic species (Georges, 2007). The sequencing of about 6 kb of the promoter and 3'UTR of the 4 casein genes allowed us to identify 29 SNPs. The SNP density across genes and regulatory regions were in the range of 1 SNP every 200 kb. According to Zhao et al. (2003), the SNP density in the human genome is much lower, with averages of 8.44, 5.28, 8.21 and 7.51 SNPs every 10 kb in intergenic, exonic, intronic and 3'UTR regions, respectively.

The variability of the ovine casein genes has been generally underestimated (Chessa et al., 2010) if compared with that of cattle and goats, probably because much less studies have been performed. Indeed, the molecular studies in sheep have been mostly focused on selection for scrapie resistance and parentage testing (Carta et al. 2009). Noteworthy, the majority of SNPs investigated in the present study were already present in the Ensembl database (http://www.ensembl.org), meaning that recent next generation sequencing experiments performed in sheep have filled this important gap.

In the present study, haplotype blocks and relative htSNPs at the sheep casein gene cluster were identified. This approach can be used to identify and characterize conserved haplotypes, which may represent functionally important genomic regions.

It may also explain the relationships existing between adjacent SNPs; and it can significantly reduce the number of markers that must be genotyped to test associations (Johnson et al., 2001; International HapMap Consortium, 2003).

Part of the casein variation co-localized with predicted transcription factor or microRNA binding sites. For instance. The T-allele of the SNP rs590028369 at *CSN3* creates binding sites for C/EBPα1p (CCAAT/enhancer bonding protein), Oct-1 (Oct-1, octamer binding factor), Pit-1a (pituitary-specific), and NF-1 (nuclear factor), whereas the alternative allele A would allow the binding of several other transcription factor as like as Oct-1, Oct-11, Pit-1a, Id3 (inhibitor of differentiation), HNF-1C (hepatocyte nuclear factor), GNC4 (general control nonderepressible) and Ftz (zinc finger). Similarly, the A-allele of the SNP rs398437598 suppressed a putative miRNA-183 binding site. This microRNA is primarily involved in the enhancement of the expression of adipogenic marker genes and it is also a positive regulator of the activity of the transcription factor C/EBPa1p (Chen et al., 2014). However, the biological meaning of the co-localizations (SNP vs putative functional motifs) identified in the present study and the *in silico* prediction of transcription factor and microRNA binding sites is plagued by high rates of false positives. In the absence of functional information, it is difficult to make inferences about the biological consequences of the identified mutations.

Variation at regulatory sites of the ovine CSN1S2 and CSN3 genes is associated with milk protein and casein contents

The variability of the *CSN1S2* gene was the one that showed the most significant associations with milk protein and casein contents. Indeed, these association were evidenced both for single SNPs (rs398328750, rs424698365 and rs404755583) as well as for Block 3, which was exclusively determined by the SNPs at *CSN1S2*. The significant association detected for rs404755583, was particularly interesting because this was also a haplotype tagging SNP (htSNP) included in LD Block 3. Our results contrast with those obtained by Corral et al. (2013). These authors analysed two missense variants of the ovine *CSN1S2* gene in Merino sheep but they did not find any significant association with protein content. These differences may emerge as a consequence of the different SNP panels and populations (linkage phases between causal and marker mutations differ amongst breeds) employed in each study.

Three out of the five *CSN3* SNPs investigated in the present study displayed a significant association with milk protein and casein contents as well as with curd firming time. Consistently, TT individuals for the rs593666385 SNP showed the highest protein and casein contents as well as the shortest k20. The main function of *CSN3* is to stabilize casein micelles and the enzymatic cleavage of this protein initiates milk renneting. Interestingly, Giambra et al. (2014) reported a significant association between *CSN3* IEF variants, computed as single or interaction effects, on protein content in East Friesian dairy ewes. In goats, an analogous influence of *CSN3* on fat and protein content has been evidenced by Hayes et al. (2006) in Norwegian goats, and in cattle the B variant of the *CSN3* gene has been consistently associated with a shorter renneting time, a higher milk protein content and a better cheese yield when compared with the A allele (Martin et al., 2002).

Variation at regulatory sites of the ovine CSN1S1 and CSN2 genes is associated with milk coagulation properties

The association analysis revealed that two out of the five investigated SNPs at the *CSN1S1* gene displayed a significant association with milk traits. In particular, rs419330297 was associated with the three milk coagulation properties RCT, k20 and a30. The shortest coagulation times and the highest values of curd firmness were recorded in CT heterozygote ewes. In goats, multiple mutations with causal effects on milk protein and casein contents and cheese yield and flavour have been identified in the *CSN1S1* gene (Martin et al., 2002). In contrast, such evidence has not been consistently found in sheep. In this way, Calvo et al. (2013) did not find any significant association between *CSN1S1* polymorphism and milk yield, protein, fat and lactose contents. In contrast, Giambra et al. (2014) reported a significant effect of *CSN1S1* genotype on milk protein content in East Friesian and Lacaune ewes. With regard to *CSN2* genotype, we found significant associations with curd firming time (k20). Interestingly, Bonfatti et al. (2010) reported that *CSN2* determines curd tension and hardening in cattle. Such hypothesis is consistent with the result of the present study because k20 is an indirect measure of the first stages of curd formation and development (Pazzola et al., 2014).

Scarcity of significant associations between whey protein genotypes and milk traits

In strong contrast with the casein genes, regulatory variation at the *LALBA* and *BLG* genes did not seem to have a significant impact on milk composition and coagulation

properties. The only association was found between SNP rs411787522 at the *LALBA* gene and RCT. Sheep with AA genotypes, in comparison with AG and GG, showed an extremely delayed coagulation time. Indeed, significant associations between the *LALBA* genotype and RCT have been recently observed in goats by Dettori et al. (2015), though differences between genotypes were smaller. Dettori et al. (2015) have proposed an indirect effect of lactose on milk renneting properties. However, in our study we have not found any relationship between *LALBA* genotype and lactose content, a feature that makes difficult to envisage how the *LALBA* genotype may affect RCT. Similarly, Zidi et al. (2014) have not found any significant association between SNPs in the promoter and 3'UTR of the goat *LALBA* gene and milk lactose content. With regard to the ovine *BLG* gene, it was not found any significant association with traits under analysis. Indeed, the biological role of this protein is still unclear. It has been theorized that it may have a relevant function in the intestinal tract of young suckling animals by facilitating the transport and metabolism of dietary lipids (Mele et al., 2007), but such hypothesis has not been fully confirmed yet.

CONCLUSIONS

The results indicated that the promoter and 3'UTR of the ovine casein genes show a remarkable level of polymorphism in Sarda sheep. Moreover, part of this variation displays significant associations with milk protein and casein contents as well as with milk coagulation properties. These findings suggest that regulatory variation at the

casein genes may affect the composition and coagulation properties of ovine milk, though functional studies will be needed to confirm this hypothesis.

ACKNOWLEDGMENTS

This research was supported by a grant from the Fondazione Banco di Sardegna (Bank of Sardinia Foundation).

The authors wish to thank the farmers for giving access to their flocks, the A.I.P.A./A.P.A.s (Provincial Farmers Associations) of Cagliari, Nuoro, Sassari and Oristano for their support in sample collection and A.R.A. Sardegna (Regional Farmer Association of Sardinia) for support in milk analysis.

REFERENCES

Agarwal V., Bell G.W., Nam J., Bartel D.P. 2015. Predicting effective microRNA target sites in mammalian mRNAs. eLife, 4:e05005.

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, S109–S123.

Barrett, J. C., Fry B., Maller J., Daly M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21, 263-265.

Belsham D.D., Mellon P.L. 2000. Transcription factors Oct-1 and C/EBPβ (CCAAT/enhancer-binding protein-β) are involved in the glutamate/nitric oxide/cyclic-guanosine 5′-monophosphate-mediated repression of

gonadotropin-releasing hormone gene expression. Molecular Endocrinology 14, 212-228.

Bonfatti V., Di Martino G., Cecchinato A., Degano L., Carnier P. 2010. Effects of β–κ-casein (CSN2-CSN3) haplotypes, β-lactoglobulin (BLG) genotypes, and detailed protein composition on coagulation properties of individual milk of Simmental cows. Journal of Dairy Science 93, 3809–3817.

Boyazoglu J., Morand-Fehr P. 2001. Mediterranean dairy sheep and goat products and their quality. A critical review. Small Ruminant Research 40, 1–11.

Calvo J.H., Dervishi E., Sarto P., González-Calvo L., Berzal-Herranz B., Molino F., Serrano M., Joy M.. 2013. Structural and functional characterisation of the αS1-casein (CSN1S1) gene and association studies with milk traits in Assaf sheep breed. Livestock Science 157, 1–8.

Carta A., Casu Sara, Salaris S. 2009. Invited review: current state of genetic improvement in dairy sheep. Journal of Dairy Science 92, 5814–5833.

Chen C. Xiang H., Peng Y.L., Peng J., Jiang S.W. 2014. Mature miR-183, negatively regulated by transcription factor GATA3, promotes 3T3-L1 adipogenesis through inhibition of the canonical Wnt/β-catenin signaling pathway by targeting LRP6. Cell Signalling 26, 1155-1165.

Chessa S., Rignanese D., Berbenni M., Ceriotti G., Martini M., Pagnacco G. 2010. New genetic polymorphisms within ovine β- and αS2-caseins. Small Ruminant Research 88, 84–88.

Corral J.M., Padilla J.A., Izquierdo M. 2010. Associations between milk protein genetic polymorphisms and milk production traits in Merino sheep breed. Livestock Science 129, 73–79.

Corral J.M., Padilla J.A., Izquierdo M., Martínez-Trancón M., Parejo J.C., Salazar J., Hernández-García F.I. 2013. Detection and genetic characterization of ovine CSN1S2*B polymorphisms and their associations with milk production traits. Livestock Science 153, 10–19.

Dettori M.L., Pazzola M., Paschino P., Pira M.G., Vacca G.M. 2015. Variability of the caprine whey protein genes and their association with milk yield, composition and renneting properties in the Sarda breed. 1. The LALBA gene. Journal of Dairy Research in press, doi:10.1017/S0022029915000461.

Echard G., Broad T.E., Hill D., Pearce P. 1994. Present status of the ovine gene map (Ovis aries); comparison with the bovine map (Bos taurus). Mammalian Genome 5, 324-332.

Gabriel S.B., Schaffnerv S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., Liu-Cordero S.N., Rotimi C., Adeyemo A., Cooper R., Ward R., Lander E.S., Daly M.J., Altshuler D. 2002. The structure of haplotype blocks in the human genome. Science 296, 2225-2229.

Georges M. 2007. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. Annual Review of Genomics and Human Genetics 8, 131-162.

Giambra I.J., Brandt H., Erhardt G. 2014. Milk protein variants are highly associated with milk performance traits in East Friesian Dairy and Lacaune sheep. Small Ruminant Research 121, 382–394.

Grabe N. . 2002. AliBaba2: context specific identification of transcription factor binding sites. In Silico Biology 2, S1–S15.

Hayes B., Hagesaether N., Adnøy T., Pellerud G., Berg P.R., Lien S. 2006. Effects on production traits of haplotypes among casein genes in Norwegian goats and evidence for a site of preferential recombination. Genetics 174, 455–464.

International HapMap Consortium. 2003. The International HapMap Project. Nature 426, 789-796.

Johnson G.C., Esposito L., Barratt B.J., Smith A.N., Heward J., Di Genova G., Ueda H., Cordell H.J., Eaves I.A., Dudbridge F., Twells R.C., Payne F., Hughes W., Nutland S., Stevens H., Carr P., Tuomilehto-Wolf E., Tuomilehto J., Gough S.C., Clayton D.G., Todd J.A. 2001. Haplotype tagging for the identification of common disease genes. Nature Genetics 29, 233–237.

Martin P., Szymanowska M., Zwierzchowski L., Leroux C. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. Reproduction Nutrition Development 42, 433–459.

Mele M., Conte G., Serra A., Buccioni A., Secchiari P. 2007. Relationship between beta-lactoglobulin polymorphism and milk fatty acid composition in milk of Massese dairy ewes. Small Ruminant Research 73, 37–44.

Ordás J.G., Rando A., Senese C., Masina P. 1997. DNA polymorphisms of casein genes in Spanish dairy sheep. Small Ruminant Research 26, 9-13.

Pazzola M., Dettori M.L., Cipolat-Gotet C., Cecchinato A., Bittante G., Vacca G.M. 2014. Phenotypic factors affecting coagulation properties of milk from Sarda ewes. Journal of Dairy Science 97, 7247-7257.

Raynal-Ljutovac K., Lagriffoul G., Paccard P., Guillet I., Chilliard Y. 2008. Composition of goat and sheep milk products: an update. Small Ruminant Research 79, 57–72.

Selvaggi M., Laudadio V., Dario C., Tufarelli V. 2014. Investigating the genetic polymorphism of sheep milk proteins: a useful tool for dairy production. Journal of the Science of Food and Agriculture 94, 3090-3099

Vacca G.M., Carcangiu V., Dettori M.L., Pazzola M., Mura M.C., Luridiana S., Tilloca G. 2008. Productive performance and meat quality of Mouflon x Sarda and Sarda x Sarda suckling lambs. Meat Science 80, 326–334.

Vacca G.M., Dhaouadi A., Rekik M., Carcangiu V., Pazzola M., Dettori M.L. 2010. Prolificacy genotypes at BMPR 1B, BMP15 and GDF9 genes in North African sheep breeds. Small Ruminant Research 88, 67–71.

Vacca G.M., Pazzola M., Dettori M.L., Pira E., Malchiodi F., Cipolat-Gotet C., Cecchinato A., Bittante G. 2015. Modeling of coagulation, curd firming, and syneresis of milk from Sarda ewes. Journal of Dairy Science 98, 2245-2259.

Zhao Z., Fu Y.-X., Hewett-Emmett D., Boerwinkle E. 2003. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene 312, 207–213.

Zidi A., Casas E., Amills M., Jordana J., Carrizosa J., Urrutia B., Serradilla J.M. 2014. Genetic variation at the caprine lactalbumin, alpha (LALBA) gene and its association with milk lactose concentration. Animal Genetics 45, 612–613.

Table 1 Milk protein SNPs typed in Sarda sheep (n = 760) and description of their diversity

SNP	Chromosome position	%Geno	ObsHET	PredHET	HWpval	MAF	Alleles[1]
BLG							
rs401368035	3:3576072	UND	NA	NA	NA	NA	T:C
rs416633597	3:3577656	97.1	0.49	0.50	0.83	0.47	G:(A)
LALBA							
rs411787522	3:137390224	97.9	0.24	0.23	0.12	0.13	G:(A)
CSN1S1							
c.-2023 A>G	6:85087462	96.8	0.0	0.0	NA	0.0	A:(G)
rs419330297	6:85087747	98.3	0.31	0.33	0.23	0.21	T:(C)
rs400042158	6:85087754	95.9	0.43	0.45	0.37	0.34	(G):A
rs416208751	6:85104364	97.5	0.03	0.03	0.87	0.02	T:(C)
rs410096855	6:85104388	97.9	0.12	0.13	0.25	0.07	G:(A)
rs398437598	6:85104690	98.7	0.16	0.16	0.98	0.09	G:(A)
CSN2							
rs425554254	6:85124767	93.2	0.04	0.04	1.00	0.02	T:(C)
rs411318553	6:85124822	97.6	0.03	0.03	1.00	0.02	T:(C)
rs405274538	6:85124935	98.7	0.03	0.03	1.00	0.02	C:(T)
rs418140899	6:85124944	98.8	0.03	0.03	1.00	0.02	T:(C)
rs403942452	6:85124964	97.0	0.03	0.03	1.00	0.02	(G):C
CSN1S2							
rs408450512	6:85182920	95.8	0.0	0.0	1.00	0.0	C:(T)
rs421588492	6:85182946	97.6	0.0	0.0	1.00	0.0	G:(A)
rs398328750	6:85183088	98.3	0.47	0.49	0.42	0.43	(G):A
rs596875693	6:85183125	98.4	0.48	0.49	0.51	0.43	(A):G
rs424698365	6:85183711	98.3	0.31	0.33	0.18	0.21	C:(A)
rs404755583	6:85183720	95.0	0.48	0.49	0.45	0.43	(G):T
rs408578846	6:85195571	UND	NA	NA	NA	NA	G:A
rs398919262	6:85195677	98.4	0.32	0.33	0.71	0.21	G:(A)
rs421554515	6:85195728	92.9	0.31	0.32	0.49	0.20	A:(G)
rs410928088	6:85195797	88.7	0.30	0.31	0.36	0.19	(G):C
CSN3							
rs410816096	6:85307269	61.2	0.44	0.50	0.02	0.47	A:(G)
rs598701385	6:85307360	79.6	0.31	0.30	0.26	0.18	(T):C
rs593666385	6:85307503	97.1	0.29	0.32	0.03	0.20	(T):A
rs403429881	6:85307735	93.9	0.0	0.0	1.00	0.0	A:(G)
rs429593714	6:85308132	97.2	0.0	0.0	1.00	0.0	(A):G
rs590028369	6:85308437	UND	NA	NA	NA	NA	T:A
rs605713973	6:85317361	98.4	0.40	0.40	1.00	0.28	(C):T
rs593210006	6:85317365	97.6	0.34	0.33	0.83	0.21	A:(C)

%Geno, percentage of genotyped samples; ObsHET, observed heterozygosity; PredHET, predicted heterozygosity; HWpval, Hardy-Weinberg test P value; ; MAF, minor allele frequency; 1, allele with minor frequency in brackets; UND, undetermined (0% genotyped samples); NA, not available data.

Table 2 Least square means of Sarda sheep milk traits according to the different genes, SNP and genotypes; n= 760.

Gene	SNP name	genotype	Fat %		Protein %		Casein %		Lactose %		RCT min		k20 min		a30 mm	
LALBA	rs411787522	AA	6.89	ns	5.67	ns	4.44	ns	4.76	ns	12.15^{a}	*	2.33	ns	48.02	ns
		AG	6.40		5.34		4.16		4.78		9.24^{b}		2.09		48.63	
		GG	6.39		5.41		4.22		4.79		8.99^{b}		2.01		49.11	
CSN1S1	rs419330297	CC	6.75	ns	5.52	ns	4.32	ns	4.70^{b}	*	9.21^{a}	*	1.99^{ab}	*	52.12^{A}	**
		CT	6.37		5.40		4.21		4.80^{a}		8.55^{b}		1.94^{b}		50.03^{A}	
		TT	6.38		5.38		4.20		4.79^{a}		9.33^{a}		2.09^{a}		48.15^{B}	
	rs416208751	CT	6.50	ns	5.32	ns	4.14	ns	4.78	ns	8.72	ns	2.32^{a}	*	47.05	ns
		TT	6.39		5.40		4.21		4.79		9.11		2.03^{b}		49.05	
CSN2	rs425554254	TT	6.39	ns	5.40	ns	4.21	ns	4.79	ns	9.05	ns	2.02^{b}	*	49.07	ns
		TC	6.51		5.31		4.14		4.79		8.66		2.32^{a}		47.07	
	rs411318553	TT	6.40	ns	5.40	ns	4.21	ns	4.79	ns	9.11	ns	2.03^{b}	*	49.06	ns
		TC	6.51		5.32		4.15		4.79		8.73		2.32^{a}		47.06	
	rs405274538	CT	6.51	ns	5.32	ns	4.14	ns	4.79	ns	8.70	ns	2.32^{a}	*	47.08	ns
		TT	6.40		5.40		4.21		4.79		9.10		2.03^{b}		49.07	
	rs418140899	TT	6.40	ns	5.40	ns	4.21	ns	4.79	ns	9.10	ns	2.03^{b}	*	49.07	ns
		TC	6.51		5.31		4.14		4.79		8.70		2.32^{a}		47.08	
	rs403942452	CG	6.50	ns	5.32	ns	4.15	ns	4.79	ns	8.68	ns	2.32^{a}	*	47.15	ns
		CC	6.39		5.40		4.21		4.79		9.06		2.02^{b}		49.12	
CSN1S2	rs398328750	AA	6.41	ns	5.46^{A}	**	4.25^{a}	*	4.77	ns	8.85	ns	1.97	ns	49.39	ns
		AG	6.39		5.39^{A}		4.21^{a}		4.79		9.28		2.07		48.80	
		GG	6.44		5.29^{B}		4.13^{b}		4.82		8.98		2.04		48.66	
	rs424698365	AA	6.44	ns	5.29^{B}	**	4.13b	*	4.82	ns	8.97	ns	2.04	ns	48.65	ns
		AG	6.38		5.39^{A}		4.21ab		4.79		9.27		2.07		48.83	
		GG	6.41		5.46^{A}		4.25a		4.77		8.84		1.97		49.42	
	rs404755583	GG	6.41	ns	5.29^{B}	**	4.13^{b}	*	4.83	ns	8.90	ns	2.03	ns	48.99	ns
		GT	6.39		5.40^{A}		4.21^{ab}		4.79		9.25		2.07		48.98	
		TT	6.41		5.46^{A}		4.26^{a}		4.77		8.84		1.98		49.57	
CSN3	rs593666385	AA	6.41	ns	5.37^{b}	*	4.19^{b}	*	4.78	ns	9.19	ns	2.08^{a}	*	48.73	ns
		AT	6.33		5.43^{b}		4.24^{ab}		4.81		8.83		1.96^{b}		49.48	
		TT	6.67		5.57^{a}		4.35^{a}		4.76		8.99		1.88^{b}		50.89	
	rs605713973	CC	6.69^{a}	*	5.51	ns	4.30	ns	4.72^{b}	*	9.45	ns	2.04	ns	49.91	ns
		CT	6.35^{b}		5.40		4.21		4.78^{a}		9.11		2.03		48.93	
		TT	6.40^{b}		5.38		4.20		4.79^{a}		9.04		2.04		48.92	
	rs593210006	AA	6.40	ns	5.40	ns	4.21	ns	4.80^{A}	**	9.07	ns	2.05	ns	48.66	ns
		AC	6.36		5.38		4.19		4.77^{A}		9.09		2.00		49.41	
		CC	6.69		5.52		4.30		4.67^{B}		9.29		2.07		50.45	

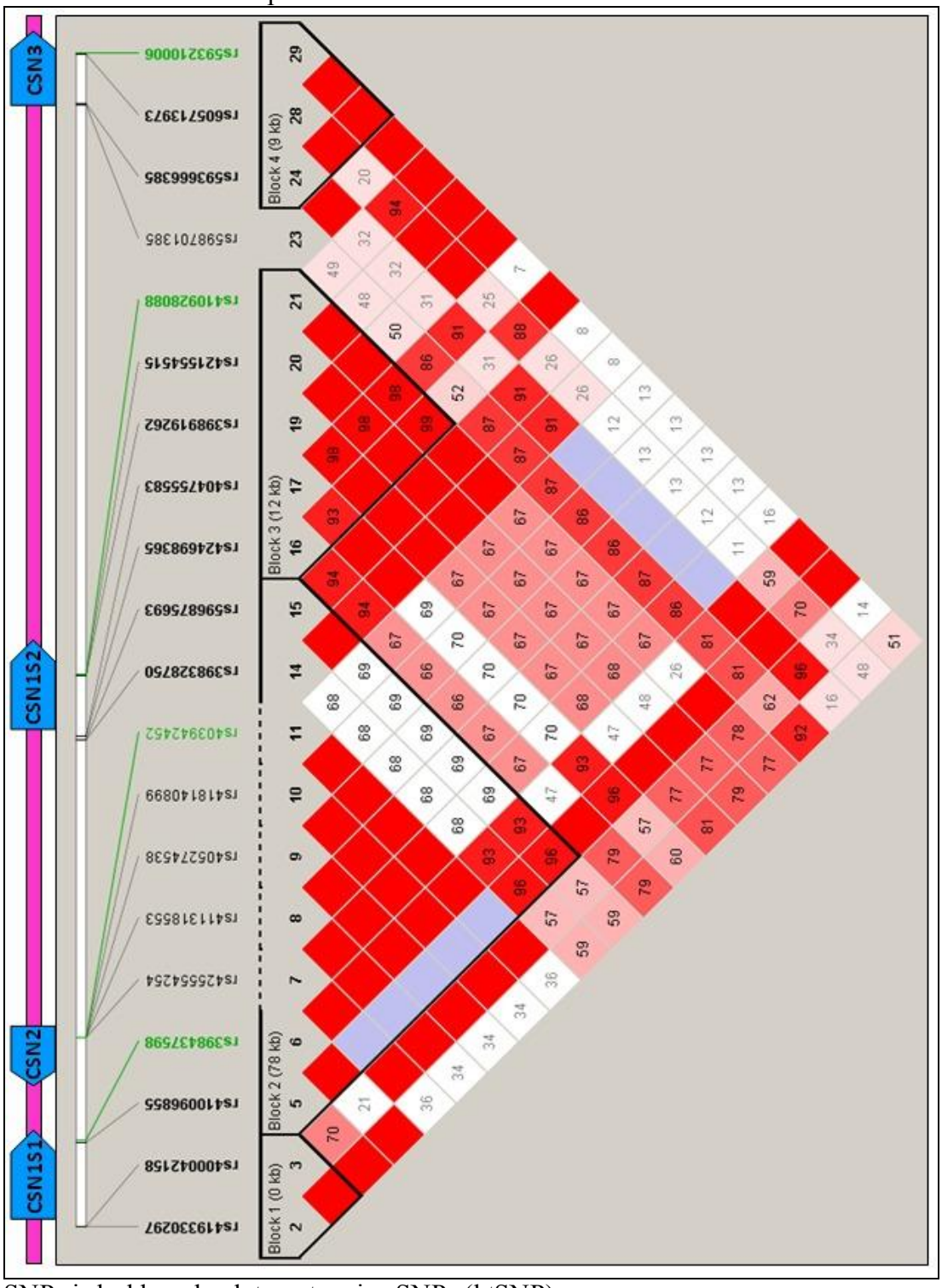
RCT, rennet coagulation time; k20, curd firming time; a30 curd firmness.
$^{A\text{-}B,\ a\text{-}b}$Means with different superscript with capital or lower-case letters in each row differ significantly in genotype comparison respectively at $P<0.01$ and $P<0.05$.
*, $P<0.05$; **, $P<0.01$; ns, non significant.

Table 3 Least square means of Sarda sheep milk traits and coagulation properties according to the different blocks and haplotypes; n= 760.

Blocks	Haplotype	(H)	n	Protein %	Casein %	RCT min
Block 3	CGGAC	(H9)	133	5.47[a] *	4.28 [a] *	8.63 ns
	CTGAC	(H10)	90	5.37[b]	4.19 [b]	8.70
Block 4	ATA	(H13)	206	5.44 ns	4.24 ns	9.22[a] *
	TTA	(H14)	39	5.46	4.25	8.48[b]

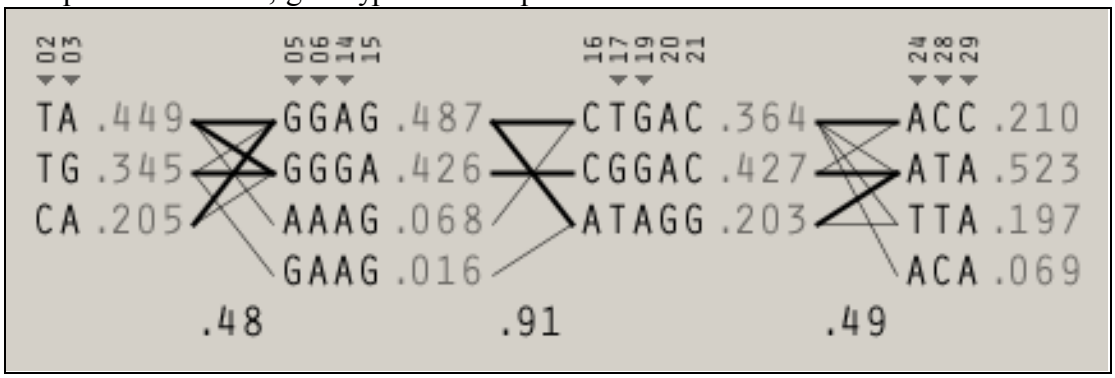
H, Haplotype numbering; RCT, rennet coagulation time.
SNPs included in Block 3: rs424698365; rs404755583; rs398919262; rs421554515; rs410928088.
SNPs included in Block 4: rs593666385; rs605713973; rs593210006.
[A-B, a-b]Means with different superscript with capital or lower-case letters in each row differ significantly in genotype comparison respectively at $P<0.01$ and $P<0.05$.
**, $P<0.01$; *, $P<0.05$; ns, non significant.

Figure 1 Linkage disequilibrium structure and blocks of SNPs of the casein gene cluster in the Sarda sheep breed.



SNPs in bold are haplotype tagging SNPs (htSNP).

Figure 2 Common casein haplotypes defined by the polymorphisms at each linkage disequilibrium block, genotypes and frequencies.



SNPs: 2 = rs419330297; 3= rs400042158; 5 = rs410096855; 6 = rs398437598; 14 = rs398328750; 15 = rs596875693; 16 = rs424698365; 17 = rs404755583; 19 = rs398919262; 20 = rs421554515; 21 = rs410928088; 24 = rs593666385; 28 = rs605713973; 29 = rs593210006.

A grey triangle (▼) under the SNP indicates this is a haplotype tagging SNP (htSNP).

Supplementary File Table S1.
Primers' names, sequences and annealing temperatures used to amplify 5' upstream and 3' downstream regions of each casein gene analysed in the present study.

		Forward primer		Reverse primer		Annealing temperature ℃
Gene	**region**	**primer name**	**sequence 5' - 3'**	**primer name**	**sequence 5' - 3'**	
CSN1S1	5' flanking	CSN1S1 fw1 _prA	CTTTCTTATAAACAAGGAGTTGCA	CSN1S1 rw1_prA	ATTTTGGGTGTACTAGTTGTGGTA	60
	5' flanking	CSN1S1 fw2 _prB	TAGTTACCACAACTAGTACACCCAA	CSN1S1 rw2 _prB	TGATCAAGGTGATGGCAGAC	60
	3' UTR	CSN1S1 fw_3'UTR	TATCTTAACAGTAGCTTCTCCTTTCA	CSN1S1 rw_3'UTR	GGTTGTGTAAATTCCTCCCAAT	60
CSN2	5' flanking	CSN2 fw1 _prA	GTACTTAGGGTAAGACTCTCCTCCTGTA	CSN2 rw1_prA	GACCTGAAGAGTGAAGAGAGTATAGACT	60
	5' flanking	CSN2 fw2 _prB	TCTCTTCACTCTTCAGGTCTAATCTAT	CSN2 rw2 _prB	AAAGTAGAGGAGAAGAAGTGAAGG	60
	3' UTR	CSN2 fw_3'UTR	CAGAATTGACTGCGACTGGA	CSN2 rw_3'UTR	TGACTGGTTAGGAAATAGATTCTTAA	60
CSN1S2	5' flanking	CSN1S2 fw1_prA	CATGCATACACATTCTCTAAATTTG	CSN1S2 rw1 _pr_A	GGAGAATTGGCAGGTGAAAATA	64
	5' flanking	CSN1S2 fw2_prB	GTTTGGTGGTTATTCTGTCAGG	CSN1S2 rw2_pr_B	ACTTCATGTTTACTTGATCCTGTGA	64
	3' UTR	CSN1S2 fw_3'UTR	CTGCTTCTACCTGGTTATGGTAAG	CSN1S2 rw_3'UTR	AGACAGAATGCCAAGATTGGA	64
CSN3	5' flanking	CSN3 fw1 _prA	GATCAGGACTTCATAGATAAAATGAA	CSN3 rw1_prA	CATTAATCAATATAAAGTAGGGATCC	60
	5' flanking	CSN3 fw2 _prB	GAAAATGGATCCCTACTTTATATTGA	CSN3 rw2_prB	GAACTGTAAGGAACACCCAAGGTA	60
	3' UTR	CSN3 fw_3'UTR	ACAACGCAGGTAAATGAGCA	CSN3 rw_3'UTR	TGGACTCTGCCATTATCCATT	64

Supplementary File Table S2.
GenBank accession numbers of *CSN1S1*, *CSN2*, *CSN1S2* and *CSN3* promoter and 3'UTR sequences obtained in the present study. Sheet 1.

Gene	Region	Sample ID	Accession Number
CSN1S1	5' flanking	2SS51.1p	KT283255
		1OR6.1p	KT283256
		5CA38.2p	KT283257
		8CA6.3p	KT283258
		5NU27.1p	KT283259
		3SS11.3p	KT283260
		1CA54.2p	KT283261
		3NU23.2p	KT283262
		6NU28.1p	KT283263
		7CA27.2p	KT283264
		6CA25.3p	KT283265
		4CA3.1p	KT283266
		5CA23.2p	KT283267
		5CA51.1p	KT283268
		887p	KT283269
		375p	KT283270
		1SS18.2p	KT283271
		1CA18.1p	KT283272
		5CA261p	KT283273
		797p	KT283274
	3' UTR	2-SS51-1	KT283275
		1-OR6-1	KT283276
		4-CA3-1	KT283277
		6CA25.3	KT283278
		5NU27.1	KT283279
		3SS11.3	KT283280
		1-CA54-2	KT283281
		3NU23.2	KT283282
		6-NU28-1	KT283283
		7CA27.2	KT283284
		1CA18.1	KT283285
		5CA261	KT283286
		5CA23.2	KT283287
		5CA51.1	KT283288
		887	KT283289
		375	KT283290
		8CA6.3	KT283291
		5CA38.2	KT283292
		797	KT283293
		1SS18.2	KT283294

Supplementary File Table S2.
GenBank accession numbers of *CSN1S1*, *CSN2*, *CSN1S2* and *CSN3* promoter and 3'UTR sequences obtained in the present study. Sheet 2.

Gene	Region	Sample ID	Accession Number
CSN2	5' flanking	1OF6p	KT283295
		4CA3p	KT283296
		6CA2p	KT283297
		5NU2p	KT283298
		3SS1p	KT283299
		1CA5p	KT283300
		5CA5p	KT283301
		1SS18_2p	KT283302
		8CA6p	KT283303
		797p	KT283304
		6NU281p	KT283305
		887p	KT283306
		4SS9p	KT283307
		3NU2p	KT283308
		1CA1p	KT283309
		5CA26p	KT283310
		5CA23p	KT283311
		2SS5p	KT283312
		5CA3p	KT283313
		8CA1p	KT283314
		3SS11	KT283315
	3' UTR	797	KT283316
		8CA63	KT283317
		6CA25	KT283318
		5NU27	KT283319
		6NU28	KT283320
		7CA272	KT283321
		1SS182	KT283322
		2CA44.2	KT283323
		1CA18	KT283324
		5CA23	KT283325
		9CA443	KT283326
		8CA11	KT283327
		3NU8.1	KT283328
		2SS51	KT283329
		4SS91	KT283330
		4NU12.2	KT283331
		1OF61	KT283332
		5CA51	KT283333
		4CA31	KT283334
		3NU23	KT283335
		2.NU28	KT283336
		1CA32.3	KT283337
		6SS16.3	KT283338
		8CA38.3	KT283339

Supplementary File Table S2.
GenBank accession numbers of *CSN1S1*, *CSN2*, *CSN1S2* and *CSN3* promoter and 3'UTR sequences obtained in the present study. Sheet 3.

Gene	Region	Sample ID	Accession Number
CSN1S2	5' flanking	2SS51p	KT283340
		2.2CA44p	KT283341
		1SS27p	KT283342
		1CA18p	KT283343
		4CA3_1p	KT283344
		5CA26p	KT283345
		797p	KT283346
		8CA38p	KT283347
		4NU12p	KT283348
		8CA11p	KT283349
		2NU28p	KT283350
		5CA23p	KT283351
		3NU8_1p	KT283352
		1CA32p	KT283353
		5CA38p	KT283354
	3' UTR	2CA44_2	KT283355
		1CA54_2	KT283356
		3SS11_3	KT283357
		3.UN.23._2	KT283358
		3NU8_1	KT283359
		1CA18_1	KT283360
		5CA51_1	KT283361
		5CA23_2	KT283362
		2NU28_1	KT283363
		6CA25_3	KT283364
		5CA26_1	KT283365
		4CA3_1	KT283366
		1CA32_3	KT283367
		8CA6_3	KT283368
		5NU27_1	KT283369
		6NU28_1	KT283370
		4NU12_2	KT283371
		8CA11_2	KT283372
		1SS27_2	KT283373
		6SS16_3	KT283374
		797	KT283375
		2SS51_1	KT283376
		8CA38_3	KT283377
		5CA38_2	KT283378
		4SS9_1	KT283379
		9CA44_3	KT283380
		7CA27_2	KT283381
		1OR6_1	KT283382

Supplementary File Table S2.
GenBank accession numbers of *CSN1S1*, *CSN2*, *CSN1S2* and *CSN3* promoter and 3'UTR sequences obtained in the present study. Sheet 4.

Gene	Region	Sample ID	Accession Number
CSN3	5' flanking	6-NU28-1p	KT283383
		5-CA38-2p	KT283384
		5-NU27-1p	KT283385
		1-SS18-2p	KT283386
		4-SS9-1p	KT283387
		797p	KT283388
		8-CA6-3p	KT283389
		1-CA54-2p	KT283390
		5-CA23-2p	KT283391
		7-CA27-2p	KT283392
		4-CA3-1p	KT283393
		FW-8-CA11-2p	KT283394
		726p	KT283395
		2-SS51-1p	KT283396
		3-SS11-3p	KT283397
		1-CA18-1p	KT283398
	3' UTR	8CA11_2	KT283399
		1SS18_2	KT283400
		797	KT283401
		2SS51_1	KT283402
		4NU12_2	KT283403
		6SS16_3	KT283404
		5CA38_2	KT283405
		4SS9_1	KT283406
		8CA38_3	KT283407
		7CA27_2	KT283408
		1OR6_1	KT283409
		9CA44_3	KT283410
		3NU23_2	KT283411
		1CA54_2	KT283412
		2CA44_2	KT283413
		3SS11_3	KT283414
		5CA51_1	KT283415
		5CA23_2	KT283416
		3NU8_1	KT283417
		1CA18_1	KT283418
		5CA26_1	KT283419
		4CA3_1	KT283420
		2NU28_1	KT283421
		6CA25_3	KT283422
		5NU27_1	KT283423
		6NU28_1	KT283424
		1CA32_3	KT283425
		8CA6_3	KT283426

Supplementary File Table S3 *In silico* predicted transcription factors (5' upstream region) and microRNA (miRNA, 3' downstream region) according to the different genes, regions, SNPs and alleles.

Gene and region	SNP	allele	transcription factors or microRNA
CSN1S1			
5' upstream region	c.-2023 A>G	A	C/EBPα1p
		G	C/EBPβ, Oct-1, GATA-1
3' downstream region	rs410096855	G	miRNA-216a
		A	-
	rs398437598	G	miRNA-183
		A	-
CSN2			
5' upstream region	rs425554254	T	C/EBPa1p, TBP, Oct-1
		C	-
CSN1S2			
	rs408450512	C	Sp1
		T	Sp1
5' upstream region	rs421588492	G	C/EBPβ
		A	-
	rs596875693	A	C/EBPα1p
		G	-
CSN3			
	rs598701385	T	C/EBPα1p
		C	-
5' upstream region	rs593666385	T	C/EBPα1p, Oct-1
		A	C/EBPα1p, Antp
	rs429593714	A	HNF-1, Oct-1
		G	YY1
	rs590028369	T	C/EBPα1p, Oct-1, Pit-1a, NF-1
		A	Oct-1, Oct-11, Pit-1a, Id3, HNF-1C, GNC4, Ftz

C/EBP, CCAAT/enhancer binding protein; Oct-1, octamer binding factor; GATA-1, globin transcription factor; TBP, TATA-box-binding protein; Sp1, specificity protein 1; Antp, antennapedia; HNF-1, hepatocyte nuclear factor; YY1, Ying-Yang factor; Pit-1a, pituitary-specific; NF-1, nuclear factor; Id3, inhibitor of differentiation; GNC4, general control nonderepressible; Ftz, zinc finger.

Table S4 Descriptive statistics of milk yield and composition, pH, total microbial count (TMC) and somatic cell count (SCC), n= 760.

Parameters	mean	SD	min	max	kurtosis	skewness
Milk yield (g/day)	1614	777	200	5760	2.111	1.253
Fat (g/100 ml)	6.42	1.19	3.48	12.52	1.885	0.787
Protein (g/100 ml)	5.37	0.66	4.04	8.24	0.642	0.839
Casein (g/100 ml)	4.19	0.55	3.04	6.67	0.747	0.838
Lactose (g/100 ml)	4.83	0.26	3.70	5.59	0.649	-0.437
pH	6.67	0.08	6.50	7.02	0.420	0.427
TMC (1000 cells/ml)	1214	2508	2	17000	12.855	3.378
LBC	5.36	0.85	3.30	7.23	-0.759	0.106
SCC (1000 cells/ml)	841	1906	21	15578	22.052	4.431
SCS	5.44	0.59	4.32	7.19	0.005	0.738

LBC, $\log_{10}$ TMC + 3; SCS, $\log_2 SCC \times 10^{-1} + 3$.

Supplementary File Table S5 (Sheet 1). *P*-values for milk parameters (daily milk yield, milk composition and coagulation properties) according to each of the 24 polymorphic SNPs out of the investigated 32 SNPs at the casein and whey protein genes. Sheet 1: Genotype effect.

Gene	SNP_name	NumDF	DenDF	Daily milk yield	Fat %	Protein %	Casein %	Lactose %	ph	SCS	LBC	RCT	k20	a30	a45	a60
BLG	rs416633597	2	602	0.369	0.076	0.560	0.608	0.078	0.100	0.968	0.984	0.658	0.427	0.117	0.115	0.306
LALBA	rs411787522	2	608	0.489	0.316	0.060	0.068	0.771	0.212	0.142	0.700	0.020	0.163	0.815	0.856	0.865
CSN1S1	rs419330297	2	611	0.107	0.068	0.260	0.216	0.027	0.935	0.174	0.766	0.014	0.016	0.006	0.229	0.921
	rs400042158	2	593	0.442	0.776	0.109	0.136	0.282	0.325	0.396	0.720	0.491	0.997	0.743	0.759	0.718
	rs416208751	1	606	0.781	0.591	0.434	0.442	0.964	0.322	0.959	0.820	0.589	0.030	0.318	0.135	0.161
	rs410096855	2	608	0.873	0.544	0.895	0.882	0.569	0.645	0.814	0.197	0.559	0.360	0.282	0.361	0.644
	rs398437598	2	614	0.697	0.499	0.944	0.973	0.670	0.458	0.822	0.271	0.511	0.130	0.118	0.095	0.263
CSN2	rs425554254	1	573	0.801	0.541	0.388	0.406	0.978	0.337	0.897	0.385	0.582	0.031	0.317	0.132	0.157
	rs411318553	1	607	0.796	0.583	0.430	0.439	0.970	0.325	0.963	0.427	0.594	0.030	0.316	0.134	0.161
	rs405274538	1	615	0.785	0.595	0.428	0.437	0.999	0.328	0.971	0.762	0.574	0.030	0.319	0.136	0.165
	rs418140899	1	616	0.785	0.596	0.427	0.435	0.998	0.328	0.972	0.669	0.574	0.030	0.318	0.136	0.164
	rs403942452	1	602	0.791	0.574	0.428	0.443	0.991	0.338	0.986	0.535	0.592	0.025	0.324	0.144	0.175
CSN1S2	rs398328750	2	612	0.244	0.837	0.009	0.025	0.083	0.377	0.447	0.374	0.261	0.151	0.677	0.754	0.928
	rs596875693	2	586	0.233	0.941	0.008	0.021	0.061	0.235	0.358	0.247	0.254	0.245	0.735	0.839	0.970
	rs424698365	2	611	0.369	0.124	0.126	0.208	0.568	0.195	0.070	0.256	0.913	0.077	0.132	0.165	0.436
	rs404755583	2	613	0.272	0.837	0.010	0.027	0.086	0.364	0.510	0.597	0.268	0.157	0.687	0.778	0.949
	rs398919262	2	612	0.407	0.135	0.456	0.617	0.535	0.127	0.094	0.673	0.918	0.067	0.176	0.260	0.610
	rs421554515	2	570	0.566	0.281	0.070	0.134	0.354	0.210	0.170	0.543	0.833	0.068	0.322	0.422	0.751
	rs410928088	2	539	0.421	0.395	0.082	0.142	0.448	0.280	0.297	0.379	0.877	0.052	0.225	0.399	0.740
CSN3	rs410816096	2	334	0.059	0.742	0.392	0.422	0.415	0.458	0.905	0.758	0.550	0.907	0.446	0.917	0.490
	rs598701385	2	469	0.134	0.786	0.879	0.870	0.821	0.845	0.122	0.787	0.779	0.375	0.064	0.024	0.055
	rs593666385	2	602	0.274	0.103	0.027	0.026	0.181	0.510	0.368	0.272	0.396	0.024	0.282	0.194	0.134
	rs605713973	2	612	0.190	0.038	0.178	0.248	0.043	0.346	0.418	0.807	0.682	0.963	0.745	0.870	0.896
	rs593210006	2	606	0.885	0.161	0.328	0.369	0.002	0.623	0.921	0.657	0.937	0.503	0.412	0.562	0.935

Significant P-values are evidenced in yellow.
NumDF: numerator degrees of freedom; DenDF: denominator degrees of freedom.
SCS: somatic cell score = log2 somatic cell count ×100,000-1+3.
LBC: logarithmic bacterial count = log10 total bacterial count + 3.
RCT, rennet coagulation time in minutes; k20, curd firming time, min; a30, a45 and a60 curd firmness in mm, respectively 30, 45 and 60 minutes after rennet addition.
DIM: days in milking as the measure of the stage of lactation.

Supplementary File Table S5 (Sheet 2). *P*-values for milk parameters (daily milk yield, milk composition and coagulation properties) according to each of the 24 polymorphic SNPs out of the investigated 32 SNPs at the casein and whey protein genes. Sheet 2: Farm, parity and DIM effects.

		Daily milk yield			Fat %			Protein %			Casein %			Lactose %			ph			SCS			LBC			RCT			k20			a30			a45			a60		
Gene	SNP_name	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM	Farm	Parity	DIM
BLG	rs416633597	0.001	0.001	0.001	0.001	0.262	0.001	0.001	0.507	0.001	0.001	0.450	0.001	0.001	0.001	0.001	0.001	0.845	0.004	0.001	0.001	0.427	0.001	0.827	0.303	0.001	0.190	0.001	0.001	0.003	0.076	0.001	0.107	0.337	0.001	0.843	0.405	0.001	0.923	0.446
LALBA	rs411787522	0.001	0.001	0.001	0.001	0.211	0.001	0.001	0.471	0.001	0.001	0.438	0.001	0.001	0.001	0.001	0.001	0.906	0.002	0.001	0.001	0.438	0.001	0.851	0.370	0.001	0.239	0.001	0.001	0.003	0.089	0.001	0.116	0.438	0.001	0.836	0.489	0.001	0.933	0.451
CSN1S1	rs419330297	0.001	0.001	0.001	0.001	0.320	0.001	0.001	0.521	0.001	0.001	0.469	0.001	0.001	0.001	0.001	0.001	0.848	0.003	0.001	0.001	0.369	0.001	0.856	0.354	0.001	0.286	0.001	0.001	0.002	0.068	0.001	0.113	0.261	0.001	0.832	0.340	0.001	0.924	0.422
	rs400042158	0.001	0.001	0.001	0.001	0.253	0.001	0.001	0.414	0.001	0.001	0.387	0.001	0.001	0.001	0.001	0.001	0.763	0.003	0.001	0.001	0.263	0.001	0.847	0.357	0.001	0.382	0.001	0.001	0.009	0.057	0.001	0.072	0.316	0.001	0.898	0.416	0.001	0.955	0.416
	rs416208751	0.001	0.001	0.001	0.001	0.228	0.001	0.001	0.475	0.001	0.001	0.445	0.001	0.001	0.001	0.001	0.001	0.757	0.003	0.001	0.001	0.375	0.001	0.858	0.340	0.001	0.222	0.001	0.001	0.002	0.090	0.001	0.106	0.407	0.001	0.848	0.474	0.001	0.922	0.499
	rs410096855	0.001	0.001	0.001	0.001	0.228	0.001	0.001	0.449	0.001	0.001	0.414	0.001	0.001	0.001	0.001	0.001	0.823	0.004	0.001	0.001	0.438	0.001	0.857	0.317	0.001	0.225	0.001	0.001	0.003	0.071	0.001	0.108	0.298	0.001	0.821	0.349	0.001	0.930	0.417
	rs398437598	0.001	0.001	0.001	0.001	0.208	0.001	0.001	0.534	0.001	0.001	0.478	0.001	0.001	0.001	0.001	0.001	0.857	0.004	0.001	0.001	0.393	0.001	0.852	0.332	0.001	0.231	0.020	0.001	0.002	0.062	0.001	0.088	0.289	0.001	0.803	0.344	0.001	0.913	0.435
CSN2	rs425554254	0.001	0.001	0.001	0.001	0.284	0.001	0.001	0.793	0.001	0.001	0.752	0.001	0.001	0.001	0.001	0.001	0.649	0.002	0.001	0.001	0.205	0.001	0.849	0.370	0.001	0.337	0.001	0.001	0.002	0.085	0.001	0.117	0.357	0.001	0.805	0.403	0.001	0.921	0.495
	rs411318553	0.001	0.001	0.001	0.001	0.242	0.001	0.001	0.427	0.001	0.001	0.395	0.001	0.001	0.001	0.001	0.001	0.842	0.003	0.001	0.001	0.421	0.001	0.855	0.307	0.001	0.229	0.001	0.001	0.002	0.090	0.001	0.099	0.382	0.001	0.847	0.459	0.001	0.918	0.508
	rs405274538	0.001	0.001	0.001	0.001	0.215	0.001	0.001	0.598	0.001	0.001	0.543	0.001	0.001	0.001	0.001	0.001	0.837	0.003	0.001	0.001	0.399	0.001	0.853	0.327	0.001	0.221	0.001	0.001	0.002	0.077	0.001	0.099	0.346	0.001	0.829	0.418	0.001	0.912	0.482
	rs418140899	0.001	0.001	0.001	0.001	0.218	0.001	0.001	0.596	0.001	0.001	0.541	0.001	0.001	0.001	0.001	0.001	0.836	0.003	0.001	0.001	0.401	0.001	0.063	0.366	0.001	0.222	0.001	0.001	0.002	0.076	0.001	0.098	0.347	0.001	0.828	0.418	0.001	0.911	0.483
	rs403942452	0.001	0.001	0.001	0.001	0.274	0.001	0.001	0.485	0.001	0.001	0.453	0.001	0.001	0.001	0.001	0.001	0.799	0.003	0.001	0.001	0.328	0.001	0.890	0.289	0.001	0.305	0.001	0.001	0.003	0.068	0.001	0.114	0.348	0.001	0.810	0.468	0.001	0.855	0.513
CSN1S2	rs398328750	0.001	0.001	0.001	0.001	0.200	0.001	0.001	0.513	0.001	0.001	0.460	0.001	0.001	0.001	0.001	0.001	0.883	0.004	0.001	0.001	0.415	0.001	0.794	0.341	0.001	0.237	0.001	0.001	0.003	0.073	0.001	0.095	0.317	0.001	0.770	0.366	0.001	0.889	0.438
	rs596875693	0.001	0.001	0.001	0.001	0.233	0.001	0.001	0.535	0.001	0.001	0.486	0.001	0.001	0.001	0.001	0.001	0.856	0.003	0.001	0.001	0.365	0.001	0.775	0.324	0.001	0.251	0.001	0.001	0.002	0.077	0.001	0.103	0.310	0.001	0.801	0.349	0.001	0.903	0.420
	rs424698365	0.001	0.001	0.001	0.001	0.232	0.001	0.001	0.510	0.001	0.001	0.463	0.001	0.001	0.001	0.001	0.001	0.815	0.002	0.001	0.001	0.394	0.001	0.847	0.378	0.001	0.221	0.001	0.001	0.003	0.085	0.001	0.106	0.343	0.001	0.825	0.395	0.001	0.937	0.401
	rs404755583	0.001	0.001	0.001	0.001	0.288	0.001	0.001	0.433	0.001	0.001	0.390	0.001	0.001	0.001	0.001	0.001	0.874	0.001	0.001	0.001	0.402	0.001	0.762	0.312	0.001	0.401	0.001	0.001	0.006	0.085	0.001	0.099	0.346	0.001	0.837	0.425	0.001	0.901	0.424
	rs398919262	0.001	0.001	0.001	0.001	0.189	0.001	0.001	0.560	0.001	0.001	0.506	0.001	0.001	0.001	0.001	0.001	0.861	0.002	0.001	0.001	0.374	0.001	0.839	0.375	0.001	0.216	0.001	0.001	0.003	0.072	0.001	0.099	0.338	0.001	0.814	0.389	0.001	0.918	0.427
	rs421554515	0.001	0.001	0.001	0.001	0.175	0.001	0.001	0.473	0.001	0.001	0.430	0.001	0.001	0.001	0.001	0.001	0.798	0.005	0.001	0.001	0.524	0.001	0.830	0.332	0.001	0.245	0.003	0.001	0.005	0.132	0.001	0.182	0.436	0.001	0.938	0.425	0.001	0.936	0.532
	rs410928088	0.001	0.001	0.001	0.001	0.293	0.001	0.001	0.434	0.001	0.001	0.412	0.001	0.001	0.001	0.001	0.001	0.744	0.004	0.001	0.001	0.154	0.001	0.845	0.440	0.001	0.288	0.001	0.001	0.007	0.251	0.001	0.122	0.338	0.001	0.924	0.376	0.001	0.973	0.278
CSN3	rs410816096	0.001	0.001	0.001	0.001	0.230	0.001	0.001	0.156	0.001	0.001	0.119	0.001	0.001	0.001	0.001	0.001	0.912	0.050	0.001	0.001	0.084	0.001	0.792	0.302	0.001	0.497	0.020	0.001	0.314	0.327	0.001	0.612	0.351	0.001	0.918	0.693	0.001	0.843	0.644
	rs598701385	0.001	0.001	0.001	0.001	0.400	0.001	0.001	0.672	0.001	0.001	0.688	0.001	0.001	0.001	0.001	0.001	0.478	0.003	0.001	0.001	0.392	0.001	0.861	0.340	0.001	0.386	0.001	0.001	0.029	0.035	0.001	0.210	0.172	0.001	0.657	0.114	0.001	0.937	0.206
	rs593666385	0.001	0.001	0.001	0.001	0.266	0.001	0.001	0.443	0.001	0.001	0.405	0.001	0.001	0.001	0.001	0.001	0.836	0.003	0.001	0.001	0.330	0.001	0.852	0.261	0.001	0.244	0.001	0.001	0.004	0.042	0.001	0.133	0.266	0.001	0.899	0.342	0.001	0.951	0.373
	rs605713973	0.001	0.001	0.001	0.001	0.178	0.001	0.001	0.482	0.001	0.001	0.443	0.001	0.001	0.001	0.001	0.001	0.879	0.003	0.001	0.001	0.370	0.001	0.780	0.334	0.001	0.177	0.001	0.001	0.003	0.072	0.001	0.120	0.301	0.001	0.862	0.359	0.001	0.921	0.435
	rs593210006	0.001	0.001	0.001	0.001	0.196	0.001	0.001	0.467	0.001	0.001	0.438	0.001	0.001	0.001	0.001	0.001	0.861	0.004	0.001	0.001	0.431	0.001	0.829	0.389	0.001	0.214	0.001	0.001	0.002	0.073	0.001	0.122	0.319	0.001	0.870	0.390	0.001	0.909	0.460

Significant P-values are evidenced in yellow.
NumDF: numerator degrees of freedom; DenDF: denominator degrees of freedom.
SCS: somatic cell score = log2 somatic cell count ×100,000-1+3.
LBC: logarithmic bacterial count = log10 total bacterial count + 3.
RCT, rennet coagulation time in minutes; k20, curd firming time, min; a30, a45 and a60 curd firmness in mm, respectively 30, 45 and 60 minutes after rennet addition.
DIM: days in milking as the measure of the stage of lactation.

Supplementary File Table S6 Common ovine casein haplotypes defined by polymorphisms mapping to each linkage disequilibrium block (n=760).

H	Blocks and SNPs					n	frequenc
	Block 1						
	rs41933029	rs40004215					
1	A	T				14	**0.191**
2	A	C				38	**0.050**
3	G	T				92	**0.121**
	Block 2						
	rs41009685	rs39843759	rs39832875	rs59687569			
4	A	A	A	G		5	0.007
5	G	A	A	G		-	0.000
6	G	G	A	G		17	**0.229**
7	G	G	G	A		13	**0.183**
	Block 3						
	rs42469836	rs40475558	rs39891926	rs42155451	rs41092808		
8	A	T	A	G	G	28	0.037
9	C	G	G	A	C	13	**0.175**
1	C	T	G	A	C	90	**0.118**
	Block 4						
	rs59366638	rs60571397	rs59321000				
1	A	C	A			3	0.004
1	A	C	C			29	0.038
1	A	T	A			20	**0.271**
1	T	T	A			39	**0.051**

H, Haplotype numbering; n, number of ewes carrying that haplotype.
Haplotype frequencies higher than 0.05 are in bold and consequently submitted to association analysis with milk traits.

Supplementary File Table S7 (Sheet 1). *P*-values for milk parameters (daily milk yield, milk composition and coagulation properties) according to each of the 4 blocks of SNPs at the casein genes. Sheet 1: Haplotype effect.

			Daily milk yield	Fat %	Protein %	Casein %	Lactose %	ph	SCS	LBC	RCT	k20	a30	a45	a60
Blocks	NumDF	DenDF	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value
Block 1	2	155	0.646	0.905	0.947	0.936	0.727	0.484	0.288	0.053	0.323	0.092	0.276	0.447	0.703
Block 2	2	193	0.163	0.819	0.156	0.224	0.058	0.053	0.571	0.568	0.069	0.071	0.512	0.283	0.659
Block 3	2	139	0.355	0.628	0.014	0.028	0.063	0.926	0.073	0.990	0.288	0.379	0.765	0.771	0.496
Block 4	3	158	0.619	0.249	0.104	0.121	0.094	0.404	0.817	0.503	0.035	0.805	0.242	0.809	0.768

Significant *P*-values are evidenced in yellow.
NumDF: numerator degrees of freedom; DenDF: denominator degrees of freedom.
SCS: somatic cell score = $\log_2$ somatic cell count $\times 100{,}000^{-1}+3$.
LBC: logarithmic bacterial count = $\log_{10}$ total bacterial count + 3.
RCT, rennet coagulation time in minutes; k20, curd firming time, min; a30, a45 and a60 curd firmness in mm, respectively 30, 45 and 60 minutes after rennet addition.
DIM: days in milking as the measure of the stage of lactation.

Supplementary File Table S7 (Sheet 2). *P*-values for milk parameters (daily milk yield, milk composition and coagulation properties) according to each of the 4 blocks of SNPs at the casein genes. Sheet 2: Farm, parity and DIM effects.

	Daily milk yield			Fat %			Protein %			Casein %			Lactose %			ph			SCS			LBC		
Blocks	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**
Block 1	0.002	0.001	0.001	0.540	0.552	0.002	0.005	0.095	0.001	0.010	0.927	0.001	0.042	0.025	0.107	0.741	0.769	0.687	0.011	0.033	0.316	0.002	0.753	0.383
Block 2	0.032	0.066	0.001	0.185	0.749	0.017	0.063	0.942	0.001	0.052	0.865	0.001	0.287	0.040	0.083	0.031	0.740	0.707	0.346	0.057	0.045	0.001	0.926	0.930
Block 3	0.035	0.012	0.001	0.197	0.851	0.007	0.079	0.272	0.003	0.054	0.308	0.003	0.492	0.405	0.258	0.008	0.950	0.209	0.230	0.233	0.186	0.001	0.803	0.543
Block 4	0.001	0.001	0.001	0.181	0.155	0.001	0.001	0.068	0.001	0.001	0.045	0.001	0.022	0.312	0.068	0.238	0.497	0.205	0.145	0.018	0.680	0.001	0.308	0.492

	RCT			k20			a30			a45			a60		
Blocks	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**	**Farm**	**Parity**	**DIM**
Block 1	0.049	0.178	0.536	0.925	0.101	0.129	0.008	0.464	0.204	0.001	0.946	0.968	0.050	0.324	0.976
Block 2	0.420	0.955	0.662	0.162	0.339	0.500	0.001	0.804	0.830	0.945	0.819	0.009	0.001	0.547	0.910
Block 3	0.711	0.769	0.106	0.154	0.710	0.293	0.001	0.639	0.545	0.378	0.353	0.001	0.002	0.545	0.724
Block 4	0.157	0.920	0.195	0.701	0.662	0.729	0.001	0.169	0.268	0.128	0.123	0.001	0.001	0.068	0.421

Significant *P*-values are evidenced in yellow.
NumDF: numerator degrees of freedom; DenDF: denominator degrees of freedom.
SCS: somatic cell score = $\log_2$ somatic cell count $\times 100{,}000^{-1}+3$.
LBC: logarithmic bacterial count = $\log_{10}$ total bacterial count + 3.
RCT, rennet coagulation time in minutes; k20, curd firming time, min; a30, a45 and a60 curd firmness in mm, respectively 30, 45 and 60 minutes after rennet addition.
DIM: days in milking as the measure of the stage of lactation.

RUNNING HEAD: Spanish sheep ovine diversity

Estimating the population structure and the amount of shared variation amongst Spanish ovine breeds

A. Noce[1,9], A. Cánovas[1], A. Manunza[1], J. Casellas[2], A. Martínez[3], A. Pons[4], L.A. Bermejo[5], V. Landi[3], A. Sànchez[1,2], J. Jordana[2], J.V. Delgado[3], S. Adán[6], J. Capote[7], O. Vidal[8], M. Pazzola[9], G.M. Vacca[9] , A. Martínez[3], and M. Amills[1,2]

[1]Department of Animal Genetics, Center for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB), Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain; [2]Departamento de Genética, Universidad de Córdoba, Córdoba 14071, Spain; [3]Unitat de Races Autòctones, Servei de Millora Agrària i Pesquera (SEMILLA), Son Ferriol 07198, Spain; [4]Departamento de Ingeniería, Producción y Economía Agrarias, Universidad de La Laguna, 38071 La Laguna, Tenerife, Spain; [5]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain; [6]Federación de Razas Autóctonas de Galicia (BOAGA), Pazo de Fontefiz, 32152 Coles. Ourense, Spain; [7]Instituto Canario de Investigaciones Agrarias, La Laguna 38108, Tenerife, Spain; [8]Departament de Biologia, Universitat de Girona, Girona 17071, Spain; [9]Dipartimento di Medicina Veterinaria,Università degli Studi di Sassari, via Vienna,2 07100 Sassari,Italy

Abstract

We have characterized the autosomal variability of six Spanish ovine breeds (Xisqueta, Ripollesa, Segureña, Gallega, Roja Mallorquina and Canaria de Pelo) by genotyping 140 individuals with the Infinium 50K Ovine SNP BeadChip (Illumina). This dataset was merged with data from the International Sheep Genomics Consortium (Churra, Ojalada, Castellana, Latxa and Rasa Aragonesa). Multidimensional scaling and Admixture analyses of these 11 ovine Spanish breeds revealed that Canaria de Pelo and, to a lesser extent, Roja Mallorquina, Latxa and Churra are clearly differentiated populations, while the remaining seven breeds are almost indistinguishable from each other. In addition, we have generated RNA-Seq data from Longissimus dorsi muscle tissue of a subset of five breeds (each one represented by a pool of 10 individuals) to infer how much of the variation found is breed-specific. The alignment of sequence reads against the sheep reference genome revealed that the amount of shared (~184,000 SNPs) vs pool-specific (~146,000 SNPs) variants happened to be quite similar. Moreover, the number of pool-specific single nucleotide polymorphisms (SNPs) in Canaria de Pelo (56,116 SNPs) and Xisqueta (45,826 SNPs) was substantially higher than in Gallega (9,384 SNPs), Ripollesa (15,817 SNPs) and Roja Mallorquina (19,143 SNPs). These results suggest that a significant part of the variation of Spanish sheep populations might be breed-specific, emphasizing the need of implementing conservation programmes that ensure the maintenance and management of these valuable and unique genetic resources.

Introduction

Sheep are an important source of milk, meat and wool, particularly in Mediterranean countries where they represent an essential agricultural asset. In Spain, there are 43 officially recognized breeds and 33 of them have an endangered status. The demographic decline of certain Spanish breeds has been caused by competition and replacement with highly productive sheep from foreign countries, the progressive abandonment of low income rural activities and geographical confinement (Taberlet et al. 2008). Genetic analyses with mitochondrial (Pedrosa et al. 2007) and microsatellite (Arranz et al. 2001, Calvo et al. 2011) markers have demonstrated that, in general, Spanish ovine breeds display a high diversity (expected heterozygosities around 0.75) and a weak to moderate population structure ($F_{ST} \sim 0.07$).

More recently, the development of high throughput SNP typing platforms and massive sequencing methods have allowed to analyse the population genetics of livestock breeds with an unparalleled resolution. Such approaches have been used to characterize the diversity of sheep at a worldwide scale, observing genomic patterns compatible with high levels of admixture and recent selection (Kijas et al. 2012). However, a detailed analysis of the autosomal variation of a representative number of Spanish ovine breeds is still lacking. In this regard, it would be particularly interesting to find out how much of this variation is shared amongst breeds, because a central argument justifying their conservation is that they may harbour unique genetic variants with significant effects on adaptation traits as well as on phenotypes of economic interest. In the current work, we have investigated the levels of

population structure and diversity of 11 Spanish ovine breeds (N=369) by using genotyping data generated with the Infinium 50K Ovine SNP BeadChip. Moreover, we have used a RNA-Seq approach to characterize the variation expressed in the *Longissimus dorsi* muscle tissue of a subset of 5 breeds (each one represented by a pool of 10 individuals) with an endangered status (Xisqueta, Ripollesa, Segureña, Gallega, Roja Mallorquina and Canaria de Pelo). Skeletal muscle is an economically relevant tissue because all breeds under consideration are specialized in meat production. Our goal was to obtain an approximate estimate of the amount of shared *vs* non-shared diversity amongst these five populations.

Materials and Methods.

Sampling and genomic DNA purification and genotyping

Blood samples from Segureña (N=12), Xisqueta (N=24), Ripollesa (N=23), Gallega (N=25), Canaria de Pelo (27) and Roja Mallorquina (N=29) sheep were extracted with Vacutainer tubes. Genomic DNA was purified from leukocytes using a phenol-chloroform method. In this way, 500 µl of whole blood were washed with TE buffer (Tris 10 mM, EDTA 1 mM, pH 8.0) and vortexed and centrifuged at 13,000 rpm for 30 seconds. This procedure was repeated until a clean white pellet was obtained. Afterwards the cell pellet was resuspended in 200 µl cell lysis buffer (50 mM KCl, 10 mM Tris, 0.5% Tween 20) with 10 µl proteinase K (10 mg/ml) and incubated for 4 hours at 56°C. One volume of phenol:chloroform:isoamyl alcohol (25:24:1) was

added to the lysate, and the resulting mixture was vortexed and centrifuged at 13,000 rpm for 15 min. Next, the aqueous upper layer was transferred to a fresh tube and NaCl 2M (0.1 volumes) and absolute ethanol (2 volumes at -20°C) were added. Subsequently, this mixture was centrifuged at 13,000 rpm for 30 min, the supernatant was discarded and the DNA pellet was washed with 500 µl 70% ethanol (15 min at 13,000 rpm) to remove salt contamination. Finally, the DNA pellet was air-dried at room temperature, and resuspended in 50 µl milli-Q water. Genomic DNA samples were genotyped with the Illumina Infinium 50K Ovine SNP BeadChip (Illumina, San Diego, CA) following the instructions of the manusfacturer (http://www.illumina.com).

Muscle biopsy collection, RNA isolation and library construction and sequencing.

Biopsies from *Longissimus dorsi* muscle tissue were retrieved from Ripollesa, Canaria de Pelo, Roja Mallorquina, Xisqueta and Gallega sheep (10 individuals per population). Muscle biopsies were placed into RNAlater (Ambion, Austin, TX). and stored at -20ºC until use. Total RNA was extracted using the RiboPure RNA Purification kit (Ambion) following the recommendations of the manufacturer. Total RNA concentration was estimated with a Nano-Drop ND-1000 spectrophotometer (NanoDrop products; Wilmington, USA). Quality of the total RNA was evaluated using the RNA Integrity Number (RIN) value in the Agilent Bioanalyzer 2100 (Agilent Technologies, Inc., Santa Clara CA, USA). All samples showed RIN values

greater than 7. Subsequently, RNA pools derived from each one of the five populations were made.

Sequencing libraries (five cDNA pools) were made using the TrueSeq RNA Sample Preparation kit (Illumina, San Diego, CA). Starting with 1 μg of total RNA, mRNA was purified, fragmented and converted to cDNA. Adapters were ligated to the ends of double-stranded cDNA and PCR amplified to create libraries (Canovas et al. 2010). RNA sequencing was performed on a HiSeq 2000 sequencer analyzer (Illumina, San Diego, CA) at the Centre Nacional d'Anàlisi Genòmica (Barcelona, Spain).

Population genetics analyses

The GenomeStudio software (Illumina) was used to generate standard ped and map files as well as to perform sample and marker-based quality control measures. We considered a GenCall score cutoff of 0.15 and an average call rate of 99%. Before the filtering step of non-informative SNPs, we merged our dataset with another 50K dataset kindly provided by the International Sheep Genomics Consortium (IGGC, http://www.sheephapmap.org) and composed by 229 sheep belonging to the Ojalada (N=24), Castellana (N=23), Rasa Aragonesa (N=22), Churra (N=120) and Latxa (N=40) breeds. The PLINK v1.07 software (Purcell et al., 2007) was used to merge the data and subsequently we proceeded to exclude unmapped markers as well as those mapping to sexual chromosomes. We also removed SNPs with minor allele frequencies (MAF) below 0.05 and those deviating significantly (*P*-value ≤ 0.001)

from the Hardy-Weinberg expectation. The PLINK v1.07 program was also used to perform a multidimensional scaling analysis (MDS) based on a matrix of genome-wide pairwise identity-by-state distances (Purcell et al. 2007). Besides, we carried out a clustering analysis with Admixture v1.23, which calculates maximum likelihood estimates of individual ancestries based on data provided by multiple loci (Alexander et al 2009). Finally, pairwise F_{ST} values were obtained with Arlequin 3.5 (Excoffier and Lischer, 2010).

Bioinformatic analysis

Sequence paired-end reads (75 bp) were assembled on the annotated ovine reference genome Oar_v3.1.75 (http://www.ncbi.nlm.nih.gov/genome/guide/sheep/index.html) using the CLC Genomics Workbench software v7.5.0 (CLC Bio, Aarhus, Denmark). Quality control analysis was performed using the application NGS quality control tool of CLC Genomics workbench (Cánovas et al. 2014). This tool assesses sequence quality indicators based on the FastQC-project (http:// www. bioinformatics. babraham. ac. uk/projects/fastqc/). Quality was measured taking into account sequence-read lengths and base-coverage, nucleotide contributions and base ambiguities, quality scores as emitted by the base caller and over-represented sequences. All samples passed the filtering criteria described by Cánovas et al. (2014).

CLC Genomics Workbench allowed us to make the automated detection of SNPs. The algorithm used is a combination of a Bayesian model and a maximum likelihood approach to calculate prior and error probabilities for the Bayesian model. The SNP

detection is based on the Neighborhood Quality Standard (NQS) algorithm described by Altschuler et al. (2000). More specifically, a statistical test is performed at each site to determine if the detection of a specific nucleotide is better explained by either a sequencing error or by the segregation of one (or more) alleles at some unknown frequency. If the latter is the case, a variant corresponding to the significant allele is called and its estimated frequency is calculated. To perform this statistical test, a threshold of required significance of 0.05 was considered. The parameters "Minimum average quality of surrounding bases" and "Minimum quality of the central base" were set to 30 (out of 42 on a Phred scale) while "Minimum coverage" and "Minimum variant frequency or count" were set to 10 reads and 2 read counts which represents the minimum number of reads that present the alternative allele, respectively (Canovas et al. 2010). In order to reduce the number of false positives, we filtered SNPs with a minimum allele frequency (MAF) below 0.05.

Pool-specific SNPs and SNPs shared across pools were studied. Afterward, SNPs which were predicted to generate a missense mutation or mapping to splice sites were selected using the "Amino acid changes analysis and predict splice sites effect" tool. This dataset was used to identify those variants that may have functional effects by employing the "Variant Effect Predictor" tool of Ensembl.

Results and Discussion

Analysis of the population structure of eleven Spanish ovine breeds

The MDS analysis of 11 Spanish sheep breeds with a wide geographic distribution (**Figure 1**) revealed that the Canaria de Pelo breed is highly differentiated from the remaining populations (**Figure 2A**). We also observed a scattered and divergent cluster represented by the Churra breed. The Roja Mallorquina and Latxa breeds also showed some level of genetic differentiation, while the remaining breeds were mixed in a single cluster and they could not be distinguished from each other. We made a second MDS analysis of this latter group of mixed breeds (**Figure 2B**), being able to distinguish the Gallega sheep from the remaining populations. These results were consistent with the Admixture analysis (**Figure 3**), which showed that the Canaria de Pelo, Roja Mallorquina, Latxa and Churra breeds have a well defined genetic identity. In contrast, the remaining breeds (Castellana, Ojalada, Rasa Aragonesa, Xisqueta, Ripollesa, Gallega and Segureña) share the same genetic background and they cannot be easily distinguished from each other. Consistent with this, the F_{ST} coefficients shown in **Supplementary Table 1** were low (around 0.04), evidencing that population structure within the Iberian Peninsula is remarkably weak. Similar estimates have been obtained in a worldwide analysis of ovine variation (Kijas et al. 2012), reflecting the existence of a common and recent ancestry amongst sheep breeds combined with frequent admixture events.

Canaria de Pelo is a hair sheep breed that, according to our study (**Figure 2A**), appears as strongly differentiated from its counterparts. Hair sheep are the most widespread race in Africa because of their excellent adaptation to the highly humid tropical forest (Blench 1993). Linguistic and genetic evidences also connect the aborigin Canarian population with the Imazighen peoples indigenous to North Africa

(Fregel et al. 2009). In consequence, we attribute the high genetic differentiation of the Canaria de Pelo sheep to the fact that it has an African rather than Iberian ancestry. Geographic isolation, until the discovery of the Canarian archipelago by the Spanish in the 15th century, combined with the occurrence of population bottlenecks may have also contributed to this accelerated process of genetic divergence (Delgado et al. 1998, Álvarez et al. 2013). Consistently, Canarian goat breeds also appear as highly differentiated when compared to caprine populations from Africa and Europe (Amills et al. 2013).

As previously said, other Spanish ovine breeds with a marked genetic identity are Roja Mallorquina, Churra and Latxa (**Figures 2 and 3**). Roja Mallorquina sheep display phenotypic features that are distinctive of certain breeds from North Africa and Asia such as a fat triangular tail and a red color. Indeed, fat-tailed sheep are particularly abundant in Lybia, Tunisia and Algeria and it is assumed that they were introduced from the Middle East (Blench 1993). Churra and Latxa also appear as clearly differentiated from the main group of Spanish breeds (Segureña, Gallega, Xisqueta, Ripollesa Ojalada, Rasa Aragonesa and Castellana). This finding agrees well with the classical phenotypic classification of Spanish ovine breeds proposed by Sanchez-Belda et al. (1989), where four main ovine lineages are distinguished: (1) Churro (Churra and Latxa breeds, that have a coarse wool), (2) Merino (not represented in our dataset), (3) Medium fine wool (Segureña, Gallega, Ripollesa, Rasa Aragonesa, Castellana and others), and (4) Iberian (Xisqueta, Ojalada and others). Our genetic data, however, do not support the existence of a substantial genetic divergence between Medium fine wool and Iberian sheep. As shown in

Figure 2B, Ojalada and Xisqueta sheep are not significantly differentiated from their Segureña, Ripollesa, Rasa Aragonesa and Castellana counterparts, suggesting that these populations belong to a single genetic lineage. It can be observed, however, that the Gallega breed shows a certain level of genetic divergence when compared to the remaining six breeds (**Figure 2B**).

Estimating the proportion of muscle transcriptome variation shared amongst breeds

In the current work, we have analysed the transcriptome variation of five endangered Spanish ovine breeds. Approximately 300 million of paired-end reads (75 bp) were produced for each population, and around 74% of them were successfully mapped to the ovine genome (**Supplementary Table 1**). The majority (80-90%) of these reads mapped to exons, though intronic reads were also identified. It is unlikely that the source of these intronic sequences could be contaminating genomic DNA because, during the process of cell lysis with the TRI Reagent (a solution of phenol, guanidine isothiocyanate and other components), RNA and DNA are efficiently separated in the aqueous and organic phases, respectively, and subsequently RNA is captured in a glass fiber filter column. Our view is that we have detected intron-derived transcripts that, according to recent studies, represent the major fraction of the non-coding RNA transcribed in mammalian cells (Laurent et al. 2012). These intronic transcripts are not necessarily a mere byproduct of the splicing machinery. Instead, a fraction of them may represent long non-coding RNAs involved in the regulation of gene

expression and other cellular processes (Louro et al. 2009). To avoid any potential bias, we have based our variation data analyses exclusively on exonic SNPs.

The main goal of the RNA-Seq experiment was to decompose the total variation observed in the five muscle cDNA pools into two components *i.e.* "variation shared amongst two or more pools" *vs* "variation exclusively present in a single pool". Approximately, 20% (36,846 SNPs) of the shared polymorphisms (184,536 SNPs) were present in all pools despite the heterogeneous origins of the analysed breeds (**Figure 4**). Probably, these findings reflect the fact that all sheep descend from a common ancestor represented by the mouflons domesticated in the Fertile Crescent 10,000 YBP (Zeder et al. 2006). While the pairwise comparison of sheep populations showed that the majority (70-90%) of SNPs of each specific breed were shared with other populations (**Table 1**), when we considered all populations globally (**Figure 4**), the amounts of shared (~184,000 SNPs) *vs* pool-specific (~146,000 SNPs) variants happened to be quite similar. These numbers are consistent with those estimated in European and Asian pigs and wild boars, where around 15 and 19.4 million of SNPs are population-specific and shared amongst populations, respectively (Bianco et al. 2015). Of course, the proportion of pool-specific SNPs estimated by us in Spanish sheep may decrease by sampling additional individuals and breeds. In this regard, the proportion of pool-specific SNPs that were present in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) fluctuated between 37% (Gallega) and 69% (Canaria de Pelo). These numbers highlight that a substantial part of the pool-specific variation detected is also present in other ovine populations and, in consequence, it cannot be considered as breed-specific. On the other hand, we

have purged from our dataset those SNPs with a MAF < 0.05. This is expected to result in the elimination of low-frequency alleles, that are much more prone to be population-specific than those with medium or high frequencies.

The weight of the pool-specific component varied substantially amongst ovine breeds. In this way, the number of pool-specific SNPs in Canaria de Pelo and Xisqueta was approximately 2-5 fold higher than in Gallega, Ripollesa and Roja Mallorquina. In humans, Choudhury et al. (2014) observed a 6-7 fold higher number of population-specific SNPs in the Kenyan Luhya population than in the Iberian and Finnish ones. Similarly, Bianco et al. (2015) identified around 8 and 0.9 million population-specific SNPs in Asian and European wild boars, respectively. Such differences can emerge for diverse reasons related with sample size, historical origin (age of the population and geographic distribution), demography (founder effects, bottlenecks etc.) and admixture (introgression of a breed with an unsampled distantly-related population may increase the population-specific component of such breed). Historical and genetic accounts indicate that Canaria de Pelo sheep may have been mixed with other African and Andalusian breeds brought by the Spanish conquerors several centuries ago, and more recently with hairless sheep from Venezuela (Delgado et al. 1998, Álvarez et al. 2012). This extensive admixture may explain the significant fraction of population-specific alleles detected. A high proportion of pool-specific SNPs in the Xisqueta sheep were also detected. This feature could be due to an introgression with an exotic breed not present in our dataset (*e.g.* Suffolk, Berrichon), although in principle selected individuals were classified as purebred. In contrast, the count of pool-specific SNPs of the Roja

Mallorquina breed, which has a mixed Iberian and African ancestry, is very similar to that observed in other peninsular ovine breeds as Ripollesa and Gallega.

It is difficult to infer how much of the variation identified as "pool-specific" is also "breed-specific". In theory, to define an allele as population-specific would imply to sample all the individuals of all potential populations to make sure that it just segregates in one particular population. Such experimental design would be completely unfeasible, so the population-specific component of variation is generally estimated in a reduced subset of populations and individuals. For instance, the population-specific component of human genetic variation was inferred in 14 populations represented by just ~ 100 individuals (Choudhury et al. 2014). In the present study, five ovine populations composed by 50 individuals were used. On a qualitative basis, our results suggest that a significant fraction of the variation characterized in the current work might be breed-specific. This is consistent with estimates obtained in humans, where 17% of low-frequency variants and 53% of rare alleles display population-specific patterns of segregation (The 1000 Genomes Project Consortium. 2012). Consistent findings have been obtained in a recent study comprising 128 pig genomes (Bianco et al. 2015), where the population-specific and the shared components of diversity appeared to have similar magnitudes. Population-specific SNPs generally correspond to rare and young variants with a restricted geographic distribution, so the recent creation of breeds may be a major reason explaining the existence of such variation in domestic animals. Indeed, the study of Bianco et al. (2015) reflects that domestic pìgs harbour 6.5 million SNPs that are not present in wild boars despite the fact that they diverged a short time ago (10,000

YBP). Other factors that may contribute to the maintenance of breed-specific SNPs are connected with the very same process of breed creation, often involving a few founder individuals and intensive artificial selection schemes focused on a few specialized breeding goals.

A relevant observation is that predictions made with the Sorting Tolerant from Intolerant (SIFT) algorithm (Kumar et al. 2009) indicate that part of the pool-specific missense variation detected may have consequences on protein function (**Table 2**). For instance, one L382P polymorphism was identified in the calpain 7 gene (*CAPN7*) that only segregates in the Gallega pool and another one (W135R) in the calpain 5 (*CAPN5*) gene that is Ripollesa pool-specific. Calpains form part of a proteolytic enzymatic system that plays a major role in meat tenderization (Lian et al. 2013). Similarly, 8 missense and potentially deleterious mutations were detected in three myosin heavy chain genes (*MYH3*, *MYH8* and *MYH11*) that segregate in a pool-specific manner. In pigs, myosin heavy chain isoform abundance and function may have broad effects on many parameters related with meat quality, such as juiciness, flavor, acidity, tenderness and drip loss (Chang et al. 2003, Kang et al. 2011). Another locus of interest is the cathepsin (*CTSO*) gene, a major player in muscle proteolysis (O'Halloran et al. 1997) that contains one L143S substitution only present in the Ripollesa pool and predicted by SIFT to have functional consequences. In pigs, 27% and 21% of the missense genetic variation of European and Asian pigs, respectively, is classified by SIFT as deleterious (Bianco et al. 2015), meaning that it may have significant phenotypic effects. Obviously, this could be the tip of the iceberg because there might be also population-specific variants with regulatory

effects that cannot be accurately predicted with current bioinformatic tools. Such variation may explain, for instance, the existence of population-specific patterns of gene expression (Li et al. 2010) and genome methylation in humans (Fraser et al. 2012).

A central argument justifying the genetic conservation of local breeds is that they may carry variants with positive effects on traits of economic or adaptative interest that would be lost upon extinction. In our view, genomic tools are providing mounting evidence that this statement is probably true. Indeed, large-scale whole genome studies performed in humans (The 1000 Genomes Project Consortium. 2012) and pigs (Bianco et al. 2015), as well as results provided in the current study, consistently indicate that the fraction of population-specific variation might be considerably large, and this would be particularly true for low frequency variants. Besides, part of this breed-specific variation is predicted to be deleterious, suggesting that it could have a significant phenotypic impact (though functional studies would be needed to confirm it). The existence of local breeds is constantly threatened by the progressive implantation of intensive management practices as well as by their substitution by more productive, but less adapted, industrial varieties. Unless urgent measures are taken, this will probably result in the irreversible erosion, or even the extinction, of unique genetic resources that, throughout the ages, have formed an essential part of the economy and culture of rural communities.

ACKNOWLEDGMENTS

This research was partially funded by a grant (RZ2011-00015-C03-01) from the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (Spain). A. Cánovas received a fellowship from the Juan de la Cierva program (JCI-2011-10804 - A. Cánovas)

REFERENCES

1000 Genomes Project Consortium, Abecasis G.R., Auton A., Brooks L.D., DePristo M.A., Durbin R.M., Handsaker R.E., Kang H.M., Marth G.T., McVean G.A. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56-65.

Alexander D.H., Novembre J., Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19: 1655-64.

Álvarez I., Capote J., Traoré A., Fonseca N., Pérez K., Cuervo M., Fernández I., Goyache F. 2013. Mitochondrial analysis sheds light on the origin of hair sheep. Anim Genet. 44: 344-7.

Amills M., Martínez A., Manunza A., Landi V., Santana D.M., Badaoui B., Ferrando A., Jordana J., Camacho M.E., Capote J., Fernandez G., Gómez M., Pons A., Vidal O. and Delgado J.V. 2013. Documenting population structure and gene flow between goats from Spain and Africa. XXI Plant and Animal Genome Conference, San Diego, California.

Altshuler D., Pollara V. J., Cowles C. R., Van Etten W. J., Baldwin J., Linton L., Lander E. S. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407: 513-516

Arranz J.J., Bayón Y., San Primitivo F. 2001. Differentiation among Spanish sheep breeds using microsatellites. Genet Sel Evol. 33: 529-42.

Bianco E., Nevado B., Ramos-Onsins S.E., Pérez-Enciso M. 2015. A deep catalog of autosomal single nucleotide variation in the pig.PLoS One 10(3): e0118867.

Blench R. 1993. Ethnographic and linguistic evidence for the prehistory of African ruminant livestock, horses and ponies. In: The archaeology of Africa: food, metals and towns (Eds. Shaw T, Sinclair P, Andah B and Okpoko A), Routledge, Abingdon UK, pp.78-81.

Calvo J.H., Alvarez-Rodríguez J., Marcos-Carcavilla A., Serrano M., Sanz A., 2011. Genetic diversity in the Churra tensina and Churra lebrijana endangered Spanish sheep breeds and relationship with other Churra group breeds and Spanish mouflon. Small Rumin Res 95: 34-39.

Canovas A., Rincon G., Trejo A.I., Wickramasinghe S., Medrano J.F. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mamm Genome. 21: 592–598.

Cánovas, A., Rincón G., Bevilacqua C., Islas-Trejo A., Brenaut P., Hovey R.C., Boutinaud M., Morgenthaler C., VanKlompenberg M.K., Martin P., Medrano J.F. 2014. Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. Sci. Rep. 4, 5297;

Chang K.C., da Costa N., Blackley R., Southwood O., Evans G., Plastow G., Wood J.D., Richardson R.I. 2003. Relationships of myosin heavy chain fibre types to meat quality traits in traditional and modern pigs. Meat Sci. 64: 93-103.

Choudhury A., Hazelhurst S., Meintjes A., Achinike-Oduaran O., Aron S., Gamieldien J., Jalali Sefid Dashti M., Mulder N., Tiffin N., Ramsay M. 2014. Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. BMC Genomics 15:437.

Delgado J.V., Fresno M.R., Camacho M.E., Rodero E., Barba C. 1998. Origin and influences of canarian sheep. Arch. Zootec. 47: 511-516.

Excoffier L., Lischer H.E.L. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res.10:564-7.

Fraser H. B., Lam L.L., Neumann S.M., Kobor M.S. 2012. Population-specificity of human DNA methylation. Genome Biol. 13(2):R8.

Fregel R., Pestano J., Arnay M., Cabrera V.M., Larruga J.M., González A.M. 2009. The maternal aborigine colonization of La Palma (Canary Islands). Eur J Hum Genet. 17: 1314-24.

Kang Y.K., Choi Y.M., Lee S.H., Choe J.H., Hong K.C., Kim B.C. 2011.Effects of myosin heavy chain isoforms on meat quality, fatty acid composition, and sensory evaluation in Berkshire pigs. Meat Sci. 89: 384-9.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B. International Sheep Genomics Consortium Members. 2012. Genome-wide analysis of the world's sheep

breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10: e1001258.

Kumar P., Henikoff S., Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 4: 1073-81.

Laurent G.S., Shtokalo D., Tackett M.R., Yang Z., Eremina T., Wahlestedt C., Urcuqui-Inchima S., Seilheimer B.,. McCaffrey T. A, Kapranov P. 2012. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. BMC Genomics. 13:504.

Li J., Liu Y., Kim T., Min R., Zhang Z. 2010. Gene expression variability within and between human populations and implications toward disease susceptibility.PLoS Comput Biol. 6(8), DOI: 10.1371/journal.pcbi.1000910.

Lian T., Wang L., Liu Y. 2013. A New Insight into the Role of Calpains in Post-mortem Meat Tenderization in Domestic Animals: A review. Asian-Australas J Anim Sci. 26: 443-54.

Louro R., Smirnova A.S., Verjovski-Almeida S. 2009. Long intronic noncoding RNA transcription: expression noise or expression choice . Genomics. 93: 291-8.

O'Halloran G.R., Troy D.J., Buckley D.J., Reville W.J. 1997. The role of endogenous proteases in the tenderisation of fast glycolysing muscle. Meat Sci. 47: 187-210.

Pedrosa S., Arranz J.J, Brito N., Molina A., San Primitivo F., Bayón Y. 2007. Mitochondrial diversity and the origin of Iberian sheep. Genet Sel Evol. 39: 91-103.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A, Bender D., et al. 2007 "PLINK: a toolset for whole-genome association and population-based linkage analysis". Am J Hum Genet. 81:559-75.

Sánchez-Belda A. 1989. Razas ovinas. Orígenes y aptitudes. Ovis 5: 9-19.

Taberlet P., Valentini A., Rezaei H.R., Naderi S., Pompanon F., Negrini R., Ajmone-Marsan P. 2008. Are cattle, sheep, and goats endangered species Mol Ecol. 17: 275-84.

Zeder M.A., Emshwiller E., Smith B.D., Bradley D.G. 2006. Documenting domestication: the intersection of genetics and archaeology. Trends Genet. 22: 139-55.

Table 1. Magnitude and nature of pool-specific variation in five Spanish ovine breeds

Breed	**Total SNPs**	**Pool-specific SNPs**				
		Total	**Genic**	**Intergenic**	**Missense**	**Splice effects**
Ripollesa	119,405	15,817	14,445	1,370	3,728	116
Xisqueta	200,271	45,826	41,902	3,394	5,874	143
Canaria de Pelo	201,215	56,116	52,943	3,173	4,276	149
Roja Mallorquina	145,981	19,143	17,488	1,655	6,383	96
Gallega	84,411	9,348	8,313	1,033	3,095	75

Table 2. Ovine pool-specific missense polymorphisms mapping to meat quality genes

Gene	Polymorphism	SIFT score	Count	Coverage	Frequency(%)	Breed
CAPN5	W135R	Deleterious(0)	3	29	10,34	Ripollesa
CAPN7	L382P	Deleterious(0.01)	2	21	9,52	Gallega
CAST	R449Q	Tolerated(0.69)	54	295	18,31	Roja Mallorquina
CTSO	105 W/*	Deleterious(0)	2	38	5,26	Canaria
	L143S	Deleterious(0)	2	34	5,88	Ripollesa
IGFBP2	S8T	Tolerated(1)	14	14	100	Ripollesa
	P10R	Tolerated(1)	19	19	100	
MYH3	G378S	Deleterious(0.01)	2	12	16,67	Roja Mallorquina
	K504R	Tolerated(0.80)	2	10	20	
	E929D	Tolerated(0.15)	48	52	92,31	Xisqueta
	A1892P	Deleterious(0.04)	2	24	8,33	Roja Mallorquina
MYH8	G740E	Deleterious(0.03)	2	26	7,69	Canaria
	A929V	Tolerated(0.57)	4	40	10	
	E1124D	Tolerated(0.40)	3	35	8,57	Roja Mallorquina
	L1125R	Deleterious(0)	3	38	7,89	
	R1185H	Tolerated(1)	2	14	14,29	Gallega
	L1207I	Tolerated(1)	3	23	13,04	Canaria
	I1228L	Tolerated(0.48)	2	31	6,45	
	E1527K	Deleterious(0.02)	3	54	5,56	
	L1635M	Tolerated (0.31)	2	11	18,18	Roja Mallorquina
MYH11	V71A	Deleterios(0)	3	53	5,66	Roja Mallorquina
	L983P	Deleterious(0)	2	37	5,41	Gallega
MYH13	L1363M	Tolerated(0.21)	80	80	100	Roja Mallorquina
RYR1	P838T	Tolerated(0.10)	57	823	6,93	Ripollesa
	P4270R	Deleterious(0)	8	12	66,67	Gallega
	G4476S	Tolerated(0.61)	279	25,62	10,89	Canaria

LEGENDS TO FIGURES

Figure 1. Geographic distribution of the 11 Spanish ovine breeds under analysis in the current work.

Figure 2A. Multidimensional scaling plot based on genome-wide identity-by-state pairwise distances inferred with PLINK. This graph displays the genetic relationships between Castellana (CASTEL), Churra (CHURRA), Ojalada (OJALAD), Rasa aragonesa (RASA AR), Xisqueta (XISQUE), Ripollesa (RIPOLL), Latxa (LATXA), Canaria de Pelo (CANARI), Roja Mallorquina (ROJA MAR), Gallega (GALLEG) and Segureña (SEGURE) sheep. 2B The same multidimensional scaling plot shown in Figure 2A, but excluding Churra, latxa, Canaria de Pelo and Roja Mallorquina breeds

Figure 3. Admixture analysis of 11 Spanish ovine breeds: Castellana (CASTEL), Churra (CHURRA), Ojalada (OJALAD), Rasa aragonesa (RASA AR), Xisqueta (XISQUE), Ripollesa (RIPOLL), Latxa (LATXA), Canaria de Pelo (CANARI), Roja Mallorquina (ROJA MAR), Gallega (GALLEG) and Segureña (SEGURE). We set the number of clusters to K = 7 (this K-value had the lowest cross-validation error).

Figure 4. Number of single nucleotide polymorphisms that segregate in one or several of the five pools representing the Xisqueta (XISQUE), Ripollesa (RIPOLL), Canaria de Pelo (CANARI), Roja Mallorquina (ROJA MAR) and Gallega (GALLEG) breeds.

Figure 1

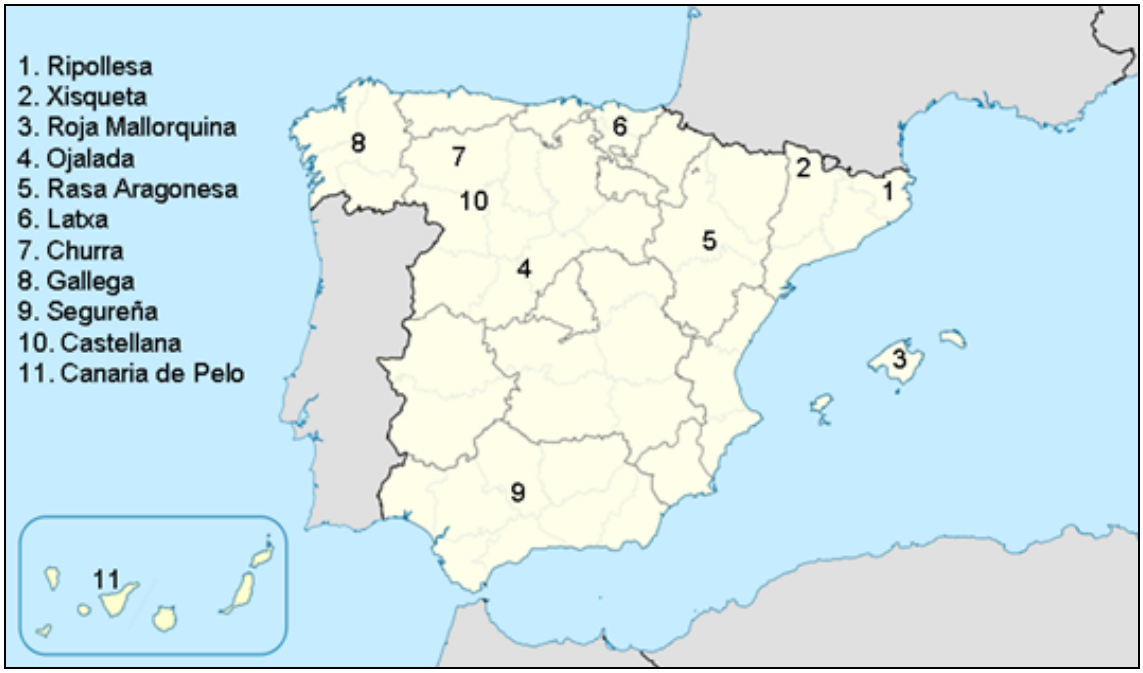

1. Ripollesa
2. Xisqueta
3. Roja Mallorquina
4. Ojalada
5. Rasa Aragonesa
6. Latxa
7. Churra
8. Gallega
9. Segureña
10. Castellana
11. Canaria de Pelo
1
2
3
4
5
6
7
8
9
10
11

Figure 2

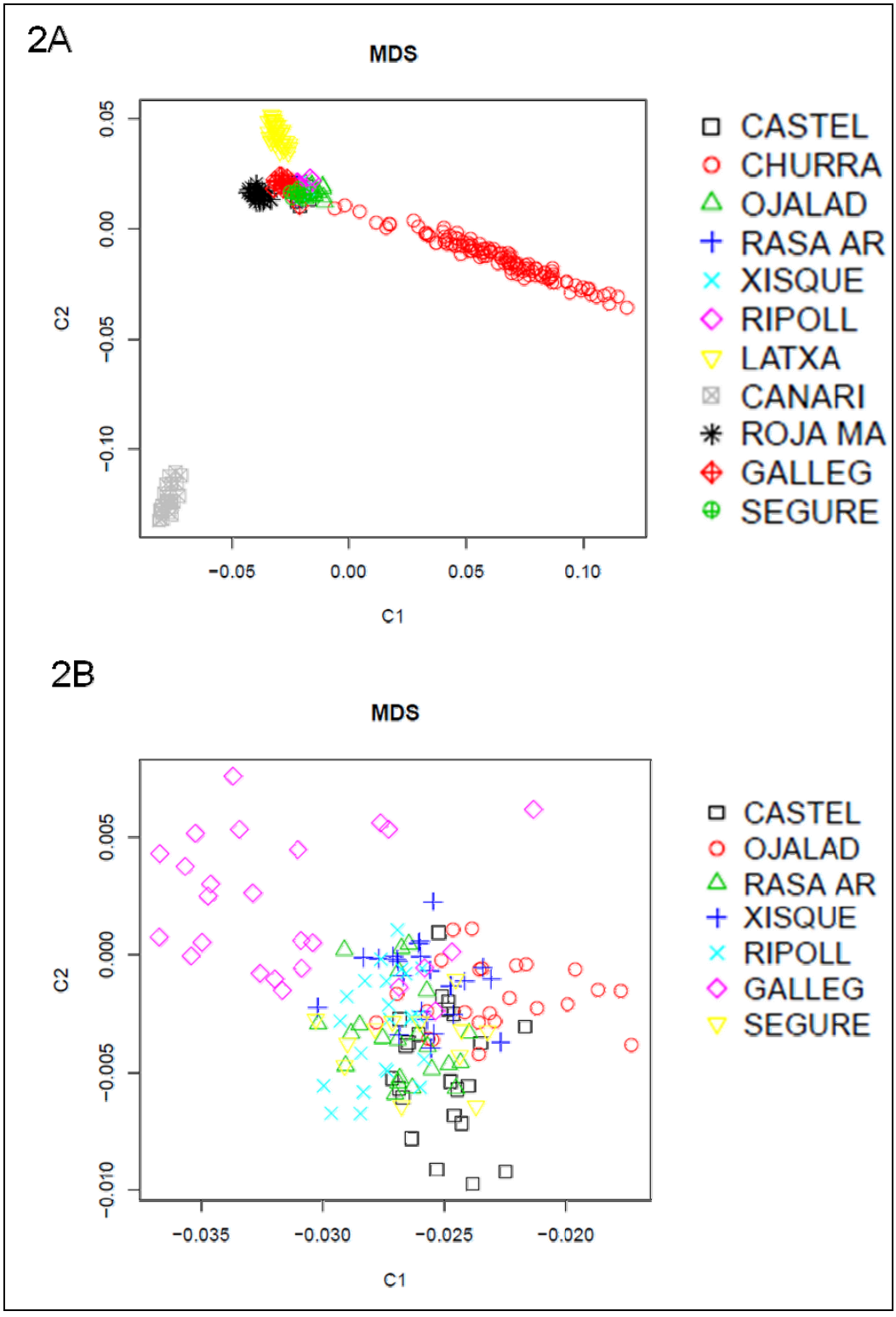

2A
MDS
C2
C1
0.05
0.00
−0.05
−0.10
−0.05
0.00
0.05
0.10
CASTEL
CHURRA
OJALAD
RASA AR
XISQUE
RIPOLL
LATXA
CANARI
ROJA MA
GALLEG
SEGURE
2B
MDS
C2
C1
0.005
0.000
−0.005
−0.010
−0.035
−0.030
−0.025
−0.020
CASTEL
OJALAD
RASA AR
XISQUE
RIPOLL
GALLEG
SEGURE

Figure 3

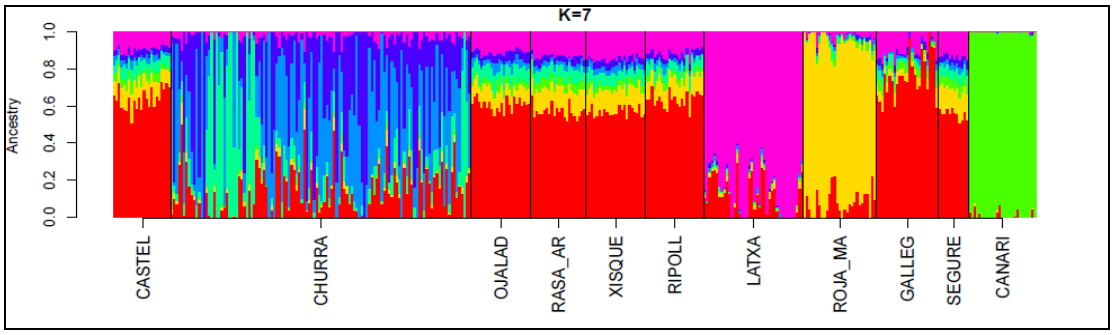

K=7
Ancestry
0.0
0.2
0.4
0.6
0.8
1.0
CASTEL
CHURRA
OJALAD
RASA_AR
XISQUE
RIPOLL
LATXA
ROJA_MA
GALLEG
SEGURE
CANARI

Figure 4

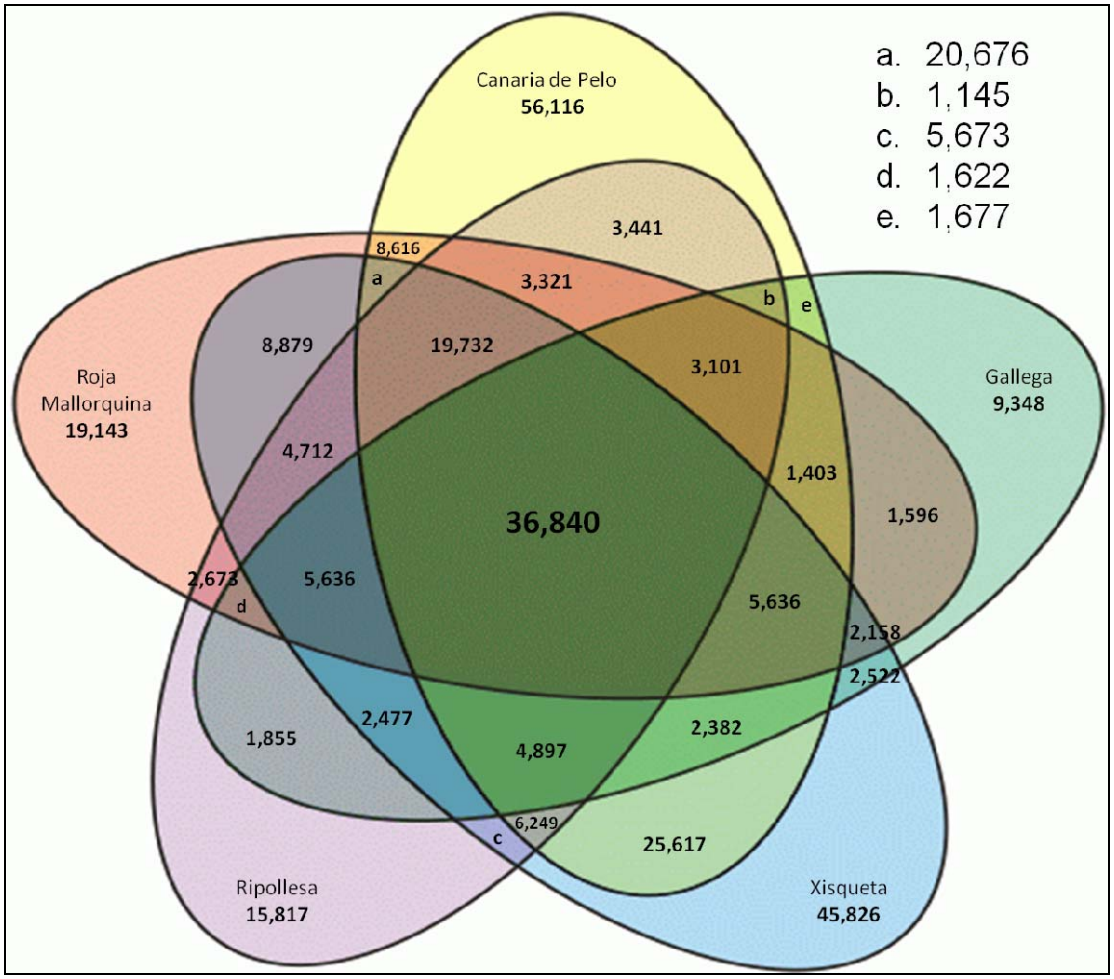

a. 20,676
b. 1,145
c. 5,673
d. 1,622
e. 1,677
Canaria de Pelo
56,116
Roja
Mallorquina
19,143
Gallega
9,348
Ripollesa
15,817
Xisqueta
45,826
8,616
3,441
3,321
8,879
19,732
3,101
4,712
1,403
1,596
36,840
2,673
5,636
5,636
2,158
2,522
2,477
1,855
2,382
4,897
6,249
25,617
a
b
e
d
c

Table S1.Population pairwise FSTs of 11 Spanish ovine breeds analyzed.

	Castellana	**Latcha**	**Gallega**	**Ripollesa**	**RMallorquina**	**Segureña**	**Churra**	**Canaria**	**Rasaaragonesa**	**Xisqueta**	**Ojalada**
Castellana	0.00000										
Latcha	0.04643	0.00000									
Gallega	0.04136	0.04992	0.00000								
Ripollesa	0.03131	0.04293	0.03793	0.00000							
RMallorquina	0.05545	0.06842	0.06300	0.05307	0.00000						
Segureña	0.02490	0.03910	0.03289	0.02257	0.04728	0.00000					
Churra	0.04526	0.05784	0.05379	0.04369	0.06809	0.03988	0.00000				
Canaria	0.09450	0.10659	0.09983	0.09373	0.10837	0.08956	0.10462	0.00000			
Rasaaragonesa	0.02130	0.03472	0.02932	0.01858	0.04402	0.01187	0.03585	0.08401	0.00000		
Xisqueta	0.02304	0.03440	0.03049	0.01665	0.04512	0.01407	0.03594	0.08602	0.01018	0.00000	
Ojalada	0.02953	0.04284	0.03790	0.02773	0.05356	0.02226	0.04042	0.09267	0.01815	0.01983	0.00000

Table S2. Number of reads for each pool of breeds analysed

Breed	Total reads (n)	Mapped reads (n)	Mapped reads (%)	Mapped reads in pair (n)	Mapped reads in pair (%)	Total exon mapped reads	Intron Mapped reads
Ripollesa	300,166,724	213,096,933	70.99	183,731,888	61.21	173,750,304	21,898,685
Xisqueta	318,330,042	236,551,207	74.31	205,874,170	64.67	171,967,105	33,888,394
Canaria de Pelo	317,402,354	233,636,066	73.61	203,786,090	64.20	170,524,459	35,700,619
Roja Mallorquina	297,594,190	229,917,584	77.26	199,667,492	67.09	171,438,893	28,306,781
Gallega	311,855,350	224,384,886	71.95	193,258,050	61.97	184,810,558	19,128,920
Average	309,069,732	227,517,335	73.61	197,263,538	63.82		

4. Discussion and conclusions

DISCUSSION

Genetic variation can be used as a source of information to determine the ancestry and relationships amongst livestock breeds as well as to identify the genetic determinants of complex milk and meat production traits. In the first part of this thesis, a resequencing strategy is used to define the genetic variation of regulatory sites of the ovine casein genes and to investigate its association with dairy and rheological traits of economic interest in Sarda sheep. The second part is basically a population genetic analysis of several Spanish breeds by using a state of the art approach based on massive genotyping and sequencing techniques.

1. Variation at regulatory regions of the ovine milk protein genes is associated with milk traits.

After millennia of continuous artificial selection, sheep have became an efficient source of milk, meat and wool. In particular, Sarda is a very successful dairy specialized breed, that produces around 5% of the whole world's ewe milk (Vacca et al., 2008). Since Sarda milk is totally transformed into cheese, selection goals are focused on increasing milk protein content (Raynal-Ljutovac et al., 2008). In this context, casein genotypes could be used as selection criteria but, unfortunately, information about genetic variation of casein genes in sheep is still scarce. This is illustrated by the fact that only one association analysis between the polymorphism of ovine casein genes and milk traits has been performed so far (Corral et al. 2010).

In this thesis we have characterized, with a Sanger resequencing approach, the promoter and the 3'UTR variation of Sarda sheep. Our hypothesis was that

regulatory polymorphisms might be one of the main sources of variation for agronomic traits, as stated by George (2007). Such effects could be achieved by modifying the transcriptional rate, structure and stability of mRNA transcripts. Our resequencing experiment allowed us to identify a total of 29 SNPs mapping to the four casein genes. This amount of polymorphism is quite high, particularly if we take into account that artificial selection may have targeted the casein genes because of their important effect on milk composition. This set of 29 SNPs, plus 3 previously reported SNPs localized in the α-lactalbumin (*LALBA*) and β-lactoglobulin (*BLG*) genes, was genotyped with a multiplex TaqMan Open Array Real-Time PCR assay in 760 Sarda sheep with records for milk composition and coagulation properties. Subsequently, an association analysis between genotypes and phenotypes was carried out. Our results showed the existence of significant associations between *CSN1S2* and *CSN3* genotypes and milk protein and casein contents. Such results do not agree with those reported by Corral et al. (2013) regarding the Merino sheep. In this study two missense variants in the *CSN1S2* gene did not display any significant association with protein content. Discrepancies between the two studies may be due to the different panel of SNP and population analyzed. As regards to *CSN3*, three SNPs displayed a significant association with milk protein and casein contents in Sarda sheep. These results are consistent with previous research showing that *CSN3* IEF variants in East Friesian dairy ewes are associated with protein content (Giambra et al., 2014) and that, in Norwegian goats, the *CSN3* genotype influences fat and protein contents (Hayes et al., 2006).

Moreover, *CSN1S1* and *CSN2* genotypes had important effects on the coagulation properties of milk in Sarda sheep. Consistently, Giambra et al. (2014) reported a significant effect of *CSN1S1* genotype on milk protein content in East Friesian and Lacaune ewes. Besides, multiple mutations in the goat *CSN1S1* gene have been shown to have dramatic effects on milk protein and casein contents and cheese yield and flavour (Martin et al., 2002). Paradoxically, Calvo et al. (2013) did not find any significant association between *CSN1S1* polymorphism and milk yield, protein, fat and lactose contents in Spanish sheep. This outcome can be explained by biological factors (differences in the gene pools of each population) as well as by technical reasons (sample size is a critical parameter to ensure an adequate statistical power).

Nucleotide substitutions, deletions and insertions in regulatory sites may increase or decrease the transcriptional rate of milk genes, thus modifying the composition of milk (Martin et al. 2002). For instance, in goats the skipping of the exons 9, 10 and 11 of the *CSN1S1* gene induce the synthesis of defective transcripts that result in a lower milk protein content (Ferranti et al., 1997, Chianese et al., 1996). Currently, we do not know if the SNPs that showed associations with milk traits in Sarda sheep are causal or not. Interestingly, we have found that a number of these SNPs co-localize with *in silico* predicted transcription factor and micro RNAs binding sites. However, the bioinformatics tools employed to make such predictions can have very high false positive rates, so these co-localizations need to be judged cautiously.

Though the analysis of the candidate genes can shed light on the determinism of milk yield and composition, high throughput genotyping tools are a much better choice because they make possible the performance of genome-wide association analyses. This latter approach is very powerful because it investigates the genetic architecture of complex traits at a whole genome scale. A future approach would be to combine this strategy with phenotyping techniques that allow to measure each casein protein fraction (*CSN1S1*, *CSN2*, *CSN1S2* and *CSN3*) instead of the total protein content.

2. Elucidating the genetic relationships amongst Spanish ovine breeds

One of the objectives of the current thesis was to investigate the population structure and genetic relationships amongst Spanish ovine breeds by using a high throughput SNP genotyping approach. With this aim, we typed six Spanish breeds (Xisqueta, Ripollesa, Roja Mallorquina, Segureña, Gallega and Canaria de Pelo) with the Illumina ovine 50K SNP BeadChip, and this dataset was merged with previous data obtained by the International Sheep Genomics Consortium (http://www.sheephapmap.org) in Latxa, Churra, Castellana, Ojalada and Rasa Aragonesa. Our findings indicate the existence of a remarkably weak population structure in ovine breeds raised in the Iberian Peninsula (F_{ST} 0,04). These results are consistent with those presented by Kijas et al (2012), who genotyped, with the ovine 50K SNP BeadChip, a total of 2,819 individuals from 74 sheep breeds with a worldwide distribution. This latter study revealed the existence of an extensive haplotype sharing amongst ovine breeds irrespective of their geographic origins.

According to our results, this poor differentiation is particularly applicable to the case of the Castellana, Ojalada, Rasa Aragonesa, Xisqueta, Ripollesa, Gallega and Segureña breeds, which share a very similar genetic background despite the fact that we have used a high resolution SNP panel. In contrast, Canaria de Pelo, Roja Mallorquina, Latxa and Churra breeds have distinctive genetic reservoirs. This is particularly true for the Canaria de Pelo breed, that appears as an outlier in all genetic analyses. This result is probably explained by the fact that this is an insular breed that has remained in geographic isolation for hundreds of years (i.e. until the Spanish colonization of the Canarian archipelago) and that has suffered a strong founder effect. Moreover, the ancestry of the Canaria de Pelo breed is probably African rather than European, a feature that contributes to exacerbate its high genetic differentiation. Consistently, the Palmera goat breed from the La Palma Island also shows a strong genetic differentiation when compared with other European caprine breeds (Manunza, unpublished results).

It is also interesting to emphasize that significant discordances may emerge when comparing the phenotypic *vs* genetic classification of livestock breeds. In our analysis, the Churra and Latxa breeds appear as clearly differentiated from the main group of Spanish breeds (Castellana, Ojalada, Rasa Aragonesa, Xisqueta, Ripollesa, Gallega and Segureña). This result agrees well with the classical phenotypic classification of Spanish ovine breeds which considers the following lineages (Sanchez, 1989 1. Churro (Churra and Latxa breeds, that have a coarse wool), 2. Merino, 3. Medium fine wool (Segureña, Gallega, Ripollesa, Rasa Aragonesa, Castellana and others), and 4. Iberian (Xisqueta, Ojalada and others). The population

genetic analysis performed by us challenges the existence of two differentiated Medium fine (Entrefino) wool and Iberian lineages. This result indicates that breed classifications based on one or a few phenotypes can be very misleading because it does not take into account the existence of genetic heterogenity and convergent selection and, more importantly, it ignores neutral variation.

A previous mitochondrial DNA study revealed the existence of a high variability in Iberian sheep (Pedrosa et al. 2007). Besides, genetic differentiation was even weaker than that detected by us. In general, mitochondrial markers display a limited resolution when investigating the population structure of livestock populations because their inheritance is exclusively maternal and they represent a very small fraction of the whole information contained in the genome of a species. Another investigation focused on exploring the genetic relationships amongst five Spanish sheep breeds (Churra, Latxa, Castellana Rasa-Aragonesa and Merino) with 18 microsatellites revealed the existence of a high variability (heterozigosity 0.77) within breeds. However, clustering analysis of individuals did not indicate the existence of genetic differentiation between sheep from the Medium fine and Churro types (Arranz et al. 2001). In this study, Churro and Latxa breeds showed different pattern of clustering. In this way, all Latxa sheep grouped together in a small cluster while Churra individuals were less clearly differentiated from the Medium fine wool sheep. This suggests a higher gene flow between Churra and the other analyzed Spanish sheep in comparison with Latxa. Findings obtained by us, however, are consistent with an independent genetic origin for Churra and Medium fine wool sheep. Probably, this discrepancy is due to the fact that the resolution of the SNP

panel used by us is much higher than that of the microsatellite panel employed by Arranz et al. (2001), making possible to detect the tiniest genetic differences amongst breeds

3. Spanish ovine breeds may harbour unique genetic variants.

In the last decade, availability of genetic information grew rapidly through the introduction of high-throughput molecular technologies. Resequencing of population samples (population genomics) has changed our understanding of essential evolutionary concepts such as natural selection, gene flow and gene expression pattern (De Wit et al., 2012).

Around 14% of worldwide sheep breeds have became extinct during the last century (Scherf, 2000). In Spain there are 43 officially recognized breeds and 33 of them have an endangered status. An important issue, investigated in this Thesis, would be to infer if these breeds harbour genetic variants that do not segregate in other ovine populations and that, in consequence, are unique. Such information would be indispensable for implementing urgent conservation measures that avoid the loss of irreplaceable farm animal genetic resources (Taberlet el al., 2011). Next generation sequencing technologies have opened new frontiers in the characterization of livestock genetic resources. In our study we have investigated the amount of shared vs breed-specific variation in five ovine populations (Xisqueta, Ripollesa, Segureña, Gallega, Roja Mallorquina and Canaria de Pelo) using a RNA-seq approach. We have investigated the 5 breeds mentioned above because muscle samples were available at the time of performing this research. We also used a pool

strategy, instead of sequencing each individual independently, because it is more cost effective.

The analysis of RNA-seq data generated from pools of 10 individuals per breed revealed that the amount of shared (~184,000 SNPs) vs pool-specific (~146,000 SNPs) variants is quite similar. It is even possible that we are underestimating the size of the pool-specific variation because in general breed-specific variants have low or very low frequencies, and such variants could be easily missed in our experimental pipeline. It is also worth to mention that the number of pool-specific SNPs in Canaria de Pelo and Xisqueta was approximately 2-5 fold higher than in Gallega, Ripollesa and Roja Mallorquina. A high number of pool-specific SNPs in Canaria de Pelo sheep is expectable because this breed has a pure North African origin and it has underwent an intense process of genetic drift. The case of the Xisqueta is less obvious, but the introgression with some ovine breed not considered in our project may explain the existence of many pool-specific variants.

It is important to highlight that our experimental design has some limitations, because we are just sampling 50 individuals from 5 breeds. This circumstance makes difficult to infer how much of the variation that we have identified as "pool-specific" is also "breed-specific". Ideally, we should have sampled all the individuals of all potential populations but this is completely unfeasible. Similar analyses performed in humans also use a limited collection of individuals and populations, and it should be understood that they just provide approximate estimates of the magnitude of the shared vs population-specific variation components (Choudhury et al., 2014). This kind of analysis is normally carried out by using a reduced subset of

populations and individuals. In the case of our study, it would be good to perform a confirmatory analysis with 100 SNPs genotyped in a larger sample of breeds to assess if the SNPs identified by us as pool-specific are also breed-specific. In any case, the molecular information generated in the current Thesis could be useful to establish prioritization criteria in the conservation of ovine genetic resources at Spain.

Another possible application of the molecular information produced in this Thesis could be the implementation of breed-specific SNP panels. Such markers can be used to confirm the source of food products with appelation of origin. For instance, Fontanesi et al. (2011) analyzed the variability of the melanocortin 1 receptor (*MC1R*) in sheep of different coat colors and genotyped the SNP (c.-31G>A) at 5' untranslated region in 14 ovine breeds, demonstrating that the A allele was fixed in the Massese sheep breed. Hence, SNP c.-31G>A was subsequently used to authenticate dairy products from the Massese breed. Using a similar approach, Badaoui et al. (2014) analyzed the variability of 4 pigmentation genes (*KIT, TYR, TYRP2 and MC1R*) in different goat breeds. They found that the gene with the highest level of variability was *MC1R* and revealed that the Palmera breed showed highest frequency for the A allele at the *MC1R* SNP c.764G>A. This SNP was subsequently used as a marker to confirm the origin of Palmero cheese. A similar approach could be implemented in order to certify the origin of meat products derived from Xisqueta, Ripollesa, Canaria de Pelo, Roja Mallorquina and Gallega sheep.

As explained before, our data are consistent with the existence of variants that segregate specifically in certain ovine breeds, converting them in unique genetic resources. In this regard, it is important to highlight that part of the pool-specific SNPs found in our work map to genes related with meat quality and involve amino acid substitutions, For instance, the L382P polymorphism of the calpain 7 gene (*CAPN7*) was exclusively found in the Gallega pool and W135R polymorphism, which maps to the calpain 5 gene (*CAPN5*), has been only identified in the Ripollesa pool. Calpains form part of a proteolytic enzymatic system which plays a major role in meat tenderization (Lian et al., 2013) and, for this reason, the variants identified by us may have an impact on meat quality. The performance of association analyses between candidate SNPs and meat quality traits would be needed to confirm such hypothesis

CONCLUSIONS

1. The level of polymorphism at the promoter and 3'UTRs of the casein genes is quite remarkable in Sarda sheep, with a density of one SNP every 200 bp. Part of this variation displays significant associations with milk protein and casein contents as well as with milk coagulation properties. This finding suggests that regulatory variation at the casein genes may affect the composition and coagulation properties of milk, though functional studies are required to formally demonstrate it.

2. The analysis of population structure of eleven Spanish ovine breeds showed a weak population structure (F_{ST} = 0.04). Canaria de Pelo breed was the most

differentiated breed. Moreover, Churra, Latxa and Roja Mallorquina sheep also showed a distinctive genetic identity. The remaining breeds (Segureña, Gallega, Ripollesa, Rasa Aragonesa, Castellana, Xisqueta and Ojalada) were poorly differentiated, contradicting the existence of two divergent Medium fine wool and Iberian lineages in Spain.

3. RNA-Seq analysis of muscle samples from five endangered Spanish breed demonstrated that the amount of shared (~184,000 SNPs) vs pool-specific (~146,000 SNPs) variants is quite similar in these populations. These results suggest that a significant part of the variation of Spanish sheep populations might be breed-specific, emphasizing the need of implementing conservation programmes that ensure the maintenance and management of these valuable and unique genetic resources. The breed-specificity of the detected SNPs, however, still needs to be validated through an independent genotyping or sequencing approach.

5. References

Al-Mamun H.A., Kwan P., Clark S.A., Ferdosi M.H., Tellam R., Gondro C. 2015. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. Genetics Selection Evolution 47, 66.

Anders S., Huber W. 2010. Differential expression analysis for sequence count data. Genome Biology 11, R106.

Animal QTLdb 2015. The Animal Quantitative Trait Loci (QTL) Database (Animal QTLdb). Online http://www.animalgenome.org/cgi-bin/QTLdb/index permanent. Last accession September 19, 2015.

Arranz J.J., Bayón Y., San Primitivo F. 2001. Differentiation among Spanish sheep breeds using microsatellites. Genetics Selection Evolution 33, 529-542.

Asso.Na.Pa. 2013. Associazione Nazionale della Pastorizia (Italian National Sheepers Association), Standard of the Sarda sheep. Online http://www.assonapa.it/norme_ecc/ovini_llgg/sarda-ovina.htm. Last accession September 1, 2013.

Badaoui B., Manunza A., Castelló A., D'Andrea M., Pilla F., Capote J., Jordana J., Ferrando A.,Martínez A., Cabrera B., Delgado J.V., Landi V., Gómez M., Pons A., El Ouni M., Vidal O., Amills M. 2014. Technical note: Advantages and limitations of authenticating Palmera goat dairy products by pyrosequencing the melanocortin 1 receptor (MC1R) gene. Journal of Dairy Science 97, 7293–7297.

Barillet F. 2007. Genetic improvement for dairy production in sheep and goats. Small Ruminant Research 70, 60–75.

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, S109–S123.

Barillet F., Arranz J.J., Carta A., Jacquiet P.,Stear M., Bishop S. 2006. Final Consolidated Report of the European Union Contract of Acronym GeneSheepSafety, QTLK5-CT- 2000-00656.

Barillet F., Boichard D. 1987. Studies on dairy production of milking ewes. II. Estimates of genetic parameters for total milk composition and yield. Genetics Selection Evolution 19, 459–474.

Becker D., Tetens J., Brunner A., Bürstel D., Ganter M., Ganter M., Kijas J. for the International Sheep Genomics Consortium, Cord Drogemuller. 2010. Microphthalmia in Texel sheep is associated with a missense mutation in the Paired-Like Homeodomain 3 (PITX3) gene. PLoS ONE 5(1): e8689.

Beriain M.J., Horcada A., Purroy A., Lizaso G., Chasco J., Mendizabal J.A. 2000. Characteristics of Lacha and Rasa Arangonesa lambs slaughtered at three live weights. Journal of Animal Science 78, 3070-3077.

Boyazoglu J., Morand-Fehr P. 2001. Mediterranean dairy sheep and goat products and their quality. A critical review. Small Ruminant Research 40, 1-11.

Calvo J.H., Alvarez-Rodríguez J., Marcos-Carcavilla A., Serrano M., Sanz A. 2011. Genetic diversity in the Churra tensina and Churra lebrijana endangered Spanish sheep breeds and relationship with other Churra group breeds and Spanish mouflon. Small Ruminant Research 95, 34-39.

Calvo J.H., Dervishi E., Sarto P., González-Calvo L., Berzal-Herranz B., Molino F., Serrano M., Joy M.. 2013. Structural and functional characterisation of the αS1-casein (CSN1S1) gene and association studies with milk traits in Assaf sheep breed. Livestock Science 157, 1–8.

Cànovas A., Rincon G., Islas-Trejo A., Wickramasinghe S., Medrano J.F. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mammalian Genome 21, 592–598.

Carta, A., Casu Sara, Salaris S. 2009. Invited review: current state of genetic improvement in dairy sheep. Journal of Dairy Science 92, 5814–5833.

Casu S. 1971. Allevamento ovino da latte in Sardegna: situazione attuale e possibilità di evoluzione. Options Méditerranéennes 7, 100-106.

Casu S., Carta R., Flamant J.C. 1975. Amélioration génétique de la production laitiére des brebis Sarde. Annales de Génétique et Sélection Animale 7, 73–90.

Cavanagh C.R, Jonas E., Hobbs M., Thomson P.C., Tammen I., Raadsma H.W. 2010. Mapping Quantitative Trait Loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. Genetics Selection Evolution 16, 42-36.

Chakravarti S. 2015. Nextgeneration sequencing technologies: a novel approach for SNP genotyping studies. Bioinformatic Approaches for Livestock Genome Analysis 11, 33.

Chepelev I., Wei G., Tang Q., Zhao K. 2009. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucleic Acids Research 37, e106.

Chianese L., Garro G., Mauriello R., Laezza P.,Ferranti P., Addeo F. 1996. Occurrence of five αs1-casein variants in ovine milk. Journal of Dairy Research 63, 49–59.

Choudhury A., Hazelhurst S., Meintjes A., Achinike-Oduaran O., Aron S., Gamieldien J., Jalali Sefid Dashti M., Mulder N., Tiffin N., Ramsay M. 2014. Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. BMC Genomics 15, 437.

Cifuni G.F., Napolitano F., Pacelli C., Riviezzi A.M., Girolami A. 2000. Effect of age at slaughter on carcass traits, fatty acid composition and lipid oxidation of Apulian lambs. Small Ruminant Research 35, 65-70.

Cloonan N., Forrest A.R.R., Kolle G., Gardiner B.B.A., Faulkner G.J., Brown M.K., Taylor D.F., Steptoe A.L, Wani S., Bethel G., Robertson A.J., Perkins A.C., Bruce S.J., Lee C.C., Ranade S.S., Peckham H.E., Manning J.M., McKernan K.J., Grimmond S.M. 2008. Nature Methods 5: 613 – 619.

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X., Bibé B., Bouix J., Caiment F., Elsen J.M., Eychenne F., Larzul C., Laville E., Meish F., Milenkovic D., Tobin J., Charlier C., Georges M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nature Genetics 38, 813-818.

Corral J.M., Padilla J.A., Izquierdo M. 2010. Associations between milk protein genetic polymorphisms and milk production traits in Merino sheep breed. Livestock Science 129, 73–79.

CRENoS 2012. Economia della Sardegna, Sintesi del 19° Rapporto 2012 (Economics of Sardinia, Report 2012).

DAD-IS 2015. Domestic Animal Diversity Information System hosted by FAO. Online http://www.faostat.fao.org. FAO, Rome, Italy, last access September 18th 2015.

Dalrymple B.P., Kirkness E.F., Nefedov M., McWilliam S., Ratnakumar A. 2007. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. Genome Biololy 8, R152.

Dario C., Carnicella D., Dario M., Bufano G. 2008. Genetic polymorphism of β-lactoglobulin gene and effect on milk composition in Leccese sheep. Small Ruminant Research 74, 270-273.

DePristo M.A., Banks E., Poplin R.E., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., del Angel G, Rivas M.A, Hanna M., McKenna A., Fennell T.J., Kernytsky A.M., Sivachenko A.Y., Cibulskis K., Gabriel S.B., Altshuler D., Daly M.J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43, 491-498.

Dettori M.L., Pazzola M., Pira E., Paschino P., Vacca G.M. 2015. The sheep Growth Hormone gene polymorphism and its effects on milk traits. Journal of Dairy Research 82, 169-176.

De Wit P., Pespeni M.H., Ladner J.T., Barshis D.J., Seneca F., Jaris H., Therkildsen N.O., Morikawa M., Palumbi S.R. 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Molecular Ecology Resources 12, 1058–1067.

Di C.Z., Liang K.Y., 2011. Likelihood ratio testing for admixture models with application to genetic linkage analysis. Biometrics 67, 1249–1259.

Di Cagno R., Gobbetti M. 2011. Cheese. Hard Italian Cheeses. In Encyclopedia of Dairy Sciences (Second Edition), eds: Roginski H, Fuquay JW, Fox PF, pp 728-736.

Djari A., Esquerré D., Weiss B., Martins F., Meersseman C., Boussaha M., Klopp C., Rocha D. 2013. Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics 14, 307.

Du Y., Liu B., Guo F., Xu G., Ding Y., Liu Y., Sun X., Guoliang X. 2012. The essential role of Mbd5 in the regulation of somatic growth and glucose homeostasis in mice. PLoS ONE 7, e47358.

EU, 1996. Commission Regulation (EC) No 1263/96 of 1 July 1996 supplementing the Annex to Regulation (CE) No 1107/96 on the registration of geographical indications and designations of origin under the procedure laid down Article 17 of Regulation (ECC) No 2081/92. Official Journal of the European Union L 163, 19–21.

EU, 2001. Commission implementing regulation (EU) 2015/793 of 19 May 2015 approving non-minor amendments to the specification for a name entered in the register of protected designations of origin and protected geographical indications (Agnello di Sardegna, PGI) Official Journal of the European Union L 127, 6.

EU, 2009. Commission Regulation (EC) No 1030/2009 of 29 October 2009 approving minor amendments to the specification of a name registered in the

register of protected designations of origin and protected geographical indications (Pecorino Romano, PDO). Official Journal of the European Union L 283, 43–46.

Fahrenkrug S.C., Freking B.A., Rexroad C.E., Leymaster K.A., Kappes S.M. 2000. Comparative mapping of the ovine CLPG locus. Mammalian Genome 11, 871–876.

FAO. 1999. The global strategy for the management of farm animal genetic resources: executive brief. Rome: Food and Agriculture Association of the United Nations. http://lprdad.fao.org/cgi-bin/getblob.cgi?sid=-1,50006152

FAOSTAT. 2013. Statistics division of the Food and Agriculture Organization of the United Nations. Online at http://faostat3.fao.org/. Last accession July 15, 2013.

Feligini M., Parma P., Aleandri R., Greppi G.F., Enne G. 1998. PCR-RFLP test for direct determination of β-lactoglobulin genotype in sheep. Animal Genetics 29, 473-474.

Ferranti P., Addeo F., Malorni A., Chianese L., Leroux C., Martin P. 1997. Differential splicing of pre-messenger RNA produces multiple forms of goat αs1-casein. European Journal of Biochemistry 249, 1–7.

Fontanesi L., Beretti F., Dall'Olio S., Portolano B., Matassino D., Russo V. 2011. A melanocortin 1 receptor (MC1R) gene polymorphism is useful for authentication of Massese sheep dairy products. Journal of Dairy Research 78, 122–128.

Fuertes J.A., Gonzalo C., Carriedo J.A., San Primitivo F. 1998. Parameters of test day milk yield and milk components for dairy ewes. Journal of Dairy Science 81, 1300–1307.

Gan Q., Chepelev I., Wei G., Tarayrah L., Cui K., Zhao K., Chen X. 2010. Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Research 20, 763–783.

Gandini G.C. , Ollivier L., Danell B., Distl O., Georgoudis A., Groeneveld E., Martyniuk E., van Arendonk J.A.M., Woolliams J.A. 2004. Criteria to assess the degree of endangerment of livestock breeds in Europe Livestock Production Science 91, 173–182.

García-Gámez E., Gutierrez-Gil B., Sahana G., Sanchez J. P., Bayon Y., Arranz J.J. 2012. GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the *LALBA* gene. PLoS ONE 7, e47782.

Garcia-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, 6059–6069.

García-Gámez E., Reverter A., Whan V., McWilliam S.M., Arranz J.J., International Sheep Genomics Consortium, Kijas J. 2011. Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in Merino sheep. PLoS ONE 6, e2115814.

George M. 2007. Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. Annual Review of Genomics and Human Genetics 8, 131-162.

Giambra I.J., Brandt H., Erhardt G. 2014. Milk protein variants are highly associated with milk performance traits in East Friesian Dairy and Lacaune sheep. Small Ruminant Research 121, 382–394.

Ginger M.R., Grigor M.R. 1999. Comparative aspects of milk caseins. Comparative Biochemistry and Physiology Part B 124, 133–145.

Goldammer T., Di Meo G.P., Lühken G., Drögemüller C., Wu C.H. 2009. Molecular cytogenetics and gene mapping in sheep (Ovis aries, 2n = 54). Cytogenetics Genome Research 126, 63–76.

Gonda Y., Sakurai H., Hirata Y., Tabata H., Ajioka I., Nakajima K. 2007. Expression profiles of Insulin-like growth factor binding protein-like 1 in the developing mouse forebrain. Gene Expression Patterns 7, 431–440.

Gutiérrez-Gil B., El-Zarei M.F., Alvarez L., Bayón Y., de la Fuente L.F., San Primitivo F., Arranz J.J. 2009. Quantitative trait loci underlying milk production traits in sheep. Animal Genetics 40, 423-434.

Harrington G., Kempster A.J. 1989. Improving lamb carcass composition to meet modern consumer demand. In: Dyrmundsson, O.R., Thorgeirsson, S. (Eds.), Reproduction, Growth and Nutrition in sheep. Dr. Halldor Palsson Memorial Publication, Agricultural Research Institute and Agricultural Society, Reykjavik, Iceland, pp. 79–90.

Hayes B., Hagesaether N., Adnøy T., Pellerud G., Berg P.R., Lien S. 2006. Effects on production traits of haplotypes among casein genes in Norwegian goats and evidence for a site of preferential recombination. Genetics 174, 455–464.

Henchion M., McCarthy M., Resconi V.C., Troy D. 2014. Meat consumption: trends and quality matters. Meat Science 98, 561–568.

Hiendleder S., Mainz K., Plante Y., Lewalski H. 1998. Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. Journal of Heredity 89, 113–120.

Hill W.G. 2013. Selective Breeding. Brenner's Encyclopedia of Genetics, 2nd edition, volume 6, pp. 371-373, doi:10.1016/B978-0-12-374984-0.01390-5.

Hopkins D.L., Mortimer S.I. 2014. Effect of genotype, gender and age on sheep meat quality and a case study illustrating integration of knowledge. Meat Science 98, 544–555.

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 postgenome. Nucleic Acids Research 41, D871–879.

Hurtaud C., Rulquin H., Vérite R. 1993. Effect of infused volatile fatty acids and caseinate on milk composition and coagulation in dairy cows. Journal of Dairy Science 76, 3011–3020.

International Sheep Genomics Consortium, Archibald A.L., Cockett N.E., Dalrymple B.P., Faraut T., Kijas J.W., Maddox J. F, McEwan J.C., Hutton Oddy V.,

Raadsma H.W., Wade C., Wang J., Wang W., Xun X. 2010 The sheep genome reference sequence: a work in progress Animal Genetics, 41, 449–453.

ISMEA, 2002. Istituto di Servizi per il Mercato Agricolo Alimentare (Italian Agency for Agricultural Food Market). Online http://www.ismea.it. Last accession March 16 2010.

IZS, 2008. Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale". Italian National Data Bank for the Animal Husbandry Register, Founded by the Ministry of Health. Online http://ww.izs.it. Last accession July 16, 2013.

Karamichou E., Nute G.R., Richardson R.I., McLean K., Bishop S.C. 2005. Identifying QTL for meat quality and carcass composition traits in Blackface sheep. Proceedings of the British Society of Animal Science Annual Conference; 4-6 April 2005; York 2005, 45.

Karamichou E., Richardson R.I., Nute G.R., McLean K.A., Bishop S.C. 2006. A partial genome scan to map quantitative trait loci for carcass composition, as assessed by X-ray computer tomography, and meat quality traits in Scottish Blackface Sheep. Animal Science 82, 301-309.

Khatkar M.S., Thomson P.C., Tammen I., Raadsma H.W. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genetics Selection Evolution 36, 163–190.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B., other members of the 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, e1001258, doi:10.1371/journal.pbio.1001258.

Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F., et al. 2009 A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS ONE 4: e4668. doi:10.1371/journal.pone.0004668.

Knief C. 2014. Analysis of plant microbe interactions in the era of next generation sequencing technologies. Frontiers in Plant Science 5, 216.

Leroy G. 2011. Genetic diversity, inbreeding and breeding practices in dogs: results from pedigree analyses. The Veterinary Journal 189, 177–182.

Li R., Li Y., Fang X., Yang H., Wang J., Kristiansen K., Wang J. 2009. SNP detection for massively parallel whole-genome resequencing. Genome Research 19, 1124–1132.

Lian T., Wang L., Liu Y. 2013. a new insight into the role of Calpains in post-mortem meat tenderization in domestic animals: a review. Asian-Australasian Journal of Animal Science 26, 443-54.

Macciotta N.P.P., Cappio-Borlino A., Pulina G. 1999. Analysis of environmental effects on test day milk yields of Sarda dairy ewes. Journal of Dairy Science 82, 2212-2217.

Maher C.A., Kumar-Sinha C., Cao X., Kalyana-Sundaram S., Han B., Jing X., Sam L., Barrette T., Palanisamy N., Chinnaiyan A.M. 2009. Transcriptome sequencing to detect gene fusions in cancer. Nature 458, 97-101.

Marcq F., Larzul C., Marot V., Bouix J., Eychenne F., Laville E., Bibe B., Leroy P. L., Georges M., Elsen J.M. 2002. Preliminary results of a whole-genome scan targeting QTL for carcass traits in a Texel × Romanov intercross. Institut National de la Recherche Agronomique (INRA) 0–4.

Mardis E.R. 2008a. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics 9, 387-402.

Mardis E.R. 2008b. The impact of next-generation sequencing technology on genetics. Trends in Genetics 24, 133-141.

Martin P., Szymanowska M., Zwierzchowski L., Leroux C. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. Reproduction Nutrition Development 42, 433–459.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. Genome Research 20, 1297-1303.

Meadows J.R.S., Li K., Kantanen J., Tapio M., Sipos W., Pardeshi V., Gupta V., Calvo J. H., Whan V., Norris B., Kijas J.W. 2005. Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. Journal of Heredity 96, 494–501.

Metzker M.L. 2010. Sequencing technologies—the next generation. Nature Reviews Genetics 11, 31-46.

Miller J.M., Moore S.S., Stothard P., Liao X., Coltman D.W. 2015. Harnessing cross-species alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (Ovis canadensis). BMC Genomics 16, 397.

Ministerio de Agricultura Españolo 2013. Ministerio de Agricultura Alimentación y Medio Ambiente, Dirección General de Producciones y Mercados Agrarios. Caracterización del sector ovino y caprino en España, año 2013.

sheep and goat milk quality. Small Ruminant Research 68, 179–192.

Moioli B., Napolitano F., Orrù L., Catillo G. 2006. Analysis of the genetic diversity between Gentile di Puglia, Sopravissana and Sarda sheep breeds using microsatellite markers. Italian Journal of Animal Science 5, 73–78.

Mortazavi A., Williams B.A,. McCue K., Schaeffer L., Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5, 621–628.

Mullen M.P., Berry D.P., Howard D.J., Diskin M.G., Lynch C.O., Giblin L., Kenny D.A., Magee D.A., Meade K.G., Waters S.M. 2011. Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF-1) gene are associated with performance in Holstein-Friesian dairy cattle. Frontiers in Genetetics 2, 3.

Ng-Kwai-Hang K.F., Hayes J.F., Moxley J.E., Monardes H.G. 1981. Environmental influences on protein content and composition of bovine milk. Journal of Dairy Science 64,1993–1998.

Notter D. R., Scherf B., Hoffmann I. 2013. Breeding of Animals. Encyclopedia of Biodiversity 1, 636-649.

Nudda A., Feligini M., Battacone G., Campus R.L., Pulina G. 2000 Effects of β-lactoglobulin genotypes and parity on milk production and coagulation properties in Sarda dairy ewes. Zootecnica e Nutrizione Animale 26, 137-143.

Oshlack A., Robinson M., Young M. 2010. From RNA-seq reads to differential expression results. Genome Biology 11, 220.

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, 2692-2698.

Pareek C.S., Smoczynski R., Tretyn A. 2011. Sequencing technologies and genome sequencing. Journal of Applied Genetics 52, 413–435.

Pazzola M., Dettori M.L., Cipolat-Gotet C., Cecchinato A., Bittante G., Vacca G.M. 2014. Phenotypic factors affecting coagulation properties of milk from Sarda ewes. Journal of Dairy Science 97, 7247–7257.

Pazzola M., Dettori M.L., Piras G., Pira E., Manca F., Puggioni O., Noce A., Vacca G.M. 2013. The effect of long-term freezing on renneting properties of Sarda sheep milk. Agricolturae Conspectus Scientificus 78, 275–279.

Pedrosa S., Arranz J.J., Brito N., Molina A., San Primitivo F., Bayón Y. 2007. Mitochondrial diversity and the origin of Iberian sheep. Genetics Selection Evolution 39, 91-103.

Pflueger D., Terry S., Sboner A., Habegger L., Esgueva R., Lin P.C., Svensson M.A., Kitabayashi N., Moss B.J., MacDonald T.Y., Cao X., Barrette T., Tewari A.K., Chee M.S., Chinnaiyan A.M, Rickman D.S., Demichelis F., Gerstein M.B, Rubin

M.A. 2011. Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. Genome Research 21, 56–67.

Pirisi A., Comunian R., Urgeghe P.P., Scintu M.F. 2011. Sheep's and goat's dairy products in Italy: Technological, chemical, microbiological, and sensory aspects. Small Ruminant Research 101, 102–112.

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, 45.

Ramos M., Juarez M. 2011. Sheep milk. In Encyclopedia of Dairy Science, (Second Edition), eds: Roginski H, Fuquay JW, Fox PF. pp 494-502.

Raynal-Ljutovac K., Lagriffoul G., Paccard P., Guillet I., Chilliard Y. 2008. Composition of goat and sheep milk products: an update. Small Ruminant Research 79, 57–72.

Rege J.E.O., Gibson J.P. 2003. Animal genetic resources and economic development: issues in relation to economic valuation. Ecological Economics 45, 319–330.

Ripoll G., Joy M., Muñoz F., Albertí P. 2008. Meat and fat colour as a tool to trace grassfeeding systems in light lambs production. Meat Science 80, 239-248.

Roberts A., Pimentel H., Trapnell C., Pachter L. 2011. Identification of novel transcripts in annotated genomes using RNA-seq. Bioinformatics 27, 2325-2329.

Robinson M.D., McCarthy D.J., Smyth G.K. 2010. edgeR: Bioconductor package for differential expression analysis of digital gene expression data. Bionformatics 26, 139-140.

Ryder M.L. 1964. The history of sheep breeds in Britain. Agricultural History Reviews 12, 65–82.

Sánchez A. 1989. Razas ovinas. Orígenes y aptitudes. Ovis 5, 9-19.

Sanna S.R., Casu S., Ruda G., Carta A., Ligios S., Molle G. 2001. Comparison between native and 'synthetic' sheep breeds for milk production in Sardinia. Livestock Production Science 71, 11–16.

Santos V.A.C., Silva S.R., Mena E.G. , Azevedo T.J.M. 2007. Live weight and sex effects on carcass and meat quality of "Borrego terrincho-PDO" suckling lambs. Meat Science 77, 654-661.

Scintu M.F., G. Piredda 2007. Typicity and biodiversity of goat and sheep milk products. Small Ruminant Research 68, 221–231.

Sheephapmap 2015. The International Sheep Genomics Consortium website. Online http://www.sheephapmap.org. Last accession September 19, 2015.

Scherf B.D. 2000. World watch list for domestic animal diversity, 3rd edn. Food and Agriculture Organization of the United Nations, Rome.

Stanford K., Jones S.D.M., Price M.A 1998. Methods of predicting lamb carcass composition: a review. Small Ruminant Research 29, 241–254.

Stinnakre M.G., Vilotte J.L., Soulier S., Mercier J.C. 1994. Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. Proceedings of the National Academy of Science USA 91, 6544–6548.

Sultan M., Schulz M.H, Richard H., Magen A., Klingenhoff A., Scherf M., Seifert M., Borodina T., Soldatov A., Parkhomchuk D., Schmidt D., O'Keeffe S., Haas S., Vingron M., Lehrach H., Yaspo M.L. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321, 956–960.

Taberlet P., Valentini A., Rezaei H.R., Naderi S., Pompanon F., Negrini R., Ajmone-Marsan P. 2008. Are cattle, sheep, and goats endangered species? Molecular Ecology 17, 275-84.

Tellam R.L., Cockett N.E., Vuocolo T., Bidwell C.A. 2012. Genes contributing to genetic variation of muscling in sheep. Frontiers in Genetics 3, 164.

Trapnell C, Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L., Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7, 562-578.

Trapnell C., Salzberg S.L. 2009. How to map billions of short reads onto genomes. Nature Biotechnology 27, 455–457.

Vacca G.M., Carcangiu V., Dettori M.L., Pazzola M., Mura M.C., Luridiana S., Tilloca G. 2008. Productive performance and meat quality of Mouflon × Sarda and Sarda × Sarda suckling lambs. Meat Science 80, 326–334.

Vacca G.M., Dettori M.L., Balia F., Luridiana S., Mura M.C., Carcangiu V., Pazzola M. 2013. Sequence polymorphism at the growth hormone GH1/GH2-N and GH2-Z gene copies and their relationship with dairy traits in domestic sheep (Ovis aries). Molecular Biology Reports 40, 5285–5294.

Vagnoni E., Franca A., Breedveld L., Porqueddu C., Ferrara R., Duce P. 2015. Environmental performances of Sardinian dairy sheep production systems at different input levels. Science of the Total Environment 502, 354–361.

Walling G.A., Visscher P.M., Simm G., Bishop S.C. 2001. Confirmed linkage for QTLs affecting muscling in Texel sheep on chromosomes 2 and 18. In: Proceedings of the 52nd Annual Meeting of the European Association for Animal Production.Budapest, Hungary, Paper G5.6.

Walling G.A., Visscher P.M., Wilson A.D., McTeir B.L., Simm G., Bishop S.C. 2004. Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. Journal of Animal Science 82, 2234-2245.

Walling G.A., Wilson A.D., McTeir B.L., Visscher P.M., Simm G., Bishop S.C. 2002. A candidate region approach allows efficient QTL detection in UK Suffolk and Texel populations. In: Proc. 7th World Cong. Genet. Appl. Livest. Prod., Montpellier. 19–23 August, Inra, Castanet-Tolosan, France, Abstract 02-16.

Wang Z., Gerstein M., Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10, 57-63.

Wang E.T., Sandberg R., Luo S., Khrebtukova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., Burge C.B. 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476.

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.

Wood N.J., Phua S.H. 1996. Variation in the control region sequence of the sheep mitochondrial genome. Animal Genetics 27, 25–33.

Zander K.K., Signorello G., De Salvo M., Gandini G., Drucker A.G. 2013. Assessing the total economic value of threatened livestock breeds in Italy: Implications for conservation policy. Ecological Economics 93, 219–229.

Zhang L., Liu J., Zhao F., Ren H., Xu L., Lu J., Zhang S., Zhang X., Wei C., Lu G., Zheng Y., Du L. 2013. Genome-wide association studies for growth and meat production traits in sheep. PLoS ONE 8, e66569.

Zhao X., Dittmer K.E., Blair H.T., Thompson K.G., Rothschild M.F., Garrick D.J. 2011. A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale Sheep. PLoS ONE 6, e21739.

6. Acknowledgements

I sincerely thank:

- the supervisors of the present thesis;
- all the co-authors of the research papers;
- the staff at the Department of Veterinary Medicine, University of Sassari;
- the staff at the Department of Animal Genetics, Center for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB), Universitat Autònoma de Barcelona;
- the Fondazione Banco di Sardegna (Bank of Sardinia Foundation);
- the Sardinian farmers for giving access to their flocks;
- the A.I.P.A./A.P.A.s (Provincial Farmers Associations) of Cagliari, Nuoro, Sassari and Oristano for their support in sample collection and A.R.A. Sardegna (Regional Farmer Association of Sardinia) for support in milk analysis;
- dottoressa Emanuela Pira for her support in experimental protocol of Paper 1 and cover's photograph.