



UNIVERSITÀ DEGLI STUDI DI PALERMO

Dottorato di Ricerca Scienze Agrarie Forestali e Ambientali

Dipartimento di Scienze Agrarie Alimentari e Forestali (SAAF)

Indirizzo: Agro-Ecosistemi Mediterranei
S.S.D. AGR17

**Comparison of Genome Wide Association Studies for milk production
traits in Valle del Belice dairy sheep**

IL DOTTORE
Anna Maria Sutera

IL COORDINATORE
Prof. Vincenzo Bagarello

IL TUTOR
Prof. Baldassare Portolano

IL CO TUTOR
Dott. Antonino di Grigoli

CICLO XXX
2018

Acknowledgments

Life makes you know people who help you grow, takes away people who were part of you, makes you find people ready to help you, people who will hinder you and others that will help you achieve your goals. During this journey many people I met both good and bad, but all “effects” that have contributed to the realization of my research, and it is well known that “environmental effects” has a strong impact...on life and...on research.

And at the end of this long journey, lasting three years, I have the duty to thank the most significant “effects”:

In primis, I want to thank my PhD tutor Prof. Baldassare Portolano. Thank you very much for giving me the opportunity to work on this project. Your knowledge and advices were of great help for me, and I am very happy for the chance to collaborate with you!

It is difficult to overstate my gratitude to my group, Ciupina, Marco, Rosali and Salvo. With their enthusiasm, their provided encouragement, “sound effect”, good teaching, good company, lots of good ideas, and, moreover, for their timely support and guidance till the completion of my project, they have been a family for me! Heartily thank you!! A special thanks to my friend Marco for his "all inclusive" B&B.

Thank you very much to my supervisor Dr. Valentina Riggio and Dr. Ricardo Pong-Wong for their support in statistical approaches for genomic data analyses and for their help in my survival far from home! Thanks to the whole Genomics group of the Roslin Institute for the proven affection, it was a pleasure to meet them.

Thank you to my colleagues Alessandra, Roberta and Gianluca for the opportunity to compare and consult me for everything related to the doctoral activity and with which I have shared this difficult path!

A special thanks to the PhD Coordinator Prof. Vincenzo Bagarello for his guidance and care to help carry out all the tasks related to the doctorate and the PhD Vice Coordinator Prof. Adriana Bonanno, and the PhD co-tutor Dr. Antonino Di Grigoli for their support and availability.

Last but not least, I wish to thank my parents for adoring me, for helping me grow, for supporting me, for teaching me how to live, and for loving me. I dedicate this thesis to them.

Abstract

The great achievements in livestock species selection during the last 50 years largely relied on quantitative genetic theory and infinitesimal genetic model. In the last 20 years, due to the application of advanced techniques in molecular genetics and statistics, several chromosomal regions that influence quantitative traits have been discovered. Combinations of molecular and classical quantitative information in a composite selection index have been proposed to increase the accuracy of selection. Nowadays, many genotyping arrays for thousands of SNPs are available for several livestock species, such as: cattle, sheep, pigs, horses, goat and chickens. The overall aim of this thesis is the comparison of different GWAS approaches to identify SNPs associated with milk production traits in Valle del Belice dairy sheep. In particular different genetic merit indices (breeding values and their deregressed and weighted values) and single test day observations will be evaluated to identify SNPs associated with milk yield (MY), fat yield (FY), fat percentage (F%), protein yield (PY) and protein percentage (P%) using different statistical approaches. The raw phenotypes data set included 5,586 observations of 481 ewes for MY, F%, FY, P% and PY traits. All animals were genotyped using the Illumina OvineSNP50K BeadChip. A single trait repeatability test-day animal model was performed to estimate the breeding values. The EBVs for MY, FY, F%, PY and P% estimated with the mixed model were deregressed (DEBV) and weighted (DEBVw) to obtain a more accurate estimate of the expected phenotype according to Garrick *et al.* (2009). Genome-wide association analysis was carried out based on regression of phenotypes (EBVs, DEBVs and DEBVw) with the genotypes of animals for one SNP at a time. For single-marker GWAS, we used a three-step approach referred to as genomic GRAMMAR-GC as implemented in GenABEL package. Other two approaches for genome wide association study were used. The first accounts the covariances between repeated measures for each individual using the RepeatABEL package was used. The last approach known as Regional Genomic Relationship Mapping or Regional Heritability Mapping (RHM) provides heritability estimates attributable to small genomic regions, and it has the power to detect regions containing multiple alleles that individually contribute too little variance to be detected by standard GWA studies. Comparison among the estimated breeding values and their deregressed and weighted values as responsible variables, respect to their influence on our GWAS results, has demonstrated that DEBVs and DEBVw allow identifying a greater numbers of SNPs than using EBVs. Several SNPs using different approaches were identified and some of these SNPs are mapped within the previously reported QTL regions and within candidate genes for milk production traits. The results confirmed

the roles of LALBA gene and AQP genes, on OAR 3, as candidate genes for milk production traits in sheep. Moreover some genomic regions identified by close SNPs associated with a specific trait should be further investigated to verify their effect on the traits. The general consistence of the significant SNPs detected herein with the reported QTL and candidate genes for milk traits allow us to be confident of the results obtained. The information generated from this thesis has important implications for the design and applications of association studies as well as for the development of selection breeding programs for the Valle del Belice sheep breeds.

Chapter 1

General Introduction

1.1. Dairy sheep

Sheep (*Ovis aries*) were one of the first livestock species to be domesticated along with goat and are believed to have undergone several domestication events (Meadows *et al.*, 2007; Pedrosa *et al.*, 2005). Following the domestication, a great variety of sheep breeds has been developed during the spread of sheep to other regions (Larson *et al.* 2014). At first, they were mostly kept for meat, but later also for wool and milk (Chessa *et al.*, 2009) and have therefore been selected for meat, fiber and milk production. At first glance, the economic importance of dairy sheep seems to be low compared to others livestock species. Sheep milk production accounts for 4.6% of the total milk production in Italy (ISTAT, 2016) and the Mediterranean basin with 60% of total world production is the most important area (FAO, 2014). Dairy sheep breeding is usually based on local breeds that are very well adapted to their production areas, systems and environments. Traditionally this activity has occupied less favored areas, using natural resources of low interest for other species, and helping to maintain the ecological equilibrium and the natural landscape. In addition, it has contributed to sustaining economic activity and the population in rural areas. In certain region and for certain breeds, the management system varies from semi-extensive to intensive according to the economic relevance of the production chain and the specific environment. Differences in management condition and good adaptability of the local sheep breeds to their specific environments lead to remarkable genetic diversity between them. Local breeds are also often linked to "high quality" products protected by quality labels. The above factors have resulted in the establishment of breeding programs designed to safeguard and improve production traits in local breeds against foreign breeds that often are not able to adapt to specific environments.

1.2 Origin and description of the Valle del Belice dairy breed

The Valle del Belice dairy sheep (Figure 1) originates from western Sicily and its name is derived from the Belice valley, delimited by the Sicilian provinces of Palermo, Agrigento and Trapani.



Figure 1. Valle del Belice dairy sheep

This breed is considered to originate from a three way cross between the Pinzirita, Comisana and Sarda dairy breeds (Portolano, 1987). The Pinzirita breed, a native Sicilian sheep found in the western part of Sicily (Portolano *et al.*, 1996), was first crossed with the Comisana dairy breed, which originated in the south-east of Sicily (Portolano, 1987). Crosses between the Pinzirita and Comisana breeds gave birth to individuals having intermediate characteristics between parental lines. These animals were crossed with sheep belonging to the Sarda dairy breed, imported to Sicily during the Arab domination (~800 A.D.) of the island (Portolano, 1987). In the nineties, data collected by the Regional Sicilian Breeders Association (ARAS) for milk recording and morphological measures collected by the University of Palermo, allowed the development of the Valle del Belice breed standard. This was submitted in 1996 to both the Dairy Sub-Committee and the Ewes Technical Committee, and in 1997 the Valle del Belice breed was given official recognition as local breed. This breed is mainly used for milk production and its average milk production is 139 ± 35

liters in the first lactation and 210 ± 62 liters for later lactations (AIA, 2006) considering a lactation length of 120 days. Fat and protein contents are 6% and 5.5%, respectively. The head is fine and extended and the trunk well developed with good transversal diameters. A white coat covers the entire body with the exception of limbs, belly and head. Typical are the reddish brown spots surrounding the eyes and on the distal part of the ears. A typical family farming system is conducted and the breed is mainly raised under semi-extensive grazing conditions. Ewes are milked twice a day (morning and evening), and are housed in old storehouses or kept in fenced after the evening milking. Most of the farmers milk ewes by hand but some of the farms use a milking machine. Furthermore, the lambing system is different from the one adopted in other Mediterranean regions (e.g., Carta *et al.*, 1995; Ligda *et al.*, 2000). The lambing season of the Valle del Belice breed is all year long, starting in July and finishing in the following June, with a reduction in May and June. The primiparous ewes usually give birth between December and March. Moreover, sheep are fed with natural pastures and fodder crops; supplementation, consisting of hay and sometimes concentrates, is occasionally supplied, for example at the end of gestation (Cappio-Borlino *et al.*, 1997). The main use of the milk from Valle del Belice breed is for the production of traditional raw milk cheeses (Pecorino Siciliano and PDO Vastedda del Belice), at farm level or by small local dairies or by cheese industries working at regional level.

1.3 The use of the genomic approach to enhance the response to selection

Most of the traits of economic interest in livestock have a complex quantitative expression coded by a large number of genes and affected by environmental factors. Statistical analysis of phenotypes and pedigree information allows estimating the genetic merit (breeding value) of the animal candidate to

selection following the Fischer's infinitesimal model, according to which observed phenotypes are determined by an infinite number of loci, each with an infinitesimal additive effect (Goddard *et al.*, 1992). Under this hypothesis, mean of a quantitative trait in a population can be modified choosing the best genotypes based on the breeding values estimated using Best Linear Unbiased Predictors (BLUP) methodology. In the best situation, all sources of information on phenotypes and additive relationships among animals are included in a BLUP model to estimate a breeding value for all the animals in the population. Estimates of genetic parameters such as heritability are needed as the basis for description and prediction. Heritability applies to a single trait measured on animals in a specific population at a given time point. Estimates of heritability for a trait can differ between dairy sheep breeds and may change slowly over time. Heritability is estimated from performance records on animals and pedigree information used to establish genetic relationships between those animals. Heritability helps explain the degree to which genes control expression of a trait and is used to calculate genetic evaluations, to predict response to selection, and to help producers decide if it is more efficient to improve traits through management or through selection. In general heritability is estimated using the formula:

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

where σ_A^2 is the additive genetic variance and σ_P^2 is the phenotypic variance. Traditional methods such as analysis of variance or regression cannot cope adequately with unbalanced data and the complex pedigrees used in the context of livestock study. An important generalization has been the development of the animal model in which the phenotype of each individual is defined in terms of effects and the genetic structure is incorporated in the variances and covariances of these effects. For example, a basic model is:

$$y = Xb + Za + e$$

where y is a vector of phenotype of the individual, b is a vector of fixed effects (e.g. herd, parity, etc), a is a vector of random animal effects (breeding values), e is a vector of random errors, X and Z are incidence matrices for fixed and random effects, respectively. Complex mixed model equations can adjust for all kind of environmental factors, maternal effects or repeated records and simultaneously release the requested individual additive genetic merit. Prediction of breeding values is a fundamental component of breeding programs as animals with the highest values should be selected for selection. For over 50 years, BLUP and related methodology have dominated genetic evaluation of dairy cattle, and models have become increasingly complex. The great achievements in livestock species selection during the last 50 years largely relied on quantitative genetic theory and infinitesimal genetic model (Figure 2).

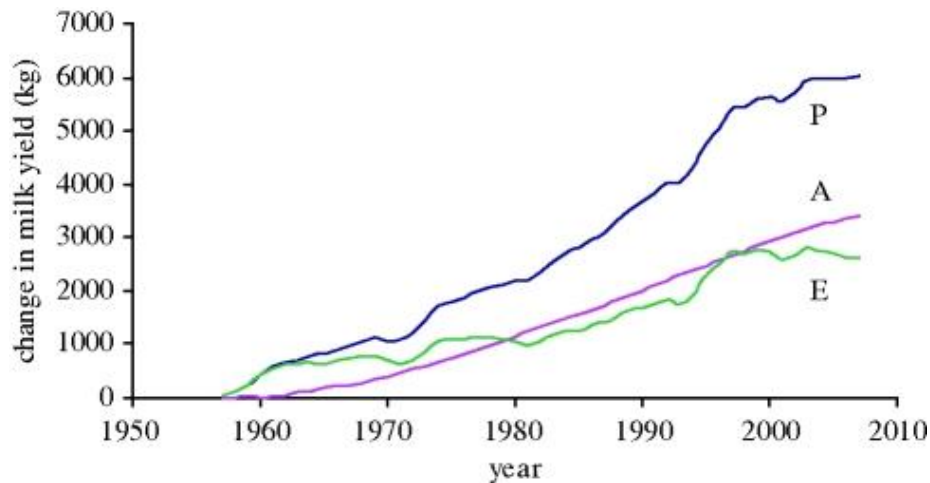


Figure 2. Changes in milk yields of US Holstein cows: phenotypic mean yields (P), mean breeding values (A) and environmental effects ($E = A - P$) derived from USDA data.

Development of methodology continues, particularly of the statistical methods required to undertake the BLUP predictions. Standard quantitative selection

requires, therefore, field data collection on individual phenotypic performance and expected covariate among animals due to relationship between them. Recorded traits so far have been mainly limited to production performances. But even for "simple" traits like milk yield, there may be regional or national differences which make comparison difficult to manage. Phenotypes may be also sex-limited (i.e. milk production) and recorded late in the life of an animal (i.e. carcass traits). Recording is also quite expensive and in many countries costs are largely covered by National Governments. In Italy, for example, the total budget for selection granted by the Ministry of Agriculture is around 93 million euro, 77 of which are dedicated for field phenotype recording. Cost of phenotype recording are also quite variable in different species and the value of phenotyped animal may not justify the expense. As an example, detection of milk composition (fat, protein, lactose, etc) in sheep milk requires the collection of a milk sample that, at the end of the lactation, may represent more than 10% of the daily released milk. Recording requires also efficient services covering a relevant part of the farms, a structure that cannot be created quickly and cheaply. Parentage ascertaining is also extremely expensive. In livestock species, where artificial insemination (AI) is largely used, the paternity of an animal is easily predictable. In species where natural mating is predominant, as in sheep, parentage definition is a crucial step in selection and the use of molecular markers could be a relevant part in the choice of the best sires. Furthermore, a relevant constraint to the genetic progress is represented by the inverse relationship between accuracy of breeding values and generation interval, kept constant the other variables in the equation of genetic gain. Therefore, the more accurate breeding value we want to estimate for a sire, the more time we need to wait to collect phenotypic information from progenies. The finite amount of DNA in the mammalian genome suggests that must be a finite number of loci that controls the expression of quantitative traits (between 20,000 and 35,000 genes) (Ewing and Green, 2000), in contrast with the

infinitesimal gene model. Thanks to the advances in the molecular techniques, a large number of genetic markers have been discovered. Possible strategies to use and integrate these new sources of information with the aim of enhancing the accuracy of selection have been extensively reviewed and proposed from different authors (Lande and Thompson, 1990; Meuwissen *et al.*, 2001; Dekkers, 2004). In the last 20 years, due to the application of advanced techniques in molecular genetics and statistics, several chromosomal regions that influence quantitative traits have been discovered. Hayes and Goddard (2001) investigated the distribution of the Quantitative Trait Loci (QTLs) effects in dairy cattle and swine, enforcing the evidence that there are few genes with large effects and many of small effects. How this relevant amount of knowledge is going to change the selection of farm animals is still an open issue. Combinations of molecular and classical quantitative information in a composite selection index have been proposed to increase the accuracy of selection (Lande and Thompson, 1990). Several approaches have been indicated to integrate molecular information in current breeding programs. The first step is to estimate the genetic merit for a candidate to selection for a specific trait to associate it with molecular information. An advantage of genetic markers is that they are available early in life, so that the accuracy of breeding values estimated for young animals can be increased and the generation interval reduced. Use of molecular data represents an opportunity to enhance the response to selection especially for low-heritability traits, or whose phenotype is difficult or expensive to measure or expressed later in age. Sex-limited traits, such as milk production in dairy breeds, can be objectives of selection based on molecular data, in order to reduce the generation interval. For such traits, the molecular based breeding value can be available early in life and for both genders (Dekkers and Hospital, 2002). Although advances in molecular genetics have been able to explain part of the genetic variances due to QTLs, the possibility of implement this information in a marker assisted

selection (MAS) program has been limited by several reasons. Firstly, only a limited number of genes have been identified. Secondly, in most cases the marker maps used in the past were sparse, so that the QTLs have been mapped with very large confidence interval.

1.4 Genome Wide Association studies

Molecular markers or genetic markers are small sequences of DNA that reveal polymorphism in genomes (Tanksley, 1983). They are powerful tools to detect genetic uniqueness of individuals and the diversity of populations (Chauhan and Rajiv, 2010). Molecular genetic markers that have been used in genetic analyses are allozymes, restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), expressed sequence tags (EST), microsatellites (or simple sequence repeats, SSR) and single nucleotide polymorphisms (SNPs) (Vignal et al., 2002). Single nucleotide polymorphisms (SNPs) are single base-pair (bp) differences which are widely used as genetic markers and have been studied in many species (Figure 3). SNPs have many advantages compared to other molecular markers such as availability in high numbers, presence in coding and non-coding regions, low-scoring error rates, relative ease of calibration between different studies and conformation to simple models of mutation (Haynes and Latch, 2012). They also represent the most abundant polymorphism in any organism's genome, adaptable to automation, and reveal hidden polymorphism not detected with other markers and methods (Chauhan and Rajiv, 2010). SNPs information for many species is available in dbSNP online database (<https://www.ncbi.nlm.nih.gov/projects/SNP/>).

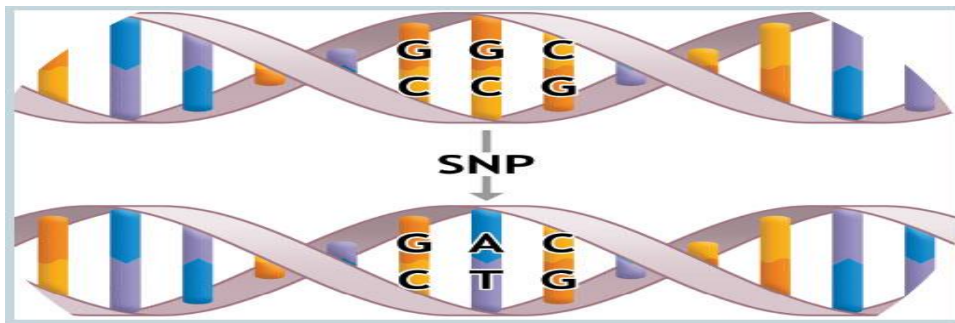


Figure 3. Single Nucleotide Polymorphism.

The availability of high-density SNP panels has given a great impulse toward the identification of genomic regions associated to complex traits and diseases in humans and, recently, in several livestock species (Yang *et al.* 2010; Hayes and Goddard, 2010). Nowadays, many genotyping arrays for thousands of SNPs are available for several livestock species, such as: cattle, sheep, pigs, horses, goat and chickens. When a dense set of polymorphic markers across a genome is genotyped in samples, it is possible to look for common genetic variants associated with a specific phenotype. These are so called genome-wide association studies (GWAS) and are mainly used to identify genetic risk factors associated with diseases in humans (Bush and Moore, 2012) and economic traits of animals (Zhang *et al.*, 2013). Therefore, GWAS allows detecting associations between markers widely distributed along the genome and production or functional traits or diseases. They are considered relatively powerful and fast compared to other methods used to identify genetic effects (Hirschhorn and Daly, 2005). For example, QTL studies have a large confidence interval so the causal genes can be hard to locate within them (Zhang *et al.*, 2013). Associations are studied by examining many common genetic variants in different individuals and then verifying if any variant is associated with a trait of interest. GWAS was first used in the analysis of human disease considering case and control group. (The Wellcome Trust Case Control Consortium, 2007). GWAS was extended to the field of domestic

animal genetics and breeding when genomic sequences were available for several domestic species and large numbers of SNPs were discovered as a result of sequencing or in subsequent re-sequencing projects. Information about phenotypes must also be available in order to find association between genotypes and phenotypes. Phenotypes can be either categorical or quantitative. Although the application of GWAS to domestic animals has only occurred relatively recently, there have been a series of results reported, especially from the analysis of the genetic mechanisms of quantitative traits. For example, in dairy cattle significant association using genome-wide data were detected for milk yield (Hayes *et al.*, 2009; Bolormaa *et al.*, 2010), for milk fat and protein contents (Pryce *et al.*, 2010; Schopen *et al.*, 2011), for somatic cell score (Meredith *et al.*, 2012). Genome wide-association studies were also carried out for carcass weight (Lee *et al.*, 2013), birth weight and size in pigs (Utsunomiya *et al.*, 2013). Besides cattle, GWA studies were also performed in other domestic animals, including pigs (Ramos *et al.*, 2011; Grindflek *et al.*, 2011), sheep (Johnston *et al.*, 2011), goat (Martin *et al.*, 2016), horses (Brooks *et al.*, 2010), chickens (Gu *et al.*, 2011; Liu *et al.*, 2011) and dogs (Mogensen *et al.*, 2011).

1.5 GWAS in Sheep

Few GWA studies have been carried out for sheep data due to limited information about the sheep genome. With the recently released assembly of the whole sheep genome, the number of GWAS on sheep is growing (Zhang *et al.*, 2013). The latest assembly of the sheep genome (Oar_v4.0) was generated by sequencing the DNA of two animals, a male and a female belonging to the Texel sheep breed using Illumina technology. Briefly, the coverage of the reference genome is 166 fold with a contig length of ~40kb and a total

assembled length of 2.61 gigabase (Gb) (Jiang *et al.*, 2014). Before the release of the sheep genome, there were only about 700 genes known in sheep (Zhang *et al.*, 2013) but the current release counts 29,110 genes in the sheep genome and 43,046 proteins.

The first study of GWAS in sheep was made on horn types by Johnston *et al.*, (2011). This study was conducted using 36,000 SNPs and determined the main genetic candidate for horns to be *RXFP2*, an autosomal gene with known involvement in determining primary sexual characteristics in humans and mice (Yuan *et al.*, 2010).

A Chinese GWAS on 329 sheep of different breeds looked for association to 11 traits related to muscle growth. The study identified 5 candidate genes for growth and meat production traits (Zhang *et al.*, 2013). Moreover, multiple QTLs using GWAS and the Illumina Ovine SNP50 BeadChip were identified containing several candidate genes (e.g., *SPPI*, *MEPE*, *IBSP*, *LCORL* and *NCAPG*) for bone-related traits and meat quality traits in sheep (Matika *et al.*, 2016). An association of genomic regions to susceptibility and control of ovine Lentivirus has been studied and few candidate genes were found (White *et al.*, 2012). A GWAS for nematode resistance and body weight was performed in Scottish Blackface lambs by Riggio *et al.* (2013) and strong evidence for association was found on chromosomes 14 and 6. The first GWAS for milk traits using an OvineSNP50 BeadChip, identified the most likely candidate gene (i.e., *LALBA*) affecting milk protein and fat contents in dairy sheep (García-Gómez *et al.*, 2012). Moreover, other strong candidate genes (e.g., *ABCG2*, *SPPI*, *SCD*, *SOCS2*, *PKD2*, *MEPE*, and *IBSP*) associated with milk production traits, have been proposed using GWAS (Rupp *et al.*, 2015).

1.6 Aim and outline of this thesis

The overall aim of this thesis is the comparison of different GWAS approaches to identify SNPs associated with milk production traits in Valle del Belice dairy sheep. In particular different genetic merit indices and single test day observations to identify SNPs associated with milk yield (MY), fat yield (FY), fat percentage (F%), protein yield (PY) and protein percentage (P%) will be evaluated.

Chapter 2

Materials and Methods

2.1 Population and phenotypes

In this study, several lactations on a total of 481 ewes belonging to the Valle del Belice dairy sheep were considered. Phenotypic data were collected by the University of Palermo between 2009 and 2015 in four flocks located in Agrigento province. The procedures involving animal sample collection followed the recommendation of directive 2010/63/EU. Milk samples were collected at approximately monthly intervals, following an A4 recording scheme (ICAR, 2016). All ewes were milked manually twice a day and milk from morning and evening milking has been weighted and collected to determine daily milk yield and composition. For each ewe some information as lambing date, number of lambs born and order of parity were registered. Milk samples were added with the preservative Bronopol and transferred under controlled temperature in the laboratory of Dipartimento di Scienze Agrarie, Alimentari e Forestali (SAAF) for chemicals analysis. Milk composition was determined by the method of infrared spectrophotometry using Combifoss 6200 apparatus (Foss Electric Hillerød, Denmark). In particular, fat (F%) and protein (P%) percentages were calculated as the weighted average of the morning and evening percentages according to the corresponding daily milk yield. Moreover, fat (FY) and protein (PY) yield were calculated considering the weighted average percentages according to the corresponding milk yield. The raw phenotypes data set included 5,586 observations of 481 ewes for MY, F%, FY, P% and PY traits. Data editing were performed using S.A.S. version 9.2 (SAS v9.2.3, 2012) to guarantee the quality of the data to be analyzed. Ewes with lactation longer than 300 days or with less than 3 observations (test-day, TD) within lactation were discarded. After editing, the data set consisted of 5,446

observations of 481 ewes. Means, standard deviations, and coefficients of variation of the test-day traits are given in Table 1 of chapter 3.

From the pedigree data available online on Associazione Nazionale della Pastorizia (<http://www.assonapa.com>), the relationship information of the individuals under study was extracted. To verify the identities of the fathers, mothers and their progenies Pedigree Viewer software was used (Kinghorn, B. and Kinghorn, S. 1999). Pedigree was composed of 5,175 animals of which 180 rams and 2,549 dams. The 481 ewes with phenotypes were distributed in 9 half-sib families with an average size of 50 daughters per ram (ranging from 11 to 173 animals per half-sib family).

2.2 Genotyping

Blood samples from the 481 ewes were collected. The use of animals was performed followed the recommendation of directive 2010/63/EU. Genomic DNA was extracted from blood samples using a *salting out* method (Miller *et al.*, 1988). The concentration of extracted DNA was assessed with NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). All animals were genotyped using the Illumina OvineSNP50K Genotyping BeadChip and the HiScanSQ platform, (Illumina Inc., San Diego, CA, USA) following standard operating procedures recommended by the manufacturer. Genotyping was performed by Dipartimento SAAF (Figure 4). The OvineSNP50 BeadChip, developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGS), is a genome-wide genotyping array for the ovine genome that has different applications such as identification of QTLs, GWAS, characterization of genetic variability among breeds and genomic selection. This chip contains 54,241 SNPs chosen for being uniformly widespread across the ovine genome (with an average gap size of around 50.9 kb) and for their evident polymorphism in the evaluated animals.

Raw signal intensities were converted into genotype calls using the Illumina Genome Studio Genotyping Module v2.0 software (Illumina Inc., San Diego, CA) by applying a call threshold of 0.15. In particular, the software was used to generate *ped* and *map* files that containing genotypes and the genomic coordinates of the markers, respectively. Several quality controls of genotypic data were applied to investigate the SNPs integrity and usefulness. Genotyping data were initially tested for quality using PLINK (Purcell, 2007) software. Chromosomal coordinates for each SNP were obtained from the latest release of the ovine genome sequence assembly Oar_v4.0. SNPs were filtered to exclude loci assigned to unmapped contigs and to sex chromosomes. After quality control, 46,827 SNPs, distributed across 26 autosomes were retained.



Figure 4. HiScanSQ platform and the Ovine50K BeadChip of Illumina.

2.3 Quality Control

Data cleaning is the first and essential step for data analysis. Whether the goal is prediction of the outcomes or to discover new biology underlying the trait of interest, the inference of GWAS depends upon the overall quality of the data (Anderson *et al.*, 2010). Even simple statistical tests of association are compromised in the context of GWAS with data that have not been properly

cleaned, potentially leading to false-negatives and false-positive associations. Hence, we followed the common steps for quality control in genome-wide association studies to prevent these issues. The SNPs' assays that failed on a large number of samples are poor assays, and are likely to result in spurious data. Hence, SNPs with low call rate must be discarded. A recommended threshold for removing SNPs with low call rate is approximately 90–99%, although this threshold may vary from study to study and it should be decided by researcher. For our data and analysis a call rate greater than 95% was chosen. Another important issue in quality control is to exclude SNPs with low variability for minor allele so called rare SNPs. This filtering step helps to improve statistical power. So, removing extremely rare SNPs including any monomorphic SNP has been recommended. The choice of threshold depends on the size of study and the impact of SNP-effects in priori (Anderson *et al.*, 2010). In our study, SNPs with minor allele frequency (MAF) > 0.02 was used. Checking for Hardy-Weinberg Equilibrium (HWE) is the final step in the quality control analysis of markers in genome-wide association studies. Under Hardy-Weinberg assumptions, allele and genotype frequencies can be estimated from one generation to the next. Typically, HWE deviations toward an excess of heterozygotes reflect a technical problem in the sample, such as non-specific amplification of the target region. If no technical errors are detected then a number of biologically plausible explanations exist for HWE deviations such as population stratification, assortative mating and inbreeding. In animal studies and some human population, Hardy-Weinberg equilibrium check, may not be as usual due to inbreeding and not random mating in the sample population. Not random mating and inbreeding are two conditions that violate crucial assumption of HWE because inbreeding increases the frequency of homozygous, and decreases the frequency of heterozygous genotypes. In our data, samples in the same farms are likely to share the same alleles, inherited from common ancestors. Therefore, their progeny has an increased chance of being autozygous that refers to inherit a copy of exactly the same ancestral

allele from both parents. In our analysis, markers and animals with no extreme deviation from Hardy-Weinberg proportions ($P > 0.001$, Bonferroni corrected) were kept in the analysis. Quality control was performed with GenABEL packages in R (GenABEL Project Developers, 2013; R core Team, 2013) using the *check.markers* function. SNPs that did not satisfy these quality criteria were discarded. After filtering, the final number of samples and SNPs retained for analyses were 469 and 37,228 respectively.

2.4 Population structure

After quality control of phenotypes and genotypes, a major practical issue for studying complex traits or disease is to identify population structure in the data as ignoring this step reduces the power of genetic studies. Populations can be divided into subpopulations which are more or less distinct breeding groups in limited areas. Allelic frequency is one of the factors used to study populations which should be representative for the whole population. If the population is divided into subpopulations then the allelic frequencies can be different between the subgroups (Hartl, 1994). Therefore, it is important to reveal all possible sub-populations before genetic analysis of population is performed. If we do not account for population structure, we will identify spurious associations due to differences in ancestry rather than true association of alleles with the traits. There are many algorithms and programs designed to study the populations' structure. Common methods are principal component analysis (PCA) (Lee et al., 2009) and multidimensional scaling (MDS) (Purcell *et al.*, 2007) which are both multivariate statistical techniques. In PCA, the principal components are constructed from a linear combination of the genotypes of genetic markers such as SNPs. Each principal component should maximize the variance between the samples used in the analysis. MDS is used to analyse genetic distance matrices and places samples on a graph so that the distances

between them represent their true genetic distances. Population structure based on genetic markers is often viewed on a two-dimensional graph plotting two components of MDS or PCA. The results from these two methods are generally quite similar to each other (Wang *et al.*, 2012). However this can be done using kinship coefficients estimated from genomic data where the genomic estimate of kinship for a pair of individuals i and j is obtained using the formula (Leutenegger *et al.*, 2003)

$$f_{i,j} = \sum_k^{1,N} \left(\frac{(x_{i,k} - p_k)(x_{j,k} - p_k)}{p_k(1 - p_k)} \right)$$

where $f_{i,j}$ is the genomic kinship (identical-by-state) between animal i and j , k ranges from 1 to N (number of autosomal SNPs), $x_{i,k}$ or $x_{j,k}$ is the genotype of i^{th} or j^{th} animal for k^{th} SNP (coded as 0, 1/2 and 1) and p_k is the allele frequency at the k^{th} SNP. The kinship matrix was transformed to a distance matrix ($0.5 - f_{i,j}$) and principal components (PCs) of variation of the genomic distance matrix were calculated using the *cmdscale* function. The first two PCs (PC1 and PC2) were used to obtain the classical multidimensional scaling (MDS) plot (Figure 5).

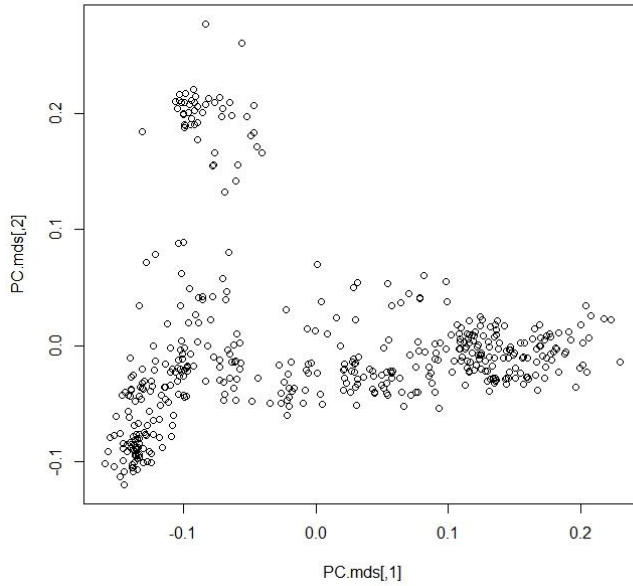


Figure 5. Multi-Dimensional Scaling plot.

2.5. Estimated breeding value (EBVs)

The most common phenotypes used in genome-wide association studies and other genomic based analyses are individual measurements. Another possibility is to use the estimated breeding values (EBVs) of the individuals (Pausch *et al.*, 2011). This has been done in several GWAS. Estimated breeding values are usually based on information about individuals, their offsprings and their relatives (depending on available information and the model used to calculate the value). In this study, the phenotypes analysed as response variables were MY, F%, FY, P% and PY. A single trait repeatability test-day animal model was performed to estimate the breeding values as follows:

$$y_{ijklmno} = \mu + htd_i + OP_j + LS_k + STG_l + FIM_m + a_n + pe_n + pe_{jn} + e_{ijklmno}$$

where $y_{ijklmno}$ was the test-day measurement for the considered traits; μ was the population's mean; htd_i was the random effect of herd by test-day interaction i (180 levels); OP_j was the fixed effect of the parity (5 levels); LS_k was the fixed effect of litter size class k (2 levels, single or multiple born lambs); STG_l was the fixed effect of season of lambing l (2 levels), where the season of lambing was coded as 1 if a ewe gave birth in the period January through June, otherwise it was coded as 2 (Riggio *et al.*, 2007); FIM_m was the fixed effect of fourteen days in milk m (22 levels); a_n was the random animal additive genetic effect n (481 levels); pe_n was the general random permanent environmental effect of individual n across lactations (481 levels); pe_{jn} was the random permanent environmental effect on the individual n within parity class j (2,405 levels); $e_{ijklmno}$ was the random residual effect. The estimating genetic variances were carried with the pedigree information available for the last three generations that included 1,304 animals of which 101 rams and 823 ewes. Variance components, genetic parameters and estimated breeding values for each trait were estimated using ASReml (Gilmour, 2009).

2.6 Deregressed and weighted breeding values

Genetic evaluation of animal population results in EBVs that are a weighted function of the parent average EBV, any information on the individual, adjusted for fixed effects, and a weighted function of the EBV of offspring, adjusted for the merit of the mates (Garrick *et al.*, 2009). In the study of Garrick *et al.* (2009), it has argued the removal of parent average effects in constructing information for genomic analyses and that information from genotyped descendants should also be removed to avoid double-counting. Simulation in the study of Garrick *et al.* (2009) suggests that the double-counting of descendants' performance has negligible impact on genomic predictions. A

logical approach was suggested by these authors which demonstrated the appropriate weights for analyzing observations with heterogeneous variance and explained the need for and the manner in which EBVs should have parent average effects removed, be deregressed and weighted. The simplest form of deregression of EBVs was obtained dividing each EBV by its reliability (r^2) (Garrick *et al.*, 2009). When DEBVs are corrected for information of relatives, this also means that weights of the DEBVs are on average expected to be lower than the corresponding weights of the original EBVs. Whether or not deregression of EBVs and computation of appropriate weights has an (large) effect on the results in subsequent analyses depends on differences between reliabilities of EBVs across animals due to different amounts of information per individual.

The EBVs for MY, FY, F%, PY and P% estimated with the mixed model were deregressed according to Garrick *et al.*(2009) to obtain a more accurate estimate of the expected phenotype as follows:

$$DEBV = \frac{EBV}{r^2}$$

where EBV is the estimated breeding value of each individual considering each milk production trait and r^2 is the reliability of that EBV.

To account for the heterogeneous variance of deregressed breeding values due to differences in breeding value accuracy of individual belonging to different populations, a weighting factor (w_i) should be used. Then following the approach of Garrick *et al.*(2009), a weighting factor (w) was estimate based on the reliability and heritability of each trait as follows:

$$w = \frac{1 - h^2}{\left[c + \frac{1 - r^2}{r^2} \right] h^2}$$

where h^2 was the heritability of the traits, r^2 was the reliability of the EBV, and c was the fraction of genetic variance not explained by markers assumed to be 0.1.

Weighted deregressed breeding values (DEBVw) were then estimated applying weighting factor (w) to DEBVs.

2.7 Genome Wide Association analysis

Genome-wide association analysis was carried out based on regression of phenotypes (EBVs, DEBVs and DEBVw) with the genotypes of animals for one SNP at a time. For single-marker GWAS, we used a three-step approach referred to as genomic GRAMMAR-GC (Genome-wide Rapid Association using Mixed Model and Regression-Genomic Control) (Aulchenko *et al.*, 2007; Amin *et al.*, 2007). The advantage of this approach especially in livestock is that it accounts for cryptic population structure caused by the presence of closely related animals (Aulchenko *et al.*, 2007) inferring relationships through genomic marker data. Following this approach in the first step, phenotypes were corrected by accounting for familiar dependence among individuals using:

$$y_i^* = y_i - (\hat{\mu} + \hat{G}_i)$$

where y_i^* is the so-called “environmental residual”, y_i is the phenotype of i^{th} animal, $\hat{\mu}$ is the overall mean, \hat{G}_i is the estimated polygenic contribution. In the second step, these familiar correlation-free residuals were used as dependent quantitative traits for association analysis of each SNP using a linear regression model:

$$y_i^* = \mu + \alpha_j g_i + e_i$$

where y_i^* is as defined before, g_i is the genotype of the i^{th} individual at the marker under study, α_j is the effect of j^{th} SNP and e_i is the random residual for the i^{th} individual. In the third step, genomic control (GC) is applied to correct the test statistic using a deflation factor (ζ) calculated by:

$$\hat{\zeta} = \text{Median}(T_1^2 + T_2^2 + \dots, T_j^2) / 0.465$$

where T_j^2 is the observed chi-squared (X^2) statistic for the j^{th} SNP and 0.465 is the expected median of $X^2_{(1)}$ distribution with a non-central variance. T^2 for each SNP is calculated by:

$$T_j^2 = \alpha_j^2 / \text{var}(\hat{\alpha}_j)$$

where α_j^2 is the effect of j^{th} SNP. $T^2/\hat{\zeta}$ is compared with $X^2_{(1)}$ to determine whether the locus is significantly associated with the quantitative trait. The deflation factor is estimated in the same way as inflation factor (λ) in conventional GC method (Bacanu *et al.*, 2002) with the difference that $\zeta < 1$ in contrast to λ that is constrained to be > 1 . This difference is due to the regression of residuals instead of original trait on n loci in step 2. *Polygenic* (Thompson *et al.*, 1990) and *qtscore* (Aulchenko *et al.*, 2007; Amin *et al.*, 2007) functions implemented in the GenABEL package of R software were used for association analysis.

2.8 Genome Wide Association analysis using repeated measures

So far, one of common denominator is that association analyses are made for phenotypic data where single record per individual is collected. This represented a problem for longitudinal data, in particular in dairy sheep the

phenotypic data for milk production traits are measured at different time points during a life cycle. Analyzing such kind of data affords us an opportunity to investigate the heterogeneity of traits over time and early prediction of longitudinal traits. It is, therefore, essential that a method for GWAS includes the process of repeated sampling. In a GWAS, the effects of thousands of SNPs need to be fitted and any model development is constrained by the computational requirements. Until now, GWAS software was not able to analyze repeated records for an individual. The only one method that accounts for repeated measures is the R package RepeatABEL (Husby *et al.*, 2015; Rönnegård *et al.*, 2016) that fits fixed SNP effects in a linear mixed model that can include both random polygenic effects and permanent environmental effects. In this way, the model can correct for population structure and model repeated measures. The covariance structure of the linear mixed model is first estimated and, subsequently, used in a generalized least squares setting to fit the SNP effects. Before carried out the analysis with RepeatABEL, we performed a quality control with the package GenABEL of R software, in this way we considered animals and markers that passed the quality criteria. RepeatABLE involves two steps: in the first one a linear model with all explanatory variables included (fixed and random) except the SNPs' effects, is fitted to estimate variance components. The linear model was performed with the function *preFitModels* that include the same random and fixed effects described previously. In the second step, these variance component estimates are used in the GWAS where, one a time, each marker was fitted with the *rGLS* function. The *rGLS* function is the main function of the package and by default fits a linear mixed model including permanent environmental effects (*p*) and polygenic effects (*g*) with correlation matrix given by the genomic relationship matrix in:

$$y = X\beta + Zg + Zp + e$$

where the vector y was the studied trait, β was the vector of fixed effect, g and p are the effects defined above, and e was the residual random effect, X and Z incidence matrix for fixed and random effects, respectively. The random effects were assumed to have a multivariate normal distribution such that $g \sim N(0, G\sigma_g^2)$ where G is the genomic relationship matrix construct using the genotypic information of animals; $p \sim N(0, I\sigma_p^2)$ where I is an identity matrix of proper order and $e \sim N(0, \sigma_e^2)$. Thus, the estimated (co)variance matrix for this model was: $\hat{V} = ZGZ^T\hat{\sigma}_g^2 + ZZ^T\hat{\sigma}_p^2 + I\hat{\sigma}_e^2$.

Finally, a linear model is fitted (using generalized least square, *GLS*) for each marker where the covariate x_{SNP} is coded as 0, 1, 2:

$$y = X\beta + x_{SNP} + e$$

Where:

$$e \sim N(0, \sigma_e^2 \hat{V})$$

2.9 Statistical Inference

Bonferroni method was adopted to adjust for multiple testing from the number of SNPs detected using the EBVs and their weighted measures (DEBVs and DEBVw) or longitudinal data as response variable. After Bonferroni correction, we considered a significant SNP at the genome-wide or suggestive levels, if a raw p-value $< 0.05/N$ and $1/N$, where N is the number of SNP loci tested in analyses, was obtained. Significance thresholds were $P < 1.34 \times 10^{-6}$ and $P < 2.68 \times 10^{-5}$ for genome-wide and suggestive levels, corresponding to $-\log_{10}(P)$ 5.87 and 4.57, respectively. The results of GWAS were plotted using Manhattan plot based on the obtained P-value of each SNP. The p-values of the association test were transformed to $-\log_{10}(P)$ for each SNP *versus* its chromosomal location. Manhattan plot shows locations of statistically significant SNPs across the chromosomes (horizontal axis) associated with their

effects (vertical axis) on the traits of interest. The black and red lines correspond to genome-wide and suggestive levels respectively. Quantile-quantile (Q-Q) plots were used to analyze the extent to which the observed distribution of the statistic test followed the expected (null) distribution. This was done to assess potential systematic bias due to population structure or analytical approach.

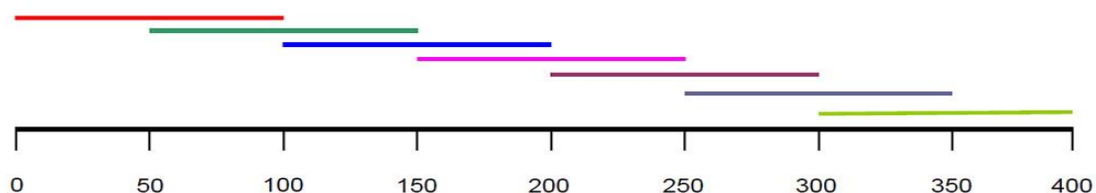
2.10 Regional Heritability Mapping

The Ovine SNP50k chip has already been demonstrated as providing the ability to map causal mutations for traits showing simple patterns of inheritance (Becker *et al.*, 2010). However, in sheep, as well as in other species including humans, answers have not been so definitive for complex traits and GWA studies have generally failed to explain most of the known genetic variation influencing complex diseases or production traits (Manolio *et al.*, 2009; Kemper *et al.*, 2011). These studies typically test each marker independently for an association with the trait. The expectation is that the variance explained by each marker is proportional to the size of the effect of the (unobserved) polymorphism on the trait, the degree of association between the marker and the polymorphism, and the experimental error associated with the measurement. Attempts to increase the power of association studies have focused either on increasing the number of markers or the number of observations for a trait. Few authors have attempted to formally estimate the distribution of marker effects with dense SNP markers (Hayes and Goddard., 2010; Kemper *et al.*, 2011) and thus the required power of experiments for complex disease traits, such as nematode resistance, is still unknown. An alternate approach exploiting dense SNP chip data, known as Regional Genomic Relationship Mapping or Regional Heritability Mapping (RHM) (Nagamine *et al.*, 2012), has been advanced as a better approach to capture more of the underlying genetic effects. This method

provides heritability estimates attributable to small genomic regions, and it has the power to detect regions containing multiple alleles that individually contribute too little variance to be detected by standard GWA studies. To do this we used genome-wide SNP data to estimate the genetic relationships between all pairs of individuals in the population, both at the level of the whole genome and for each region within the genome. We then employed these relationships to estimate the trait variance contributed both by the genome as a whole (the genomic heritability) and by short regions of the genome (the regional heritability). The genomic heritability provides an estimate of the overall heritability but also controls for population structure. Studies by other authors have demonstrated that using the pedigree or genomic relationship matrix in a mixed model to estimate single SNP effects in pedigree structured data proves a powerful and unbiased analysis (Aulchenko *et al.*, 2007; Kenny *et al.*, 2011). Thus inclusion of the genomic relationship means that the regional heritability is unbiased by overall population structure and hence provides a metric that indicates local genomic regions contributing to trait variation.

2.11. Procedures

To handle with repeated measurements a modification of the methodology was applied. Each chromosome (OAR) was divided into windows of a pre-defined number of SNPs, and the variance attributable to each window estimated. In this study, a window size of 100 adjacent SNPs was used to construct a regional relationship matrix and the window was shifted every 50 SNPs. Therefore, first, second and third regional matrix, for example, used from 1th to 100th, 51th to 150th and 101th to 200th.



A linear mixed model accounting for the same fixed and random effects, as reported in the previous analyses was used:

$$y = X\beta + Zu + Zv + e$$

In this case the additive genetic effects were fitted as both regional genomic and whole genomic effects.

$$\text{var}(u) = G\sigma_u^2, \text{Var}(v) = Q\sigma_v^2, \text{Var}(e) = I\sigma_e^2$$

where the vector y represents the phenotypic values, X is the design matrix for the fixed effects, and Z is the design for random effects. The remaining vectors are, u : whole genomic additive effect, v : regional genomic additive genetic effect, e : residual, and β : fixed effect. Matrices G and I are whole genomic relationship matrix using all SNPs for whole genomic additive effects and a unit matrix for residual, respectively. Q is regional genomic relationship matrix obtained using 100 SNPs for regional genomic additive effects. Whole genomic, regional genomic and residual variances are σ_u^2, σ_v^2 and σ_e^2 respectively. Phenotypic variance, σ_p^2 is $\sigma_u^2 + \sigma_v^2 + \sigma_e^2$. Whole genomic heritability and regional heritability are $h_u^2 = (\sigma_u^2/\sigma_p^2)$, and $h_v^2 = (\sigma_v^2/\sigma_p^2)$, respectively.

2.12 Statistical inference for RHM analysis

A likelihood ratio test (LRT) was used to test for the differences in regional variance, comparing a model fitting variance in a specific window (fitting both whole-genome and region-specific additive variance) against the null

hypothesis of no variance in that window (whole-genome additive variance only). The test statistic was assumed to follow a mixture of $\frac{1}{2}\chi^2_{(1)}$ and $\frac{1}{2}\chi^2_{(0)}$ distributions (Self and Liang, 1987). In total 827 windows were tested, of which half was used in the Bonferroni correction, to account for the overlapping windows. Hence, after Bonferroni correction to account for multiple testing, the LRT thresholds were 13.48 and 9.20, for genome-wide and suggestive significance levels, respectively. As for the other methods, Manhattan plots were used to graph the results.

2.13 Significant SNPs location on OAR4.0 genome assembly

To investigate the gene location of the significant SNPs associated with milk production traits, detected in this study, we interrogated dbSNP in NCBI Database (<https://www.ncbi.nlm.nih.gov/projects/SNP/>). Gene content of significant regions of heritability was assessed using Sheep RefSeq in the Genome Data Viewer genome browser in NCBI Database (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?acc=GCF_000298735.2&context=genome). Finally, both for significant SNPs and regions of heritability, SheepQTL database (<https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>) was interrogated and QTL for milk production traits reported.

Chapter 3

Results and Discussion

3.1 Results using EBv, DEBV, and DEBVw

Quality control for SNPs was described in Chapter 2. Classical Multi Dimensional Scaling identified no outlier in our sheep breed indicating the absence of genetic substructure. After filtering, the final number of samples and SNPs retained for analyses were 469 and 37,228, respectively. Moreover, the final dataset for GWAS contained more than 5,400 records for MY, FY, F%, PY and P% (Table 1).

Table 1. Milk production traits descriptive statistics.

Traits	N	Mean±SD	CV(%)	Min-Max
MY (g)	5,446	1,367±532	38.91	100-3,924
FY (g)	5,437	94.80±33.91	35.76	6-277
F% (%)	5,437	7.08±1.06	14.94	2.53-15.78
PY (g)	5,436	78.93±29.32	37.15	6-239
P% (%)	5,436	5.82±0.65	11.20	2.32-11.60

MY: Milk Yield; FY: Fat Yield; F%: Fat percentage; PY: Protein Yield; P%: Protein percentage; SD: standard deviation; CV: coefficient of variation; Min-Max: minimum and maximum value

Among the traits MY, FY and PY showed the highest variability with a coefficient of variation of 38.91%, 35.76% and 37.15%, respectively. In some cases, milk production and fat and protein content were very high (3,924 g, 277 g, and 239 g, respectively) revealing a good performance of this breed for milk production. The mean phenotypic values were in agreement with previous studies on the same sheep breed (Riggio *et al.*, 2007). Additive genetic variance, residual variance, phenotypic variance, repeatability within and across lactation, and heritability for milk production traits, estimated using an animal model with REML algorithm, are showed in Table 2.

Table 2. Genetic parameters estimates and heritability for milk production trait.

Traits	σ_a^2	σ_r^2	σ_p^2	$r_{wit}\pm SE$	$r_{acr}\pm SE$	$h^2\pm SE$
--------	--------------	--------------	--------------	-----------------	-----------------	-------------

MY	15,356.2	254,794.8	256,330	0.266±0.039	0.169±0.021	0.060±0.037
F%	0.0107	1.1150	1.1250	0.065±0.022	0.177±0.017	0.009±0.021
FY	70.0003	1,003.6	1,073.6	0.293±0.040	0.162±0.021	0.065±0.039
P%	0.0124	0.3676	0.3800	0.100±0.032	0.240±0.020	0.033±0.032
PY	54.4695	753.96	808.43	0.276±0.038	0.148±0.020	0.067±0.037

σ_a^2 : additive genetic variance; σ_r^2 : residual variance; σ_p^2 : phenotypic variance; r_{wit} : repeatability within lactation, r_{acr} : repeatability across lactation; h^2 : heritability; SE: standard error.

Heritability estimates for milk yield and milk composition traits were low and ranged between 0.009 to 0.067 with standard error ranging from 0.021 to 0.039. In a previous study, Riggio *et al.* (2007) considering a single lactation of primiparous ewes of the Valle del Belice breed, reported slightly higher estimates for these traits. These differences are due probably to different data set and model used. In fact in our data set, several lactations of the same individual are considered and a second permanent effect within lactation in the animal model was added. The heritability of milk production in sheep is estimated to range from 0.13 (Gutiérrez *et al.*, 2007) to 0.35 (Baro *et al.*, 1994), but other studies estimate the value of heritability of test-day milk production data from 0.18 (e.g. El-Saied *et al.*, 1998a) to 0.25 (e.g. Barillet and Boichard, 1994). Heritability values for milk yield, fat and protein content were lower than those reported by Othmane *et al.* (2002) in Churra sheep breed. Heritability estimates depends on the population sampled and statistical model used. These estimates were lower than those frequently reported in dairy cows where selection programs are more advanced and the production levels are higher. Furthermore, our results were influenced by the differences of milking practices and management for Valle del Belice breed respect to dairy cows (e.g., hand vs machine milking, feeding practices). Repeatability within lactation estimates for all milk production traits ranged between 0.06 and 0.29, while repeatability across lactation estimate ranged between 0.16 and 0.24. The standard errors were around 0.3 for the repeatability across lactation and 0.2 for the one within lactation. The lowest estimates for heritability and repeatability within lactation were found for F%. Repeatability estimates for milk composition traits were lower than those reported for dairy ewes (Riggio *et al.*, 2007; Othmane *et al.*, 2007).

3.1.1 Significant SNPs

The total number of SNPs associated with milk production traits considering different breeding value estimations are reported in Table 3. As described above, p values (in term of $-\log_{10}(P)$) of tested SNPs for the five traits are shown using a Manhattan plot with a suggestive and a genome-wide threshold levels specific for each test statistics association (Figures 6-8). The quantile-quantile (Q-Q) plots (scatter plot of the observed *vs.* expected values of test statistics derived from the GWAS analyses) were also inspected providing extra evidence for true associations of the GWAS analyses (Supplementary Figure S1).

Table 3. Total number of SNPs associated with milk production traits.

Traits	EBV	DEBV	DEBV _w	OVERLAP _{DEBV_{vs}DEBV_w}
MY	-	30	11	1
F%	1	6	1	-
FY	1	19	30	16
P%	4	22	22	21
PY	-	3	3	3

EBV: Estimated breeding value; DEBV: deregressed breeding value; DEBV_w: weighted deregressed breeding value. OVERLAP_{DEBV_{vs}DEBV_w}: N of SNPs overlaps between DEBV and DEBV_w

A total of 153 SNPs using EBV, DEBV and DEBV_w as response variables for the five milk production traits were found. DEBV and DEBV_w allowed identifying the largest number of markers associated with the traits. Since some of these SNPs are associated with more than one trait, the total number of distinct identified SNPs was 6, 80, and 67, for EBV, DEBV, and DEBV_w, respectively. No SNPs found with EBV approach overlaps with the two other approaches (DEBV and DEBV_w). Between DEBV and DEBV_w, a total of 41 SNPs overlapped and a total of 14 SNPs were in common among traits, therefore, 28 and 23 distinct SNPs were detected by DEBV and DEBV_w, respectively. Thirty-four SNPs (42.5%) reached the genome-wide threshold

level of significant using DEBV as response variable, whereas 24 SNPs (35.8%) using DEBVw. Among the identified SNPs only one (rs422391756) located on chromosome (OAR) 2 was found at genome-wide threshold level associated with P% trait using EBV.

The details of these SNPs associated with milk production traits, including their raw P values, the position in the genome and the known genes within they laid, are given in Tables 4 through 8.

3.1.2 Milk Yield

As showed in Table 4 and Figure 6, significant associations for MY were identified on 14 different chromosomes (1, 2, 3, 4, 5, 7, 8, 10, 11, 14, 16, 17, 18, 20). Most of these SNPs were mapped on OAR2. SNPs identified on OAR2, located between 162.16 and 178.47 Mb (rs424490890, rs403122982, rs419432879 and rs429723758) laid into the region that have been reported to harbour QTL (ID=13992) by Raadsma *et al.* (2009) for milk production. Moreover, other two SNPs (rs416541416 and rs424204447) very close to each other (0.03 Mb) were located in a genomic region which have been reported to harbour QTL (ID=14147) for MY (Maatescu *et al.*, 2010). In the same study two QTLs (ID=14149 and ID=14150) on OAR18, in the same region of the identified SNP rs424842019 for MY, were reported. Garcia-Gamez *et al.* (2013) identified several QTLs on OAR 2, 16, 17 for MY. In our study, SNPs rs403120738 and rs408758615 on OAR2 between 55.60 and 61.13 Mb, where QTL for MY (ID=57737) is located, have been detected. On OAR16 we identified SNP rs405662186 that laid in the region of QTLs IDs 57730 and 57752, both for MY. The SNP rs424648601e rs423013684 on OAR17 between 19.38 and 25.08 Mb reported a QTL (ID=57756) for the same trait.

Markers associated with milk yield are mapped within 15 known potential candidate genes identified in different species. Studies have shown that lipin proteins, particularly *LPINI*, play crucial roles during adipose tissue

development and triacyl-glycerol accumulation (Phan and Reue, 2005). Furthermore, the expression levels of the lipin genes have been shown to influence lactation, with *LPINI* predominating during lactation. Thus, this gene is clearly involved in modifying the composition of milk during lactation. Finck *et al.* (2006) demonstrated that *LPINI* is essential for PPAR (peroxisome proliferator activated receptor) activation, suggesting that *LPINI* may be involved in regulating the transcription of other genes involved in milk fat synthesis.

Table 4. Genome-wide and suggestive significant SNPs for MY

SNPs	Chr	Position	P-value	Genotype	Gene
rs423676435 ^c	1	239181622	8.72E-07	A/G	
rs407841431 ^b	1	265454228	1.88E-09	A/G	
rs422253848 ^c	1	54156664	2.40E-05	A/C	
rs425325353 ^c	1	57356494	4.93E-07	A/G	
rs429625897 ^b	1	15875818	2.90E-06	A/G	<i>HIVEP3</i>
rs422638951 ^b	2	28678758	8.28E-06	C/T	<i>ASPN/CENPP</i>
rs424490890 ^b	2	162163064	6.85E-06	A/G	
rs403122982 ^b	2	162207791	1.31E-05	A/G	
rs419432879 ^{bc}	2	177265054	1.33E-05	A/G	<i>NCKAP5</i>
rs429723758 ^c	2	178476225	2.19E-06	C/T	
rs403120738 ^b	2	55605381	4.39E-08	C/T	
rs416541416 ^b	2	78960640	5.18E-06	C/T	
rs424204447 ^b	2	78996919	1.97E-06	A/G	
rs418285319 ^b	2	177378334	3.44E-06	A/G	<i>GPR39</i>
rs408758615 ^b	2	61134963	8.40E-06	A/T	
rs418431341 ^b	3	102980403	2.56E-11	C/T	<i>TMEM131</i>
rs415464905 ^b	3	114017582	1.33E-05	C/T	
rs422404167 ^b	3	20508250	7.01E-07	A/G	<i>LPIN1</i>
rs420925403 ^b	4	101796199	2.98E-07	C/T	
rs423430191 ^b	5	672391	2.27E-07	A/G	<i>MAPK9</i>
rs160014503 ^c	5	5308939	6.44E-06	A/G	<i>SLC27A1</i>
rs430152343 ^b	7	5066465	1.55E-05	A/G	<i>CPLX2</i>
rs423026263 ^c	7	96755821	1.09E-05	C/T	<i>KCNK10</i>
rs422503523 ^b	8	9741235	7.36E-06	A/G	
rs404623015 ^b	8	11199454	2.54E-07	C/T	<i>RSPO3</i>
rs426563184 ^c	10	35067247	2.75E-05	A/G	
rs415562533 ^c	10	37081632	4.03E-06	A/G	
rs407468867 ^b	11	54447900	1.91E-05	A/G	<i>RNF157</i>
rs430744508 ^b	11	45685462	6.69E-07	A/G	
rs413395758 ^b	14	2788578	1.73E-07	A/G	<i>CNTNAP4</i>
rs404689394 ^b	14	3812997	5.88E-07	C/T	
rs411845712 ^b	14	1783532	1.36E-06	C/T	<i>ZNRF1</i>
rs405662186 ^b	16	11964131	9.40E-06	C/T	
rs424648601 ^c	17	19385183	1.86E-05	C/T	
rs423013684 ^b	17	25087697	1.35E-06	C/T	
rs424557837 ^b	17	58827476	1.71E-06	C/T	
rs414091996 ^b	17	59065347	9.49E-08	G/T	
rs424842019 ^b	18	23087969	2.68E-06	A/G	<i>ADAMTSL3</i>
rs405048880 ^b	20	44900188	3.97E-11	G/T	

SNPs with superscript *a* are detected by EBVs, SNPs with superscript *b* are detected by DEBVs, SNPs with superscript *c* are detected by DEBVw, SNPs with superscript *bc* are detected by both DEBVs and DEBVw; SNPs in italic are located within QTL regions. Names of SNPs are in standard reference name format from the NCBI database (<http://www.ncbi.nlm.nih.gov>).

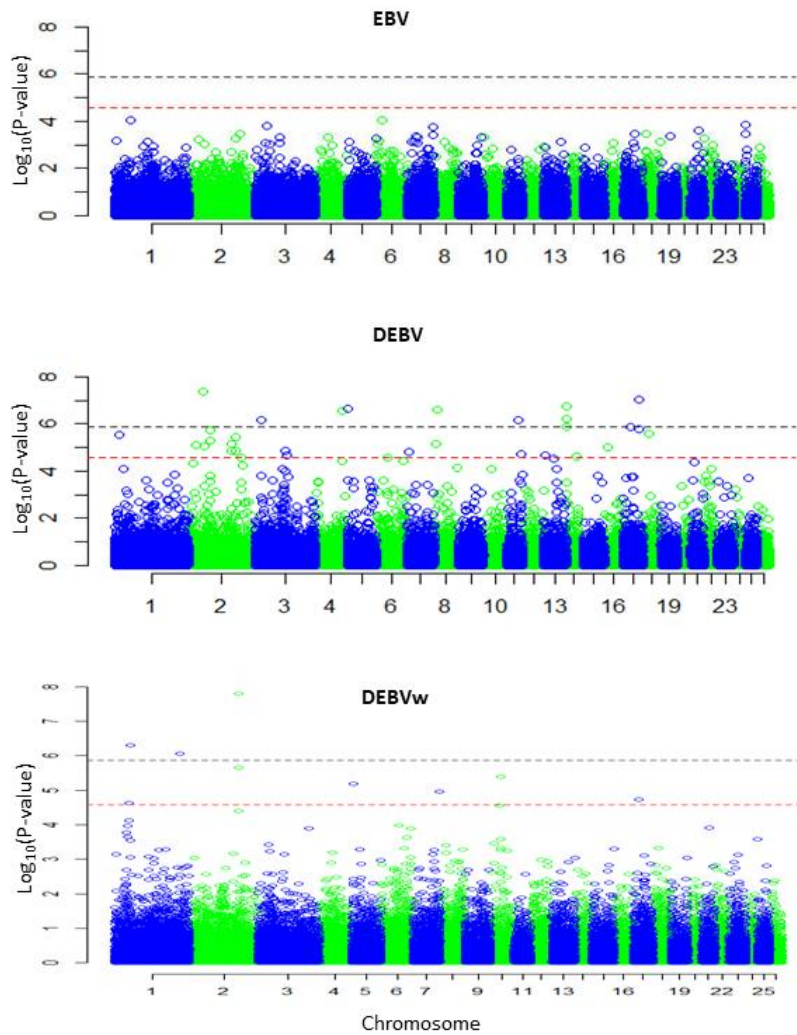


Figure 6 Manhattan plot for MY using EBV, DEBV and DEBVw

NCKAP5 on OAR2 is a protein coding gene interacting with a SH3 (Src homology region 3) domain of the adaptor protein NCK. This gene, in cattle, is associated with productivity traits, animal welfare and labor safety. (Valente, T. D. S. 2016). *SLC27A1* (solute carrier family 27 member 1) is a member of the fatty acid transport protein family. It is the transmembrane protein that facilitates long chain fatty acid (LCFA) transport across the cytoplasmic membrane. The study conducted on mice purified *SLC27A1* protein revealed its long-chain and very long-chain acyl-CoA synthetase activity (Hall *et al.*, 2003). *SLC27A1* encoding gene was mapped to bovine chromosome 7 where

QTLs for milk production traits have been identified (Ordovás *et al.*, 2005, Ogorevc *et al.*, 2009).

3.1.3 Fat percentage and yield

Tables 5,6 and Figure 7 showed the SNPs identified for F% and FY. SNP rs405045517 located on OAR25, laid in the same region that reports a QTL (ID=14010) for F% (Raadsma *et al.*, 2009). Furthermore, 7 of the SNPs associated with F% are mapped within genomic regions with known genes (Table 5). SNPs associated with FY, were identified on several chromosomes (1, 2, 3, 4, 5, 6, 7, 9, 11, 13, 17, 18, 20, 21). In OAR3 are located the highest number of significant SNPs between 97.58 and 120 Mb. Marker rs403120738 on OAR2 is mapped within a genomic region in which a QTL (ID=57736) for FY discovered by Garcia-Gamez *et al.* (2013) has been mapped. Marker rs406975522 located on OAR9 were identified in the same region that reported a QTL (ID=16015) for FY (Jonas *et al.*, 2011). Thirteen markers associated with FY were within 13 know genes (Table 6)

Table 5. Genome-wide and suggestive significant SNPs per F%

SNPs	Chr	Position	P-values	Genotype	Gene
rs402192268 ^b	3	117695610	1.68E-06	A/C	
rs417152368 ^a	4	37664073	2.62E-05	A/G	<i>PCLO</i>
rs399443712 ^c	9	59167089	1.24E-05	A/C	<i>EXT1</i>
rs55628574 ^b	13	17612673	2.27E-05	A/C	<i>PAR3</i>
rs429599628 ^b	19	29064142	2.01E-06	A/C	
rs411206905 ^b	21	29438769	5.22E-12	C/T	<i>KIRREL3</i> <i>UEVLD</i>
rs405002813 ^b	21	25433976	1.04E-08	C/T	<i>RHOBTB1</i>
rs405045517 ^b	25	16469719	3.36E-06	A/G	<i>BTB</i>

SNPs with superscript *a* are detected by EBVs, SNPS with superscript *b* are detected by DEBVs, SNPS with superscript *c* are detected by DEBVw, SNPs with superscript *bc* are detected by both DEBVs and DEBVw; SNPs in italic are located within the QTL regions. Names of SNPs are in standard reference name format from the NCBI database (<http://www.ncbi.nlm.nih.gov>).

Table 6. Genome-wide and suggestive significant SNPs for FY

SNPs	Chr	Position	P-value	Genotype	Gene
rs407841431 ^{bc}	1	265454228	1.12E-09	A/G	
rs417007329 ^b	1	266964438	1.31E-05	C/T	<i>HLCS</i>
<i>rs424490890^c</i>	2	162163064	8.67E-06	A/G	
<i>rs403120738^c</i>	2	55605381	3.05E-05	C/T	
rs420112859 ^b ^c	3	97589857	2.15E-05	G/T	
rs418431341 ^b ^c	3	102980403	3.99E-08	A/G	<i>TMEM131</i>
rs415464905 ^b ^c	3	114017582	2.15E-07	C/T	
rs416315048 ^c	3	114194065	1.11E-05	A/G	
rs402727086 ^c	3	120239095	1.27E-05	A/G	
rs401873093 ^c	4	101672990	2.71E-05	A/G	
rs420925403 ^c	4	101796199	4.75E-06	C/T	
rs401226229 ^c	5	95932217	1.32E-05	C/T	
rs423430191 ^b ^c	5	672391	4.45E-06	A/G	<i>MAPK9</i>
rs430111891 ^a	6	24257929	1.00E-05	A/G	<i>EMCN</i>
rs429361683 ^{bc}	6	65844069	4.54E-06	A/G	<i>GABRB1</i>
rs429981329 ^c	7	7961927	2.73E-05	C/T	
rs193638457 ^{bc}	7	33318065	4.74E-06	C/T	<i>DNAJC17, GCHFR, C7H15orf62</i>
<i>rs406975522^{bc}</i>	9	87132817	8.20E-08	A/G	
rs407723884 ^{bc}	11	18974119	1.01E-06	C/T	
rs407468867 ^c	11	54447900	9.25E-07	A/G	<i>RNF157</i>
rs420717702 ^c	11	19274974	3.34E-06	A/G	<i>NLK</i>
rs417619187 ^b ^c	13	5065704	2.02E-05	C/T	
rs411524173 ^c	13	45107038	2.21E-05	A/G	
rs406819520 ^c	13	55326126	6.27E-06	C/T	
rs424557837 ^{bc}	17	58827476	5.05E-07	C/T	<i>LOC105606343</i>
rs414091996 ^{bc}	17	59065347	1.20E-09	G/T	
rs409260825 ^c	17	64741368	3.39E-06	A/G	<i>SGPSM1</i>
rs424842019 ^{bc}	18	23087969	4.57E-06	C/T	<i>ADAMTSL3</i>
rs413552380 ^{bc}	18	30764193	4.22E-07	A/G	<i>SCAPER</i>
rs405048880 ^{bc}	20	44900188	1.47E-06	A/G	<i>LOC101117887</i>
rs419768704 ^c	21	9360101	9.96E-06	A/G	<i>DLG2</i>
rs424021641 ^{bc}	21	1802139	2.94E-06	A/G	<i>FAT3</i>

SNPs with superscript *a* are detected by EBVs, SNPs with superscript *b* are detected by DEBVs, SNPs with superscript *c* are detected by DEBVw, SNPs with superscript *bc* are detected by both DEBVs and DEBVw; SNPs in italic are located within the QTL regions. Names of SNPs are in standard reference name format from the NCBI database (<http://www.ncbi.nlm.nih.gov>).

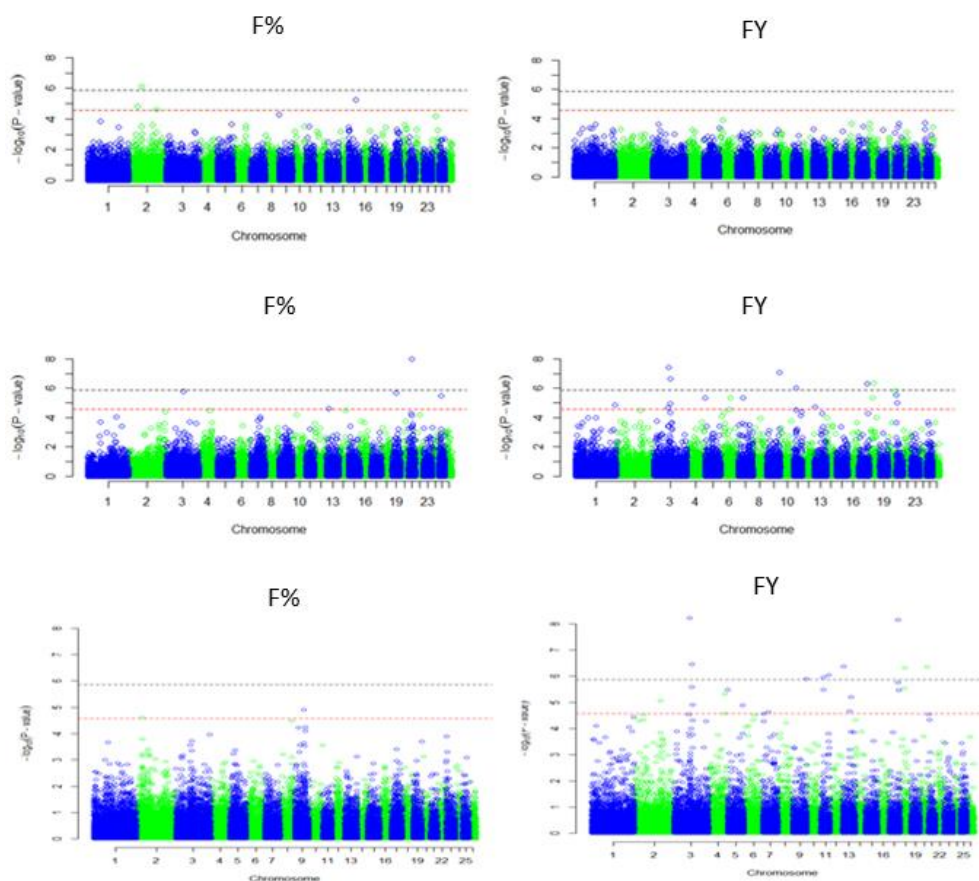


Figure 7 Manhattan plot for F% and FY using EBV, DEBV and DEBVw

3.1.4 Protein percentage and yield

Tables 7,8 and Figure 8 showed the SNPs identified for P% and PY. Most of markers associated for P% were located on OAR3. However, only the SNPs on OAR2 laid in a QTL for the considered trait (ID=57738) (Garcia-Gamez *et al.*, 2013; Maatescu *et al.*, 2010; Raadsma *et al.*, 2009). Ten markers related with P% were mapped within know genes (Table 7). In Table 8, the three SNPs associated with PY were reported. The variant rs406975522 on chromosome OAR9, was discovered located in the intronic region of MMP16 (matrix metallopeptidase 16) gene.

Table 7. Genome-wide and suggestive significant SNPs per P%.

SNP rs	Chr	Position	P-value	Genotype	Gene
rs403126323 ^{bc}	1	223905490	3.39E-10	A/G	
rs423919863 ^{bc}	2	14351231	2.08E-05	A/G	<i>IKBKAP</i>
rs417043797 ^{bc}	2	184393569	2.49E-05	C/T	<i>EPB41L5</i>
rs413985466 ^a	2	57368578	1.59E-05	A/G	
rs422391756a	2	87846505	7.66E-07	C/T	
rs413820801 ^{bc}	3	191528809	7.49E-06	A/G	
rs421071422 ^{bc}	3	191544417	1.35E-07	C/T	
rs424888367 ^{bc}	3	200068982	4.42E-10	A/G	<i>ATF7IP</i>
rs414580044 ^{bc}	3	206167771	4.94E-07	C/T	<i>CLEC4A</i>
rs408402130 ^{bc}	3	117079611	3.56E-06	A/G	<i>PPFIA2</i>
rs424276469 ^c	3	4735779	2.48E-05	A/G	
rs416914270 ^{bc}	4	29207729	2.84E-06	C/T	
rs426928771 ^{bc}	6	45108417	2.47E-06	C/T	<i>SEL1L3</i>
rs420418036 ^{bc}	6	46458951	9.89E-09	G/T	
rs417798754 ^{bc}	6	41745918	5.49E-07	A/G	
rs399420286 ^{bc}	9	31431961	1.21E-06	A/G	
rs418199431 ^{bc}	10	14719926	1.33E-07	A/G	
rs410586719 ^{bc}	11	24224499	2.24E-07	A/C	<i>ZZEF1</i>
rs422865208 ^{bc}	11	8385995	1.18E-06	A/G	
rs402691388 ^{bc}	11	45194161	1.66E-07	A/G	
rs414067603 ^{bc}	15	79902589	3.09E-07	G/T	<i>TCN1</i>
rs402908273 ^a	15	49771118	5.65E-06	C/T	<i>NUMA1</i>
rs400223670 ^{bc}	15	76158452	2.48E-05	G/T	
rs414091996 ^b	17	59065347	2.41E-05	G/T	
rs405883702 ^{bc}	18	59466973	3.34E-06	A/G	
rs421246582 ^{bc}	18	66191615	1.55E-06	C/T	<i>TRAF3</i>
rs426561933 ^{bc}	19	26430979	1.53E-06	C/T	

Table 8. Genome-wide and suggestive significant SNPs per PY.

SNP	Chr.	Position	P-value	Genotype	Gene
rs406975522 ^{bc}	9	87132817	3.87E-06	A/G	MMP16
rs409260825 ^{bc}	17	64741368	1.21E-05	A/G	
rs424021641 ^{bc}	21	1802139	1.42E-05	A/G	

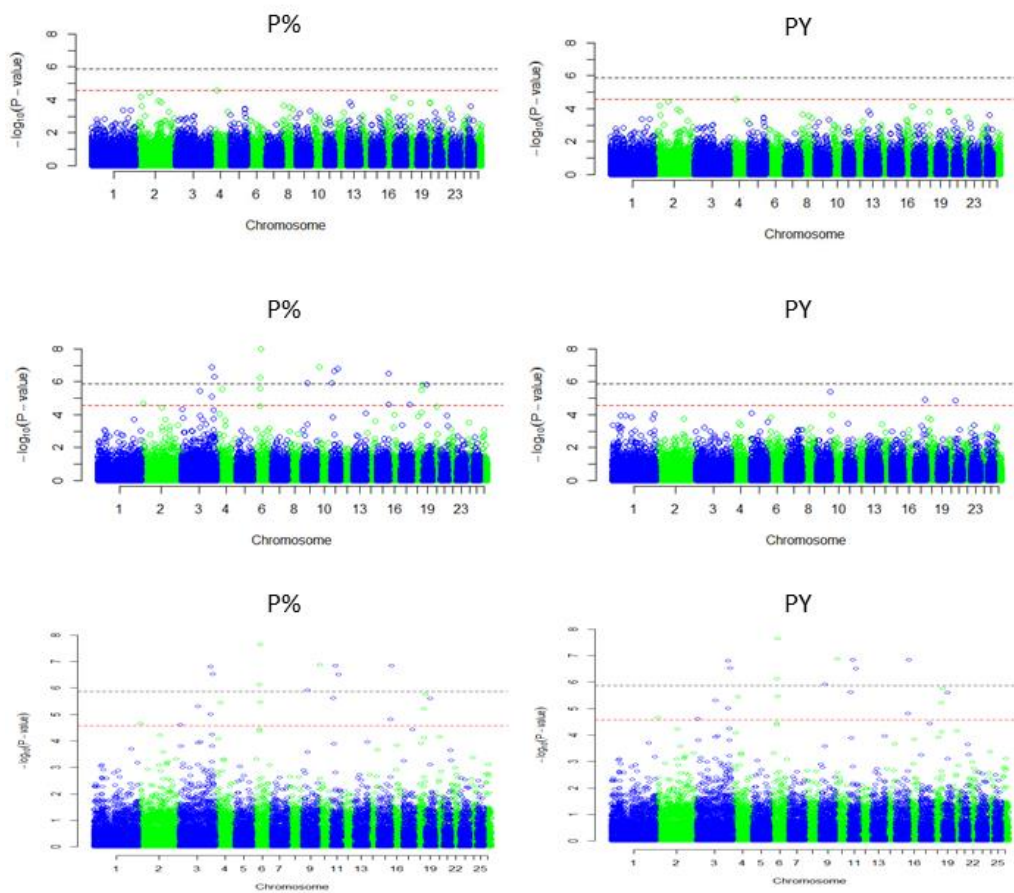


Figure 8 Manhattan plot for P% and PY using EBV, DEBV and DEBVw

3.1.5 Considerations using EBV, DEBV and DEBVw for GWAS

In GWAS of domesticated animals, often, the researchers may have available raw phenotypes of individuals and also their estimated breeding values (EBVs) from pedigree-based analyses of historical data. Some animals may have genotypes and EBVs based on information from their relatives, but no direct phenotypes. Although EBVs have been used as dependent variables in GWAS (Becker *et al.*, 2013; Johnston *et al.*, 2011), this approach has a high false positive rate (Ekine *et al.*, 2013). Consequences of using EBVs include varying levels of precision and ‘shrinkage effect’ among the values used to represent phenotypes of different individuals, a reduction in the sample variance of the

phenotypes, and double-counting of information from relatives (Garrick *et al.*, 2009; Ostersen *et al.*, 2011). As an alternative, the EBVs can be ‘deregressed’ (Garrick *et al.*, 2009; Ostersen *et al.*, 2011) to standardize the variance and influence of the individuals’ EBVs while still accounting for information from relatives. The use of deregressed EBVs (DEBVs) as dependent variables can improve the power of GWAS (Sevillano *et al.*, 2015; Sell-Kubiak *et al.*, 2015). To account for the heterogeneous variance of DEBVs due to differences in breeding value accuracy of individual belonging to different populations, the weighting factor was used and deregressed weighted estimated breeding values (DEBVw) was also used in for our GWA study. Comparisons among these three approaches with respect to their influence on our GWAS results demonstrate that DEBVs and DEBVw presented advantages over EBVs. We also compared DEBVs and DEBVw approaches and the results are more or less overlapped. In total, the present study revealed 153 genome-wise significant SNPs for milk production traits in Valle del Belice dairy sheep breed using GRAMMAR-GC approach. Several of these SNPs (n=19) are located within the previously reported QTL regions, and some within to the reported candidate genes. The significant SNPs detected within previously reported QTLs and candidate genes demonstrated that our results are in agreement with other GWAS studies on dairy sheep breeds.

3.2 Results using repeated measures

Incorporation of repeated measures may increase power to detect associations, but also requires specialized analysis methods to consider the correlated nature of data. This is the first analysis on sheep that consider longitudinal data to identify SNPs along the genome. In Table 9, the results considering all milk production traits are showed. The GWAS results for milk production traits showed the presences of a total of 6 significantly associated markers localized

within or near functional genes (Kominakis *et al.*, 2017) The SNPs associated with milk traits were identified on chromosomes 1, 7, 10, 21, and 26.

Table 9. Genome-wide and suggestive significant SNPs for milk production traits considering repeated measures

Trait	SNPs	Chr	Position	P-value	Genotype	Gene	Distances
MY	rs398340969	21	28853635	1.73E-05	A/G	<i>DCPS</i>	
MY	rs417920940	26	9913215	1.47E-05	C/T	<i>TENM3</i>	
F%	rs425417915	7	98977544	2.00E-09	C/T	<i>TTC7B</i>	
P%	rs417079368	1	233594672	4.34E-07	C/T	<i>SUCNR1</i>	0.37
P%	rs425417915	7	98977544	2.97E-06	C/T	<i>TTC7B</i>	
P%	rs419987770	10	28275437	4.68E-06	A/G	<i>KL</i>	0.01
PY	rs400055578	7	72674247	1.59E-05	A/G	<i>WDR89</i>	0.03
PY	rs398340969	21	28853635	2.10E-05	A/G	<i>DCPS</i>	

The GWAS analysis identified two SNPs with genome-wide significant association: rs425417915 for F% on OAR7 and rs417079368 for P% on OAR1, while other 6 SNPs reached the suggestive levels. Manhattan plots displaying the GWAS results of observed P-values against expected P-values for the traits considered are showed in Figure 9. In particular, the SNP rs425417915, associated at the same time with F% and P%, was found close to the *TTC7B* gene (tetratricopeptide repeat domain 7B) that play a crucial role in the lipid metabolism in cattle (Macleod *et al.*, 2016) and it was also reported as candidate gene for obesity in mice (Morton *et al.*, 2011). SNP rs398340969 is associated with MY and PY and it is located close to the *DCPS*, a gene differentially expressed in ovine mammary gland. Very interesting is that two SNPs (rs425417915 and rs398340969) were associated at the same time with different quantitative traits. Moreover *SUCNR1* gene identified at 0.37 Mb of SNP rs417079368 associated with P%, was a gene related to the pathways initially selected with regard to cheese trait. It is well known that genetic correlation exists between milk production traits and would be interesting to investigate the allelic substitution effect of these SNPs on MY and PY, and F% and P%. As a supporting the identified associations, the quantile-quantile (Q-Q) plots visualizing the distribution of the observed test statistics derived from the GWAS analyses, were checked providing extra evidence for true associations of the GWAS analyses (Supplementary Figure S2). This analysis has identified a smaller number of variants than using the breeding values as response

variables for the traits considered and this could be due to the reduced number of observations per animal or to the analytical method used here.

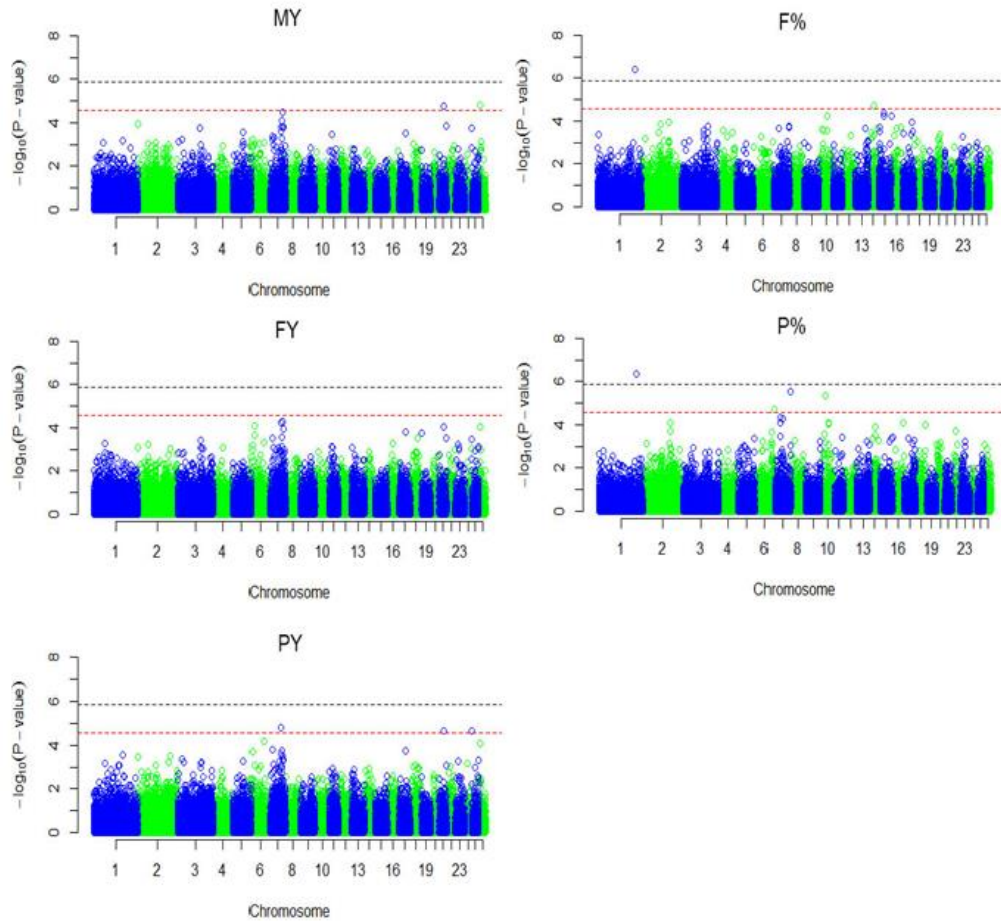


Figure 9 Manhattan plot for milk production traits using repeated measures.

4.3 Regional heritability mapping (RHM)

The results of the RHM analysis using 100 SNPs windows size for milk production traits are showed in Table 10 and Figure 10. A region on OAR2 was found significant ($LRT=16.25$) at the genome-wide level for F%, with a

partially overlapped window reaching the suggestive significance threshold (LRT=10.98). Three more regions (on OAR3 and OAR20) reached the suggestive significant threshold for the same trait. However, the same two regions on OAR3 reached the genome-wide significance threshold for P%. No other regions were found significant at either the genome-wide or the suggestive thresholds for the other traits. The genomic regions detected on OAR3 reported several QTLs identified in different studies for the considered traits. In particular, a QTL (ID=13905) for P% was found by Gutiérrez-Gil *et al.* (2009) on OAR3 in a commercial population of Spanish Churra sheep. The same authors reported QTLs (ID=13915 and 13917) on OAR2 and OAR20, respectively, which showed suggestive significant associations with F%) for F% (Gutiérrez-Gil *et al.*, 2009). Several GWAS studies for milk production traits on Spanish Churra sheep reported several QTLs (IDs=57707, 57708, 57739, 57741, 17200) on OAR 3 which overlapped with our regions significant associated with F% and P%. (Garcia-Gamez *et al.*, 2012; Garcia-Gamez *et al.*, 2012b; Garcia-Gamez *et al.*, 2013)

On OAR2, within the two partially overlapped regions, we found 132 protein-coding genes, 15 pseudogenes, 8 tRNA genes, and 15 unknown genes. On OAR3, within the two partially overlapped regions, we found 106 protein-coding genes, 16 pseudogenes, 9 tRNA genes, and 8 unknown genes. On OAR20, we found 243 protein-coding genes, 28 pseudogenes, 80 tRNA genes, and 28 unknown genes.

In particular, within the partially overlapped regions on OAR3, LALBA gene (α -lactalbumin) is present as a strong functional candidate gene affecting the traits (Supplementary Table S1). Alpha-lactalbumin is a major whey protein that forms a subunit of the lactose synthase binary complex. Because lactose synthase is necessary for the production of lactose and the subsequent movement of water into the mammary secretory vesicles, this enzyme is critical in the lactation control and secretion of milk. Previous studies in LALBA-deficient mice have shown the influence of this enzyme on the protein and fat concentration in milk. Homozygous mutant mice produce highly viscous milk

that is very rich in fat and protein and devoid of alpha-lactalbumin and lactose (Stinnakre *et al.*, 1994). Polymorphisms in the LALBA gene were studied in the 1990s as possible markers related to milk production in dairy species. However, to our knowledge, there are no reported studies on the influence of the ovine LALBA polymorphisms on milk traits. The candidacy initially suggested by Garcia-Gamez *et al.*, in (2010), for the LALBA gene was strongly supported by the GWA study reported by the same authors in 2013. This GWA study identified marker rs399070200, located in the third intron of this gene, as the SNP with the most highly significant association detected on OAR3 both for P% and F%.

Moreover, on OAR3 we found Aquaporins (AQPs) genes, in particular AQP6, AQP5 and AQP2 (in bold in Supplementary Table S1). This is a family of ubiquitous membrane proteins involved in the transport of water and wide range of solutes (Gomes *et al.*, 2009). A functional role for some members of this family during the production and secretion of bovine milk was confirmed in an immunohistochemical study conducted by Mobasheri *et al.* (2011).

Finally, we observed on OAR 3 and OAR20 genomic regions containing olfactory receptor (OR) family genes (19 on OAR3 and 62 on OAR20 indicated as “LOC” in bold in Supplementary Table S1). Olfactory receptors detect and identify a wide range of odors and chemosensory stimuli, are necessary to find food, detect mates and offspring, to recognize territories and avoid dangers.

The use of regional heritability mapping can improve the detection rate of variants. Nagamine *et al.* (2012) showed that RHM is more powerful for detecting rare variants than the other methods. This may contribute to an improved ability to detect variance associated with the local region when regional effects are due to only the additive effects of multiple segregating alleles. The approach to the analysis of genome-wide SNP has the potential to capture some of the heritable variance that escapes the standard SNP by SNP analysis. The use of regional windows and estimation of variance in a mixed model framework integrates over the gametic variance in a region and escapes from reliance on the association between single causative alleles and single

SNP alleles. It thus has the ability to integrate effects over several causative variants providing a joint estimate of the combined effects of common and rare variants in a region. The results of this analysis suggest that regions known to harbour effects large enough to be detected by standard SNP by SNP analyses may yield some additional variance when analyzed by this approach. Furthermore, regions where no single associated SNP has a large enough effect to be detected as significant at the genome-wide level may explain sufficient variance to be detected by this approach.

Table 10. Genome-wide and suggestive significant SNPs for milk production traits considering regional heritability mapping approach

Trait	OAR	window	SNP and position (in bp)		LRT	h^2_w
			start	end		
Fat %	2	83	rs413324492 234715088	rs401097503 240388878	16.25	0.03
	2	84	rs407871693 237161006	rs412038888 243300137	10.98	0.02
	3	48	rs412220800 134166768	rs419412283 139724363	11.71	0.02
	3	49	rs407496519 137065870	rs418178732 142443170	12.94	0.02
	20	9	rs429947991 23043970	rs399189470 30066632	9.61	0.02
Protein%	3	48	rs412220800 134166768	rs41941228 139724363	13.86	0.03
	3	49	rs407496519 137065870	rs399189470 142443170	16.27	0.03

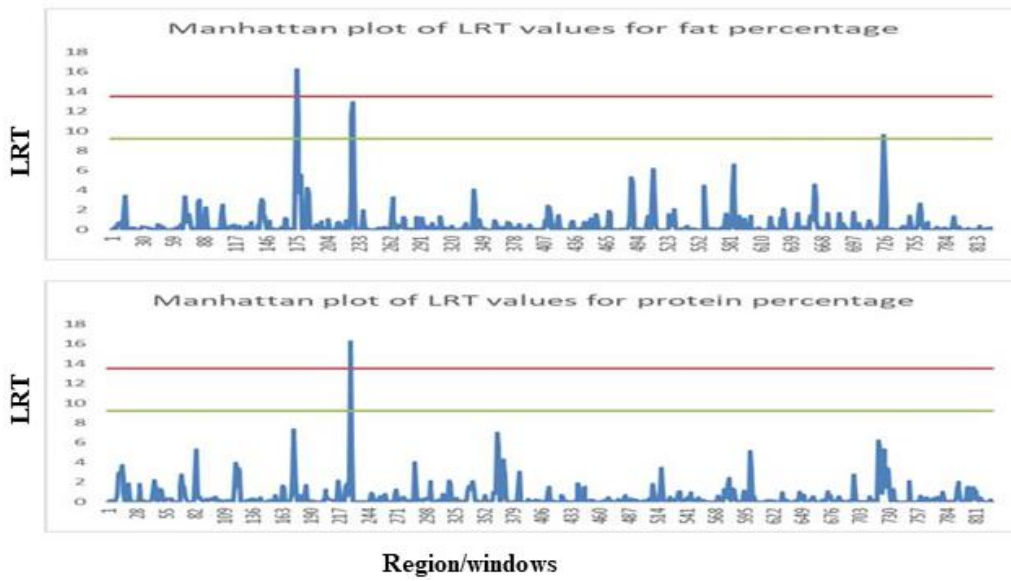


Figure 10 Manhattan plot for F% and P% traits using regional heritability approach.

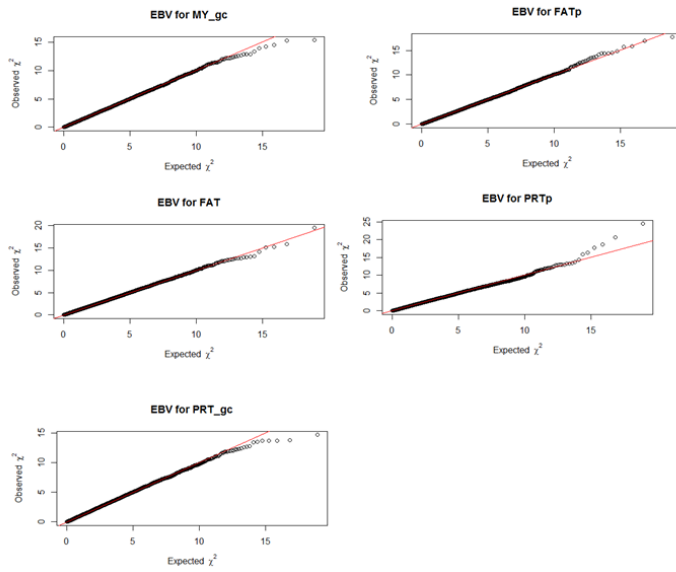
Conclusions

The application of recently developed genomic technology, such as high-density single nucleotide polymorphism (SNP) arrays, has great potential to increase our understanding on the genetic architecture of complex traits, to improve selection efficiency in domestic animals through genomic selection, and to conduct association studies. In this thesis, GWAS was used for the first time with the aim to identify SNPs associate with the milk production traits, to compare different response variables and to apply different analysis methodologies in Valle del Belice sheep breed.

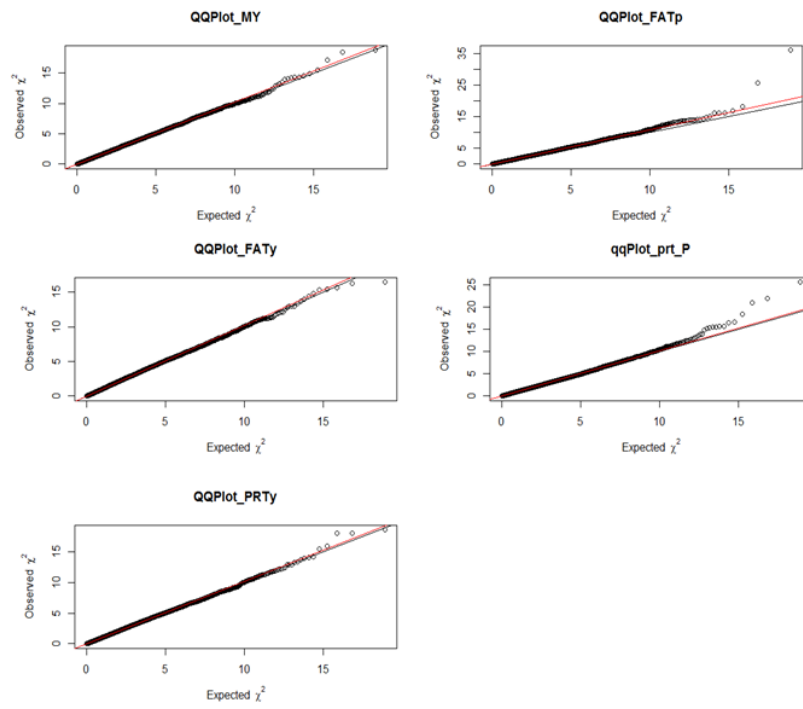
Comparison among the estimated breeding values and their deregressed and weighted values as responsible variables, respect to their influence on our GWAS results, has demonstrated that DEBVs and DEBVw allow identifying a greater numbers of SNPs than using EBVs. Several of these SNPs (n=19) are mapped within the previously reported QTL regions and within candidate genes for milk production traits. The general consistence of the significant SNPs detected herein with the reported QTL and candidate genes for milk traits allow us to be confident of the results obtained. Moreover some genomic regions identified by close SNPs associated with a specific trait should be further investigated to verify their effect on the traits. In particular on OAR2 at position 177.26-178.47 Mb (rs419432879-rs429723758) and on OAR 14 at 1.7-3.8 Mb (rs411845712-rs404689394) markers associated with MY were found. On OAR3 we found an interest region at 114 to 120 Mb for FY (rs416315048-rs402727086), and on OAR 6 at 41,7 and 46,45 Mb for P% (rs426928771-rs420418036). Another approach considering repeated measures for each animal was performed. This analysis has identified a smaller number of SNPs probably due to the reduced number of observations per animal or to the analytical method used. Regional Heritability Mapping approach provides heritability estimates attributable to small genomic regions, and it has the power to detect genomic regions containing multiple alleles that individually contribute too little variance to be detected by more standard GWAS approaches. Even taking into account the limitations imposed by sample size

and the number of SNPs analyzed some of the results are consistent with previous studies. The results confirmed the roles of LALBA gene and AQP genes, on OAR 3, as candidate genes for milk production traits in sheep. These genomic regions should be reproduced in future studies, favorably of a larger scale. Replication of the GWAS will also help to clear the picture of the weak signals detected in our study. To improve the statistical power, it would be possible to replicate the study, preferably including more genotyped samples. However, the results create opportunities for changing milk production traits through breeding by selecting individuals based on their genetic merit for milk production traits, which can consider the possibility of implementation the GWAS aspect into genomic selection in Valle del Belice dairy sheep. Therefore, the information generated from this thesis has important implications for the design and applications of association studies as well as for the development of selection breeding programs for the Valle del Belice sheep breeds.

Supplementary Material



Supplementary Figure S1. Quantile-quantile (Q-Q) plots of the observed test statistics of the genome wide association studies using EBV.



Supplementary Figure S2. Quantile-quantile (Q-Q) plots of the observed test statistics of the genome wide

References

- AIA, Bollettino dei controlli funzionali del latte 2006. Associazione Italiana Allevatori AIA, Rome, Italy (2006)
- Amin N, van Duijn CM, Aulchenko YS (2007) A genomic background based method for association analysis in related individuals. *PLoS ONE* **2**:e1274.
- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, *et al* (2010) Data quality control in genetic case-control association studies. *Nat Protoc* **5**(9):1564-1573.
- Aulchenko YS, de Koning DJ, Haley C (2007) Genomewide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* **177**:577- 585.
- Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**:1294-1296.
- Bacanu SA, Devlin B, Roeder K (2002) Association studies for quantitative traits in structured populations. *Genet epidemiol* **22**(1): 8-93.
- Barillet F, Boichard D (1994) Use of first lactation test-day data for genetic evaluation of the Lacaune dairy sheep. In Proc. 5th World Congr. *Genet Appl Livest Prod* Guelph, ON, Canada **18**:111-114.
- Barillet F, Arranz JJ, Carta A, Jacquiet P, Stear M, *et al* S (2006) Final Consolidated Report of the European Union Contract of Acronym *genesheepsafety* QTLK5-CT-2000-00656: p.145.
- Baro JA, Carriedo JA, San Primitivo F (1994) Genetic parameters of test day measures for somatic cell count, milk yield, and protein percentage of milking ewes. *J Dairy Sci* **77**(9):2658-2662.

- Becker D, Tetens J, Brunner A, Burstel D, Ganter M *et al* (2010) Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain3 (PITX3) gene. *PLoS One* **5**:e8689.
- Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, (2011) Genome-wide association studies for feedlot and growth traits in cattle. *J Anim Sci* **89**:1684-1697.
- Bolormaa S, Neto LR, Zhang YD, Bunch RJ, Harrison BE, *et al* (2011) A genome-wide association study of meat and carcass traits in Australian cattle. *J Anim Sci* **89**:2297-2309.
- Bolormaa S, Pryce JE, Hayes BJ, Goddard ME (2010) Multivariate analysis of a genome-wide association study in dairy cattle. *J Dairy Sci* **93**(8):3818-3833.
- Brooks SA, Gabreski N, Miller D, Brisbin A, Brown HE, *et al* (2010) Whole-genome SNP Association in the Horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS genet.* **6**(4):e1000909.
- Bush WS, Moore JH (2012) Genome-wide association studies. *PLoS Comput Biol.* **8**(12):e1002822.
- Cappio-Borlino A, Portolano B, Todaro M, Macciotta NPP, Giaccone P *et al* (1997) Lactation curves of Valle del Belice dairy ewes for yields of milk, fat and protein estimated with test day models. *J. Dairy Sci* **80**:3023-3029.
- Carta A, Sanna SR, Casu S (1995) Estimating lactation curves and seasonal effects for milk, fat and protein in Sarda dairy sheep with a test day model. *Livest Prod Sci* **44**:37-44.
- Chauhan T, Rajiv K (2010) Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology* **1**(04): 281-291.

- Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, *et al* (2009) Revealing the history of sheep domestication using retrovirus integrations. *Science* **324** (5926):532-536.
- Dadousis C, Biffani S, Cipolat-Gotet C, Nicolazzi EL, Rossoni A, *et al* (2016) Genome-wide association of coagulation properties, curd firmness modeling, protein percentage, and acidity in milk from Brown Swiss cows. *J Dairy Sci* **99**(5):3654-3666.
- Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *J. Anim Sci* **82**:E313-328.
- Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet* **3**:22-32.
- Duijvesteijn N, Knol EF, Merks JW, Crooijmans RP, Groenen MA, *et al* (2010) A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. *BMC Genet* **11**:42.
- Ekine CC, Rowe SJ, Bishop SC, de Koning DJ (2014) Why breeding values estimated using familial data should not be used for genome-wide association studies. *G3: Genes, Genomes, Genetics* **4**(2):341-347.
- El-Saied UM, Carriedo J, San Primitivo F (1998) Heritability of test day somatic cell counts and its relationship with milk yield and protein percentage in dairy ewes. *J Dairy Sci* **81**(11):2956-2961.
- Ewing B, Green P (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet* **25**:232-234.
- Feugang JM, Kaya A, Page GP, Chen L, Mehta T, *et al* (2009) Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. *BMC Genomics* **10**:176.

- Finck BN, Gropler MC, Chen Z, Leone TC, Croce MA, *et al* (2006) Lipin 1 is an inducible amplifier of the hepatic PGC-1 α /PPAR α regulatory pathway. *Cell Metab* **4**(3):199-210.
- Finlay EK, Berry DP, Wickham B, Gormley EP, Bradley DG (2012) A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. *PLoS One* **7**(2):e30545.
- Fisher R (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans Roy SocEdin* **52**:399-433.
- Food and Agriculture Organization (2014) The State of Food Insecurity in the World 2014: Strengthening the Enabling Environment for Food Security and Nutrition. Rome: *Food and Agriculture Organization of the United Nations* Available at: www.fao.org.
- Gao H, Lund MS, Zhang Y, Su G (2013) Accuracy of genomic prediction using different models and response variables in the Nordic Red cattle population. *J Anim Breed Genet* **130**(5):333-340.
- García-Fernández M, Gutiérrez-Gil B, García-Gámez E, Sánchez JP, Arranz JJ (2010) The identification of QTL that affect the fatty acid composition of milk on sheep chromosome 11. *Anim Genet* **41**(3):324-328.
- García-Gámez E, Gutiérrez-Gil B, Sahana G, Sánchez JP, Bayón Y *et al* (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**(10):e47782.
- García-Gámez E, Gutiérrez-Gil B, Sánchez JP, Arranz JJ (2012b) Replication and refinement of a quantitative trait locus influencing milk protein percentage on ovine chromosome 3. *Anim Genet* **43**(5): 636-641.
- García-Gámez, E, Gutiérrez-Gil B, Suarez-Vega A, de la Fuente LF, Arranz JJ (2013) Identification of quantitative trait loci underlying milk traits in

- Spanish dairy sheep using linkage plus combined linkage disequilibrium and linkage analysis approaches. *J Dairy Sci* **96**(9):6059-6069.
- Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* **41**:55.
- GenABEL project developers (2013) GenABEL: Genome-wide SNP association analysis. R package version 1.8-0. <http://CRAN.Rproject.org/package=GenABEL>
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml User Guide Release 3.0 VSN International Ltd, Hempstead, HP1 1ES, UK.
- Goddard ME (1992) A mixed model for analyses of data on multiple genetic markers. *Theor appl Genet* **83**(6-7):878-886.
- Gomes D, Agasse A, Thiébaud P, Delrot S, Gerós H *et al* (2009) Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. *BBA-Biomembranes* **1788**(6):1213-1228.
- Grindflek E, Aasmundstad T, Hamland H, Hansen MHS, Nome T, *et al* (2011) Revealing genetic relationships between compounds affecting boar taint and reproduction in pigs. *J Anim Sci* **89**(3):680-692.
- Gu X, Feng C, Ma L, Song C, Wang Y, *et al* (2011) Genome-wide association study of body weight in chicken F2 resource population. *PLoS One* **6**:e21872.
- Guo G, Lund MS, Zhang Y, Su G (2010) Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. *J Anim Breed Genet* **127**:423-432.
- Gutiérrez JP, Goyache F, Fernández I, Álvarez I, Royo LJ (2007) Genetic relationships among calving ease, calving interval, birth weight, and

- weaning weight in the Asturiana de los Valles beef cattle breed. *J Anim Sci* **85**(1):69-75.
- Gutiérrez-Gil B, El-Zarei MF, Alvarez L, Bayón Y, De La Fuente LF, *et al* (2009) Quantitative trait loci underlying milk production traits in sheep. *Anim Genet* **40**(4):423-434.
- Hall AM, Smith AJ, Bernlohr DA (2003) Characterization of the Acyl-CoA synthetase activity of purified murine fatty acid transport protein 1. *J Biol Chem* **278**(44):43008-43013.
- Hartl (1994) Population genetics and evolution. *In Genetics* (3 ed., pp. 197-224). Boston:Jones and Bartlett Publishers.
- Hayes BJ, Goddard ME (2008) Technical note: Prediction of breeding values using marker-derived relationship matrices. *J Anim Sci* **86**:2089-2092.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* **92**:433-443.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Savin K, van Tassell CP, *et al* (2009) A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS One* **4**:e6676.
- Hayes BJ, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. *Genome* **53**:876-883.
- Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk fat percentage, and type in Holstein cattle as contrasting model traits. *Plos Genet* **6**:e1001139.
- Hayes B, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. *Genet Sel Evol* **33**(3):209.

- Haynes GD, Latch EK (2012) Identification of Novel Single Nucleotide Polymorphisms (SNPs) in Deer (*Odocoileus* spp.) Using the BovineSNP50 BeadChip. *Plos One* **7**(5):11.
- Henderson CR (1975) Rapid method for computing inverse of a relationship matrix. *J Dairy Sci* **58**:1727-1730.
- Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**(2):95-108.
- Husby A, Kawakami T, Rönnegård L, Smeds L, Ellegren H, *et al* (2015) Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proc. R. Soc. B* **282**:20150156.
- ICAR (International Committee for Animal Recording) (2016) International agreement of recording practices. Available online: [http://www.icar.org/wp-content/uploads/2015/11/Guidelines 2016](http://www.icar.org/wp-content/uploads/2015/11/Guidelines%202016).
- ICAR (2016) International Committee for Animal Recording. International Regulations for milk recording in sheep. Institut de l'élevage, Département Génétique et Contrôle des Performances, Paris, 20 pp.
- Illumina (2010). OvineSNP50 DNA analysis kits. Viewed 28th February 2010
- ISTAT Annuario Statistico Italiano (2016). Istituto Nazionale di statistica.
- Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, *et al* (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**(6188):1168-1173.
- Johnston SE, McEwan JC, Pickering NK, Kijas JW, Beraldi D, *et al* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Mol Ecol* **20**:2555–2566.

- Jonas E, Thomson PC, Hall EJ, McGill D, Lam MK, *et al* (2011) Mapping quantitative trait loci (QTL) in sheep. IV. Analysis of lactation persistency and extended lactation traits in sheep. *Genet Sel Evol* **43**(1):22.
- Kemper KE, Palmer DG, Liu SM, Greeff JC, Bishop SC, *et al* (2010) Reduction of faecal worm egg count, worm numbers and worm fecundity in sheep selected for worm resistance following artificial infection with *Teladorsagia circumcincta* and *Trichostrongylus colubriformis*. *Vet Parasitol* **171**:238-246.
- Kenny EE, Kim M, Gusev A, Lowe JK, Salit J, *et al* (2011) Increased power of mixed models facilitates association mapping of 10 loci for metabolic traits in an isolated population. *Hum Mol Genet* **20**:827-839.
- Kinghorn B, Kinghorn S (1999) PEDIGREE VIEWER Program Version 6,5. University of New England. Australia.
- Kominakis A, Hager-Theodorides AL, Zoidis E, Saridaki A, Antonakos G *et al* (2017) Combined GWAS and ‘guilt by association’-based prioritization analysis identifies functional candidate genes for body size in sheep. *Genet Sel Evol* **49**(1):41.
- Lande R, Hompson RT (1990) Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genetics* **124**:743-756.
- Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, *et al* (2014) Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences of the United States of America* **111**(17): 6139–6146
- Lee C, Abdool A, Huang CH (2009) PCA-based population structure inference with generic clustering algorithms. *BMC bioinformatics* **10**(Suppl 1):S73.
- Lee JW, Kil DY, Keever BD, Killefer J, McKeith FK, *et al* (2013) Carcass fat quality of pigs is not improved by adding corn germ, beef tallow, palm

- kernel oil, or glycerol to finishing diets containing distillers dried grains with solubles. *J Anim Sci* **91**(5):2426-2437.
- Leutenegger AL, Prum B, Genin E, Verny C, Lemainque A *et al* (2003) Estimation of the inbreeding coefficient through use of genomic data. *Am. J. Hum. Genet.* **73**:516-523.
- Ligda C, Gabriilidis G, Papadopulos T, Georgoudis A (2000) Estimation of genetic parameters for production traits of Chios sheep using a multitrait animal model. *Livest Prod Sci* **66**:217-222.
- Lv Y, Wei C, Zhang L, Lu G, Liu K *et al* (2011) Association between polymorphisms in the SLC27A1 gene and milk production traits in Chinese Holstein cattle. *Anim biotechnol* **22**(1):1-6.
- MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, (2016) Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**(1):144.
- Mai MD, Sahana G, Christiansen FB, Guldbbrandtsen B (2010) A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *J AnimSci* **88**:3522-3528.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, *et al* (2009) Finding the missing heritability of complex diseases. *Nature* **461**:747-753.
- Martin PM, Palhière I, Ricard A, Tosser-Klopp G, Rupp R (2016) Genome Wide Association Study Identifies New Loci Associated with Undesired Coat Color Phenotypes in Saanen Goats. *PloS one* **11**(3):e0152426.
- Mateescu RG, Thonney ML (2010) Genetic mapping of quantitative trait loci for milk production in sheep. *Anim Genet* **41**(5):460-466.
- Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, *et al* (2016) Genome-wide association reveals QTLfor growth, bone and in vivo carcass

- traits as assessed by computed tomography in Scottish Blackface lambs. *Genet Sel Evol* **48**(1):11.
- Meadows JRS, Cemal I, Karaca O, Gootwine E, Kijas JW (2007) Five ovine mitochondrial lineages identified from sheep breeds of the near East. *Genetics* **175**(3):1371-1379.
- Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, *et al* (2012) Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genetics* **13**(1):21.
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**:1819-1829.
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**:1215.
- Minozzi G, Williams JL, Stella A, Strozzi F, Luini M, *et al* (2012) Meta-analysis of two genome-wide association studies of bovine paratuberculosis. *PLoS One* **7**(3):e32578.
- Minozzi G, Buggiotti L, Stella A, Strozzi F, Luini M, *et al* (2010) Genetic loci involved in antibody response to *Mycobacterium avium* ssp. paratuberculosis in cattle. *PLoS One* **5**(6):e11117.
- Mobasher A, Kendall B, Maxwell J, Sawran A, German A, *et al* (2008) Aquaporins along the secretory pathway of the lactating bovine mammary gland. *The FASEB Journal* **22**(1 Supplement):1159.21.
- Mogensen MS, Karlskov-Mortensen P, Proschowsky HF, Lingaas F, Lappalainen A, *et al* (2011) Genome-wide association study in Dachshund: identification of a major locus affecting intervertebral disc calcification. *J Hered* **1**:81-86.
- Morton NM, Nelson YB, Michailidou Z, Di Rollo EM, Ramage L, *et al* (2011) A stratified transcriptomics analysis of polygenic fat and lean mouse

- adipose tissues identifies novel candidate obesity genes. *PLoS One* **6**(9):e23944.
- Murdoch BM, Murdoch GK, Settles M, McKay S, Williams JL, *et al* (2011) Genome-wide scan identifies loci associated with classical BSE occurrence. *PLoS One* **6**(11):e26819.
- Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I *et al* (2012) Localising loci underlying complex trait variation using regional genomic relationship mapping. *PLoS ONE* **7**:e46501.
- Ogorevc J, Kunej T, Razpet A, Dovc P (2009) Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim Genet* **40**(6):832-851.
- Ordovas L, Roy R, Zaragoza P, Hayes H, Eggen A, *et al* (2005) Assignment of the solute carrier family 27 member 1 (SLC27A1) gene to bovine. *Anim Genet* **36**:352-353.
- Ostersen T, Christensen OF, Henryon M, Nielsen B, Su G, Madsen P (2011) Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet Sel Evol* **43**(1):38.
- Othmane MH, De La Fuente LF, Carriedo JA, 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. *J Dairy Sci* **85**(10): 2692-2698.
- Pausch H, Flisikowski K, Jung S, Emmerling R, Edel C, *et al* (2011) Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. *Genetics* **187**(1):289-297.

- Pedrosa S, Uzun M, Arranz JJ, Gutiérrez-Gil B, San Primitivo F, *et al* (2005) Evidence of three maternal lineages in near eastern sheepsupporting multiple domestication events. *P R Soc B-Biol Sci* **272**(1577):2211-2217.
- Phan J, Reue K (2005) Lipin, a lipodystrophy and obesity gene. *Cell Metab* **1**(1):73-83.
- Portolano B, Giaccone P, Truscelli A, Todaro M, Alabiso M, *et al* (1996) Milk production in the Pinzirita breed ewes. *Agric Medit* **126**:194-199.
- Portolano N (1987) Pecore e capre Italiane. Edagricole Bologna.
- Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME, Hayes BJ (2010) A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J Dairy Sci* **93**:3331-3345.
- Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**(1):149-150.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, *et al* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American J Hum Genet* **81**(3):559-575.
- R Core Team 2013 R: A language and environment for statistical computing. <http://www.R-project.org/>.
- Raadsma HW, Jonas E, McGill D, Hobbs M, Lam MK, *et al* (2009) Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. *Genet Sel Evol*, **41**(1), 45.
- Ramos AM, Duijvesteijn N, Knol EF, Merks JW, Bovenhuis H, *et al* (2011) The distal end of porcine chromosome 6p is involved in the regulation of skatole levels in boars. *BMC Genet.*; **12**:35

- Riggio V, Finocchiaro R, Van Kaam JBCHM, Portolano B, Bovenhuis H (2007) Genetic parameters for milk somatic cell score and relationships with production traits in primiparous dairy sheep. *J dairy Sci*, **90**(4), 1998-2003.
- Riggio V, Matika O, Pong-Wong R, Stear, MJ, Bishop SC (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity*, **110**(5), 420-429.
- Ron M, Weller JI (2007) From QTL to QTN identification in livestock – winning by points rather than knock-out: a review. *Anim Genet*, **38**: 429–439.
- Rönnegård L, McFarlane SE, Husby A, Kawakami T, Ellegren H, *et al* (2016) Increasing the power of genome wide association studies in natural populations using repeated measures–evaluation and implementation. *Methods Ecol Evol*. **7**(7), 792-799
- Rupp R, Senin P, Sarry J, Allain C, Tasca C, Ligat L, *et al* (2015). A point mutation in suppressor of cytokine signalling 2 (Socs2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genetics*, **11**(12): e1005629
- Sahana G, Gulbrandsen B, Bendixen C, Lund MS (2010) Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Anim Genet*, **41**:579–588.
- SAS Institute Inc (2012) Bases SAS® 9.2.4 Procedures Guide, Second Edition, Volumes 1, 2, 3, and 4. Cary, NC: SAS Institute Inc.
- Schopen GC, Visker MH, Koks PD, Mullaart E, van Arendonk JA, *et al* (2011). Whole-genome association study for milk protein composition in dairy cattle. *J Dairy Sci*, **94**:3148–3158.

- Self SG, Liang KY Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under non standard conditions. (1987). *J Amer Stat Assoc* **82**: 605–610.
- Sell-Kubiak E, Duijvesteijn N, Lopes MS, Janss LL, G, Knol EF, *et al* (2015). Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC genomics*, **16**(1), 1049.
- Sevillano CA, Lopes MS, Harlizius B, Hanenberg EH, Knol EF, *et al* (2015). Genome-wide association study using deregressed breeding values for cryptorchidism and scrotal/inguinal hernia in two pig lines. *Genet Sel Evol*, **47**(1), 18.
- Snelling WM, Allan MF, Keele JW, Kuehn LA, McDanel T, *et al* (2010) Genome-wide association study of growth in crossbred beef cattle. *J Anim Sci* ; 88:837–848. doi: 10.2527/jas.2009-2257.
- Stinnakre MG, Vilotte JL, Soulier S, Mercier JC (1994) Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. *Proceedings of the National Academy of Sciences*, **91**(14), 6544-6548.
- Suárez-Vega A, Gutiérrez-Gil B, Arranz JJ (2016) Transcriptome expression analysis of candidate milk genes affecting cheese-related traits in 2 sheep breeds. *J dairy Sci*, **99**(8), 6381-6390.
- Tanksley SD (1983) Molecular markers in plant breeding. *Plant Mol Biol Reporter*, **1**(1), 3-8.
- The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*.; **447**:661–678.
- Thompson EA, Shaw RG (1990) Pedigree analysis for quantitative traits variance-components without matrix-inversion. *Biometrics* **46**: 399–413.

- Uemoto Y, Abe T, Tameoka N, Hasebe H, Inoue K, *et al* (2011) Whole genome association study for fatty acid composition of oleic acid in Japanese Black cattle. *Anim Genet* **42**: 141–148
- Utsunomiya YT, Do Carmo AS, Carvalheiro R, Neves HH, Matos MC *et al* (2013) Genome-wide association study for birth weight in Nelore cattle points to previously described orthologous genes affecting human and bovine height. *BMC genetics*, **14**(1), 52.
- Valente TDS (2016) Estudo de associação de polimorfismos de nucleotídeo único (SNP) com características de temperamento em bovinos da raça Nelore.
- VanRaden PM, Wiggans GR (1991) Derivation, calculation, and use of national animal model information. *J Dairy Sci* **74**: 2737–2746.
- Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol*, **34**(3), 275-306.
- Wang C, Zöllner S, Rosenberg NA (2012) A Quantitative Comparison of the Similarity between Genes and Geography in Worldwide Human Populations. *PLoS Genet*, **8**(8).
- White SN, Mousel MR, Herrmann-Hoesing LM, Reynolds JO, Leymaster K *et al* (2012). Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus. *PloS one*, **7**(10), e47829.
- Wiener P, Edriss MA, Williams JL, Waddington D, Law A, *et al* (2011) Information content in genome-wide scans: concordance between patterns of genetic differentiation and linkage mapping associations. *BMC Genomics*, **12**: 65
- Wright S (1921) Correlation and causation *J Agric res*, **20**(7), 557-585.

- Xie L, Luo C, Zhang C, Zhang R, Tang J, *et al* (2012) Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. *PLoS One*; **7**(2):e30910. doi: 10.1371/journal.pone.0030910
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, *et al* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**: 565-569.
- Yuan FP, Li X, Lin J, Schwabe C, Büllsbach EE, *et al* (2010) The role of RXFP2 in mediating androgen-induced inguinoscrotal testis descent in LH receptor knockout mice. *Reproduction.*; 139:759–769. doi: 10.1530/REP-09-0518.
- Zhang L, Liu J, Zhao F, Ren H, Xu L, *et al* (2013) Genome-wide association studies for growth and meat production traits in sheep. *PloS one*, **8**(6), e66569.

RESEARCH ARTICLE

Open Access



Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep

Salvatore Mastrangelo^{1*}, Marco Tolone¹, Maria T. Sardina¹, Gianluca Sottile², Anna M. Sutura¹, Rosalia Di Gerlando¹ and Baldassare Portolano¹

Abstract

Background: Because very large numbers of single nucleotide polymorphisms (SNPs) are now available throughout the genome, they are particularly suitable for the detection of genomic regions where a reduction in heterozygosity has occurred and they offer new opportunities to improve the accuracy of inbreeding (F) estimates. Runs of homozygosity (ROH) are contiguous lengths of homozygous segments of the genome where the two haplotypes inherited from the parents are identical. Here, we investigated the occurrence and distribution of ROH using a medium-dense SNP panel to characterize autozygosity in 516 Valle del Belice sheep and to identify the genomic regions with high ROH frequencies.

Results: We identified 11,629 ROH and all individuals displayed at least one ROH longer than 1 Mb. The mean value of F estimated from ROH longer than 1 Mb was 0.084 ± 0.061 . ROH that were shorter than 10 Mb predominated. The highest and lowest coverages of *Ovis aries* chromosomes (OAR) by ROH were on OAR24 and OAR1, respectively. The number of ROH per chromosome length displayed a specific pattern, with higher values for the first three chromosomes. Both number of ROH and length of the genome covered by ROH varied considerably between animals. Two hundred and thirty-nine SNPs were considered as candidate markers that may be under directional selection and we identified 107 potential candidate genes. Six genomic regions located on six chromosomes, corresponding to ROH islands, are presented as hotspots of autozygosity, which frequently coincided with regions of medium recombination rate. According to the KEGG database, most of these genes were involved in multiple signaling and signal transduction pathways in a wide variety of cellular and biochemical processes. A genome scan revealed the presence of ROH islands in genomic regions that harbor candidate genes for selection in response to environmental stress and which underlie local adaptation.

Conclusions: These results suggest that natural selection has, at least partially, a role in shaping the genome of Valle del Belice sheep and that ROH in the ovine genome may help to detect genomic regions involved in the determinism of traits under selection.

Background

Autozygosity is the homozygous state of identical-by-descent (IBD) alleles, which can result from several

phenomena such as genetic drift, population bottlenecks, mating of close relatives, natural and artificial selection [1, 2]. The increase in inbreeding (F) leads to different negative effects such as reduction in genetic variance, higher frequency of homozygous genotypes for deleterious alleles with reduction in individual performance (inbreeding depression) and lower population viability

*Correspondence: salvatore.mastrangelo@unipa.it

¹ Dipartimento Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy
Full list of author information is available at the end of the article



RESEARCH ARTICLE



Full-length sequencing and identification of novel polymorphisms in the *ACACA* gene of Valle del Belice sheep breed

ROSALIA DI GERLANDO, SALVATORE MASTRANGELO, LINA TORTORICI, MARCO TOLONE, ANNA MARIA SUTERA, MARIA TERESA SARDINA* and BALDASSARE PORTOLANO

Dipartimento Scienze Agrarie, Alimenti e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy

*For correspondence. E-mail: mariateresa.sardina@unipa.it.

Received 18 October 2016; revised 22 November 2016; accepted 15 December 2016; published online 16 August 2017

Abstract. The essential role of the acetyl-CoA carboxylase (*ACACA*) enzyme in milk fatty acid (FA) synthesis suggests that it may be responsible for the phenotypic variability observed in milk. Before attempting association analyses between this gene and/or enzyme and phenotypic traits, a study on the genetic variability within this locus is required. The aim of this work was to sequence the entire coding region of *ACACA* gene in Valle del Belice sheep breed to identify polymorphic sites. A total of 51 coding exons of *ACACA* gene were sequenced in 32 individuals of Valle del Belice sheep breed. Sequencing analysis and alignment of obtained sequences showed the presence of 23 polymorphic sites. The most polymorphic was exon 53 which showed presence of 12 single-nucleotide polymorphisms (SNPs), of which eight were missense mutations, caused amino acid changes and therefore may affect protein function or stability causing variation in phenotype. The identified polymorphisms showed high variability of the *ACACA* gene. Sequences analysis allowed to find six new SNPs in exon 53 (6832C>T; 6835C>A; 6840G>A; 6847G>T; 6852C>T and 6860G>C). A total of 31 haplotypes were inferred. Although this study could not provide association study with production traits, it shows finding of novel SNPs that might be important in future studies and laid the basis for further association analyses needed to evaluate the potential use of these SNPs as genetic markers for fat content and FAs composition in milk of Valle del Belice sheep breed.

Keywords. *ACACA* gene; single-nucleotide polymorphisms; sheep; Valle del Belice breed.

Introduction

Acetyl-coA carboxylase (*ACACA*) is the flux-determining enzyme in the regulation of fatty acid (FA) synthesis in animal tissues. The eukaryotic *ACACA* enzymes are multidomain and contain the biotin carboxylase (BC), biotin carboxyl carrier protein (BCCP), and carboxyltransferase (CT) domains (Wakil *et al.* 1983; Abu-Elheiga *et al.* 2001; Cronan and Waldrop 2002). *ACACA* is a complex multifunctional enzyme system that resides in the cytosol, which catalyses through a pathway of carboxylation ATP-dependent of acetyl-CoA to form malonyl-CoA, which is the substrate for the synthesis of palmitic acid and long-chain FAs (acyl-CoA>C22:0) (Smith *et al.* 2003) by the FA synthase (FAS) enzyme.

Based on *Ovis aries* v3.1 genome release, the ovine *ACACA* gene reported in Ensembl database (www.ensembl.org, ENSOARG00000000829) is located on chromosome 11 (OAR11) and consists of 53 exons of which 51 encode

a protein of 2257 amino acids (ENSOARG00000000829, UniProtKB-W5NRT6).

Sequence and organization of *ACACA* gene is similar in ruminant species. In bovine species, *ACACA* gene is located on chromosome 19 (NM_174224.2/ENSBTAT00000023364) and is organized in 56 exons of which only 54 (7381 bp) encoded for 2346 amino acids. In caprine species, this gene is located on the same orthologous chromosome and its mRNA is 8498-bp long of which 7041 bp translated in 2346 amino acids.

This enzyme is ubiquitously expressed but the highest levels of *ACACA* are found in lipogenic tissues such as the liver, adipose tissue and the mammary gland during lactation. The expression of the mammary gland isoform is regulated by the acetyl-CoA carboxylase- α (*ACACA*) gene (Barber and Travers 1995). During lactation, *ACACA* facilitates the recruitment of FA precursors to the mammary gland that uses to synthesize fat in milk. This occurs due to the repression and induction of *ACACA* activity



Contents lists available at ScienceDirect

Small Ruminant Research

journal homepage: www.elsevier.com/locate/smallrumres

Short communication

Association study between β -defensin gene polymorphisms and mastitis resistance in Valle del Belice dairy sheep breed



Marco Tolone*, Salvatore Mastrangelo, Rosalia Di Gerlando, Anna M. Sutera, Giuseppina Monteleone, Maria T. Sardina, Baldassare Portolano

Dipartimento Scienze Agrarie e Forestali, Università degli Studi di Palermo, Viale delle Scienze, 90128 Palermo, Italy

ARTICLE INFO

Article history:

Received 28 September 2015
 Received in revised form
 23 December 2015
 Accepted 29 December 2015
 Available online 4 January 2016

Keywords:

β -Defensin genes
 Mastitis resistance
 Single nucleotide polymorphisms
 Valle del Belice dairy sheep breed

ABSTRACT

Mastitis is generally caused by bacteria, and it is the most common disease in livestock species. Defensins are peptides with a broad spectrum of antimicrobial activity and β -defensin genes have been studied in several livestock species due to their important role in the innate immune response. The aim of this study was to establish an association between polymorphisms in the β -defensin 1 and 2 genes and mastitis resistance in the Valle del Belice dairy sheep. Data consisted of 1855 and 2804 observations for case and control group, respectively. Six single nucleotide polymorphisms and seven haplotypes were selected for association studies with mastitis. In particular, polymorphism G1747A on β -defensin 1 gene was associated with susceptibility to mastitis, while polymorphism G1659A on β -defensin 2 gene was associated with resistance to mastitis. Haplotypes ACAGGG and GCAGGG were associated with resistance to mastitis, whereas haplotype ACGGGG was associated with susceptibility to mastitis. The present study has firstly suggested the possible associations of β -defensin gene polymorphisms with mastitis resistance traits and showed the presence of interesting haplotypes in Valle del Belice dairy sheep breed. Results from association analysis provided preliminary evidence that β -defensins could be used as candidate genes or molecular markers for the improvement of ovine mastitis resistance traits in Valle del Belice dairy breed.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Mastitis is the most common and costly infectious disease of mammary gland affecting dairy animals caused by bacteria. Mastitis alters the state of well-being and health of the animals, and it leads to economic loss mainly due to reduced milk yield and quality, veterinary treatments, milk disposal due to antibiotic treatments, and early culling. Moreover, milk and dairy products from animals affected by mastitis represent a significant risk for the consumers not only for the presence of the pathogens, in particular for dairy products made with raw milk, but also for the presence of bacterial toxins and the antibiotic residues (Mitchell et al., 1998). The contagious infection causes an increase in total somatic cell count (SCC) as a consequence of both leukocyte and epithelial cell numbers increasing, with or without clinical signs of mastitis. Mastitis remains a major challenge to the worldwide dairy industry despite the widespread implementation of mastitis control

strategies (Bradley, 2002). Selection for genetic resistance to mastitis can be done directly or indirectly. Direct selection relates to the diagnosis of the infection, whereas indirect methods have been widely applied based on the evaluation of the degree of inflammation or of internal mammary lesions (Riggio and Portolano, 2015).

Antimicrobial peptides are important and effective components of innate immunity and are being evaluated as possible alternatives to conventional antibiotics due to the fact that bacteria have not developed resistance against antimicrobial peptides because they target components that are within to bacterial structures (Lai and Gallo, 2009). These endogenous host-defense molecules are encoded by distinct genes and translated from mRNA templates (Ramanathan et al., 2002). Based on their common features, two major families of antimicrobial peptides have been characterized in mammals, defensins and cathelicidins, which are the part of the antimicrobial arsenal of the leucocytes (Kościcuzuk et al., 2014). Defensins are expressed in a variety of epithelial tissues, which serve as primary microbial interface sites (Kaiser and Diamond, 2000). They are cationic peptides, 18–45 amino acids residues in length, whose structure is stabilized by three intramolecular disulfide bonds formed by six strongly conserved cysteine residues

* Corresponding author.

E-mail address: marco.tolone@unipa.it (M. Tolone).

Manuscript Number: Rumin-D-17-8453R1

Title: Determination of milk production losses and variations of fat and protein percentages according to different levels of somatic cell count in Valle del Belice dairy sheep

Article Type: Research Paper

Keywords: Valle del Belice sheep; mastitis; somatic cell count; milk production traits.

Corresponding Author: Dr. Anna Maria Sutera, Dr.

Corresponding Author's Institution: University of Palermo

First Author: Anna Maria Sutera

Order of Authors: Anna Maria Sutera; Baldassare Portolano; Rosalia Di Gerlando; Maria Teresa Sardina; Salvatore Mastrangelo; Marco Tolone

Abstract: The somatic cell count (SCC) of milk is widely used to monitor udder health and the milk quality and because of its positive genetic correlation with mastitis this trait was included in breeding schemes of dairy sheep. The aim of this study was to estimate the loss in milk yield and related composition resulting from different levels of somatic cell count in Valle del Belice dairy sheep. Data were collected between 2006 and 2016 in 15 flocks following an A4 recording scheme. Somatic cell count, fat and protein percentage were determined using mid-infrared spectroscopy. To evaluate loss in test day milk yield, fat and protein percentage, five different classes of somatic cell count were arbitrarily defined: $SCC1 \leq 500 \times 10^3$, $500 \times 10^3 < SCC2 \leq 1,000 \times 10^3$, $1,000 \times 10^3 < SCC3 \leq 1,500 \times 10^3$, $1,500 \times 10^3 < SCC4 \leq 2,000 \times 10^3$ and $SCC5 > 2,000 \times 10^3$. To estimate the loss of milk production and quality a linear model, with test day milk production traits as dependent variable, was used. Furthermore, the effect of order of parity and season of lambing were investigated to study the effects on milk production traits. Least squares means were computed for milk production traits and the differences between means were determined by Fischer's least significant difference. The estimated losses in MY according to the level of used somatic cell count were approximately 16% whereas there was an increase of 0.08% and 0.03% for fat and protein percentage, respectively. This study confirms that high levels of somatic cell count in sheep milk are associated with milk yield losses and changes in milk composition. Results suggest that it is necessary to implement a program aimed to reduce the milk somatic cell count in ewes' milk, with the aim of improving the quality of ewes' milk and dairy products.

Contents

Chapter 1	7
General Introduction	7
1.1. Dairy sheep	8
1.2 Origin and description of the Valle del Belice dairy breed	9
1.3 From quantitative genetic to genomic approach to enhance the response to selection	10
1.4 Genome Wide Association studies	15
1.5 GWAS in Sheep.....	17
1.6 Aim and outline of this thesis	19
Chapter 2.....	20
Materials and Methods.....	20
2.1 Population and phenotypes	21
2.2 Genotyping.....	22
2.3 Quality Control	23
2.4 Population structure	25
2.5. Estimated breeding value (EBVs).....	27
2.6 Deregressed and weighted breeding values	28
2.7 Genome Wide Association analysis	30
2.8 Genome Wide Association analysis using repeated measures	31
2.9 Statistical Inference.....	33
2.10 Regional Heritability Mapping	34
2.11. Procedures.....	35
2.12 Statistical inference for RHM analysis	36
2.13 Significant SNPs location on OAR4.0 genome assembly	37
Chapter 3.....	38
Results and Discussion	38
3.1 Results using EBv, DEBV, and DEBVw	39
3.1.1 Significant SNPs	41
3.1.2 Milk Yield.....	42

3.1.3 Fat percentage and yield	46
3.1.4 Protein percentage and yield.....	48
3.1.5 Considerations using EBV, DEBV and DEBVw for GWAS	50
3.2 Results using repeated measures.....	51
4.3 Regional heritability mapping (RHM).....	54
Conclusions.....	59
Supplementary Material.....	62
References.....	66