UNIVERSITÀ DEGLI STUDI DI PALERMO

Dottorato di Ricerca in Scienze Agrarie, Alimentari, Ambientali e Forestali (D028)

curriculum in Agroecosistemi Mediterranei

Dipartimento Scienze Agrarie, Alimentari e Forestali

Settore Scientifico Disciplinare AGR/17 – Zootecnia generale e miglioramento genetico

OPERATIONAL STRATEGIES FOR IMPROVING THE COMPETITIVENESS OF THE VALLE DEL BELÌCE SHEEP BREED

IL DOTTORE
DOTT. ANGELO MOSCARELLI

IL COORDINATORE
PROF. VINCENZO BAGARELLO

IL TUTOR
PROF. BALDASSARE PORTOLANO

IL CO TUTOR
DOTT.SSA MARIA TERESA SARDINA

CICLO XXXIII
ANNO CONSEGUIMENTO TITOLO 2021

Sommario

CHAPTER 1 3

General introduction 3

1.1 Historical origins of Valle del Belìce dairy sheep breed 4

1.2 Phenotypic, productive, and reproductive traits, and management of Valle del Belìce sheep breed 6

1.3 Molecular markers and their applications 9

1.4 Aim and outline of thesis 12

References 14

CHAPTER 2 21

Parentage reconstruction using microsatellite markers in Valle del Belìce dairy sheep breed 21

2.1 Summary 22

2.2 Introduction 22

2.3 Materials and methods 23

2.4 Result and discussion 27

2.5 Conclusion 36

References 37

CHAPTER 3 39

Genetic traceability of Sicilian sheep dairy production Vastedda della Valle del Belìce PDO and Pecorino Siciliano PDO 39

3.1 Summary 40

3.2 Introduction 40

3.3 Materials and methods 43

3.4 Results and discussion 51

3.4 Conclusion 55

CHAPTER 4 57

Genome-Wide association studies for phenotypic and productive traits in the Valle del Belìce dairy sheep breed 57

4.1 Genome-wide association study reveals the locus responsible for microtia in Valle del Belìce sheep breed 58

4.1.1 Summary 58

4.1.2 Introduction 58

4.1.3 Materials and methods 59

4.1.4 Result and discussion 60

References 66

4.2 Genome-wide association study for somatic cells count in Valle del Belìce dairy sheep 68

4.2.1 Summary 68

4.2.2 Introduction 68

4.2.3 Materials and methods 70

4.2.4 Results 74

4.2.5 Discussion 77

4.2.6 Conclusion....80
References....81
CHAPTER 5....87
General conclusion....87

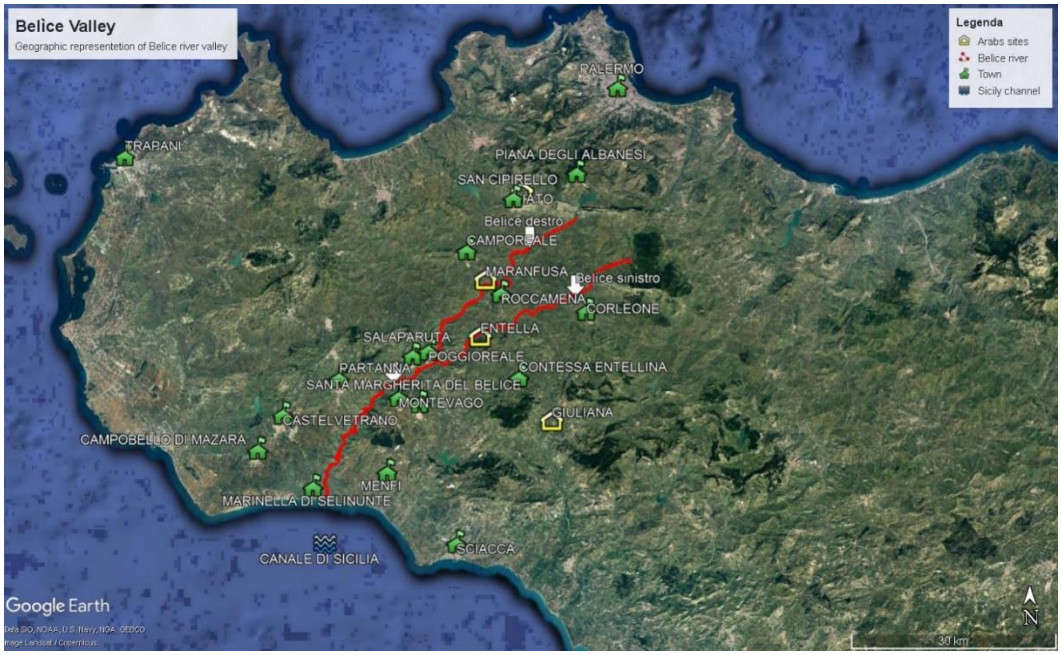
CHAPTER 1

General introduction

1.1 Historical origins of Valle del Belìce dairy sheep breed

In the western Sicily, among Palermo, Trapani and Agrigento provinces, the Belìce river is flowing from the Corleone's mountains (North-East) to the meridional Sicilian coast (South-West), exactly in the Canale di Sicilia between Marinella di Selinunte and Portopalo di Menfi (Fig.1.1.1.), in Mediterranean Sea (Protection plan for the waters of Sicily, 2007).

Figure 1.1.1. Geographic representation of Belìce valley



On the correct diction between "Belice" (incorrect) and "Belìce" (correct) it is better to clarify immediately, since there is often a lot of confusion among authors. In fact, the territory in which this important stream flows, during the Arabs domination was called "U-Bilìk", due to a castle that stood in the area, confirming the correct positioning of the accent on the name "Belìce", furthermore used in Italy until the 1968's earthquake, when journalists and Italian television mistakenly shifted the accent (Rumiz, 2009), broadcasting the improper diction "Belice".

In the valley in which the Belìce river's branches (named "right" and "left") are flowing, there are important archaeological finds dating back from prehistory to the time of the Arabs and Normans. The Belìce valley it has been frequented and populated since prehistoric times, as testified by graffiti and archaeological finds

linked to Sicani, Elimi, Phoenicians and Greeks. Findings attested the presence of rupestrian breeding activities in *Maranfusa* and *Sticca* mountains by Elima population (Spatafora & Calascibetta, 1986). This site, called *Calatrasi* on middle age, was one of the strongholds of the Muslim resistance in Sicily, with *Entella* and *Iato* sites (Fig. 1.1.1), before the definitive expulsion which took place in 1246 (Corretti *et al*., 2004). This aspect was relevant for breeding activities because, as reported below, the Arabs played an important role in historical sheep crossbreed. In fact, the sheep breeding activities in western Sicily were practiced since prehistoric times, passing through the Elimi tribe, Greek and Phoenician populations, up to the Arab domination, as evidenced by the archaeological findings (Leighton, 1999). In that historical period (VIII B.C.), those animals were crossed with Sarda sheep breed imported by Arabs, originating the Valle del Belìce sheep breed (Portolano *et al.*, 1996).

In these important historical and geographical contexts, the Belìce valley represented an important basin for agriculture development and livestock breeding activities. From this valley, naturally rich and always disputed by the civilizations during history, the "Valle del Belìce" dairy sheep breed took its name. This breed is considered to originate from the ancestors Comisana (Fig. 1.1.2A), Pinzirita (Fig. 1.1.2B) and Sarda (Fig. 1.1.2C) dairy breeds (Portolano, 1987). The Pinzirita breed was considered a native Sicilian sheep and, it was crossed with Comisana breed, originated from Comiso area, generating individuals with intermediate characteristics between parental lines (Portolano *et al.*, 1996).

Figure 1.1.2. The Valle del Belìce sheep breed ancestors*

A. Comisana **B.** Pinzirita **C.** Sarda

**images from the web*

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 Belìce breed standard. This was submitted in 1996 to both the Dairy Sub-Committee and the Ewes Technical Committee, and in 1997 the Valle del Belìce breed was given official recognition as local breed (Ministerial Decree of 1 December 1997).

Nowadays, the Valle del Belìce sheep breed is diffused throughout the territory of the Sicilian region. This breed has inherited from Pinzirita and Sarda breeds the medium-small ears shape, the ability to resist the damp cold and indoor housing, just in the coldest periods of the year. From Comisana sheep breed it has inherited the udder conformation, the high milk production aptitude and the ability to adapt the climatic conditions in which it is raised (Portolano, 1987).

1.2 Phenotypic, productive, and reproductive traits, and management of Valle del Belìce sheep breed

The individuals of Valle del Belìce breed present an open and white fleece covers the entire body with the exception of limbs, belly and head. The body is well developed whit good transversal diameters; the head generally is slim and extended, with the nasal bone generally curved (convex). Some individuals typically have reddish brown spots surrounding the eyes and on the last part of the ears (Fig.1.2.1).

Figure 1.2.1. Valle del Belìce sheep breed: A) Ram; B) Ewe with lamb; C) Young lamb.

A B C

Another peculiarity of Valle del Belìce breed, probably due to excessive reproduction in consanguinity, is the presence of some cases of congenital

hypotrichosis, consisting in complete absence of hair at birth (but not wool) (Fig.1.2.2), caused by mutations on hairless gene (Finocchiaro *et al.*, 2003).

Figure 1.2.2. Individuals with congenital hypotrichosis, so called "monkey sheep".

The Valle del Belìce sheep breed have a mean milk production of about 139 ± 35 liters on first lactation, and 210 ± 35 liters for the following lactations (AIA, 2006). The milk characteristics present fat content of about 6.5% and protein content about 5.5%. The mainly use of this breed is for milk production, even if the production of lambs (birth weight about 3.2-4.5 Kg) represents a not indifferent source of secondary income for farmers. Generally, the Valle del Belìce ewes present the first ovulation at 5-6 months old, with and estrous cycle of about 20-24 days; ewes at first parity are between 12 and 24 months of age (Portolano, 1987; Micalizzi, 1995). The lambs can weigh about 15-23 kg at 90 days old. Nowadays, it is usual to program synchronization of estrus to have suckling lambs (with a weight of about 8-9 kg) to be destined for the slaughterhouse during Easter and Christmas periods (Fig. 1.2.3). It is also true, however, that the lambing season is all year long, starting in July and ending in the next June (even if there are generally few lambs in May and June). This lambing system differs from the one adopted in others Mediterranean regions (Ligda *et al.*, 2000). The sheep are fed with natural pastures or fodder crops. The most common and preferred pasture for Valle del Belìce breed is *sulla* grass. Sulla (*Hedysarum coronarium L.*) forage was valued for its positive impact on ruminant production thanks to its palatability and content in nutraceutical substances, like condensed tannin (Bonanno *et al.*, 2016; Tibe *et al.*, 2011).

Sometimes, *i.e.* at the end of gestation, it is necessary to give additional food consisting in hay and concentrates (Cappio-Borlino *et al.*, 1997).

Figure 1.2.3. Valle del Belìce lambs: A) About 1 month old; B) About 6 months old.

A B

The annual fertility of Valle del Belìce sheep, defined as the percentage ratio between the number of sheep gave birth and the number of sheep capable of giving birth, is 95%, while the prolificacy, given by the percentage ratio between the lambs born and the number of sheep gave birth, is 145%. The annual fecundity, *i.e.*, the percentage ratio between the lambs born and the number of sheep capable of giving birth, is 138%. These values are due to high aptitude for twin births (about 80% in multiparous and 20% in primiparous).

A peculiar characteristic of Valle del Belìce breed ewes is linked to their body size and subcutaneous fat thickness on summer. In this period, they manage to significantly decrease their body weight, up to 30% less than recorded at the time of gave birth, in order to be able to feed the lambs with nutrient milk (rich in fat).

Generally, the breeders adopt family management system with semi-extensive grazing conditions, usually milk animals' morning and evening (twice a day), so they bring animals to pasture after morning milking, while keep them in the

sheepfold after the evening milking. In cold periods they adopt an indoor housing. Unfortunately, still today, sheepfold are often precarious structures or mediocre rural buildings like old storehouse or ruin, in which the farmers milk by hand. The main use of Valle del Belìce sheep breed milk is for traditional raw milk cheese productions. The most important are "*Vastedda della Valle del Belìce*" PDO, an Italian cheese produced only with milk obtained by the Valle del Belìce sheep breed (Ministry of Agriculture, Food and Forestry Policies - MIPAAF and Consorzio Vastedda del Belìce) and "*Pecorino Siciliano*" PDO, produces only with Sicilian raw sheep milk (Ministry of Agriculture, Food and Forestry Policies - MIPAAF and Consorzio Pecorino Siciliano). They were produced at farm level by small local diaries or by milk supplies to cheese industries working at regional level. Recently, for quality dairy products (such as cheeses mentioned), the genetic traceability system for the authentication of production has started to play an important role.

1.3 Molecular markers and their applications

Today, molecular genetics technologies are widely used to study and improve the breeds of livestock species or their productions. These techniques involve the use of molecular markers, necessary to deep study the genome of the individuals belonging to different populations/breeds/species.

The molecular markers are usually analyzed using the amplification of target sequences of DNA by PCR (Polymerase Chain Reaction). This method allows exponentially amplifying a target DNA fragment thanks to specific oligonucleotides or primers, reaction buffers and a specific enzyme (Taq DNA polymerase). Microsatellite markers, and the most recent SNPs, are widely used for different purposes and have been studied in many species (Peelman *et al.*, 1998; Sancristobal-Gaudy, *et al.*, 2000; Arana *et al.*, 2002; Herraeza *et al.*, 2005; Dalvit *et al.*, 2007b; Orrù *et al.*, 2006; Fernandez *et al.*, 2013; Mastrangelo *et al.*, 2020; Moscarelli *et al.*, 2020).

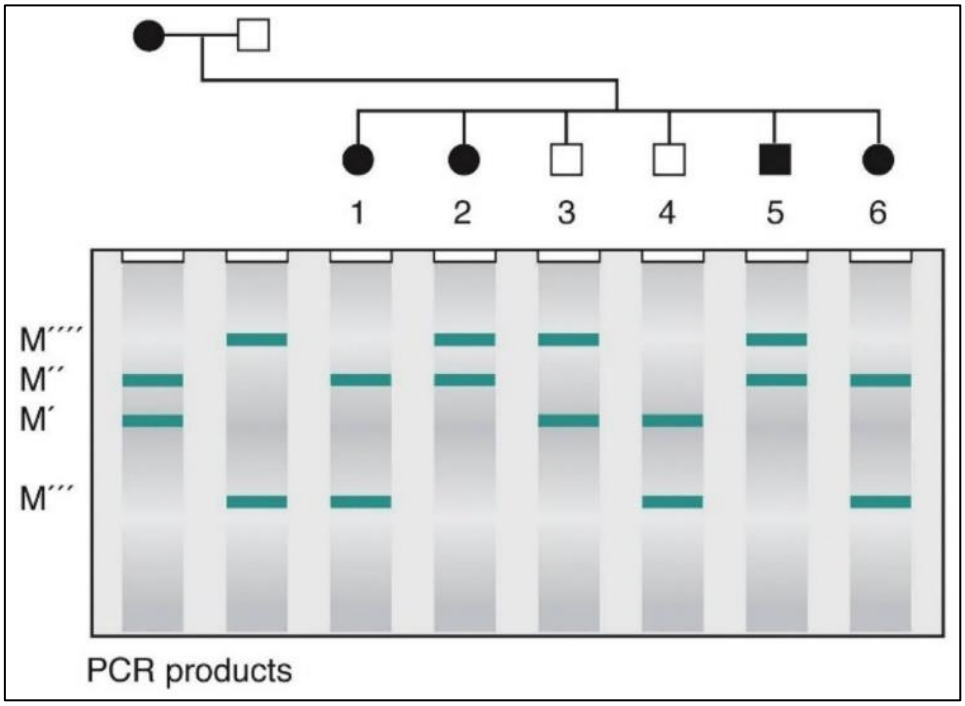
- **Microsatellite markers**

The VNTRs (Variable Number of Tandem Repeats), including microsatellites, consist of repeated sequences of non-coding DNA. They are characterized by tandem repeats of short nucleotide sequences, *e.g.*:

ACACACACACACACAC	$(AC)^n$
GTGTGTGTGTGTGTGTGT	$(GT)^n$
CAC CAC CAC CAC CAC	$(CAC)^n$
GATA GATA GATA GATA	$(GATA)^n$

They are highly distributed in the genome and are often found in non-coding regions. The polymorphism is very high and derives from the basic motif of the microsatellite number of repetitions (n). The microsatellite markers were widely used with different aims, such as the identification of diseases (Hearne *et al.*, 1992; Naidoo & Chetty, 1998), forensic investigations (Urquhart *et al.*, 1994; Lindqvist *et al.*, 2010), genetic diversity studies (Agha *et al.*, 2008; Iamartino *et al.*, 2005), population structure (Tolone *et al.*, 2012; Bordonaro *et al.*, 2012), genetic traceability (Sardina *et al.*, 2015; Dalvit *et al*., 2007a) and parentage verification (Seyedabadi *et al.*, 2006; Rosa *et al*., 2013).

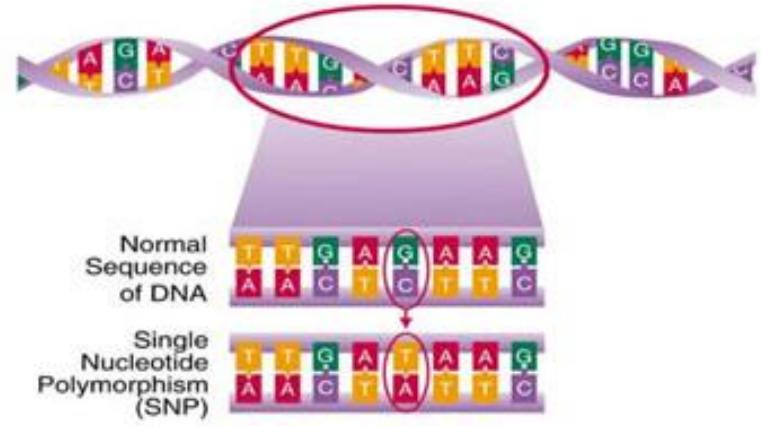
Fig. 1.3.1. Microsatellite markers Mendelian segregation scheme.



- **Single nucleotide polymorphisms (SNPs)**

A single-nucleotide polymorphism (SNP) is a substitution of a single nucleotide at a specific position in the genome, that is present in a sufficiently large fraction of the population. For example, at a specific base position, the "C" nucleotide may appear in most individuals, but in a minority of individuals, the position is occupied by an "A". This means that there is a SNP at this specific position, and the two possible nucleotide variations – "C" or "A" – are said to be the alleles for this specific position (Fig.1.3.2).

Fig. 1.3.2. Single nucleotide polymorphism graphic representation.



SNPs are the smallest unit of genetic variation, and they represent the most common type of sequence polymorphism in animal genomes. They have been used to construct high-resolution genetic maps or to trace evolution or population structures, because they are available in high number, present both in coding and non-coding regions, and dispersed throughout the genome (Gurgul *et al*., 2014).
So far, the most widely used tool in studies on livestock genomes are genotyping microarrays. They allow a relatively quick, reliable and inexpensive genotypes determination of a large single nucleotide polymorphisms (SNPs) number, being a primary source of genetic variation. The arrays are designed to describe the genetic variation within a genome of interest in the best possible way, owing to the use of linkage disequilibrium phenomenon (Matukumalli *et al*., 2009; Kranis *et al*., 2013). Currently, the most advanced genotyping tools in animal genomics are becoming available for cattle and allow analysis of about 770,000 SNPs in parallel; it means

that, in the range of each Mb of genomic sequence, the genetic variation is described by about 300 markers (Rincon *et al.,* 2011). It gives a detailed insight into the genome of the species and allows for a detailed analysis in aspects of its variability, rearrangements and structure (Gurgul *et al*., 2014).

The SNPs are used in livestock species for different studies such as to investigate the animal genome to make population structures studies, to search selection signatures, to investigate about divergences and similitudes within and between populations thanks, to the application of different statistical approaches. The SNPs were also applied in Genome-wide Association Studies (GWAS) to look for common genetic variants associated with a specific phenotype (Sharma *et al*., 2015; Mastrangelo *et al*., 2018c; Sutera *et al*., 2019). In these studies, the knowledge of linkage disequilibrium (LD) between markers, defined as the non-random association of alleles at two or more loci (Gurgul *et al*., 2014), is important in order to define the number of markers required for genomic selection, efficient association studies and fine mapping of genetic diseases (Pritchard and Przeworski 2001; Espigolan *et al*. 2013).

Several studies used SNPs to detect traits of economic interest, thanks to their easy application and their high density within genome (Viale *et al*., 2017; Palombo *et al*., 2018; Dadousis *et al*., 2016; Gutiérrez-Gil *et al*., 2018). Other application provides to use the SNPs in livestock products traceability because the detection of their polymorphism is based on the amplification of very small fragments, even smaller than SSR alleles (Zhao *et al*., 2019; Scarano and Rao, 2014; Goffaux *et al*., 2005).

1.4 Aim and outline of thesis

To carry out the operational strategies for improving the competitiveness of the Valle del Belìce sheep breed, the thesis work was divided in three chapters, due to their different kind of application necessary for each one. The common thread on the experiments carried out is certainly the genetic improvement of the Valle del Belìce sheep breed, with particular regard to the direct effects on the business economy and on the protection of biodiversity, which can derive from a correct study of all the variables that contribute to this improvement.

The overall aim of this thesis is the improvement of competitiveness of the Valle del Belìce sheep breed through different points of view:

- ✓ Parentage reconstruction using microsatellite markers in Valle del Belìce dairy sheep breed.
- ✓ Genetic traceability of Sicilian sheep dairy productions *Vastedda della Valle del Belìce* DOP and *Pecorino Siciliano* DOP.
- ✓ Genome-Wide association studies for phenotypic and productive traits in the Valle del Belìce dairy sheep breed.

References

Abdalla, B. A., Chen, J., Nie, Q., & Zhang, X. (2018). Genomic insights into the multiple factors controlling abdominal fat deposition in a chicken model. Frontiers in genetics, 9, 262.

Agha, S. H., Pilla, F., Galal, S., Shaat, I., D'andrea, M., Reale, S., ... & Li, M. H. (2008). Genetic diversity in Egyptian and Italian goat breeds measured with microsatellite polymorphism. Journal of Animal Breeding and Genetics, 125(3), 194-200.

Al-Atiyat R. M. (2015). The power of 28 microsatellite markers for parentage testing in sheep. Electronic Journal of Biotechnology 18, 116-121.

Arana, A., Soret, B., Lasa, I., & Alfonso, L. (2002). Meat traceability using DNA markers: application to the beef industry. Meat Science, 61(4), 367-373.

Azhar P. M., Chakraborty D., Iqbal Z., Malik A. A., Qaudir A., Asfar A., Bhat I. A. (2018). Microsatellite Markers as a Tool for Characterization of Small Ruminants: A Review. International Journal of Current Microbiology and Applied Sciences, 7(1), 1330-1342.

Bonanno, A., Di Grigoli, A., Mazza, F., De Pasquale, C., Giosuè, C., Vitale, F., & Alabiso, M. (2016). Effects of ewes grazing sulla or ryegrass pasture for different daily durations on forage intake, milk production and fatty acid composition of cheese. animal, 10(12), 2074-2082.

Bordonaro, S., Guastella, A. M., Criscione, A., Zuccaro, A., & Marletta, D. (2012). Genetic diversity and variability in endangered Pantesco and two other Sicilian donkey breeds assessed by microsatellite markers. The Scientific World Journal, 2012.

Bruno-de-Sousa C., Martinez A. M., Ginja C., Santos-Silva F., Carolino M. I., Delgado J. V., Gama L. T. (2011). Genetic diversity and population structure in Portuguese goat breeds. Livestock Science 135, 131-139.

Cappio-Borlino, A., Portolano, B., Todaro, M., Macciotta, N. P. P., Giaccone, P., & Pulina, G. (1997). Lactation curves of Valle del Belìce dairy ewes for yields of milk, fat, and protein estimated with test day models. Journal of Dairy Science, 80(11), 3023-3029.

Cesarani, A., Sorbolini, S., Criscione, A., Bordonaro, S., Pulina, G., Battacone, G., ... & Macciotta, N. P. P. (2018). Genome-wide variability and selection signatures in Italian island cattle breeds. Animal genetics, 49(5), 371-383.

Corretti, A., Gargini, M., Michelini, C., & Vaggioli, M. A. (2004). Tra Arabi, Berberi e Normanni: Entella ed il suo territorio dalla tarda Antichità alla fine dell'epoca sveva. Mélanges de l'école française de Rome, 116(1), 145-190.

Dadousis, C., Biffani, S., Cipolat-Gotet, C., Nicolazzi, E. L., Rossoni, A., Santus, E., ... & Cecchinato, A. (2016). Genome-wide association of coagulation properties, curd firmness modeling, protein percentage, and acidity in milk from Brown Swiss cows. Journal of Dairy Science, 99(5), 3654-3666.

Dalvit, C., De Marchi, M., Targhetta, C., Gervaso, M., Cassandro, M. (2007a). Genetic traceability of meat using microsatellite markers. Food Research International 41, 301-307.

Dalvit, C. (2008). Genetic traceability: a tool for the authentication and valorisation of animal products.

Dalvit, C., Dal Zotto, R., De Marchi, M., & Cassandro, M. (2007b). Genetic and productive characterization of the Burlina cattle breed. Poljoprivreda, 13(1), 124-127.

Ebrahimi, M. T. V., Mohammadabadi, M., Esmailizadeh, A. (2017). Using microsatellite markers to analyze genetic diversity in 14 sheep types in Iran. Archives Animal Breeding 60, 183-189.

Espigolan, R., Baldi, F., Boligon, A. A., Souza, F. R., Gordo, D. G., Tonussi, R. L., ... & Albuquerque, L. G. (2013). Study of whole genome linkage disequilibrium in Nellore cattle. BMC genomics, 14(1), 305.

Fernández, M. E., Goszczynski, D. E., Lirón, J. P., Villegas-Castagnasso, E. E., Carino, M. H., Ripoli, M. V., ... & Giovambattista, G. (2013). Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Genetics and molecular biology, 36(2), 185-191.

Finocchiaro, R., Portolano, B., Damiani, G., Caroli, A., Budelli, E., Bolla, P., & Pagnacco, G. (2003). The hairless (hr) gene is involved in the congenital hypotrichosis of Valle del Belìce sheep. Genetics selection evolution, 35(1), 1-10.

Ganal, M. W., Altmann, T., & Röder, M. S. (2009). SNP identification in crop plants. Current opinion in plant biology, 12(2), 211-21.

Geng, T., Guan, X., & Smith, E. J. (2015). Screening for genes involved in antibody response to sheep red blood cells in the chicken, Gallus gallus. Poultry science, 94(9), 2099-2107.

Goffaux, F., China, B., Dams, L., Clinquart, A., & Daube, G. (2005). Development of a genetic traceability test in pig based on single nucleotide polymorphism detection. Forensic science international, 151(2-3), 239-247.

Gurgul, A., Semik, E., Pawlina, K., Szmatoła, T., Jasielczuk, I., & Bugno-Poniewierska, M. (2014). The application of genome-wide SNP genotyping methods in studies on livestock genomes. Journal of applied genetics, 55(2), 197-208.

Gutiérrez-Gil, B., Esteban-Blanco, C., Suarez-Vega, A., & Arranz, J. J. (2018). Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50K SNP chip and whole-genome sequencing. Journal of dairy science, 101(10), 9072-9088.

Hayes, B.J., Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. Genome 53:876-883.

Hearne, C. M., Ghosh, S., & Todd, J. A. (1992). Microsatellites for linkage analysis of genetic traits. Trends in Genetics, 8(8), 288-294.

Herraeza, D. L., Schafer, H., Mosner, J., Fries, H. R., & Wink, M. (2005). Comparison of microsatellite and single nucleotide polymorphismmarkers for the genetic analysis of a Galloway cattle population.Zeitschrift Naturforschung, C60(7–8), 637–643

Iamartino, D., Bruzzone, A., Lanza, A., Blasi, M., & Pilla, F. (2005). Genetic diversity of Southern Italian goat populations assessed by microsatellite markers. Small Ruminant Research, 57(2-3), 249-255.

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., ... & Burt, D. W. (2013). Development of a high density 600K SNP genotyping array for chicken. BMC genomics, 14(1), 59.

Leighton, R. (1999). Sicily before history: an archaeological survey from the Palaeolithic to the Iron Age. Cornell University Press.

Ligda, C., Gabriilidis, G., Papadopoulos, T., & Georgoudis, A. (2000). Investigation of direct and maternal genetic effects on birth and weaning weight of Chios lambs. Livestock Production Science, 67(1-2), 75-80.

Lindqvist, A. K. B., Magnusson, P. K. E., Balciuniene, J., Wadelius C., Lindholm, E., Alarcòn-Riquelme, M. E., Gyllensten, U. B. (2010). Chromosome-specific Panels of Tri- and Tetranucleotide Microsatellite Markers for Multiplex Fluorescent Detection and Automated Genotyping: Evaluation of Their Utility in Pathology and Forensics. Genome Research 6, 1170-1176.

Ling, Y. H., Zhang, X. D., Yao, N., Ding, J. P., Chen, H. Q., Zhang, Z. J., Zhang, Y. H., Ren, C. H., Ma, Y. H., Zhang, X. R. (2012). Genetic Differentiation of Chinese Indigenous Meat Goats Ascertained Using Microsatellite Information. College of Animal Science and Technology, Anhui Agricultural University, Hefei 230036, China, 25, 177-182.

Macgregor, S., Zhao, Z. Z., Henders, A., Martin, N. G., Montgomery, G. W., & Visscher, P. M. (2008). Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucleic acids research, 36(6), e35.

Mariani, P., Panzitta, F., Nardelli Costa, J., Lazzari, B., Crepaldi, P., Marilli, M., ... & Ajmone Marsan, P. (2005, April). Metodi molecolari per la tracciabilita dei prodotti di origine animale. In Proceedings of the 4th World Italian Beef Cattle Congress (pp. 297-302).

Mastrangelo, S., Bahbahani, H., Moioli, B., Ahbara, A., Al Abri, M., Almathen, F., ... & Ciani, E. (2019). Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia. PLoS One, 14(6), e0209632.

Mastrangelo, S., Ben Jemaa, S., Ciani, E., Sottile, G., Moscarelli, A., Boussaha, M., ... & Cassandro, M. (2020). Genome-wide detection of signatures of selection in three Valdostana cattle populations. Journal of Animal Breeding and Genetics.

Mastrangelo, S., Ciani, E., Marsan, P. A., Bagnato, A., Battaglini, L., Bozzi, R., ... & Ciampolini, R. (2018a). Conservation status and historical relatedness of Italian cattle breeds. Genetics Selection Evolution, 50(1), 35.

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., Portolano, B., & Bi. Ov. Ita Consortium. (2018b). Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Animal genetics, 49(1), 71-81.

Mastrangelo, S., Di Gerlando, R., Tolone, M., Tortorici, L., Sardina, M. T., Portolano, B., & International Sheep Genomics Consortium. (2014). Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. BMC genetics, 15(1), 108.

Mastrangelo, S., Portolano, B., Di Gerlando, R., Ciampolini, R., Tolone, M., Sardina, M. T., & International Sheep Genomics Consortium. (2017). Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. Animal, 11(7), 1107-1116.

Mastrangelo, S., Sottile, G., Sutera, A. M., Di Gerlando, R., Tolone, M., Moscarelli, A., ... & Portolano, B. (2018c). Genome-wide association study reveals the locus responsible for microtia in Valle del Belìce sheep breed. Animal genetics, 49(6), 636-640.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., ... & Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. PloS one, 4(4), e5350.

Micalizzi, M. (1995). Effetti dell'integrazione alimentare sulla produzione quanti-qualitativa di latte di pecore "Valle del Belìce". Tesi di Laurea, Facoltà di Agraria, Università degli Studi di Palermo: 69.

Moscarelli, A., Sardina, M. T., Cassandro, M., Ciani, E., Pilla, F., Senczuck, G., ... & Mastrangelo, S. (2020). Genome-wide assessment of diversity and differentiation between original and modern Brown cattle populations. Animal Genetics.

Naidoo, R., & Chetty, R. (1998). The application of microsatellites in molecular pathology. Pathology & Oncology Research, 4(4), 310.

Orrù, G., Ferrando, M. L., Meloni, M., Liciardi, M., Savini, G., & De Santis, P. (2006). Rapid detection and quantitation of Bluetongue virus (BTV) using a Molecular Beacon fluorescent probe assay. Journal of virological methods, 137(1), 34-42.

Palombo, V., Milanesi, M., Sgorlon, S., Capomaccio, S., Mele, M., Nicolazzi, E., ... & D'Andrea, M. (2018). Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. Journal of dairy science, 101(12), 11004-11019.

Peelman, L. J., Mortiaux, F., Van Zeveren, A., Dansercoer, A., Mommens, G., Coopman, F., ... & Portetelle, D. (1998). Evaluation of the genetic variability of 23 bovine microsatellite markers in four Belgian cattle breeds. Animal Genetics, 29(3), 161-167.

Portolano, B., Chiodi, M., Di Gerlando, R., Sottile, G., Mastrangelo, S., Sardina, M., & Tolone, M. (2016). Metodo per la tracciabilità/autenticazione di prodotti lattiero-caseari mono-razza.

Portolano, B., Todaro, M., Spatafora, F., Bono, G., Margiotta, S., Giaccone, P., & Ortoleva, V. (1996). Confronto fra due differenti modelli della curva di lattazione in pecore da latte di razza Comisana (Comparison between two different lactation curve models in dairy Comisana ewes). Zootec. Nutr. Anim, 22, 323.

Portolano, N. (1987). Pecore e capre Italiane. Edagricole.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. BMC genetics, 13(1), 70.

Qanbari, S., Eskandari, Nasab, M. P., Osfoori, R., Hagh Nazari, A. (2007). Power of Microsatellite Markers for Analysis of Genetic Variation and Parentage Verification in Sheep. Pakistan Journal of Biological Sciences 10, 1632-1638.

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. Current opinion in plant biology, 5(2), 94-100.

Rincon, G., Weber, K. L., Van Eenennaam, A. L., Golden, B. L., & Medrano, J. F. (2011). Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys. Journal of dairy science, 94(12), 6116-6121.

Rosa, A. J. M., Sardina, M. T., Mastrangelo, S., Tolone, M., Portolano, B. (2013). Parentage verification of Valle del Belìce dairy sheep using multiplex microsatellite panel. Small Ruminant Research 113, 62- 65.

Rumiz, P. (2009). La scossa che spostò un accento. La Repubblica.

Sancristobal-Gaudy, M., Renand, G., Amigues, Y., Boscher, M. Y., Levéziel, H., & Bibé, B. (2000). Individual traceability of bovine meat using molecular markers. Productions Animales, 13(4), 269-276.

Sardina, M. T., Tortorici, L., Mastrangelo, S., Di Gerlando, R., Tolone, M., Portolano, B. (2015). Application of microsatellite markers as potential tools for traceability of Girgentana goat breed dairy products. Food Research International 74, 115-122.

Scarano, D., & Rao, R. (2014). DNA markers for food products authentication. Diversity, 6(3), 579-596.

Seyedabadi, H. R., AMIRINIA, S., BANA, B. M., & Emrani, H. (2006). Parentage verification of Iranian Caspian horse using microsatellites markers.

Shackell, G. H., Mathias, H. C., Cave, V. M., & Dodds, K. G. (2005). Evaluation of microsatellites as a potential tool for product tracing of ground beef mixtures. Meat science, 70(2), 337-345.

Sharma, A., Lee, J. S., Dang, C. G., Sudrajad, P., Kim, H. C., Yeon, S. H., ... & Lee, S. H. (2015). Stories and challenges of genome wide association studies in livestock—a review. Asian-Australasian journal of animal sciences, 28(10), 1371.

Spatafora, F., & Calascibetta, A. M. G. (1986). MONTE MARANFUSA (PA). LA PRIMA CAMPAGNA DI SCAVO. Annali della Scuola Normale Superiore di Pisa. Classe di Lettere e Filosofia, 16(4), 1105-1115.

Stevanovic, J., Stanimirovic, Z., Dimitrijevic, V., Maletic, M. (2010). Evaluation of 11 microsatellite loci for their use in paternity testing in Yugoslav Pied cattle (YU Simmental cattle). Czech Journal of Animal Science 55, 221–226.

Sutera, A. M., Riggio, V., Mastrangelo, S., Di Gerlando, R., Sardina, M. T., Pong-Wong, R., ... & Portolano, B. (2019). Genome-wide association studies for milk production traits in Valle del Belìce sheep using repeated measures. Animal genetics, 50(3), 311-314.

Tanksley, S. D. (1983). Molecular markers in plant breeding. Plant Molecular Biology Reporter, 1(1), 3-8.

Tibe, O., Meagher, L. P., Fraser, K., & Harding, D. R. (2011). Condensed tannins and flavonoids from the forage legume sulla (*Hedysarum coronarium*). Journal of agricultural and food chemistry, 59(17), 9402-9409.

Tolone, M., Mastrangelo, S., Rosa, A. J. M., Portolano, B. (2012). Genetic diversity and population structure of Sicilian sheep breeds using microsatellite markers. Small Ruminant Research 102, 18-25.

Tortorici, L., Di Gerlando, R., Tolone, M., Mastrangelo, S., Sardina, M. T. (2016). 12S rRNA mitochondrial gene as marker to trace Sicilian mono-species dairy products. Livestock Science 193, 39-44.

Urquhart, A., Kimpton, C. P., Downes, T. J., & Gill, P. (1994). Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. International journal of legal medicine, 107(1), 13-20.

Viale, E., Tiezzi, F., Maretto, F., De Marchi, M., Penasa, M., & Cassandro, M. (2017). Association of candidate gene polymorphisms with milk technological traits, yield, composition, and somatic cell score in Italian Holstein-Friesian sires. Journal of dairy science, 100(9), 7271-7281.

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.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., *et al* (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565-569.

Yilmaz, O. (2016). Power of different microsatellite panels for paternity analysis in sheep. Animal Science Papers and Reports 34, 155-164.

Zhao, J., Xu, Z., You, X., Zhao, Y., He, W., Zhao, L., ... & Yang, S. (2019). Genetic traceability practices in a large-size beef company in China. Food chemistry, 277, 222-228.

CHAPTER 2

Parentage reconstruction using microsatellite markers in Valle del Belìce dairy sheep breed

2.1 Summary

The ovine breeding sector presents several criticalities in the monitoring of couplings and in their planning, due to the flock's numerosity and the natural mating system used, sometimes favorite by the lack of breeders' formation and lack of state-of-the-art infrastructure for livestock farming. Additionally, for sheep breeders is very difficult to record correctly pedigrees, because they have difficult to monitor every single mating and every single birth. Consequently, pedigrees mistakes cause problems on animal breeding programs and make it hard to achieve the objectives of the breeding organization, due to frequently errors that can complicate genetic improvement programs. Using the molecular markers is possible to ascertain the declared parentage for selected rams and mares, perfecting the animal breeding programs. The aim of this study was properly to attribute correctly the parentage, performing the parentage analysis test to reconstruct appropriately the declared pedigrees, indispensable for improving the competitiveness of the Valle del Belìce sheep breed. For the analysis were used microsatellite markers, to reduce pedigrees errors and to set up a markers' panel useful for parentage analysis in Valle del Belìce sheep breed. All the used microsatellite markers showed high polymorphisms and the mean PIC value was similar to previously used panels, made it suitable for our purpose. Our results confirm the critical point on pedigrees recording by sheep breeder, sometimes deceived by adoption phenomena and heteropaternal superfecundation: these two phenomena were considered plausible in our study, based on the results obtained, confirming the high potential of parentage test to assist the sheep breeders in pedigree recording and contributing to explain the high frequencies of pedigrees errors.

2.2 Introduction

Pedigree information, confirmed by parentage molecular test, is essential for genetic evaluation (Yilmaz *et al.*, 2018). Pedigree errors reduce genetic progress, due to incorrect estimation of breeding values as reported in previous studies in livestock, where errors of 10-23% were detected (Weller *et al.*, 2004; Harder *et al.*,

2005; Sanders *et al.*, 2006; Vandeputte *et al.*, 2006). Pedigree mistakes cause problems on animal breeding programs and make it hard to achieve the objectives of the breeding organization (Geldermann *et al.*, 1986; Weller *et al.*, 2004; Heaton *et al.*, 2014).

The Valle del Belìce is the most diffused sheep breed reared in Sicily for milk production, and nowadays about 167,468 animals are enrolled in the National database of Teramo (www.vetinfo.it – detection of 31/12/2020). The most diffused mating system in Valle del Belìce breed flocks is natural mating. Moreover, the interchange of rams among flocks is quite rare and the reproductive management is characterized by unrecorded mating with multiple sires (Tolone *et al.*, 2012). The farmers usually divided the ewes into different groups with homogeneous estrus period, and then, introduced the selected rams in each group. In the last days of the pregnancy period, the farmers grouped all the ewes close to calving in the same closed stable waiting for the births. Usually, the ewes suckled only their own lamb/lambs and the records "mother-offspring" are based on this assumption. Therefore, there is a need to generate reliable pedigree for inbreeding control beside engendering a breeding program. The parentage analysis conducted using microsatellites markers could improve the animal breeding programs reducing pedigree errors.

This method could help to define parentage test to improve mating plans aimed to reduce inbreeding avoiding the use of consanguineous individuals as reproducers and maintaining genetic variability within breed. The aim of this work was to evaluate a PCR based microsatellite markers multiplex system for parentage verification of Sicilian Valle del Belìce dairy sheep.

2.3 Materials and methods

Sampling

A total of 151 individual blood samples, of Valle del Belìce breed, were collected from jugular vein using vacutainer tube containing EDTA as anticoagulant. Individual samples belonged to 4 flocks sited in Agrigento province. Fathers, mothers and offspring collected in each flock are detailed in Table 2.3.1.

Table 2.3.1 Detailed information about samples collected in 4 farms of Valle del Belìce breed.

Farms	Total individuals	Rams	Ewes	Progeny	Twins pairs
A - (PAL)	55	3	25	27	6
B - (BAN)	29	5	11	13	2
C - (LAO)	47	7	20	20	0
D - (VIT)	19	8	5	6	1

All starting information on relationship within each flock was based on farmer's indications and pedigrees records.

On farm A, we set up the entire protocol from microsatellite markers panel amplification till parentage test. On farms B, C, and D, we applied and validate the entire procedure.

DNA extraction.

The DNA extractions were performed starting from 1 ml of whole blood following a salting-out method (Miller *et al.*, 1988).

DNA quantification

The DNA samples were quantified with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), able of giving, through absorption at various wavelengths (230-260-280 nm), an exact measure of concentration and quality of DNA, therefore, of the degree of purity of the extracted sample.

For good quality DNA the absorption ratio values 260/280 should be 1.8-2.2. If value is different probably there is contamination of proteins or other compounds. In this case, the DNA will have to be further purified to avoid that these contaminations can interfere with the enzymatic reactions used in molecular biology.

After quantification, the DNA samples were diluted to the working concentration of 50 ng/µl and stored at 4 °C for subsequent analyses.

Microsatellites amplification by multiplex PCR.

The Polymerase Chain Reaction (PCR) were performed in a final volume of 10 µl containing 100 ng of DNA, 1X QIAGEN PCR Master Mix, 0.1 µM of each primer pair and ultrapure H2O.

Each microsatellite molecular marker was identified by name, listed on Probe section of NCBI (https://www.ncbi.nlm.nih.gov/), selectively amplified by primer pair labeled by fluorophore, and with known physical location on ovine chromosome. We used the four fluorophores VIC, FAM, PET and NED, respectively corresponding at green, blue, red and yellow colors, that allowed us to analyze, in the same automated electrophoretic run, microsatellites with identical size range in bp (base pairs), and to use in the same multiplex PCR more than four of microsatellites with different size range.

We selected 24 microsatellites from literatures (Yilmaz, 2016; Azhar *et al.*, 2018) and from international source such as ISAG (ISAG_Sheep and Goat Test Markers 2017.pdf) and FAO (ISAG_MoDAD, New Microsatellite marker sets – Recommendation of joint ISAG/FAO Standing Committee.pdf). The 24 markers were amplified in four multiplex PCRs, each one from a minimum of 4 to a maximum of 8 markers (Tables from 2.3.1 to 2.3.4).

Table 2.3.1. PCR Multilplex 1 information

MULTIPLEX I			
Marker name	**Fluorochrome**	**Range**	**Chromosome**
BM8125	FAM	100-130 bp	17
HSC	FAM	255-300 bp	20
HH47	VIC	110-150 bp	18
MAF33	NED	110-140 bp	9
MAF214	NED	170-220 bp	16
CSRD247	PET	210-260 bp	14
OarVH72	PET	120-150 bp	25

Table 2.3.2. PCR Multilplex 2 information.

MULTIPLEX II			
Marker name	**Fluorochrome**	**Range**	**Chromosome**
MAF65	FAM	115-145 bp	15
BM827	FAM	200-220 bp	3
MB099	VIC	186-200 bp	1
OarAE54	VIC	118-146 bp	25

INRA104	PET	140-180 bp	22
OarCP34	PET	95-130 bp	3
OarFCB193	NED	98-122 bp	11
OarFCB304	NED	155-191 bp	19

Table 2.3.3. PCR Multilplex 3 information.

MULTIPLEX III			
Marker name	**Fluorochrome**	**Range**	**Chromosome**
INRA063	FAM (blu)	150-195 bp	14
OarAE129	VIC (verde)	130-168 bp	5
MCM527	PET (rosso)	160-190 bp	5
OarJMP58	NED (giallo)	135-175 bp	26

Table 2.3.4. PCR Multilplex 4 information

MULTIPLEX IV			
Marker name	**Fluorochrome**	**Range**	**Chromosome**
DYMS1	FAM (blu)	160-210 bp	20
BM1824	VIC (verde)	160-180 bp	1
OarFCB128	VIC (verde)	90-135 bp	2
OarJMP29	NED (giallo)	105-160 bp	24
INRA132	PET (rosso)	140-180 bp	20

The amplification conditions are identical for each multiplex PCR and indicated on Table 2.3.5.

Table 2.3.5. amplification condition thermal cycler settings.

Step	**Temperature (°C) and time**
Initial denaturation	95°C x 15 min
30 cycles of	94°C x 30 sec 58°C x 90 sec 72°C x 60 sec
Final extension	60°C x 30 min

Capillary electrophoresis using ABI3500xl Genetic Analyzer

The fragments analysis was performed using the ABI 3500xl Genetic Analyzer instrument (Applied Biosystem), by capillary electrophoresis.

For the instrument analysis 1.0 µl of PCR product (dilution 1:25) was mixed with:

- 8.7 µl of Hi-Di Formamide (Applied Byosystems)
- 0.3 µl of GeneScan 500 LIZ Size Standard (Applied Biosystems)

The fragments were separate by capillary electrophoresis basing on their size in bp and transformed in alleles by GeneMapper v5.0 software (Applied Biosystems). After allele assignment to each microsatellite Genotypes Table, containing the genotypes to all markers of all analyzed individuals, was exported and used for further analyses. Using CERVUS v3.0.7 software (Kalinowski *et al.*, 2007), we analyzed genotypic data to obtain allele frequencies and descriptive statistics on microsatellite markers and then performed paternity assignment.

2.4 Result and discussion

Microsatellites descriptive statistics

Summary statistics of Farm A obtained using Cervus v3.0.7, are showed in Tab. 2.4.1., and contained the total number of individuals, the total number of loci, the mean number of alleles per locus, the mean observed (HObs) and expected (HExp) heterozygosity, the mean polymorphic information content (PIC), and the combined non-exclusion probabilities of first parent, second parent, parent pair, identity, and sib identity.

Tab. 2.4.1. Average values of the statistical parameters related to microsatellite markers.

Number of individuals:	55
Number of loci:	24
Mean number of alleles per locus:	7.042
Mean proportion of loci typed:	0.9917
Mean observed heterozygosity:	0.7072
Mean expected heterozygosity:	0.7208
Mean polymorphic information content (PIC):	0.6773
Combined non-exclusion probability (first parent):	0.00003118
Combined non-exclusion probability (second parent):	0.00000002
Combined non-exclusion probability (parent pair):	8.151E-0014
Combined non-exclusion probability (identity):	5.549E-0024
Combined non-exclusion probability (sib identity):	8.425E-0010

Moreover, the following report (Tab. 2.4.2) presented the summary statistics for each analyzed microsatellite marker, showing the number of alleles for each locus (k), the number of genotyped individuals (N), the observed (HObs) and expected (HExp) heterozygosity, the polymorphic information content (PIC), the mean exclusion probabilities (NE-1P, …, NE-SI) and, if required, the Hardy-Weinberg equilibrium (HW) and the frequency of null allele (Fnull).

Tab. 2.4.2. Descriptive statistics of each analyzed microsatellite marker.

Locus	k	N	HObs	HExp	PIC	NE-1P	NE-2P	NE-PP	NE-I	NE-SI	HW	F(Null)
BM8125	7	55	0.782	0.757	0.717	0.639	0.458	0.266	0.096	0.399	NS	-0.0304
CSRD247	9	55	0.836	0.836	0.806	0.513	0.340	0.165	0.052	0.349	ND	-0.0045
HSC	10	55	0.909	0.816	0.785	0.541	0.366	0.182	0.060	0.361	ND	-0.0619
MAF214	4	55	0.600	0.638	0.559	0.790	0.647	0.488	0.208	0.486	NS	0.0278
MAF33	6	55	0.764	0.785	0.742	0.615	0.436	0.255	0.085	0.382	ND	0.0081
OarHH47	10	55	0.855	0.837	0.810	0.502	0.332	0.154	0.049	0.347	ND	-0.0133
VH72	6	55	0.818	0.669	0.601	0.755	0.598	0.424	0.175	0.462	NS	-0.1171
BM827	7	50	0.280	0.674	0.627	0.737	0.561	0.368	0.151	0.454	***	0.4172
INRA104	7	55	0.691	0.752	0.715	0.640	0.456	0.259	0.095	0.401	NS	0.0446
MAF65	5	55	0.636	0.674	0.621	0.742	0.570	0.384	0.157	0.455	NS	0.0392
MB099	4	55	0.255	0.271	0.245	0.964	0.869	0.775	0.559	0.755	ND	0.0417
OarAE54	8	55	0.709	0.712	0.667	0.692	0.515	0.319	0.126	0.428	NS	-0.0027
OarCP34	6	55	0.764	0.771	0.725	0.636	0.458	0.277	0.094	0.391	NS	-0.0011
OarFCB193	5	55	0.564	0.505	0.468	0.864	0.704	0.531	0.282	0.570	NS	-0.0998
OarFCB304	6	55	0.600	0.595	0.522	0.814	0.672	0.510	0.236	0.514	NS	-0.0025
INRA063	12	54	0.796	0.807	0.775	0.554	0.378	0.190	0.065	0.366	NS	0.0049
JMP58	9	54	0.815	0.859	0.834	0.462	0.298	0.130	0.040	0.334	ND	0.0204
MCM527	9	52	0.904	0.845	0.818	0.490	0.321	0.147	0.046	0.343	ND	-0.0412
OarAE129	3	54	0.556	0.636	0.556	0.801	0.658	0.508	0.211	0.487	NS	0.0671
BM1824	5	55	0.745	0.800	0.760	0.595	0.415	0.237	0.076	0.372	ND	0.0316
DYMS1	10	55	0.745	0.725	0.681	0.680	0.501	0.307	0.117	0.420	NS	-0.0114
INRA132	7	55	0.673	0.754	0.718	0.637	0.453	0.256	0.093	0.400	NS	0.0569
JMP29	8	55	0.836	0.816	0.785	0.546	0.370	0.187	0.061	0.361	NS	-0.0202
OarFCB128	6	55	0.836	0.766	0.718	0.646	0.468	0.288	0.099	0.395	NS	-0.0494

The mean number of alleles per locus results 7.042, lower value than other similar studies previously conducted on sheep (Qanbari *et al.*, 2007; Rosa *et al.*, 2013; Al-Atiyat, 2015; Yilmaz, 2016) and cow (Stevanovic *et al.*, 2010). This difference could be due to the analyzed breed, considering that the same microsatellite could be more or less polymorphic in different breeds. Rosa *et al.* (2013) worked on Valle del Belìce sheep breed, reporting a mean number of alleles per locus of 11.21; anyway, they used a different microsatellites panel and analyzed individuals belonging from several different flocks. After this consideration we could define our result (mentioned summary statistics based only on Fam A) as quite reliable. Moreover, considering that more polymorphic a microsatellite is more optimal its use is in terms of parentage attribution, among all analyzed microsatellites only

OarAE129 presented 3 alleles, while the marker with the highest k value was INRA063. In the panel, only two microsatellites (MB099 and OarFCB193) presented PIC values lower than 0.5, while the mean values in this farm was 0.6773. This result agreed with previous studies on sheep breeds, in which PIC values ranged from a minimum of 0.67 to a maximum of 0.80 (Qanbari *et al.*, 2007; Rosa *et al.*, 2013; Al-Atiyat, 2015; Yilmaz, 2016). Usually, this value should be greater than or equal to 0.5 according with Botstein *et al.* (1980).

The mean observed (0.7072) and expected (0.7208) heterozygosity presented similar values to others previous studies such as Qanbari *et al.* (2007) which reported values of 0.74 (Hobs) and 0.72 (HExp); Rosa *et al.* (2013) 0.708 and 0.772; and Yilmaz (2016) 0.81 and 0.82, respectively.

In particular, the HObs and HExp values for INRA063 marker in Valle del Belìce breed reported by Rosa *et al.* (2013) were 0.738 and 0.813, respectively, and are similar to our results (0.796 and 0.807), even though they analyzed samples belonging to 12 farms.

Basing on these results, we deduced that this microsatellite marker panel was suitable for our purpose and we proceeded to analyze the other three farms.

The table 2.4.3 showed the summary statistics of the four analyzed farms, confirming that this panel showed notable results in terms of mean PIC values (>0.5), mean number of alleles, and heterozygosity.

Tab. 2.4.3. Summary statistics of each analyzed farm

Farms	Total individuals	Rams	Ewes	Progeny	N° of alleles	HObs	HExp	PIC
A	55	3	26	26	7.042	0.7072	0.7208	0.6773
B	29	5	11	13	5.375	0.6118	0.6084	0.5596
C	47	7	20	20	5.083	0.5368	0.5417	0.5033
D	19	8	5	6	5.708	0.7063	0.6667	0.6138

The Table 2.4.4 showed the overall summary statistics useful to evaluate the marker panel for further parentage analysis. The mean number of alleles per locus was high (10.167) with some microsatellite in the panel with really high polymorphic level such as CSRD247 and OarHH47 with 20 alleles and JMP29 with 18 alleles. The mean proportion of loci typed was higher than 90% and the mean PIC value was ~

0.70; so again, these results agreed with previous studies on different sheep breeds (Qanbari *et al.*, 2007; Rosa *et al.*, 2013; Al-Atiyat, 2015; Yilmaz, 2016).

Table. 2.4.4. Summary statistics of analyzed microsatellite marker panel

Locus	k	N	HObs	HExp	PIC	NE-1P	NE-2P	NE-PP	NE-I	NE-SI	HW	F(Null)
BM8125	7	145	0.676	0.713	0.675	0.685	0.504	0.306	0.119	0.425	NS	0.0233
CSRD247	20	147	0.769	0.887	0.875	0.368	0.225	0.074	0.022	0.314	*	0.0728
HSC	13	145	0.869	0.879	0.864	0.401	0.249	0.095	0.028	0.319	NS	0.0050
MAF214	7	147	0.497	0.582	0.532	0.814	0.654	0.476	0.224	0.516	NS	0.0859
MAF33	8	147	0.762	0.760	0.718	0.644	0.467	0.284	0.098	0.396	NS	-0.0016
OarHH47	20	147	0.776	0.888	0.876	0.367	0.225	0.076	0.023	0.313	NS	0.0683
VH72	9	146	0.610	0.589	0.543	0.805	0.640	0.455	0.214	0.510	NS	-0.0162
BM827	7	141	0.468	0.698	0.647	0.715	0.544	0.358	0.141	0.438	***	0.2010
INRA104	12	146	0.760	0.835	0.811	0.504	0.333	0.159	0.050	0.346	NS	0.0479
MAF65	7	146	0.760	0.793	0.761	0.584	0.406	0.221	0.074	0.373	NS	0.0204
MB099	4	147	0.272	0.301	0.271	0.955	0.855	0.754	0.519	0.730	ND	0.0556
OarAE54	10	147	0.633	0.697	0.644	0.714	0.546	0.361	0.144	0.438	NS	0.0441
OarCP34	7	147	0.769	0.763	0.723	0.635	0.458	0.273	0.095	0.393	NS	-0.0060
OarFCB193	5	147	0.388	0.371	0.340	0.930	0.804	0.672	0.427	0.672	NS	-0.0478
OarFCB304	10	146	0.644	0.641	0.600	0.759	0.583	0.388	0.169	0.473	NS	-0.0127
INRA063	14	102	0.784	0.787	0.764	0.565	0.384	0.184	0.066	0.375	NS	-0.0004
JMP58	11	103	0.816	0.885	0.870	0.387	0.239	0.087	0.025	0.316	ND	0.0378
MCM527	10	101	0.871	0.802	0.772	0.562	0.385	0.199	0.067	0.368	NS	-0.0447
OarAE129	5	103	0.524	0.736	0.684	0.687	0.513	0.336	0.120	0.414	***	0.1690
BM1824	8	144	0.667	0.747	0.719	0.633	0.447	0.245	0.091	0.401	NS	0.0612
DYMS1	12	145	0.697	0.732	0.706	0.647	0.460	0.252	0.097	0.409	NS	0.0317
INRA132	12	147	0.748	0.837	0.814	0.499	0.329	0.156	0.048	0.345	NS	0.0562
JMP29	18	147	0.776	0.891	0.879	0.362	0.221	0.073	0.022	0.312	**	0.0706
OarFCB128	8	146	0.671	0.736	0.698	0.664	0.483	0.291	0.107	0.410	NS	0.0462

Number of individuals:	151
Number of loci:	24
Mean number of alleles per locus:	10.167
Mean proportion of loci typed:	0.9186
Mean observed heterozygosity:	0.6753
Mean expected heterozygosity:	0.7313
Mean polymorphic information content (PIC):	0.6994
Combined non-exclusion probability (first parent):	0.00000417
Combined non-exclusion probability (second parent):	1.198E-0009
Combined non-exclusion probability (parent pair):	5.384E-0016
Combined non-exclusion probability (identity):	2.768E-0026
Combined non-exclusion probability (sib identity):	4.075E-0010

Paternity assignment with CERVUS software

The first analysis performed with CERVUS v3.0.7 software was the allele frequency estimate using genotypes table. The output file contained the previously described summary statistics, considering the overall samples (Table 2.4.2) and details of each analyzed microsatellite marker.

The second analysis was simulation of parentage verification performed both for estimating the power of a codominant loci series given their allelic frequencies, and for estimating the critical probability values to establish that the assigned parents (by breeder) are the true parents, giving to the assignment test a statistical significance value. The probability values could be indicated with LOD score, defined as the ratio between the probability that candidate parent is the real parent and the probability that it is not; or with Delta value, defined as the LOD score difference between the most probable parents and the second one. The paternity

simulation analysis considers 10,000 offspring, the number of candidate fathers (collected samples), the proportion of candidate fathers sampled in the farm, the proportion of loci typed (from allele frequency file) and the proportion of loci mistyped (0.01 by default) as parameters. Moreover, the minimum number of typed loci was set to 12 (half the number of markers in the panel).

Finally, for paternity analyses, we prepared two different files for each farm. The first contained the rams considered as "candidate fathers" of all offspring (lambs collected in each farm), and the second one contained the pairs "offspring-candidate mother" based on information recorded by the farmers on pedigrees.

Results for paternity analysis obtained from Cervus v3.0.7 software was showed in the following tables and were performed in each farm separately. Tables 2.4.2-2.4.5 showed results of offspring-mother pairs containing pair loci mismatching present in each farm.

The farm A presented the highest percentage of offspring-mother pairs containing mismatches (48%) respect to the other farms which showed lower values: 30% farm C, 16.7% farm D, and 7.7% farm B.

Table 2.4.2 Results of offspring-mother pairs containing loci mismatching in Farm A (PAL).

Offspring ID	Loci typed	Mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score
PAL3F	23	PAL3	24	23	1	5.56E+00
PAL8F	24	PAL8	23	23	13	-4.95E+01
PAL10F1	24	PAL10	24	24	1	1.21E+01
PAL10F2	24	PAL10	24	24	1	9.87E+00
PAL11F1	24	PAL11	24	24	9	-3.25E+01
PAL11F2	24	PAL11	24	24	4	-7.36E+00
PAL14F	24	PAL14	24	24	1	1.43E+01
PAL15F2	24	PAL15	24	24	5	-1.64E+01
PAL19F	24	PAL19	24	24	1	1.41E+01
PAL21F2	24	PAL21	24	24	1	8.36E+00
PAL23F	24	PAL23	24	24	2	1.33E+00
PAL24F2	24	PAL24	24	24	1	1.38E+01

Table 2.4.3. Results of offspring-mother pairs containing loci mismatching in Farm B (BAN).

Offspring ID	Loci typed	Mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score
BAN10F1	24	BAN10	24	24	1	8.49E+14

Table 2.4.4. Results of offspring-mother pairs containing loci mismatching in Farm C (LAO).

Offspring ID	Loci typed	Mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score
LAO11F	20	LAO11	20	20	1	1.06E+01
LAO12F	20	LAO12	20	20	1	8.49E+00
LAO17F	20	LAO17	16	16	2	-5.04E+00
LAO22F	20	LAO22	20	20	1	6.09E+00
LAO26F	20	LAO26	20	20	5	-2.15E+01
LAO27F	20	LAO27	20	20	1	5.55E+00

Table 2.4.5. Results of offspring-mother pairs containing loci mismatching in Farm D (VIT).

Offspring ID	Loci typed	Mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score
VIT10F	24	VIT10	24	24	1	4.35E+00

The LOD scores presented negative values when the number of mismatches was higher than the proportion of loci mistypes (1%), nevertheless, we consider for subsequent *paternity assignment test* only the offspring-mother pairs without mismatches.

Table 2.4.6 showed all the twin pairs and relative mothers (declared by the farmers) and analyzed with markers panel. Out of nine twin pairs only three (indicated in underlined bold on table 2.4.6) could be genetically real, while all the other information reported by the farmers were partially wrong. In fact, it was possible to note that in four different twin pairs, showed in bold in Table 2.4.6, only one offspring could be the real one, while the other one presented mismatch with assigned mother.

Table 2.4.6. Twins pairs and relative known mothers presented in the four analyzed farms. The real twin pairs are reported in bold and in bold underlined.

Offspring ID	Loci typed	Mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score
PAL5F1	**24**	**PAL5**	**24**	**24**	**0**	**8.32E+00**
PAL5F2	**24**	**PAL5**	**24**	**24**	**0**	**8.81E+00**
PAL10F1	24	PAL10	24	24	1	1.21E+01

PAL10F2	24	PAL10	24	24	1	9.87E+00
PAL11F1	24	PAL11	24	24	9	-3.25E+01
PAL11F2	24	PAL11	24	24	4	-7.36E+00
PAL15F1	**24**	**PAL15**	**24**	**24**	**0**	**1.32E+01**
PAL15F2	24	PAL15	24	24	5	-1.64E+01
PAL21F1	**24**	**PAL21**	**24**	**24**	**0**	**1.12E+01**
PAL21F2	24	PAL21	24	24	1	8.36E+00
PAL24F1	**24**	**PAL24**	**24**	**24**	**0**	**1.57E+01**
PAL24F2	24	PAL24	24	24	1	1.38E+01
BAN10F1	24	BAN10	24	24	1	8.49E+14
BAN10F2	**24**	**BAN10**	**24**	**24**	**0**	**1.43E+15**
BAN12F1	**24**	**BAN12**	**24**	**24**	**0**	**5.94E+14**
BAN12F2	**24**	**BAN12**	**24**	**24**	**0**	**6.09E+14**
VIT09F1	**24**	**VIT09**	**24**	**24**	**0**	**8.98E+00**
VIT09F2	**24**	**VIT09**	**24**	**24**	**0**	**8.87E+00**

in bold are indicated the real one offspring of indicated mother; **in underlined bold** are indicated the genetically real twins.

These results led us to consider that in Farm A (PAL) the number of incorrect offspring-mother pairs recorded by the farmer was high. Moreover, the four twin pairs in bold in Table 2.4.6 led us to hypothesize the adoption phenomena. To the best of our knowledges, literatures regarding adoption phenomena in Valle del Belìce sheep breed are absent even if, as reported by several farmers, it could be happened especially when in the last days of the pregnancy period, the farmers grouped all the ewes close to calving (primiparous and pluriparous) in the same closed stable waiting for the births. After birth, suckling of colostrum established a physical link between ewe and lamb even when they are not genetically related; in facts, in the first few hours the mothers are highly responsive to any neonate during the so called "sensitive period" (Hersher *et al*., 1963); it is during this sensitive period that the ewe shows intense licking activity (Gonzalez & Goddard, 1998).

The paternity assignment analyses were conducted for all offspring-mother pairs with no mismatches, present in the four farms. Results were reported in the following Tables (2.4.7-2.4.10). In particular, in Farm A (PAL), out of 13 correct offspring-mother pairs (52%), 11 trio (offspring-mother-father) were found, and it is possible to highlight that the three collected rams indicated as candidate fathers are the assigned ones (Table 2.4.7).

Table 2.4.7. Paternity assignment in Farm A (PAL) for offspring-mother pairs with no mismatching.

Offspring ID	Mother ID	Pair loci mismatch	Pair LOD score	Candidate father ID	Pair loci compared	Pair loci mismatch	Pair LOD score	Trio loci mismatch	Trio LOD score
PAL4F	PAL4	0	1.94E+01	PAL28P	23	0	1.10E+01	0	1.94E+01
PAL5F1	PAL5	0	8.32E+00	PAL28P	23	0	9.69E+00	0	1.82E+01
PAL5F2	PAL5	0	8.81E+00	PAL28P	23	0	1.21E+01	0	1.83E+01
PAL6F	PAL6	0	1.87E+01	PAL29P	24	0	1.56E+01	0	2.55E+01
PAL7F	PAL7	0	1.08E+01	PAL28P	22	0	8.96E+00	0	1.39E+01
PAL15F1	PAL15	0	1.32E+01	PAL30P	23	0	9.39E+00	0	1.39E+01
PAL17F	PAL17	0	1.20E+01	PAL28P	19	0	5.75E+00	0	1.12E+01
PAL18F	PAL18	0	1.48E+01	PAL30P	23	0	5.74E+00	0	1.77E+01
PAL21F1	PAL21	0	1.12E+01	PAL30P	23	0	1.10E+01	0	1.92E+01
PAL24F1	PAL24	0	1.57E+01	PAL28P	23	0	9.23E+00	0	1.90E+01
PAL26F	PAL26	0	1.31E+01	PAL30P	23	0	1.39E+01	0	2.13E+01

In Farm 2 (BAN), where the correct offspring-mother pairs represented the 92.3%, the number of trio was really low and only four “families” were present (Table 2.4.8).

Table 2.4.8. Paternity assignment in Farm B (BAN) for offspring-mother pairs with no mismatching.

Offspring ID	Mother ID	Pair loci mismatch	Pair LOD score	Candidate father ID	Pair loci compared	Pair loci mismatch	Pair LOD score	Trio loci mismatch	Trio LOD score
BAN6F	BAN6	0	9.30E+14	BAN2P	24	0	1.09E+15	0	2.00E+15
BAN8F	BAN8	0	1.64E+15	BAN2P	24	0	2.52E+14	0	1.19E+15
BAN12F1	BAN12	0	5.94E+14	BAN2P	24	0	1.18E+15	0	1.96E+15
BAN12F2	**BAN12**	**0**	**6.09E+14**	**BAN3P**	**24**	**4**	**-1.43E+15**	**7**	**-2.30E+15**
BAN15F	BAN15	0	9.99E+14	BAN2P	24	0	8.13E+14	0	1.21E+15

*__in bold__ possible lamb born from heteropaternal superfecundation.

In this farm, out of five collected rams only one was assigned as candidate father by the software with no mismatches. Moreover, in Table 2.4.8 we reported in bold an interesting result on the only one detected correct twin pair. In this particular case, both BAN12F1 and BAN12F2 lambs were assigned to BAN12 ewe but only

one lamb was included in a trio family while for the other one no candidate fathers were found among the five collected ones. This result could be probable assuming the heteropaternal superfecundation that in Valle del Belìce sheep breed was observed in the past, with very low frequency and not supported by literatures. Again, only four trio (Table 2.4.9) were reported in Farm C (LAO) even though the 70% of offspring-mother pairs were correct, and only two out of seven collected rams were assigned as candidate fathers by the software.

Table 2.4.9. Paternity assignment in Farm C (LAO) for offspring-mother pairs with no mismatching.

Offspring ID	Mother ID	Pair loci mismatch	Pair LOD score	Candidate father ID	Pair loci compared	Pair loci mismatch	Pair LOD score	Trio loci mismatch	Trio LOD score
LAO8F	LAO8	0	9.27E+00	LAO1P	20	0	1.15E+01	0	1.81E+01
LAO9F	LAO9	0	1.57E+01	LAO1P	20	0	9.83E+00	0	1.90E+01
LAO14F	LAO14	0	1.20E+01	LAO4P	20	0	3.36E+00	0	9.57E+00
LAO16F	LAO16	0	8.56E+00	LAO4P	20	0	1.07E+01	0	1.64E+01

Finally, results on Farm D (VIT) are showed in Table 2.4.10 where only one trio was assigned even though the 83.3% of offspring-mother pairs were correct and where, again, we highlighted (in bold) the possible heteropaternal superfecundation for the twin pair.

Table 2.4.10. Paternity assignment in Farm D (VIT) for offspring-mother pairs with no mismatching.

Offspring ID	Mother ID	Pair loci mismatch	Pair LOD score	Candidate father ID	Pair loci compared	Pair loci mismatch	Pair LOD score	Trio loci mismatch	Trio LOD score
VIT09F1	VIT09	0	8.98E+00	VIT02P	24	0	8.48E+00	0	1.45E+01
VIT09F2	**VIT09**	**0**	**8.87E+00**	**VIT05P**	**24**	**3**	**-1.04E+01**	**6**	**-1.74E+01**

***in bold** possible lamb born from heteropaternal superfecundation.

The heteropaternal superfecundation, defined as the fertilization of two or more ova during the same oestrus cycle as a result of more than one coital act from different sires, was previously observed in several poly-ovulatory specie (including sheep), often mob-mated with several rams concurrently, with an overall incidence of 35% (Berry *et al.*, 2020). This peculiarly aspect, differing from bovine sector, considerably complicates the pedigrees recording, deceiving even the most careful

of breeders. For this reason, the paternity test carried out in our study is fundamental for improving the competitiveness of the Valle del Belìce sheep breed.

Parentage test should be performed in flocks belonging to the breeding program to confirm or reconstruct the pedigree of progeny generated through multiple sires' technique, due to the detrimental effect of pedigree errors on inbreeding control, population parameters estimation, and breeding values prediction. DNA-based parentage, therefore, has the potential to assist the breeders to improve pedigree recording and selection accuracy, resulting in an increase in the rate of genetic improvement (Rosa *et al.*, 2013).

2.5 Conclusion

In this study, we set up microsatellite markers panel useful for parentage analysis in Valle del Belìce sheep breed. All the analyzed microsatellite markers showed high polymorphisms in the studied breed and the mean PIC value was similar to other used panels, made it suitable for our purpose. Obtained results highlighted that the most critical point within the analyzed farms was the farmers recorded information which conducted to several pedigree errors. Interesting were the results that evidenced the adoption phenomena and heteropaternal superfecundation in the analyzed farms of Valle del Belìce breed, demonstrating the difficult reliability of pedigrees drawn up on a documentary basis.

Application of DNA-based parentage test has the potential to assist the farmers to improve pedigree recording, resulting in an increase of genetic improvement. All young sires and a subset of young ewes should also be tested every year to evaluate the consistency of pedigree data supplied by the farmers and control the pedigree errors. Further analyses will be conducted in other farms in order to evaluate the recorded information and to provide parentage verification. Concluding, the verification of relationships with molecular markers is fundamental and preparatory for any genetic improvement program in Valle del Belìce sheep breed, and it is important to verify both the paternity and the maternity of the subjects intended for reproduction, to avoid errors deriving from the observed biological and behavioral phenomena.

References

Al-Atiyat, R. M. (2015). The power of 28 microsatellite markers for parentage testing in sheep. Electronic Journal of Biotechnology 18, 116-121.

Azhar, P. M., Chakraborty D., Iqbal Z., Malik A. A., Qaudir A., Asfar A., Bhat I. A. (2018). Microsatellite Markers as a Tool for Characterization of Small Ruminants: A Review. International Journal of Current Microbiology and Applied Sciences, 7(1), 1330-1342.

Berry, D. P., Bohan, A., O'Brien, A. C., Campion, F. C., McHugh, N., & Wall, E. (2020). Heteropaternal superfecundation frequently occurs in multiple-bearing mob-mated sheep. Animal Genetics.

Botstein, D., White, R. L., Skolnick, M., Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. American Journal of Human Genetics 32, 324-331.

Geldermann, H., Pieper, U., & Weber, W. E. (1986). Effect of misidentification on the estimation of breeding value and heritability in cattle. Journal of animal science, 63(6), 1759-1768.

Gonzalez, S. G., & Goddard, P. J. (1998). The provision of supplementary colostrum to newborn lambs: effects on post-natal lamb and ewe behaviour. Applied Animal Behaviour Science, 61(1), 41-50.

Harder, B., Bennewitz, J., Reinsch, N., Mayer, M., & Kalm, E. (2005). Effect of missing sire information on genetic evaluation. Archives Animal Breeding, 48(3), 219-232.

Heaton, M. P., Leymaster, K. A., Kalbfleisch, T. S., Kijas, J. W., Clarke, S. M., McEwan, J., ... & International Sheep Genomics Consortium. (2014). SNPs for parentage testing and traceability in globally diverse breeds of sheep. PloS one, 9(4), e94851.

Hersher, L., Richmond, J. B., & Moore, A. U. (1963). Modifiability of the critical period for the development of maternal behavior in sheep and goats. Behaviour, 311-320.

Kalinowski, S. T., Taper, M. L., Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecolar Ecology 16, 1099-1106.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A salting out method for DNA preparation. Nucleic Acids Res, 16, 1215.

Qanbari, S., Eskandari Nasab, M. P., Osfoori, R., Hagh Nazari, A. (2007). Power of Microsatellite Markers for Analysis of Genetic Variation and Parentage Verification in Sheep. Pakistan Journal of Biological Sciences 10, 1632-1638.

Rosa, A. J. M., Sardina, M. T., Mastrangelo, S., Tolone, M., Portolano, B. (2013). Parentage verification of Valle del Belìce dairy sheep using multiplex microsatellite panel. Small Ruminant Research 113, 62- 65.

Sanders, K., Bennewitz, J., & Kalm, E. (2006). Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. Journal of dairy science, 89(1), 315-321.

Stevanović, J., Stanimirović, Z., Dimitrijević, V., & Maletić, M. (2010). Evaluation of 11 microsatellite loci for their use in paternity testing in Yugoslav Pied cattle (YU Simmental cattle). Czech Journal of Animal Science, 55(6), 221-226.

Tolone, M., Mastrangelo, S., Rosa, A. J. M., & Portolano, B. (2012). Genetic diversity and population structure of Sicilian sheep breeds using microsatellite markers. Small Ruminant Research, 102(1), 18-25.

Vandeputte, M., Mauger, S., & DUPONT-NIVET, M. (2006). An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Molecular Ecology Notes, 6(1), 265-267.

Weller, J. I., Feldmesser, E., Golik, M., Tager-Cohen, I., Domochovsky, R., Alus, O., ... & Ron, M. (2004). Factors affecting incorrect paternity assignment in the Israeli Holstein population. Journal of Dairy Science, 87(8), 2627-2640.

Yilmaz O. (2016). Power of different microsatellite panels for paternity analysis in sheep. Animal Science Papers and Reports 34, 155-164.

Yilmaz, O., Cemal, I., COŞKUN, B., OĞRAK, Y. Z., Ata, N., & Karaca, O. (2018). Comparison of different paternity test panels in sheep. Turkish Journal of Veterinary and Animal Sciences, 42(6), 633-641.

CHAPTER 3

Genetic traceability of Sicilian sheep dairy production Vastedda della Valle del Belìce PDO and Pecorino Siciliano PDO

3.1 Summary

Genetic traceability is based on the identification of both animals and their products through the study of DNA. In Sicily, the Vastedda della Valle del Belìce PDO is the only cheese realized with mono-breed milk (only with Valle del belìce sheep breed), while the Pecorino PDO cheese is made exclusively with Sicilian sheep breeds. With the goal of developing a genetic traceability system for dairy products, the aim of this study was to identify specific markers able to discriminate among the most important Sicilian dairy sheep breeds, in order to detect possible adulteration in Valle del Belìce dairy products. This genetic traceability system was divided in different steps. The method was able to discriminate dairy products realized exclusively with milk from specific breeds (VDB for Vastedda) from dairy products realized with mixed milk with different origins; the discrimination were realized in term of quality (presence or absence of adulteration) and quantity (percentage level of adulteration). The results revealed mean probability to belong to the considered breeds in way directly proportional to the percentage of milk contained in the sample, allowing to quantitatively calculate, by subtraction, the adulteration level of the sample in percentage terms.

3.2 Introduction

In the last three years in Sicily there has been a gradual reduction of the livestock of the sheep sector, of about 7% in general, corresponding to about 50,000 fewer heads (National database – www.vetinfo.it). In the general panorama, the Sicilian autochthonous breeds see their numbers decrease with increases negatively higher than the general average, while the cosmopolite sheep breeds have undergone an increase in terms of heads raised in contrast to the general trend. Consequently, on the sheep's milk market the presence of cosmopolite breeds milk is increasingly felt; it can also be used sometimes improperly, through mixing with milk from local breeds. In this context, in which indigenous breeds suffer greatly from the expansion of cosmopolite breeds, it is important to enhance local productions, and it is possible applying the genetic traceability system, especially to protect mono-breed productions.

Breed genetic traceability is becoming an important issue for the authentication of their products, as there is an increasing interest in marketing mono-breed labeled lines of meat as well as dairy products, which in some cases have obtained the Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI). This interest derives from the fact that a marketing link between breeds and their originated products can contribute to improve breed profitability and sustainability of such farm animal production with significant impact on the rural economy of particular geographic areas and on breed conservation and biodiversity (Russo *et al*., 2007). This means that breed genetic traceability is important both to preserve and valorize some livestock products and breeds.

In Sicilian territory, the only mono-breed cheese realized with sheep milk is the Vastedda della Valle del Belìce PDO (VS), a fresh stretched curd cheese; while the most diffused sheep cheese is the Pecorino Siciliano PDO (PS), realized with milk from several Sicilian breeds. The growing increase of ewes' consistency belonging to the cosmopolite sheep breeds undermines the authenticity of productions such as VS, with easy adulterations if the control takes place exclusively on a documentary basis. For this reason, it is important to protect these productions through a system of genetic traceability, that is based on the identification of both animals and their products through the study of DNA. Breed and species discrimination are interesting to detect fraud and to protect and valorize typical productions. Nowadays, species-specific PCR protocols have shown to be reliable methods to control the authenticity of dairy products at species level (Galimberti *et al*., 2013), because specific target sequences (*e.g*. 12S rRNA, 16S rRNA, cytochrome b, and cox1 gene suggested as DNA barcode) can be detected in matrices containing pools of heterogeneous genomic DNA, such as milk and cheese (Mafra *et al*., 2008; Tortorici *et al*., 2016). On the contrary, the identification of the breed of origin of animal products still remains a challenge in all farms (Fontanesi *et al*., 2011). Two approaches using DNA marker information have been proposed for breed authentication: a probabilistic approach, mainly based on the use of highly variable microsatellites, AFLP markers or dozen of single nucleotide polymorphisms (SNPs) combined with different computational analyses that assign individuals to a particular breed with a certain probability (Negrini *et al*., 2009); a deterministic

approach, based on the use of few breed specific or exclusive markers, whose usefulness is due only to their fixation or absence in the considered breeds and, as a consequence, in the analyzed products (Russo *et al*., 2007). Nowadays, the probabilistic approach is difficult to apply for breed authentication of dairy products, because they usually derive from milk of more than one animal. The recently developed genomic technologies, such as medium and high-density single nucleotide polymorphism (SNP) arrays, represents important tools that can be used for these purposes. Dense genome-wide data is valuable but is relatively costly and time-consuming or computationally expensive to analyze. Therefore, it is often desirable to reduce the number of markers according to their information content, in order to create reduced panels for population genetic analysis (Paschou *et al*., 2007). Several statistical methods were used to determine which genetic markers contain the most information to discriminate among populations (Wilkinson *et al*., 2011), such as the combined approach of Principal Component Analysis (PCA) and Random Forest (RF) (Bertolini *et al*., 2015), multivariate canonical discriminant analysis (Dimauro *et al*., 2013), the statistic delta (Shriver *et al*., 1997), Wright's F_{ST} (Bowcock *et al*., 1994). While all these methodologies yielded reduced marker panels useful for breed identification, the power of assignment varied among analysis methods.

With the goal of developing a genetic traceability system for dairy products, the aim of this study was to identify specific markers able to discriminate among the most important Sicilian dairy sheep breeds, in order to detect possible adulteration in Valle del Belìce dairy products.

This genetic traceability system was divided in different steps:

i. in the first step we worked at individual DNA level trying to develop a discrimination rule able to differentiate, in Step-1a, TARGET breeds, *i.e.* Sicilian sheep breeds Valle del Belìce [VDB], Comisana [COM], and other Sicilian sheep [OTHERS] *vs.* FOREIGN breed *i.e.* not Sicilian breed such as Lacaune; and, in Step-1b, Valle del Belìce [VDB] *vs.* the other Sicilian breeds [COM+OTHERS];
ii. in the second step, we worked at pool DNA level applying the discrimination rule to differentiate, in Step-2a, known quantity of DNA

from TARGET breeds *vs.* FOREIGN one, in Step-2b, known quantity of DNA from VDB breed *vs.* [COM+OTHERS] ones, and in Step-2c, known quantity of DNA from VDB breed *vs.* FOREIGN one;

iii. in the third step, we verified and validated the discrimination rule, working at cheeses DNA level and trying to differentiate, in Step-3a, TARGET breeds *vs.* FOREIGN one, in two different cheeses (*Primo sale* and *Stagionato 4 mesi*), and, in Step-3b, VDB breed *vs.* [COM+OTHERS] ones, in the Vastedda della Valle del Belìce PDO cheese. The experimental cheeses were realized with known mixed milk quantity.

3.3 Materials and methods

Samples collection

From 2018, we started to screen 58 ovine Sicilian flocks which delivered their milk to a diary industry MiSE project partner, to have a global idea on the farms' distribution and availability of milk. After the screening, we selected eight flocks to collect TARGET blood samples (VDB, COM, OTHERS) and two additional flocks to collect FOREIGN blood samples (Lacaune; Fig.3.3.1).

Fig. 3.3.1 – Geographic representation of sampled flocks.

We collected a total of 420 individual blood samples belonging to:

iv. 300 TARGET individuals:
    - 168 Valle del Belìce breed (VDB);
    - 63 Comisana breed (COM);
    - 69 Sicilian sheep (OTHERS);

v. 120 FOREIGN individuals:
    - 120 Lacaune (LAC)

From the above reported flocks, we collected bulk milk samples and we used Combifoss FT+ system (Foss Italia srl, Fig. 3.3.2) for qualitative analyses, to monitor the milk quality during cheesemaking and reduce manual errors.

Fig.3.3.2 – Combifoss FT+ used for milk quality analysis.

DNA extractions

Individual whole blood samples (n=420) were collected from jugular vein using vacutainer tubes containing EDTA as anticoagulant. The DNA extraction was performed starting from 2 ml of blood with a salting-out method (Miller *et al.* 1988). All the 420 individuals were genotyped with Illumina BeadChip array and analyzed with statistical approach to develop the discrimination rule able to distinguish among TARGET *vs.* FOREIGN breeds (Step-1a of the MiSE project), and VDB *vs.* COM+OTHER breeds (Step-1b).

The DNA was extracted also from experimental cheeses using the QIAamp DNA Investigator Kit (QIAGEN; Fig. 3.3.3).

Fig. 3.3.3 – QIAamp DNA Investigator Kit, QIAGEN, used for cheese DNA extraction.

DNA quantification.

The DNA samples were quantified with NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA, Fig. 3.3.4), as already described in Chapter 2.

After quantification, the DNA samples were normalized to the working concentration of 50 ng/µl and stored at 4 °C for subsequent analyses.

Fig. 3.3.4. – NanoDrop ND-1000 spectrophotometer.

DNA pools preparation

Starting from the DNA samples belonging to the 420 individuals (300 TARGET + 120 FOREIGN), different DNA pools were prepared to apply the discrimination rule (Step-2a, -2b, and -2c of the MiSE project). For this reason, each pool contained known percentage of DNA of different breeds, as reported in Table 3.3.1.

Table 3.3.1 – DNA pools containing known percentages (%) of DNA belonging to the different breeds.

Step-2a		Step-2b		Step-2c	
TARGET **(% DNA)**	*FOREIGN* (% DNA)	*VDB* (% DNA)	*COM + OTHERS* (% DNA)	*VDB* (% DNA)	*FOREIGN* (% DNA)
100	0	100	0	100	0
80	20	80	20	80	20
60	40	60	40	60	40
50	50	50	50	50	50
40	60	40	60	40	60
20	80	20	80	20	80
0	100	0	100	0	100

All the 21 DNA pools were analyzed on BeadChip array, as the individual samples, in 3 replicates loading on different BeadChip arrays.

Experimental cheeses preparation

As already mentioned above, to verify and validate the genetic traceability system, we realized different experimental cheeses containing different percentages of milk belonging to different breeds. The milk used for these dairy products belonged to the flocks which delivered their milk to the dairy industry MiSE project partner; therefore, the milk of the collected and genotyped individuals was mixed with milk from other animals from the same breeds.

Using the routinely dairy production technique, three types of cheeses were prepared: the primo sale (PS, Fig.3.3.5a),the stagionato 4 mesi (S4M, Fig. 3.3.5b) and the Vastedda della Valle del Belìce PDO production (VS, Fig.3.3.5c).

Fig. 3.3.5 – Experimental cheeses

a. Primo sale **b. Stagionato 4 mesi** **c. Vastedda**

For each cheese type, known percentages of milk from TARGET, FOREIGN, VDB, and COM+OTHERS breeds were mixed as reported in Table 3.3.2. For PS and S4M productions, milk from TARGET breeds was used with concentrations

that were gradually lower than the ones of FOREIGN breeds. In the same way for VS production milk from VDB was used with concentrations that were gradually lower than the once of COM+OTHERS.

The experimental cheeses (n = 21), after their respective different ripening times, were gradually stored at -80°C, as they were ready, waiting for DNA extraction.

Table 3.3.2 – Experimental cheeses.

Step-3a				Step-3b	
PS		**S4M**		**VS**	
TARGET **(% milk)**	*FOREIGN* (% milk)	*TARGET* (% milk)	*FOREIGN* (% milk)	*VDB* (% milk)	*COM+OTHERS* (% milk)
100	0	100	0	100	0
95	5	95	5	95	5
90	10	90	10	90	10
85	15	85	15	85	15
80	20	80	20	80	20
60	40	60	40	60	40
50	50	50	50	50	50

From each experimental cheese (n= 21), three samples were collected for DNA extraction and genotyping analysis on Illumina BeadChip arrays.

Genotyping Analysis

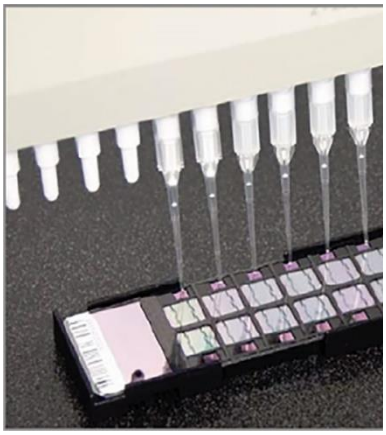
The genotyping was performed on the OvineSNP50 Genotyping BeadChip array (Fig. 3.3.6a), containing 54,241 SNPs (Illumina).

After preparation, the BeadCHip (Fig. 3.3.6a) were loaded and scanned on Illumina Hi-Scan SQ (Fig. 3.3.6b) platform.

Fig. 3.3.6. Illumina OvineSNP50 Genotyping BeadChip and Hi-Scan SQ used for sample genotyping.

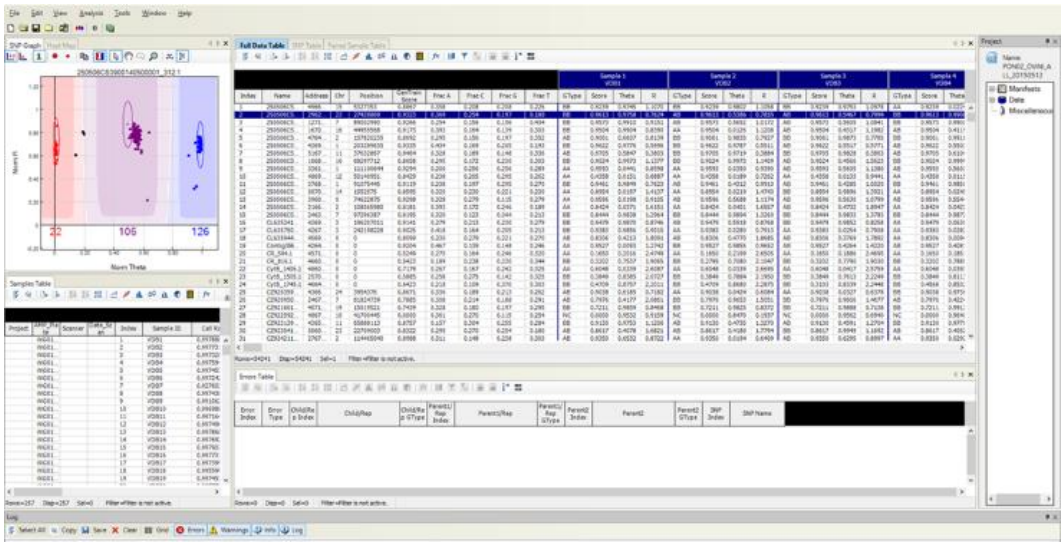
a. OvineSNP50 Genotyping BeadChip

b. Illumina Hi-Scan SQ



The GenomeStudio v.2.0 software (Fig.3.3.7) was used to transform the intensity data and the mapping data from the arrays in genotyping data contained in two different files: the **.map* file containing 54,241 SNPs genomic coordinates (Chromosome, SNP name, position in bp) and the **.ped* file containing for each row the animal ID following by genotypes of all SNPs.

Fig. 3.3.7. Genotyping result showed on GenomeStudio software.



Bioinformatic and statistical analysis

Statistical analyses started from a dataset with 420 individuals and 54,241 SNPs. From this dataset, the *m1* subset with SNPs with the same genotypes in TARGET breeds, and the *m2* subset with SNPs with the same genotypes in VDB breed, were extracted. More specifically, the genotypes of all the SNPs were transformed in *theta* values and used in a calibration step with Penalized Logistic Regression (PLR) to define the discrimination rule. The PLR was applied: in Step-1a, to a final dataset of *n1* x *m1* dimensions, where *n1* was represented by 420 individuals (TARGET+FOREIGN) and *m1* was the *theta* values of the selected SNPs; in Step-1b, to a final dataset of *n2* x *m2* dimensions, where *n2* was represented by 300 individuals (VDB + [COM + OTHERS]) and *m2* was the *theta* values of the selected SNPs. Through the calibration step, the β coefficients of *m1* and *m2* SNPs were estimated and only the informative β coefficients (*i.e.* ≠ 0) were used to define the discrimination rule based on an objective function of probability that any other analyzed individual belong to TARGET breeds (Step-1a) and VDB breed (Step-2a). The discrimination rule was then applied to DNA pools and finally validated on experimental cheese samples. The *theta* values are highly correlated with genotypes and even when the genotypes are not assigned, the *theta* values are different from zero value.

In particular, a homozygous individual for B allele (BB genotype) presents a *theta* value closer to 1, a homozygous individual for A allele (AA genotype) presents a *theta* value closer to 0, and the heterozygous individual (AB genotype) presents a *theta* value closer to 0.5. The use of the *theta* values instead of genotypes allowed us to work with a continuous variable, more informative and without missing values.

To build the discrimination rule a Penalized Logistic Regression (PLR) was applied to the dataset. The PLR is a classification method that allows modeling the probability of belonging to a specific class based on *m* explanatory variables, even when *m*>>*n*, where *n* is the sample dimensions.

So, supposing that the *y* response variable has values in $G = \{1,2\}$, we denote $y_i = I(g_i = 1)$, were $I(\cdot)$ is the indicator function. Therefore, we have:

$$\Pr(G = 2|X = x) = \frac{e^{x\beta}}{1 + e^{x\beta'}},$$

where X is the matrix of the model with n x m dimension, and β is the m-dimensional vector of coefficients. This equation can also be rewritten either in the form of logistics or log-odds in the following form:

$$\log\frac{\Pr(G=2|X=x)}{\Pr(G=1|X=x)}=X\boldsymbol{\beta},$$

The target function for the PLR is:

$$\min_{\boldsymbol{\beta}}-\left[\frac{1}{n}\sum_{i=1}^{n}y_i\cdot x_i^T\boldsymbol{\beta}-\log\left(1+e^{x_i^T\boldsymbol{\beta}}\right)\right]+\lambda\cdot J(\boldsymbol{\beta}),$$

where $J(\beta)$ is the penalty term for the β coefficients, that allows to negatively weigh models with many variables and to reward more parsimonious models, λ is the regularization parameter. To our aim we use the L_1 penalty, where

$$J(\beta)=\sum_{j=1}^{m}|\beta_j|.$$

The penalty L_1 or Least Absolute Shrinkage and Selection Operator (LASSO) was proposed by Tibshirani (1996).

When we will obtain the β coefficients, minimizing the target function, we will be able to predict the probability of belonging at the target breed for any other analyzed individual. In this way, the discrimination rule can be used both for DNA pools (Step-2a, -2b, and -2c) and experimental cheeses, just to predict the probability of belonging to TARGET breeds (Step-3a) or VDB breed (Step-3b).

To establish the best threshold, useful to confirm that each dairy products belong to TARGET or VDB breeds, the ROC (Receiving Operator Curve) analysis was used, thought a simulation study in which the pure sample was the reference standard (Gold Standard; in our case 100% TARGET for PS and S4M, or 100% VDB for VS), and the test represents the probability of the analyzed sample to belong at considered breed (or breeds). The scenarios presented different impurity percentages simulating the different milk percentages used to obtain the experimental cheeses (Step-3a and -3b). For each scenario, 1,000 cheese samples were simulated considering the breeds used to set up the discrimination rule, reflecting the different milk percentages used in each experimental cheese (Table 3.2.3). The ROC threshold values for each cheese (PS, S4M, and VS) and for each *t* percentage milk mixture, were compared with the probability values obtained by

statistical analysis. If each probability value is higher than corresponding threshold value, we can confirm the statistical result *i.e.*, the sample belonging to the TARGET or to the VDB breeds.

3.4 Results and discussion

The used statistical methods allowed to define the discrimination rule. The PLR was applied to the dataset with n x m dimensions, with $n=n1+n2$, where $n1$ was the target breed (TARGET breeds for Step-2a or VDB for Step-2b and -2c), and $n2$ was the other breed (FOREIGN for Step-2a and -2c or COM+OTHERS for the Step-2b).

The obtained results allowed:

i. to verify the entire procedure comparing with the theorical probability values obtained on DNA pools;
ii. to validate the theorical results with the experimental cheeses.

The applied statistical method allowed to identify the subsets $m1$ and $m2$, *i.e.* SNPs with homozygous genotype for TARGET and VDB breeds, which resulted in 272 and 1308 SNPs, respectively. Furthermore, the discriminant SNPs in the comparison TARGET vs FOREIGN, VDB vs COM+OTHER, and VDB vs FOREIGN were identified and consisted in 14, 688, and 299 SNPs, respectively.

The statistical method allowed to estimate the probability of belonging to TARGET or VDB breeds, both for DNA pools and cheeses samples. The mean probability values, considering the replicates, of belonging to TARGET or VDB breeds, and the assigned groups were reported in the following Tables. The mean probability of belong to TARGET (or VDB for steps 2b and 2c) was directly proportional to DNA percentage indicate in the first column, almost perfectly reflecting what was present in each DNA pool (Table 3.3.1). Moreover, this aspect is relevant because it makes us understand how the applied method, in addition to returning a certain class of assignment (Assigned group, last column), can report more precise information about the quantity of DNA present in the sample belonging to the target.

Table 3.4.5. – Mean probability values of each DNA pool to belong to TARGET breeds in Step-2a.

Step-2a			
TARGET **(% di DNA)**	***FOREIGN*** **(% di DNA)**	**Mean probability to belong to *TARGET***	**Assigned group**
100	0	0,990	*TARGET*
80	20	0,830	*TARGET*
60	40	0,611	*TARGET*
50	**50**	**0,520**	***TARGET***
40	60	0,389	*FOREIGN*
20	80	0,170	*FOREIGN*
0	100	0,010	*FOREIGN*

***** the sample which may undergo incorrect assignment classes due to approximation errors is shown **in bold**.

The results obtained for Step-2b (Table 3.4.6) were quite satisfactory even though the DNA pool containing 50% of VDB and COM+OTHERS was assigned to the COM+OTHERS group. This result can be considered reliable considering that the comparison was made using only DNA from Sicilian breeds which shared not only the same production area but also part of their genetic heritage/history (Portolano, 1987). Also in this case, if we do not limit ourselves to observing only the assignment class returned by the approach, but verify the mean probability to belong to VDB value in absolute numerical terms, we can deduce important information about the composition of the analyzed sample. For example, even if the sample contains the 50% of VDB DNA, the mean probability to belong to VDB is however approximately close to this value, determining a different assignment class only by way of approximation.

Table 3.4.6 – Mean probability values of DNA pool to belong to VDB breed in Step-2b.

Step-2b			
VDB **(% di DNA)**	***COM + OTHERS*** **(% di DNA)**	**Mean probability to belong to VDB**	**Assigned group**
100	0	0,993	*VDB*
80	20	0,794	*VDB*
60	40	0,592	*VDB*
50	**50**	**0,480**	***COM + OTHERS***
40	60	0,408	*COM + OTHERS*
20	80	0,206	*COM + OTHERS*
0	100	0,007	*COM + OTHERS*

***** the sample which may undergo incorrect assignment classes due to approximation errors is shown **in bold**.

Moreover, the results obtained for Step_2c were satisfactory (Table 3.4.7), almost perfectly reflecting the DNA percentages within the pools. The values of mean probability of assignment were higher than in Step-2a probably because the VDB and FOREIGN breeds are genetically very distant. Anyway, as reported in the precedent two cases, the assigned group in the samples containing exactly 50% of DNA for each breed group considered, can be undergo incorrect assignment classes due to approximation errors (considering 2-4% range observed); therefore it is not correct to read only the data relating to the assignment class, but to accumulate this information with the numerical information relating to mean probability to belong to VDB (or TARGET).

Table 3.4.7 – Mean probability values of DNA pool to belong to VDB breed in Step-2b.

Step-2c			
VDB **(% di DNA)**	*FOREIGN* (% di DNA)	**Mean probability to belong to VDB**	**Assigned group**
100	0	0,997	*VDB*
80	20	0,853	*VDB*
60	40	0,627	*VDB*
50	**50**	**0,540**	***VDB***
40	60	0,373	*FOREIGN*
20	80	0,147	*FOREIGN*
0	100	0,003	*FOREIGN*

***** the sample which may undergo incorrect assignment classes due to approximation errors is shown **in bold**.

Finally, the results of Step-3a (Table 3.4.8) and 3b (Table 3.4.9) were satisfactory considering that the experimental cheeses were prepared with milk mixtures belonging to different animals, with an amount of DNA extracted from each cheese on depending of the somatic cell content of the milk. Considering the simulation, the exceeding of the threshold values made it possible to confirm (with YES or NOT) if each dairy product belonged to TARGET or VDB breeds.

As expected from previously analysis realized with DNA pools, also in the cheese samples the methods used reveals two different kind of information:

- The qualitative one, indicated in the columns “Assignment group” and “Confirmation of belonging to TARGET” (or VDB for step-3b), resulting from a mathematical assignment that occurs when the mean probability to belong to the breed (or breeds) is higher or lower than the threshold value (50%);

- The quantitative one, indicated in the column "Mean probability to belong to TARGET" (or VDB for step-3b), more interesting than the precedent, as it can also give us information about the counterfeiting level of a particular product.

Table 3.4.8. – Mean probability values of experimental cheeses to belong to TARGET breeds in Step-3a.

Step-3a				
Primo sale and Stagionato 4 mesi				
TARGET **(% di latte)**	*FOREIGN* (% di latte)	**Mean probability to belong to *TARGET***	**Assignment group**	**Confirmation of belonging to TARGET**
100	0	0,949	*TARGET*	YES
90	10	0,875	*TARGET*	YES
80	20	0,793	*TARGET*	YES
70	30	0,702	*TARGET*	YES
60	40	0,603	*TARGET*	YES
50	**50**	**0,494**	***FOREIGN***	**NO**

***** the sample which may undergo incorrect assignment classes due to approximation errors is shown **in bold**.

Table 3.4.9. – Mean probability value of experimental cheeses to belong to VDB breed in Step-3b.

Step-3b				
Vastedda				
VDB **(% di latte)**	*COM+OTHERS* (% di latte)	**Mean probability to belong at *VDB***	**Assignment GROUP**	**Confirmation of belonging to VDB**
100	0	0,912	*VDB*	YES
90	10	0,891	*VDB*	YES
80	20	0,768	*VDB*	YES
70	30	0,708	*VDB*	YES
60	40	0,557	*VDB*	YES
50	**50**	**0,380**	***COM+OTHERS***	**NO**

***** the sample which may undergo incorrect assignment classes due to approximation errors is shown **in bold**.

The results show with good approximation a numerical correspondence between the percentage of breed (VDB or TARGET) milk and the mean probability to belong to the same breed (or breeds). This important correspondence can find application in the dairy industrial process and in the counterfeiting of livestock productions sector, where the phenomenon of mixing local milk with foreign milk (often reconstituted on site) is increasingly, as reported by press. This mechanism allows some unscrupulous industrialists (few actually) to dilute local milk with low-cost foreign milk, but at the same time greatly harms producers. In this way our results can represent a first deterrent against frauds and an important tool for the valorization on first of Valle del Belìce breed: the authentication of cheese obtained

from Valle del Belìce milk only. Moreover, the genetic traceability approach applied in this work can be used by counterfeiting authorities (*e.g.* ICQRF) to discriminate the level of milk mixing in the principal Sicilian dairy sheep products.

3.4 Conclusion

Conventional traceability system based on paper documents could be counterfeited while genetic traceability is based on the identification of both animal and their products through the study of DNA and therefore is more reliable. In this study, a genetic traceability system was defined using SNPs information as input data. The method was able to discriminate dairy products realized exclusively with milk from specific breeds (VDB for Vastedda or TARGET breeds for Primo sale or Stagionato 4 mesi) from dairy products realized with mixed milk with different origins (FOREIGN or COM+OTHERS). The possibility of certifying origin and identity of dairy products, through breed characterization, could provide the development of marginal areas in which these products are made, as well as, more in general, the conservation of Sicilian local breeds. The present study can represent a first deterrent against fraud and an important tool for the valorization of Valle del Belìce breed.

References

Gaglio, R., Cruciata, M., Scatassa, M. L., Tolone, M., Mancuso, I., Cardamone, C., ... & Settanni, L. (2019). Influence of the early bacterial biofilms developed on vats made with seven wood types on PDO Vastedda della valle del Belìce cheese characteristics. International journal of food microbiology, 291, 91-103.

Gaglio, R., Francesca, N., Di Gerlando, R., Cruciata, M., Guarcello, R., Portolano, B., ... & Settanni, L. (2014). Identification, typing and investigation of the dairy characteristics of lactic acid bacteria isolated from "Vastedda della valle del Belìce" cheeses. Dairy Science & Technology, 94(2), 157-180.

Scatassa, M. L., Gaglio, R., Macaluso, G., Francesca, N., Randazzo, W., Cardamone, C., ... & Settanni, L. (2015). Transfer, composition and technological characterization of the lactic acid bacterial populations of the wooden vats used to produce traditional stretched cheeses. Food microbiology, 52, 31-41.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A salting out method for DNA preparation. Nucleic Acids Res, 16, 1215.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, 81(3), 559-575.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

Portolano, B., Chiodi, M., Di Gerlando, R., Sottile, G., Mastrangelo, S., Sardina, M., & Tolone, M. (2016). Metodo per la tracciabilità/autenticazione di prodotti lattiero-caseari mono-razza.

Portolano, B., Todaro, M., Gioccone, P., Alicata, M. L., & Bonanno, A. (1996). Effect of supplementary feeding on milk yield and quality in Valle del Belìce ewes [Sicily]. Zootecnica e Nutrizione Animale (Italy).

Pianaccioli, L., Acciaioli, A., Malvezzi, R., & Giustini, L. (2007). Effect of season on characteristics of Pecorino cheese and Ricotta of Pistoiese Appennine: First results. Italian Journal of Animal Science, 6(sup1), 585-587.

Boselli, C., Bianchi, L, Delfini, P., Campagna, M., Proietti, A., Bitonti, G., ... & Amatiste, S. (2016). Composizione e qualità tecnologica del latte ovino di razza sopravissana prodotto in 6 allevamenti della regione lazio. xxii, 157.

Duranti, E., Bolla, P., Caroli, A., Chiofalo, L., Stasio, L. D., Fortina, R., ... & Zullo, A. (2003). Problems concerning ovine milk clotting aptitude. Italian Journal of Animal Science, 2(1), 89-95.

CHAPTER 4

Genome-Wide association studies for phenotypic and productive traits in the Valle del Belìce dairy sheep breed

4.1 Genome-wide association study reveals the locus responsible for microtia in Valle del Belìce sheep breed

4.1.1 Summary

Microtia is a congenital deformity of the outer ear with phenotypes varying from a small auricle to total absence (*anotia*). The genetic basis is still poorly understood, and very few studies have been performed in sheep. Valle del Belìce sheep is a breed showing *microtia*. The aim of this study was to identify the potential genomic regions involved in *microtia* in sheep. A total of 40 individuals, 20 with *microtia* and 20 normal, were genotyped with the *Illumina OvineSNP50 BeadChip*. The comparison among the results from a genome-wide association study, Fisher's exact test and F_{ST} analysis revealed a single strong association signal: *rs419889303* on chromosome 1, located within intron 3 of the CLRN1 gene. Our study suggests for the first time that this novel candidate gene is responsible for *microtia* in sheep. Additional analysis based on the sequencing would help confirm our findings and allow for the proposal of a precise genetic basis for microtia in sheep.

4.1.2 Introduction

The external ear is part of the auditory system and plays a vital role in collecting sound as the first step in hearing (Ren *et al.* 2011). *Microtia* is a congenital deformity of the outer ear with phenotypes varying from a small auricle to total absence (*anotia*) (Alasti *et al.* 2008). This hereditary underdevelopment of the ear occurs in many species. In fact, diversity in external ear size is apparent in humans as well as in livestock species. It has been described in rabbits (Tanchev *et al.* 2011), swine (Ren *et al.* 2011) and sheep (Jawasreh *et al.* 2016). According to Elias & Bennet (1992), *microtia* is more common in ovine in comparison with other ruminants. Although both genetic and environmental components have been implicated in *microtia*, the genetic causes of this innate disorder are poorly understood (Qiao *et al.* 2015). Local animal genetic resources can be investigated to disclose genetic factors affecting particular phenotypes (Schiavo *et al.* 2018). Although there are large differences in the mechanisms determining ear-size diversity and ear disease (e.g. *microtia*), genetic research on ovine ear size can contribute to understanding these mechanisms. The Valle del Belìce sheep is a dairy

breed from southern Italy showing *microtia*. There are two distinct phenotypes in this breed —normal and small (short-eared or microtia) (Fig. 4.1.2.1)— therefore, the breed provides a model for investigating the genetic basis of *microtia*.

Fig. 4.1.2.1 – Example of Valle del Belìce subject with *microtia* (A, case) and normal (B, control).

The development of single nucleotide polymorphism (SNP) panels has provided the opportunity to narrow down the confidence interval and to identify novel associated variants via genome-wide association studies (GWASes) (Hidalgo *et al.* 2016). However, very few studies have been performed to identify the related genes and mutations on a genome-wide scale for ear size in sheep (Jawasreh *et al.* 2016; Gao *et al.* 2018). In the present study, to identify the potential causative genes involved in *microtia* in sheep, we performed a case–control GWAS and a genome F_{ST} analysis in the Valle del Belìce breed. Blood samples were collected from sheep at six different farms.

4.1.3 Materials and methods

A total of 40 individuals (all females) with unknown relationships, 20 with microtia (short-eared; case) and 20 normal (control), were selected. The animals were genotyped with the *Illumina OvineSNP50 BeadChip*. Chromosomal coordinates for each SNP were obtained from the latest release of the ovine genome sequence assembly: *Oar_v4.0*. We used PLINK v1.07 software (Purcell *et al.* 2007) to perform quality control. Only SNPs located on autosomes were considered for further analyses. Moreover, the following filtering parameters were adopted to

exclude certain loci and animals and to generate the pruned input file: SNPs with a call rate less than 90%, minor allele frequency less than 1%, Hardy–Weinberg equilibrium $P < 0.001$ and animals with more than 2% of missing genotypes were removed. A total of 45 836 autosomal SNPs were used for further analyses. The data have been submitted at https://www.animalgenome.org/share/tmp/QAR1527161498.rar.gz. All animals had high quality genotyping and were included in the analysis. PLINK v1.07 (Purcell *et al.* 2007) was also used to calculate pairwise identity-by-state distances between the two case–control groups, graphically represented by multidimensional scaling (MDS) analysis.

The GWAS was carried out using the filtered SNPs by applying the univariate case–control model of the SNPASSOC R package (Gonzalez *et al.* 2007). In addition, Fisher's exact test was conducted using PLINK v1.07 (Purcell *et al.* 2007) to perform a standard case–control association analysis. Bonferroni correction was applied to account for multiple testing ($P < 0.05$, genome-wide).

An F_{ST} case–control analysis was also performed using PLINK 1.9 (Chang *et al.* 2015). The top 0.9999 SNPs of the percentile distribution were considered the most divergent across the comparison. Moreover, we scanned the genome to identify runs of homozygosity (ROH) islands shared by all animals with microtia, following the parameters reported by Mastrangelo *et al.* (2018). Genomic regions showing significant results were further explored to identify candidate genes underlying the loci

4.1.4 Result and discussion

The obtained MDS plot showed some structures not well defined and did not clearly separate the two Valle del Belìce groups (Fig. 4.1.4.1).

Fig. 4.1.4.1. – MDS plot of studied subjects.

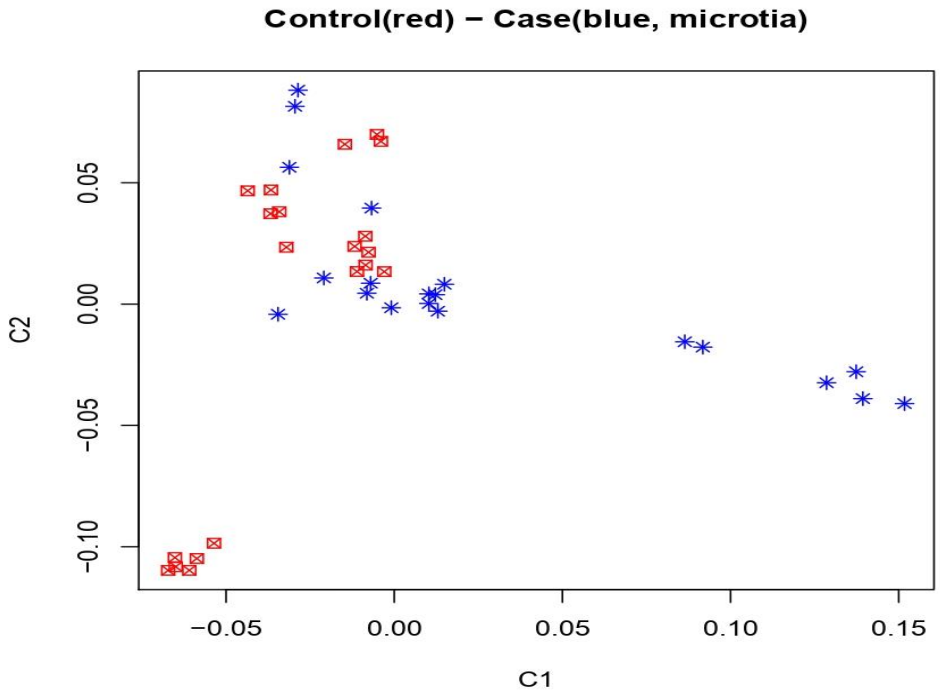


Table 4.1.4.1. - Suggestively significant (0.05 < P < 0.10; Bonferroni corrected) and significant (P < 0.05) SNPs obtained in the GWAS and Fisher's exact test respectively, and the top 0.9999 SNPs detected in the FST analysis in the Valle del Belìce sheep breed.

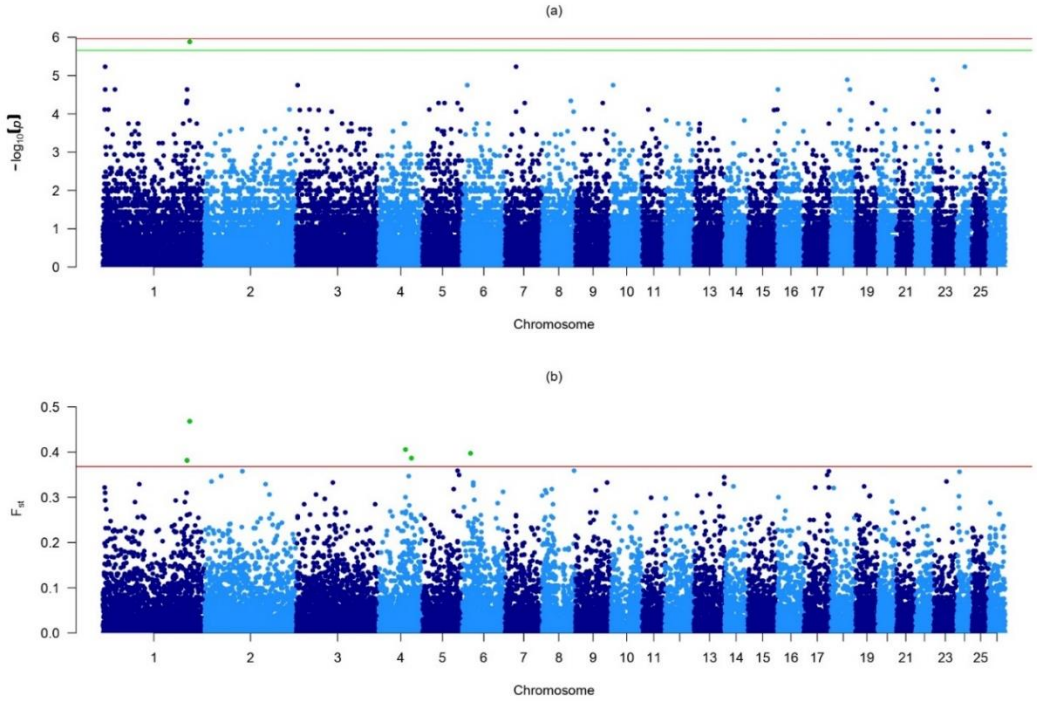
CHR	SNP	Posizion bp	GWAS *P*-value	Fisher test	F_{ST} value	Gene
1	*rs409714664*	228 034 242			0.382	
1	***rs419889303***	235 105 286	2.181691e-07	0.03123	0.468	*CLRN1*
4	*rs418996372*	71 895 715			0.406	
4	*rs399891450*	88 874 748			0.387	
6	*rs417253163*	23 207 265			0.398	

bp, base pair; GWAS, genome-wide association study; OAR, Ovis aries chromosome number; SNP, single nucleotide polymorphism. The annotated gene is reported only for the SNP (in bold) for which both P and FST values exceeded the indicated threshold.

No SNP reached the Bonferroni corrected (P < 0.05) genome-wide significance threshold (Pnominal value = 1.090846e-06). However, at the P < 0.1 Bonferroni corrected threshold (Pnominal value = 2.181691e-06) a single SNP, *rs419889303*,

located at 235,105,286 bp on *Ovies aries* chromosome 1 (OAR1), exceeded the limit of suggestive significance with a P-value of 1.311062e-06 (Table 4.1.4.1). The Manhattan plot produced in this GWAS is shown in Fig. 4.1.4.2a. The Fisher's exact test case–control association analysis showed only one statistically genome-wide significant marker (*rs419889303*) (P-value = 0.031), the same as the one identified above as suggestively significant in the GWAS (Table 4.3.1). No other SNP had an association with significance greater than P < 0.1, and one SNP (*rs418996372*) on OAR4 had an association with significance greater than P < 0.15 (P-value = 0.123).

Fig. 4.1.4.2. – (a) Manhattan plot of the P-values in the genome-wide association study analysis for microtia. The horizontal lines represent the genomewide significance (red; P < 0.05) and suggestively significant (green; P < 0.10) single nucleotide polymorphism (SNPs). (b) Manhattan plot of the F_{ST} showing the top 0.9999 SNPs above the red line (FST = 0.368).



In the F_{ST} case–control analysis, a total of five SNPs were above the selected threshold (F_{ST} = 0.368) (Table 4.3.1): two on OAR1, two on OAR4 and one on OAR6. The Manhattan plot of the F_{ST} analysis is shown in Fig. 4.1.4.2b. Finally, the scan of the genome failed to identify private ROH islands shared by all animals with microtia. We identified a common ROH island on OAR 2 (from 133,409,662

to 134,201,068 bp) shared by five individuals with microtia (25%) but at the same time reported by a normal individual.

Combining the results obtained with the GWAS, Fisher's exact test and F_{ST} analysis, we revealed a single strong association signal, *rs419889303*, associated with microtia. This variant, g.235105286G>A, displayed markedly different genotypic frequencies between groups (Table 4.1.4.2). The animals with microtia had a high frequency of GG homozygotes compared with normal animals, which were mostly heterozygous or AA homozygotes.

Table 4.1.4.2 - Genotype frequencies of the best associated SNP rs419889303 in cases and controls groups.

Group	AA	AG	GG
Case (*microtia*)	-	0.25	0.75
Control	0.40	0.55	0.05

This newly detected association signal in Valle del Belìce breed was located within intron 3 of the Clarin 1 (CLRN1) gene mapped on OAR1. CLRN1 is the causative gene in Usher syndrome type 3A, an autosomal recessive disorder characterized by progressive vision and hearing loss (Phillips *et al.* 2013). The CLRN1 gene provides information for making a protein called clarin 1. This protein is probably involved in normal hearing and vision. It is a four-pass membrane domain protein with homology with the tetraspanin family of proteins (Adato *et al.* 2002), which include connexins and Claudins, known to cause deafness in humans when mutated (Duman & Tekin 2012). It has been found in several areas of the body, including sensory cells in the inner ear called hair cells (Geller *et al.* 2009). These cells help transmit sound and motion signals to the brain. This protein is also active in the retina, which is the light-sensing tissue that lines the back of the eye. Although the function of clarin 1 has not been determined, studies suggest that it plays a role in communication between nerve cells (neurons) in the inner ear and in the retina (Zallocchi *et al.* 2009). In the Valle del Belìce breed, the *microtia* would seem to be not associated with another defect and appears to be simply inherited, as also

reported for Awassi sheep (Jawasreh *et al.* 2016). Several interviewed breeders believe that sheep with *microtia* have higher milk production. A previous GWAS for microtia in Awassi sheep revealed a statistically significant association with a SNP mapped on OAR23 (OAR_23_34647499.1). This marker is adjacent to the gene encoding transcription factor GATA-6 (Jawasreh *et al.* 2016). Recently, Gao *et al.* (2018), in a GWAS using animals with different ear size (from 102.55 to 179.95 cm2), showed a strong association with SNP *rs402740419* (on OAR23), located within the DCC gene, and two additional SNPs located within two genes—PTPRD (on OAR2) and SOX5 (on OAR3)—in the Duolang Chinese native sheep breed. The authors did not observe association with the GATA6 gene. The two studies agreed only on the chromosome (OAR23) that hosted the mutations significantly associated with ear development, with a distance of about 18 Mb between them. When we compared our results with those obtained in the aforementioned studies, differences among the identified candidate genes were evident. A possible reason for the lack of correspondence may be the different breeds used (*e.g.* Awassi fat-tailed *vs.* Valle del Belìce thin-tailed). Therefore, it is possible to hypothesize that the genetic basis of variable ear size in sheep is different among breeds and that the sharing of the mutation is more likely among sheep breeds with low genetic differentiations. Moreover, it should be highlighted that the different investigated phenotypes among studies—anotia (Jawasreh *et al.* 2016), ears with different sizes (Gao *et al.* 2018) and microtia (this study)— could also be responsible for these discrepancies. Because the genetic basis of variable ear size is still not understood, further studies are needed to evaluate if *microtia* derives from different genetic factors in other sheep breeds. However, although there are large differences in the mechanisms determining ear-size diversity and ear disease (Zhang *et al.* 2014), genetic research on sheep can contribute to the understanding of human ear development and abnormalities (Jawasreh *et al.* 2016; Gao *et al.* 2018). Therefore, it can be hypothesized that variants of the CLRN1 gene could result in an underdevelopment of the ear. Despite the limited number of genotyped case–control animals, the low-density array used, the unknown inheritance model of microtia and the effect of population heterogeneity, the combination of different approaches was useful for providing insights into the genetic basis of ear size,

excluding possible genomic regions derived by factors that could not be better managed in our study (Schiavo *et al.* 2018). An increase in the number of genotyped sheep and using the new ovine high-density SNP chip can be expected to further enhance the power to identify additional SNPs and genes related to this congenital deformity. In conclusion, our study suggests for the first time that a novel candidate gene (CLRN1) is responsible for *microtia* in sheep, though the precise function of this gene is currently not well established and has to be studied in detail. Additional analysis based on sequencing would help confirm our findings and allow the proposal of a precise biological basis for *microtia*.

References

Adato, A., Vreugde, S., Joensuu, T., Avidan, N., Hamalainen, R., Belenkiy, O., ... & Lancet, D. (2002). USH3A transcripts encode clarin-1, a four-transmembrane-domain protein with a possible role in sensory synapses. European Journal of Human Genetics, 10(6), 339-350.

Alasti, F., Sadeghi, A., Sanati, M. H., Farhadi, M., Stollar, E., Somers, T., & Van Camp, G. (2008). A mutation in HOXA2 is responsible for autosomal-recessive microtia in an Iranian family. The American Journal of Human Genetics, 82(4), 982-991.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4(1), s13742-015.

Duman, D., & Tekin, M. (2012). Autosomal recessive nonsyndromic deafness genes: a review. Frontiers in bioscience: a journal and virtual library, 17, 2213.

Elias, E., & Bennett, R. (1992). Congenital defects in Awassi fat-tailed lambs. Small Ruminant Research, 8(1-2), 141-150.

Gao, L., Xu, S. S., Yang, J. Q., Shen, M., & Li, M. H. (2018). Genome-wide association study reveals novel genes for the ear size in sheep (Ovis aries). Animal genetics, 49(4), 345-348.

Geller, S. F., Guerin, K. I., Visel, M., Pham, A., Lee, E. S., Dror, A. A., ... & Flannery, J. G. (2009). CLRN1 is nonessential in the mouse retina but is required for cochlear hair cell development. PLoS Genet, 5(8), e1000607.

González, J. R., Armengol, L., Solé, X., Guinó, E., Mercader, J. M., Estivill, X., & Moreno, V. (2007). SNPassoc: an R package to perform whole genome association studies. Bioinformatics, 23(5), 654-655.

Hidalgo, A. M., Lopes, M. S., Harlizius, B., & Bastiaansen, J. W. M. (2016). Genome-wide association study reveals regions associated with gestation length in two pig populations. Animal genetics, 47(2), 223-226.

Jawasreh, K., Boettcher, P. J., & Stella, A. (2016). Genome-wide association scan suggests basis for microtia in Awassi sheep. Animal genetics, 47(4), 504-506.

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., Portolano, B., & Bi. Ov. Ita Consortium. (2018). Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Animal genetics, 49(1), 71-81.

Phillips, J. B., Västinsalo, H., Wegner, J., Clément, A., Sankila, E. M., & Westerfield, M. (2013). The cone-dominant retina and the inner ear of zebrafish express the ortholog of CLRN1, the causative gene of human Usher syndrome type 3A. Gene Expression Patterns, 13(8), 473-481.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, 81(3), 559-575.

Qiao, R., He, Y., Pan, B., Xiao, S., Zhang, X., Li, J., ... & Ren, J. (2015). Understanding the molecular mechanisms of human microtia via a pig model of HOXA1 syndrome. Disease models & mechanisms, 8(6), 611-622.

Ren, J., Duan, Y., Qiao, R., Yao, F., Zhang, Z., Yang, B., ... & Huang, L. (2011). A missense mutation in PPARD causes a major QTL effect on ear size in pigs. PLoS Genet, 7(5), e1002043.

Schiavo, G., Bertolini, F., Utzeri, V. J., Ribani, A., Geraci, C., Santoro, L., ... & Fontanesi, L. (2018). Taking advantage from phenotype variability in a local animal genetic resource: identification of genomic regions associated with the hairless phenotype in Casertana pigs. Animal genetics, 49(4), 321-325.

Tanchev, S., Zhelyazkov, E., Philipov, J., Semerdjiev, V., Paskaslev, M., Sotirov, L., & Georgieva, S. (2011). Phenotypic traits of skeletal anomalies observed in inbred rabbits. Revue de Medecine Veterinaire, 162(3), 150.

Zallocchi, M., Meehan, D. T., Delimont, D., Askew, C., Garige, S., Gratton, M. A., ... & Cosgrove, D. (2009). Localization and expression of clarin-1, the Clrn1 gene product, in auditory hair cells and photoreceptors. Hearing research, 255(1-2), 109-120.

Zhang, L., Liang, J., Luo, W., Liu, X., Yan, H., Zhao, K., ... & Wang, L. (2014). Genome-wide scan reveals LEMD3 and WIF1 on SSC5 as the candidates for porcine ear size. PLoS One, 9(7), e102085.

4.2 Genome-wide association study for somatic cells count in Valle del Belìce dairy sheep

4.2.1 Summary

In the Mediterranean basin countries, the dairy sheep production is usually based on local breeds, which are very well adapted to their production systems and environments and can indeed guarantee income, employment and economic viability in areas where production alternatives are scarce or non-existent. Mastitis is still one of the greatest problems affecting commercial milk production. However genetic evaluation of mastitis is particularly difficult because of the low heritability and the categorical nature of the trait. The aim of this study was to identify the genomic regions putatively associated with somatic cells count (SCC) in the local economically important Valle del Belìce sheep breed using of deregressed breeding values (DEBV) as response variables. All the samples were genotyped using the Illumina OvineSNP50K BeadChip. Genome-wide association analysis was carried out based on regression of DEB. A total of 8 markers were found to be significantly associated with log-transformed SCC. Several candidate genes associated with SCC were identified related to immunity system and udder conformation. The results can help improving the competitiveness of the local Valle del Belìce breed. Anyway, further studies considering a higher sample size or independent population will be needed to confirm our results.

4.2.2 Introduction

In the Mediterranean basin countries, the dairy sheep production is usually based on local breeds, which are very well adapted to their production systems and environments and can indeed guarantee income, employment and economic viability in areas where production alternatives are scarce or absent. Mastitis is a serious problem for the milk industry due to the reformed quality of milk and increased cost of flock regeneration. It can be induced, for example, by a lack of hygiene, by pushed manual milking or feed disorder. In dairy sheep, generally, the most important agents involved in mastitis are the bacterial infections, and the most frequently isolated pathogens are coagulase-negative staphylococci (CNS) that are present on and around the udder skin (Leitner *et al*., 2008) with a different

pathogenicity, causing clinical and subclinical mastitis (Contreras *et al*., 2007; Riggio and Portolano, 2015). The udder infection determines the increase of the Somatic Cell Count (SCC) in milk (Leitner *et al*., 2008; Raynal-Ljutovac *et al*., 2007) that causes significant damage of curd and cheese yields. Since the heritability of mastitis is low, genetic improvement on anti-mastitis by traditional selection is not very effective. SCC or log transformed SCC (*i.e.* somatic cell score, SCS) have relatively higher heritability compared to mastitis and are used as the first trait to improve mastitis resistance (Shook and Schutz, 1994). Kelly *et al*. (2000) found that an elevated SCC can alter the protein fractions distribution; decrease casein and lactose levels in milk; increase rennet clotting time, cheese moisture, and losses of fat and proteins in whey, and reduce curd firmness and cheese yielding. A study conducted by Sutera *et al*. (2018) confirmed that high levels of SCC in sheep milk are associated with milk yield losses and variations of fat and protein percentages. In particular, the estimated losses in milk yield were approximately 16% and ranged from 1,052 g for SCC $\leq 500 \times 103$ to 883 g for SCC $\leq 2{,}000 \times 103$, whereas there was an increase of 0.06% and 0.29% for fat and protein percentage, respectively. The negative effects of mastitis are provoked by a combination of animal characteristics (age, lactation stage, etc.), genetic (breed, inbreeding, etc.) and environmental factors (season, management, etc.) (Oget *et al*., 2019). Therefore, different individuals may have a different susceptibility to the disease, depending on their genetic heritage. In fact, there are several studies about the mastitis in dairy sheep confirming a genetic basis for mastitis resistance (Tolone *et al*., 2013; Oget *et al*., 2019), but no assumption had been made about the genes and the relative mechanisms. Selection of animals of higher breeding values than average, to be parents of the next generation, is the basis of genetic improvement programs.

The emergence of high-throughput genotyping technologies allowed routine genome-wide association studies (GWAS) to be performed in livestock populations. GWAS allow screening of the genome utilizing a large number of genetic markers spread across the entire genome to detect genetic variants associated with a particular disease or trait. The estimated breeding values (EBVs) were generally used to perform the GWAS. 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). An advantage of GWAS is that we can overcome the candidate gene approach through which sometimes significant results were not obtained due to the wrong or incomplete choice of candidate genes. In the last decades, several GWASs were conducted in sheep for milk production related traits (Li *et al*., 2020; Sutera *et al*., 2019), for fatty acids profile (Rovadoscki *et al*., 2018), for body weight (Tao *et al*., 2020; Ghasemi *et al*., 2019), for wool production (Wang *et al*., 2014), for nematode resistance (Becker *et al*., 2020) and ovine lentivirus resistance (White *et al*., 2012). To date, few GWAS have been conducted for SCC or SCS in dairy sheep (Oget *et al*., 2019), especially in local adapted breeds.

In Sicily, dairy sheep production represents an important resource for the local economy, and the Valle del Belìce is the main local breed reared on the island for the production of traditional raw milk cheeses, at farm level by small local dairies. The breed is subjected to limited breeding selection programs for milk production traits, but shows excellent adaptability to local environments, sometimes with harsh conditions (Mastrangelo *et al*., 2017). Therefore, the aim of this study was to identify the genomic regions putatively associated with SCC in Valle del Belìce sheep breed using of DEBVs as response variables.

4.2.3 Materials and methods

Data and estimation of breeding value

Between 2006 and 2016 the University of Palermo collected phenotypic data from 15 Valle del Belìce flocks, for a total of 1,813 individuals. The milk samples were collected aseptically from each individual from the two udder halves in sterile containers following an A4 recording procedure (ICAR, 2014), stored at 4 °C and transferred to the laboratory to determine daily SCC using Fossomatic 6000 (Foss Electric Hillerød, Denmark) equipment. The phenotypic data set originated by these sampling works was composed of 15,008 observations. Animals with less than 3 test-day measurements within lactation were discarded. For each individual the

following information were registered: order of parity, number of born lambs, lactation days, age, birth season and somatic cell count. Birth season was classified in three classes: 1 if the lambing was from August to November; 2 from December to March; 3 from April to July. SCC was normalized through a logarithmic transformation into somatic cell score (SCS) according to the formula of Ali and Shook (1980):

$$SCS = log_2\left(\frac{SCC}{100{,}000}\right) + 3$$

Preliminary analyses using the general linear model of ASReml R (Butler *et al.*, 2009) were performed to determine the significance of the fixed effects where the Wald tests are implemented in the form of the ANOVA method. A single trait repeatability test day (TD) animal model was performed to estimate the breeding values (EBV) as follows:

$$y = X\beta + Z_{htd} + Z_a + Z_{pe} + e$$

where y is the observation vector for SCS TD; β is the vector of fixed effects that includes order of parity (op: 4 classes), age at first lambing (age: 4 classes, 1 when first lambing occurred at 10–14 mo of age, 2 at 15–19 mo of age, 3 at 20-24 mo of age and 4 at 25–29 mo of age); birth season (bs: 3 classes), interaction between herd and birth season (hbs: 74 classes) and days of lactation (dim) modeled with a Legendre polynomial of order three. Htd is the vector of interaction between herd and test day random effect; a is the vector of direct additive genetic effects (breeding values); pe is the vectors of permanent environmental effect between lactations; e is the residual vector. X and Z are the corresponding incidence matrices relating records to fixed, animal, and permanent environmental between lactations effects, respectively. The pedigree file included 5,534 animals with 178 sires and 2,548 dams.

The assumptions regarding the components of the model were:

$$E\begin{bmatrix} y \\ b \\ htd \\ a \\ pe \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and $V_a = A\sigma_a^2;\ V_{htd} = I\sigma_{htd}^2;\ V_{pe} = I\sigma_{pe}^2;\ V_e = I\sigma_e^2$ where A is the numerator relationship matrix based on pedigree and I are the identity matrix with orders equal to numbers of dams for htd and pe effects and equal to the records for residuals e. Variance components and breeding values for SCS were estimated based on REML method using ASReml R (Butler *et al.*, 2009). In addition, EBVs were also deregressed according to Garrick *et al.* (2009) as follows:

$$DEBV = EBV/r^2$$

where EBV is the estimated breeding value and r^2 is the reliability of that EBV.

Blood sampling and DNA extraction

A total of 476 sheep of Valle del Belìce breed were sampled. About 10 mL of blood was collected from the jugular vein using vacutainer tubes containing EDTA as anticoagulant. The procedures involving animal sample collection followed the recommendation of directive 2010/63/EU. Sampling was carried out by trained veterinarians within the frame of vaccination Campaigns, hence no permission from the animal research ethics committee was necessary. Veterinarians adhered to standard procedures and relevant national guidelines to ensure appropriate animal care. Genomic DNA was extracted from each blood sample with a salting-out method (Miller *et al.* 1988). The DNA sample was quantified with a NanoDropND-1000 spectrophotometer (NanoDropTechnologies, Wilmington, DE, USA), diluted to a final concentration of 50 ng/mL (as required by the Illumina Infinium protocol), and stored at 4°C until use.

Genotyping and quality control

All the samples were genotyped using the Illumina OvineSNP50K BeadChip v2. Position and chromosomal coordinates for each SNP were obtained from the ovine genome sequence assembly (Oar 4.0) (https://www.ncbi.nlm.nih.gov/genome/?term=ovis+aries). Quality control and association analyses were performed using GenABEL package (Aulchenko *et al.*, 2007) in R environment (http://www.r-project.org). Only SNPs located on autosomes were extracted and considered for further analyses. Animals and markers that fulfilled the following criteria were kept in the analysis: (i) call rate per individuals and per SNPs >95%; (ii) minor allele frequency >2%; (iii) no extreme deviation from Hardy-Weinberg equilibrium ($P<10^{-6}$).

GWAS analyses

Genome-wide association analysis was carried out based on regression of DEBV with the genotypes of animals for one SNP at a time. We used the three-step approach referred to as genomic GRAMMAR-GC (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. After Bonferroni correction, significant thresholds were $P < 1.34x10\text{-}6$ for genome-wide ($P < 0.05$) and $P < 2.69x10\text{-}5$ for suggestive ($P < 0.10$) (*i.e.* one false positive for genome scan), corresponding to -log10(P) equal to 5.87 and 4.56, 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, in order to assess potential systematic bias due to population structure or analytical approach. The least square means of DEBV for the three genotypes affecting somatic cell count of significant SNP were also calculated by a general linear model (GLM) using R package lsmeans (Lenth and Lenth, 2018) and the significant threshold was set at $P < 0.05$.

Annotation

Genomic regions showing significant results were further explored to identify candidate genes underlying the loci. In particular, the gene contents located at ±250 kb distances from the significant SNP were annotated using Genome Data Viewer genome browser at the National Center for Biotechnology Information Database (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_000298735.2). The presence of Quantitative Traits Loci (QTLs; http://www.animalgenome.org/QTLdb) related with the considered trait was also checked. Finally, to investigate the biological function and the phenotypes that are known to be affected by each annotated gene, we conducted a comprehensive literature search, including information from other species.

4.2.4 Results

Genetic parameters and estimated breeding values

Descriptive statistics and genetic merit for SCS in the sampled animals are presented in Table 4.2.4.1. About 15,000 TD observations for SCS were considered to estimate EBVs then, the DEBVs of 5,534 individuals were estimated. Heritability and repeatability estimates for SCS in the studied population were 0.045 (standard error= 0.02) and 0.40 (standard error= 0.01), respectively.

Table 4.2.4.1. Descriptive statistics for somatic cell score.

Variable	N	Mean	SD	CV	Min-Max
SCS	15,008	2.67	0.72	0.27	1-5.31
DEBV	5,534	-0.19	0.52	2.75	-5.59-4.16

SCS: somatic cell score; DEBV: deregressed breeding value N: number of records; SD: standard deviation; CV: coefficient of variation; Min-Max: minimum and maximum values

Quality control for genotyping data

Among the 54,241 SNPs, 7,414 SNPs are located on sex chromosomes and thus were withdrawn from the analysis. A total of 3,999 SNPs were removed due to genotype rate <0.05, 2,037 SNPs due to minor allele frequency <0.02 and 3,651 SNP due to Hardy-Weinberg disequilibrium ($P<10^{-6}$). Moreover, 12 individuals were also excluded for a low (<95%) call rate. Then, after quality control we consider in total of 37,140 SNPs and 464 individuals for further analyses.

Genome-wide association analyses

In total we detected 8 significant SNPs for SCS, and among these, only one marker reached the genome-wide significant threshold ($P<4.72x10^{-7}$). The details of these SNPs including P-values, the positions on *Ovis aries* v4.0 genome assembly, the chromosomes and the closest known genes are given in Table 4.2.4.2.

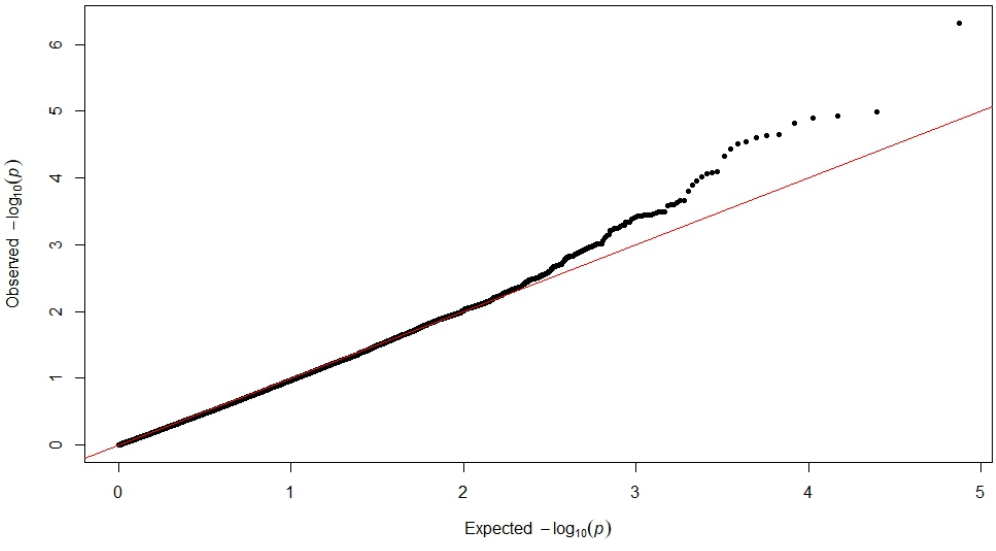
Table 4.2.4.2. Single nucleotide polymorphisms *(*SNP) significantly associated with somatic cell score at genome-wide ($P<1.34x10^{-6}$) and suggestive ($P<2.69x10^{-5}$) thresholds.

OAR	SNP	Position (bp)	P-value	Genes
1	rs401598547	46,865,607	1.16E-05	NEGR1
1	rs403091159	49,692,787	2.23E-05	LRRIQ3, LOC105605157, FPGT, LOC101120030
1	rs161717499	235,497,703	4.72E-07	SIAH2, ERICH6, LOC101119269, EIF2A, SERP1, TSC22D2, TRNAR-UCG
3	rs422960374	24,797,321	1.03E-05	FAM49A, TRNAC-GCA
7	rs406841304	57,592,284	2.52E-05	ATP8B4, LOC105607291, DTWD1, FAM227B, FGF7
8	rs420334414	67,510,451	2.34E-05	HIVEP2, AIG1, ADAT2
8	rs426621433	82,781,340	1.25E-05	SOD2, WTAP, ACAT2, TCP1, MRPL18, PNLDC1, MAS1, IGF2R, LOC106991323, LOC106991303, SLC22A1, SLC22A2
10	rs422370366	4,119,025	1.51E-05	-

OAR: *Ovis aries* chromosome; Genes: the closest genes to the significant SNP found within ±250 kb region surrounding it.

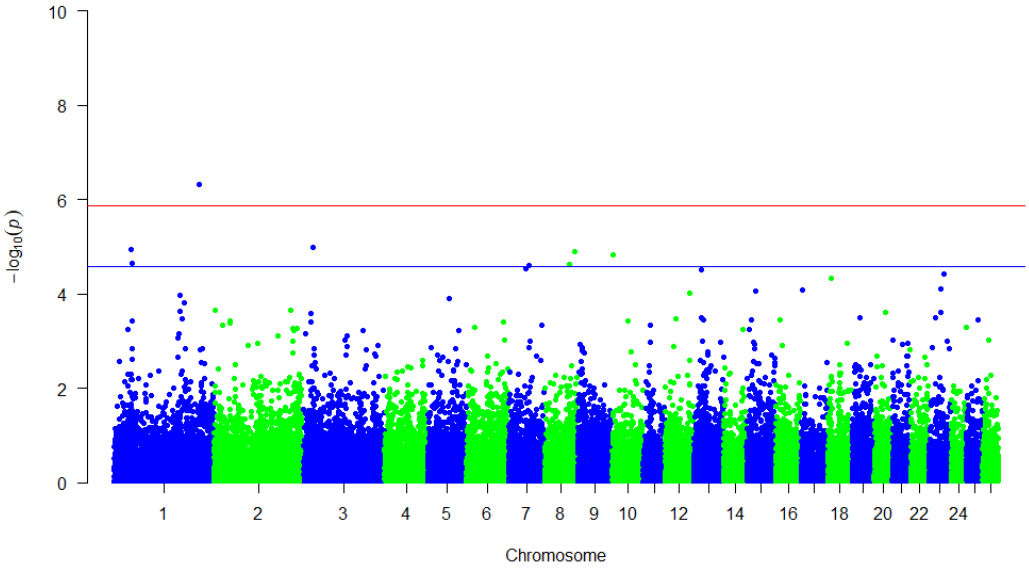
Manhattan plot, showing the profiles of the P-values (in terms of –log(P)) of all tested SNPs, is showed in Figure 4.2.4.1. The QQ-plot in Figure S1 shows the observed and expected P-values of the GWAS for SCS.

Fig. 4.2.4.1 – QQ-plot of the expected null distribution of the P-values vs. the observed null distribution of the P-values for somatic cell score in Valle del Belìce sheep.



The genomic inflation factor (lambda) was lower than one indicating some population stratification. However, departure from this line is also expected for a really polygenic trait, as many causal SNPs may not yet have reached genome-wide significance owing to a lack of power (Power *et al.*, 2017). The eight SNPs were located on five different chromosomes: three SNPs on OAR1, one SNP on OAR3, one SNP on OAR7, two SNPs on OAR8 and one SNP on OAR10. Considering the range of ±250 kb surrounding each significant SNP, a total of 34 genes (Table 4.2.4.2) were found. The most significant SNP (rs161717499) was located within the coding region of the Stress Associated Endoplasmic Reticulum Protein 1 (SERP1) gene on OAR1.

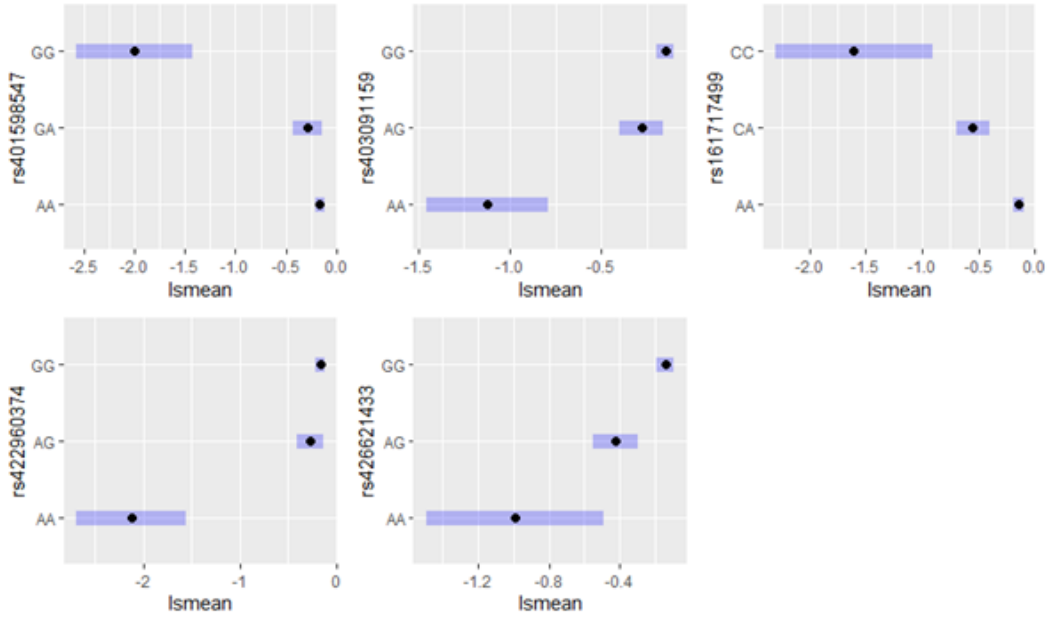
Figure 4.2.4.2. - Genome-wide plot of log10(P-values) for association of SNPs with somatic cell score. Blue and red lines represent suggestive and genome-wide thresholds, respectively.



For each of the eight significant SNPs (Figure 4.2.4.2), we also calculated the LSM of the DEBV for the three genotypes affecting the trait to investigate their genetic contribution (Figure 4.2.4.3). Five out of the eight above reported SNPs (rs401598547, rs403091159, rs161717499, rs422960374, rs426621433) reached the significance threshold ($P < 0.05$). Individuals with homozygous genotypes GG for rs401598547, CC for rs161717499 and AA for rs403091159, rs422960374 and rs426621433, showed lower somatic cells content among all three genotypes

(Figure 4.2.4.3). After checking on SheepQTLdb tool, no one of the eight detected SNPs was located within a known QTLs related at SCS or mastitis.

Figure 4.2.4.3. - Least squares means (at 95% CI) of DEBV for the three genotypes affecting the trait of significant SNP detected from GWAS study.



4.2.5 Discussion

Mastitis is still one of the greatest problems affecting commercial milk production. However genetic evaluation of mastitis is particularly difficult because of the low heritability and the categorical nature of the trait. As a consequence, SCC has been promoted as an indirect method of predicting mammary infections due to the positive correlation between these two traits (Boettcher, 2005). It is worth to mention that collecting information on SCC is easier, cheaper and less time demanding for farmers compared to use bacteriological status as direct measure of mastitis (Riggio 2012). In this study we estimated the breeding value for SCS and identified the genomic regions putatively involved in mastitis resistance in the local economically important Valle del Belìce dairy sheep breed.

The mean SCS (Table 1) was lower than those reported in previous studies (Tolone *et al.*, 2013; Riggio *et al.*, 2010) in the same breed, and by Ariznabarreta *et al.* (2002) in Churra sheep and Leitner *et al.* (2003) in Isaraeli-Assaf and Awassi sheep. These differences in SCS in the Valle del Belìce breed were due to different

sampled population. The heritability estimate for SCS in this study falls within the range (0.04-0.16) reported in literature for sheep (e.g., Barillet *et al.*, 2001; Hamann *et al.*, 2004 and Tolone *et al.*, 2013).

In this study, DEBVs of the SCS were used as trait scores for the association analysis. The estimated breeding values (EBVs) were generally used (like as pseudo-phenotypes) to perform the GWAS. Although EBVs have been used as dependent variables in GWAS (Becker *et al.*, 2013; Johnston *et al.*, 2011), this approach gave high false positive rate (Ekine *et al.*, 2014). 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). The DEBV make good use of available information from genotyped animals as well as from their relatives, which can appropriately avoid bias introduced by simply pooling or averaging data information and account for heterogeneous variance (Garrick *et al.*, 2009).

A total of 8 SNPs were found to be significantly associated with SCS in Valle del Belìce sheep. In particular, for five significant SNPs results suggest that individuals with the GG genotypes at *rs401598547*, CC at *rs161717499* and AA at *rs403091159*, *rs422960374* and *rs426621433* could be selected to reduce the somatic cells content in milk, although these genotypes had a low frequency in the breed. The lack of selection pressure in Valle del Belìce dairy sheep may also contribute to the low frequency of the favorable alleles and genotypes. Therefore, the effect of these alleles for somatic cells content trait should be verified in a larger population or by testing them in an independent sample.

The most significant SNP associated with SCS was located in the intronic region of *SERP1*. This gene encodes the stress-associated endoplasmic reticulum protein 1 and was associated with immune system (Moravčíková *et al.*, 2018). Another relevant gene close to the most significant SNP was *SIAH2*, involved to apoptosis and programmed cell death (Crisà *et al.*, 2016). On the same chromosome, the other two markers were close to *NEGR1*, a gene involved with medium white blood cell count (a leukocyte trait) in Yak (Ma *et al.*, 2019), *LRRIQ3* related to the innate immune system upon recognition of pathogens (Pablo-Maiso *et al.*, 2018) and

FPGT, which is part of the L-fucose pathway, a key sugar in complex carbohydrates involved in cell-to-cell recognition, inflammation, and immune processes (Becker and Lowe, 2003). As above reported, mastitis is a persistent, inflammatory response of mammary tissue attributed to intramammary invasion of a mastitis-causing pathogen. Therefore, according to their role and function, these aforementioned genes can be considered as candidate involved in mastitis resistance and SCS. The SNP rs422960374 on OAR03, was close (~70 Kb) to the *FAM49A* gene. Marete *et al.* (2018), in a GWAS for milking speed in French Holstein cows, reported the *FAM49A* as candidate gene for this trait. This gene was also associated to rear udder height in Holstein cattle (Gonzalez *et al.*, 2020). The genetic correlation between the SCS and udder attachment in sheep was observed by Casu *et al.* (2010); De la Fuente *et al.* (1996) reported the indirect selection for subclinical mastitis resistance due to the inclusion of udder morphology traits in selection objectives. Moreover, Gutiérrez-Gil *et al.* (2018) suggested that sheep with udders and high degree of suspension or shallow udders close to the abdominal wall should be associated to lower SCS. Despite specific functions of this gene are not known yet, all the aforementioned aspects suggested a possible involving of *FAM49A* gene in our trait. Similarly, on OAR7, near to SNP rs406841304, two close genes are related with udder conformation (*FAM227B*) (Scienski *et al.*, 2019) and with epithelial cell proliferation and differentiation (*FGF7*) (Bazer *et al.*, 2008; Yang *et al.*, 2020), suggesting their possible role in the epithelial mammary cell proliferation. Moreover, the *FGF7* has been reported as putative target gene in bovine mammary tissue infected with *Streptococcus uberis* (Naeem *et al.*, 2012). Another significant SNP was located on OAR8 (rs426621433) at position 82,781,340 bp. This SNP mapped within a QTL for SCC (81.4-83.5 Mb) reported in a commercial French dairy sheep population (Rupp *et al.* 2015), and near a QTL for SCC (ID number 160869) on OAR 8 (80.5-80.6 Mb) in Churra sheep. Among the closest annotated genes in the region of ±250 kb surrounding it, the *SOD2* gene seems to be the most plausible candidate affecting the SCS. In fact, the expression of SOD2 at mRNA and protein levels has been reported up-regulated in the mammary glands of ewes with clinical mastitis compared to healthy ewes (Gao *et al.*, 2019). Mitterhuemer *et al.* (2010) showed an increase of SOD2 gene level in mammary tissue from mastitis

cows inoculated with *E. coli* 24 h after infection as compared to controls. Finally, another candidate gene mapped near SNP rs426621433 on OAR8, was *IGF2R*, with a crucial role for the regulation of cell proliferation, growth, differentiation and survival, and associated with milk production traits. In fact, Dehoff *et al.* (1998) showed that lactation in the bovine mammary gland is associated with increased *IGF2R* concentration.

4.2.6 Conclusion

In this study, we estimated the breeding value for SCS in Valle del Belìce sheep. DEBVs of the SCS were used as phenotype for the association analysis. Several candidate genes associated with SCS were identified related to immunity system and udder conformation. These candidate genes provide valuable information for future functional characterization. Therefore, our results may contribute to increase knowledge on the role the genes play in the genetically determined mechanisms involved in mastitis in sheep. The results can help improving the competitiveness of the local Valle del Belìce breed, through the development of genetic improvement programs directed towards reducing the incidence of mastitis, also considering the udder conformation into selection objectives, with planned mating between subjects carrying favorable alleles. Anyway, further studies considering a higher sample size or independent population will be needed to confirm our results.

References

Ali, A., and Shook, G.E. (1980). An optimum transformation for somatic cell count in milk. J Dairy Sci. 63, 487–90.

Amin, N., Van Duijn, C.M., Aulchenko, Y.S.A. (2007). Genomic background-based method for association analysis in related individuals. PLOS ONE, 2(12).

Ariznabarreta, A., Gonzalo, C., San Primitivo, F. (2002). Microbiological quality and somatic cell count of ewe milk with special reference to staphylococci. J. Dairy Sci. 85(6), 1370-1375.

Aulchenko, Y.S., de Koning, D.J., Haley, C. (2007). Genome wide rapid association using mixed model and regression: A fast and simple method for genome wide pedigree-based quantitative trait loci as-sociation analysis. Genetics. 177, 577- 585.

Barillet, F., Rupp, R., Mignon-Grasteau, S., Astruc, J.M., Jacquin, M. (2001). Genetic analysis for mastitis resistance and milk somatic cell score in French Lacaune dairy sheep. Genet. Sel. Evol. 33, 397–415.

Bazer, F.W., Slayden, O.D. (2008). Progesterone-induced gene expression in uterine epithelia: a myth perpetuated by conventional wisdom. Biol. Reprod. *79*(5), 1008-1009.

Becker, D., Wimmers, K., Luther, H., Hofer, A. and Leeb, T. (2013). A genome-wide association study to detect QTL for commercially important traits in Swiss Large White boars. PLOS ONE. 8.

Becker, D.J. and Lowe, J.B. (2003). Fucose: biosynthesis and biological function in mammals. Glycobiology 13(7), 41R-53R.

Becker, G.M., Davenport, K.M., Burke, J.M., Lewis, R.M., Miller, J.E., Morgan, J.L.M., *et al.* (2020). Genome-wide association study to identify genetic loci associated with gastrointestinal nematode resistance in Katahdin sheep. Anim. Genet. 51(2), 330-335.

Boettcher, P. (2005). Breeding for improvement of functional traits in dairy cattle. Ital. J. Anim. Sci. 4, 7–16.

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J. (2009). "ASReml-R Reference Manual." Technical report, Queensland Department of Primary Industries. http://www.vsni.co.uk/software/asreml/

Casu, S., Sechi, S., Salaris, S. L., Carta, A. (2010). Phenotypic and genetic relationships between udder morphology and udder health in dairy ewes. Small Rumin. Res. 88(2-3), 77-83

Contreras, A., Sierra, D., Sánchez, A., Corrales, J. C., Marco, J. C., Paape, M. J. (2007). Mastitis in small ruminants. Small Rumin. Res. 68(1-2), 145-153.

Crisà, A., Ferrè, F., Chillemi, G., Moioli, B. (2016). RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. BMC Vet. Res. 12(1), 264.

De la Fuente, L.F., Fernandez, G., San Primitivo, F. (1996). A linear evaluation system for udder traits of dairy ewes. Livest. Prod.Sci. 45(2-3), 171-178.

Dehoff, M.H., Elgin, R.G., Collier, R.J., Clemmons, D.R. (1998). Both type I and II insulin-like growth factor receptor binding increase during lactogenesis in bovine mammary tissue. Endocrinology 122, 2412–2417.

Ekine, C.C., Rowe, S.J., Bishop, S.C., de Koning, D.J. (2014). Why breeding values estimated using familial data should not be used for genome-wide association studies. G3-Genes Genom. Genet. 4(2), 341-347. doi.org/10.1534/g3.113.008706

Gao, J., Li, T., Lu, Z., Wang, X., Zhao, X., Ma, Y. (2019). Proteomic analyses of mammary glands provide insight into the immunity and metabolism pathways associated with clinical mastitis in meat sheep. Animals 9(6), 309.

Garrick, D.J., Taylor, J.F., Fernando, R.L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41(1), 55.

Ghasemi, M., Zamani, P., Vatankhah, M., Abdoli, R. (2019). Genome-wide association study of birth weight in sheep. Animal, 13(9), 1797-1803.

Gonzalez, M., Villa, R., Villa, C., Gonzalez, V., Montano, M., Medina, G., *et al.* (2020). Inspection of real and imputed genotypes reveled 76 SNPs associated to rear udder height in Holstein cattle. Adv. Vet. Anim. Res. 7(2), 234-241.

Gutiérrez-Gil, B., Esteban-Blanco, C., Suarez-Vega, A., Arranz, J.J. (2018). Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50K SNP chip and whole-genome sequencing. J. Dairy Sci. 101(10), 9072-9088.

Hamann, H., Horstick, A., Wessels, A., Distl, O. (2004). Estimation of genetic parameters for test day milk production, somatic cell score and litter size at birth in East Friesian ewes. Livest. Prod.Sci. 87(2-3), 153-160.

ICAR (2014). International Committee for Animal Recording, International Agreement of Recording Practices. Available online: http://www.icar.org/documents/Rules

Johnston, S.E., McEwan, J.C., Pickering, N.K., Kijas, J.W., Beraldi, D., Pilkington, J.G., *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.

Kelly, A.L., Tiernan, D., O'sullivan, C., Joyce, P. (2000). Correlation between bovine milk somatic cell count and polymorphonuclear leukocyte level for samples of bulk milk and milk from individual cows. J Dairy Sci. 83(2), 300-304.

Leitner, G., Silanikove, N., Merin, U. (2008). Estimate of milk and curd yield loss of sheep and goats with intrammamary infection and its relation to somatic cell count. Small Rumin. Res. 74, 221–225.

Leitner, G., Chaffer, M., Caraso, Y., Ezra, E., Kababea, D., Winkler, M., ... & Saran, A. (2003). Udder infection and milk somatic cell count, NAGase activity and milk composition—fat, protein and lactose—in Israeli-Assaf and Awassi sheep. Small Ruminant Research, 49(2), 157-164.

Lenth, R., and Lenth, M. R. (2018). Package 'lsmeans'. The American Statistician, 34(4), 216-221.

Li, H., Wu, X.L., Tait Jr, R.G. Bauck, S., Thomas, D.L., Murphy, T.W., *et al.* (2020). Genome-wide association study of milk production traits in a crossbred dairy sheep population using three statistical models. Anim. Genet. 51, 624-628.

Ma, X., Jia, C., Fu, D., Chu, M., Ding, X., Wu, X., *et al.* (2019). Analysis of Hematological Traits in Polled Yak by Genome-Wide Association Studies Using Individual SNPs and Haplotypes. Genes, 10(6), 463.

Marete, A., Sahana, G., Fritz, S., Lefebvre, R., Barbat, A., Lund, M.S., *et al.* (2018). Genome-wide association study for milking speed in French Holstein cows. J. Dairy Sci. 101(7), 6205-6219.

Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., & Portolano, B. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belìce sheep. Genetics Selection Evolution, 49(1), 1-10.

Miller, S.A., Dykes, D.D., Polesky, H.F. (1988). A salting out method for DNA preparation. Nucleic Acids Res. 16, 1215.

Mitterhuemer, S., Petzl, W., Krebs, S., Mehne, D., Klanner, A., Wolf, E., ... & Blum, H. (2010). Escherichia coli infection induces distinct local and systemic transcriptome responses in the mammary gland. BMC genomics, 11(1), 138.

Moravčíková, N., Simčič, M., Mészáros, G., Sölkner, J., Kukučková, V., Vlček, M., *et al.* (2018). Genomic response to natural selection within alpine cattle breeds. Czech J. Anim. Sci. 63(4), 136-143.

Naeem, A., Zhong, K., Moisá, S.J., Drackley, J.K., Moyes, K.M., Loor, J.J. (2012). Bioinformatics analysis of microRNA and putative target genes in bovine mammary tissue infected with Streptococcus uberis. J. Dairy Sci. 95(11), 6397-6408.

Oget, C., Tosser-Klopp, G., Rupp, R. (2019). Genetic and genomic studies in ovine mastitis. Small Rumin. Res. 176, 55-64.

Ostersen, T., Christensen, O.F., 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.

Ostersen, T., Christensen, O.F., 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.

Pablo-Maiso, D., Doménech, A., Echeverría, I., Gómez-Arrebola, C., De Andrés, D., Rosati, S., *et al.* (2018). Prospects in innate immune responses as potential control strategies against non-primate Lentiviruses. Viruses, 10(8), 435.

Power, R. A., Parkhill, J., & de Oliveira, T. (2017). Microbial genome-wide association studies: lessons from human GWAS. Nature reviews genetics, 18(1), 41.

Raynal-Ljutovac, K., Pirisi, A., de Crémoux, R., Gonzalo, C., (2007). Somatic cells of goat and sheep milk analytical, sanitary, productive and technological aspects. Small Rumin. Res. 68, 126–144.

Riggio, V. (2012). Genetic aspects of somatic cell count and udder health in the Italian Valle del Belìce dairy sheep. Wageningen University. ISBN 9789461732460 - 137.

Riggio, V., and Portolano, B. (2015). Genetic selection for reduced somatic cell counts in sheep milk: A review. Small Rumin. Res. 126, 33-42.

Riggio, V., Portolano, B., Bovenhuis, H., Bishop, S.C. (2010). Genetic parameters for somatic cell score according to udder infection status in Valle del Belìce dairy sheep and impact of imperfect diagnosis of infection. Genet.Sel. Evol. 42(1), 30.

Rovadoscki, G.A., Pertile, S.F.N., Alvarenga, A.B., Cesar, A.S.M., Pértille, F., Petrini, J., *et al.* (2018). Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. BMC Genomics, 19(1), 375.

Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., ... & Tosser-Klopp, G. (2015). A point mutation in suppressor of cytokine signalling 2 (Socs2) increases the susceptibility to inflammation

of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. PLoS genetics, 11(12), e1005629.

Scienski, K., Ialacci, A., Bagnato, A., Reginelli, D., Durán-Aguilar, M., Strillacci, M.G. (2019). Genetic variability in a Holstein population using SNP markers and their use for monitoring mating strategies. Revista Mexicana de Ciencias Pecuarias, 10(3), 643-663.

Sell-Kubiak, E., Duijvesteijn, N., Lopes, M.S., Janss, L.L.G., Knol, E.F., Bijma, P., *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, C.A., Lopes, M.S., Harlizius, B., Hanenberg, E.H., Knol, E.F., Bastiaansen, J. WM. (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.

Shook, G.E., and Schutz, M.M. (1994). Selection on somatic cell score to improve resistance to mastitis in the United States. J Dairy Sci. *77*(2), 648-658.

Sutera, A.M., Portolano, B., Di Gerlando, R., Sardina, M.T., Mastrangelo, S., Tolone, M. (2018). Determination of milk production losses and variations of fat and protein percentages according to different levels of somatic cell count in Valle del BelìceBelìce dairy sheep. Small Rumin. Res. 162, 39-42.

Sutera, A.M., Riggio, V., Mastrangelo, S., Di Gerlando, R., Sardina, M.T., Pong-Wong, R., *et al.* (2019). Genome-wide association studies for milk production traits in Valle del Belìce sheep using repeated measures. Anim. Genet. 50(3), 311-314.

Tao, L., He, X.Y., Pan, L.X., Wang, J.W., Gan, S.Q., Chu, M.X. (2020). Genome-wide association study of body weight and conformation traits in neonatal sheep. Anim. Genet. 51(2), 336-340.

Tolone, M., Riggio, V., Portolano, B. (2013). Estimation of genetic and phenotypic parameters for bacteriological status of the udder, somatic cell score, and milk yield in dairy sheep using a threshold animal model. Livest. Sci. 151(2-3), 134-139.

Wang, Z., Zhang, H., Yang, H., Wang, S., Rong, E., Pei, W., *et al.* (2014). Genome-wide association study for wool production traits in a Chinese Merino sheep population. PloS one, 9(9).

White, S.N., Mousel, M.R., Herrmann-Hoesing, L.M., Reynolds, J.O., Leymaster, K.A., Neibergs, H.L., *et al.* (2012). Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus. PloS one, 7(10).

Yang, C., Lan, W., Ye, S., Zhu, B., Fu, Z. (2020). Transcriptomic Analyses Reveal the Protective Immune Regulation of Conjugated Linoleic Acids in Sheep Ruminal Epithelial Cells. Front. Physiol. *11*, 1368.

CHAPTER 5

General conclusion

General conclusion

This thesis presented different approaches aiming to improve the competitiveness of Valle del Belìce breed. In particular, different molecular technologies were applied for characterization and valorisation of Valle del Belìce breed and for genetic traceability of its typical dairy products.

Among the most important molecular markers, we chose microsatellite markers to set up an informative panel for parentage assignment test in Valle del Belìce breed, aiming to verify and correct pedigree errors due to recorded information by the farmers. Moreover, SNPs were used to develop traceability protocol for two important Sicilian PDO dairy products, Vastedda della Valle del Belìce and Pecorino Siciliano, and to conduct genome-wide association analysis, useful to detect genomic regions associated with morphological and productive traits.

In chapter 2, individual samples of Valle del Belìce breed collected in four farms were analyzed by 24 microsatellite markers in order to verify the offspring-mother pairs recorded by the farmers and to assign paternity, reconstructing trio families. The results of this chapter showed that only 68.8% of the total offspring-mother pairs were correctly declared on pedigrees; moreover, we were able to reconstruct only 45.5% of trio families. That aspect highlights the need to use this molecular tool in as many farms as possible, with the intent to improve the ovine pedigrees recording, while maintaining the natural mating system usually adopted in Sicilian flocks. Interesting were the results on adoption and heteropaternal superfecundation phenomena, which even though present with low frequency, caught our attention on other possible pedigrees errors that could affect Sicilian farming systems, with consequent difficulties in Valle del Belìce sheep breed genetic improvement programs.

In chapter 3, we used SNP markers to develop a genetic traceability procedure for two important Sicilian dairy products: Vastedda della Valle del Belìce PDO and Pecorino Siciliano PDO.

This method was set up on genotyping data and applied with successful results to DNA pools prepared mixing different percentage of DNA belonging to involved breeds. Finally, the procedure was validate on experimental dairy products and verified on industrial dairy productions giving us the possibility to discriminate

dairy products realized exclusively with milk from specific breeds (VDB or TARGET breeds) from dairy products realized with mixed milk with different origins (FOREIGN or COM+OTHERS). Current certification system is based on several information flows that define production processes rather than product itself. The possibility of certifying origin and identity of dairy products could provide the conservation and valorisation of local breeds.

Finally, in Chapter 4, SNPs marker information were used in two GWA studies aiming to find genomic regions involved in morphological microtia trait and productive SCC trait.

The results on *microtia* trait showed that despite some limitation due to sample size, array density, and unknown inheritance model of this trait, the combination of different approaches was useful to provide insights into the genetic basis of ear size and let us to hypothesized that variants of the CLRN1 could affect the ear development in sheep.

The estimated breeding values for somatic cell score in Valle del Belice sheep were used as phenotype for the association analysis aiming to identify several candidate genes associated with immune response and udder conformation. Our results may contribute to increase knowledge on the role the genes play in the genetically determined mechanisms involved in mastitis resistance in sheep.

Ringraziamenti

Durante il triennio di dottorato, tante sono state le persone che a vario titolo mi hanno permesso di crescere, attraverso collaborazioni, critiche costruttive, insegnamenti preziosi. Tra queste certamente tutto il gruppo di ricerca AGR/17 coordinato dal Prof. Portolano, il quale fin dal primo momento si è mostrato come una persona dedita al mondo della zootecnia ed appassionato al proprio lavoro. Per cui voglio ringraziarlo non solo per il supporto che mi ha fornito personalmente come tutor, ma anche per come riesce a trasmettere la passione per il mondo zootecnico. Anna, Ciupy, Marco, Rosalì e Salvo sono stati di grande aiuto per la mia crescita scientifica, con insegnamenti pratici di laboratorio e con supporto nell'attività di stesura dei lavori scientifici e della tesi di dottorato. Poco per volta mi hanno dato sempre più strumenti formativi, rendendomi autonomo. Inoltre, mi sopportano ogni giorno, e non è poco.

Un ringraziamento va anche a tutti i colleghi iscritti ai corsi di Dottorato di Ricerca dell'Università degli Studi di Palermo, grazie al cui sostegno elettorale ho avuto l'opportunità di essere eletto per un biennio Senatore Accademico. Una bellissima ed impegnativa esperienza, che sicuramente non dimenticherò mai e che mi ha permesso di accrescere notevolmente il mio bagaglio culturale e formativo.

Un doveroso ringraziamento va fatto alla mia famiglia, che mi ha sostenuto durante tutto il percorso universitario con tanti sacrifici, ed oggi certamente con grande gioia ed orgoglio festeggia questo traguardo e ai tanti amici ai quali questa esperienza ha sottratto del tempo. Sarebbero tanto orgogliosi anche i miei nonni, Angelo e Maria, che erano entusiasti ad ogni traguardo accademico. Li ringrazio perché certamente da lassù hanno contribuito a darmi una mano con il loro sorriso.

Infine, il ringraziamento più importante va ad Ilaria, che davvero tanto ha fatto e continua a fare per me ogni giorno, a partire dal fatto che è riuscita a sopportarmi nei periodi più frenetici ed intensi di questo triennio, ed allo stesso tempo è stata indispensabile in ogni momento, dentro e fuori l'Università; senza il suo aiuto, il suo tempo, la sua pazienza e la sua vicinanza non sarei riuscito a raggiungere questo importante risultato.

List of publications

Peer reviewed publications

- Mastrangelo, S., Sottile, G., Sutera, A. M., Di Gerlando, R., Tolone, M., **Moscarelli, A**., ... & Portolano, B. (2018). Genome-wide association study reveals the locus responsible for microtia in Valle del Belice sheep breed. Animal genetics, 49(6), 636-640.
- Mastrangelo, S., Ben Jemaa, S., Ciani, E., Sottile, G., **Moscarelli, A.**, Boussaha, M., ... & Cassandro, M. (2020). Genome-wide detection of signatures of selection in three Valdostana cattle populations. Journal of Animal Breeding and Genetics.
- Di Gerlando, R., Mastrangelo, S., **Moscarelli, A.**, Tolone, M., Sutera, A. M., Portolano, B., & Sardina, M. T. (2020). Genomic Structural Diversity in Local Goats: Analysis of Copy-Number Variations. Animals, 10(6), 1040.
- **Moscarelli, A.**, Sardina, M. T., Cassandro, M., Ciani, E., Pilla, F., Senczuk, G., ... & Mastrangelo, S. (2020). Genome-wide assessment of diversity and differentiation between original and modern Brown cattle populations. Animal Genetics.
- Sutera, A. M., **Moscarelli, A.**, Mastrangelo, S., Sardina, M. T., Di Gerlando, R., Portolano, B., & Tolone, M. (2021). Genome-Wide Association Study Identifies New Candidate Markers for Somatic Cells Score in a Local Dairy Sheep. Frontiers in Genetics, 12, 409.

Papers under review or in preparation

- Genome-wide patterns of homozygosity reveal population history of dairy local goat population. (2020). Mastrangelo, S., Di Gerlando, R., Sardina, M.T., Sutera, A.M., **Moscarelli, A.**, Tolone, M., Crepaldi, P. and Portolano, B.