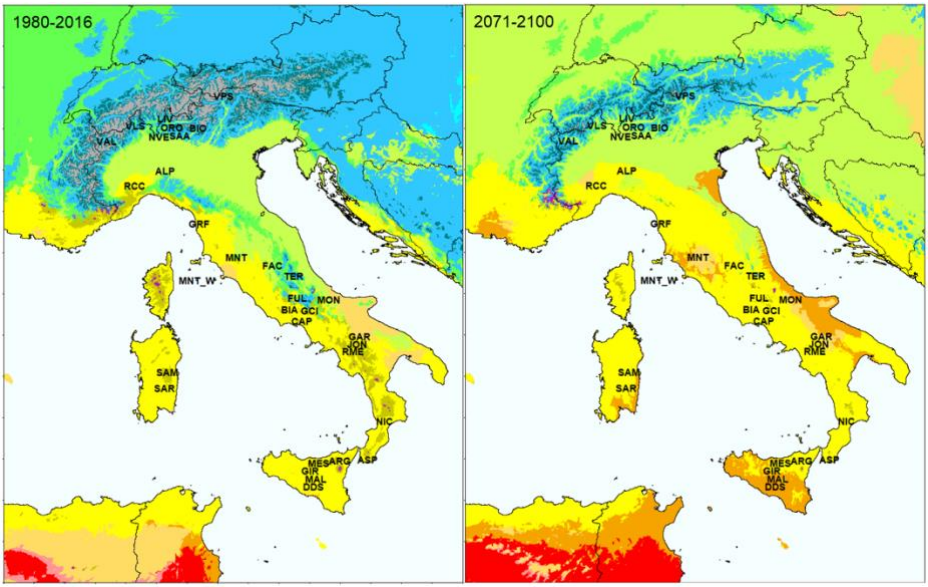
UNIVERSITA' DEGLI STUDI DI MILANO

PhD Course in Veterinary and Animal Science
Class XXXIV

Genomic investigation on the effects of the selection in small ruminant species



PhD Candidate: Dott. Matteo Cortellari

Tutor: Prof. Paola Crepaldi

Academic Year: .2020-2021

Index

Figures

Tables

CHAPTER 1 Abstract

1.1 Italian

Il cambiamento climatico previsto per i prossimi decenni nel nostro paese pone delle sfide sempre più pressanti alla gestione della biodiversità che caratterizza le razze autoctone italiane. Alla luce di queste sfide diventa quindi fondamentale andare a investigare quelli che sono i segni distintivi che selezione e adattamento all'ambiente hanno lasciato nel patrimonio genetico delle nostre razze. In particolare, le specie di piccoli ruminanti come capre e pecore, data la loro ampia variabilità e presenza in ambienti differenti, rappresentano una risorsa fondamentale per comprendere le basi genetiche che stanno dietro la capacità di adattamento alle diverse tipologie di climi e di allevamento. Uno strumento sempre più cruciale per questo scopo è l'introduzione dei dati genomici anche nel settore dei piccoli ruminanti. Infatti, grazie ad essi abbiamo la possibilità di studiare il genoma delle varie specie e razze e la loro relazione con l'ambiente, ma anche fornire informazioni sempre più utili a livello gestionale ad allevatori e associazioni di razza.

In questo lavoro di tesi abbiamo quindi utilizzato gli strumenti genomici per investigare i segni lasciati dall'adattamento e dalla selezione nell'ampio panorama della biodiversità dei piccoli ruminanti italiani, con un particolare focus sulle capre. Infatti siamo partiti dall'analisi della biodiversità di 33 diverse razze e popolazioni del nostro paese, identificando le interazioni tra i loro genotipi e l'ambiente in cui sono allevate tramite la Landscape Genomics (Lavoro 1), per poi andare a vedere come i diversi sistemi di allevamento e management utilizzati le differenzino e quali segni lascino nel genoma utilizzando la tecnica delle Runs Of Homozygosity (lavoro 2). Infine, nel

terzo ed ultimo lavoro presentato, abbiamo valutato i rapporti tra valori di inbreeding genomico e da pedigree in sei diverse razze caso studio al fine di fornire stime e dati sempre più utili e affidabili ad allevatori e associazioni e di gettare così le basi per la futura costruzione di indici genomici anche nei piccoli ruminanti italiani.

1.2 English

Climate change as predicted for the upcoming decades in Italy poses increasing challenges for the biodiversity management of native Italian breeds. Consequently, it is vital to investigate the genetic impact of climate change on these breeds in terms of selection and adaptation. Given their variability and presence in different habitats, small ruminant species, such as goats and sheep, represent ideal subjects for studying the genetics that lays behind the adaptation to different environments. By introducing genomic data in the investigation of small ruminants we can observe how the genome of these various species and breeds has adapted to artificial selection and environment as well as provide useful information to breeders and their associations .

In this thesis genomic tools were used to investigate the signs that adaptation and selection left on Italian small ruminants population with a particular focus on goats. We analyzed the biodiversity of 33 different Italian goat breeds and populations by identifying the interactions between their genotypes and the environment where they were raised by means of Landscape Genomics (the 1st paper). Then we examined how the different breeding systems and management differentiate them and what signatures are left in their genome by means of the Runs of Homozygosity (the 2nd paper).

In the final paper we evaluated the relationships between values for genomic and pedigree inbreeding coefficients for six different case-study breeds of sheep and goats in order to provide breeders and their associations with more

accurate estimates as well as useful and reliable data, thereby laying the foundation for the construction of genomic indexes for small ruminants in Italy.

CHAPTER 2 Introduction

2.1 Evolution of Small Ruminants

Small ruminants were among the first species domesticated and since that first event of domestication they have accompanied man's migration all around the globe. Among the reasons for this association are their small size, gregariousness, versatility, but above all, their adaptability to different environments (Clutton-Brock, 1999). The two species of small ruminants we are most interested in are sheep (*Ovis aries*) and goats (*Capra hircus*).

Since the Miocene, when their wild ancestors, the Asiatic mouflon (*Ovis orientalis*) and the Bezoar ibex (*Capra aegagrus)* diverged, their domestication history has been very similar (McHugo et al., 2019). Both were domesticated in the Fertile Crescent in the vicinity of Southeastern Anatolia and the Zagros Mountains in Iran at approximately at the same time. Goats began to be domesticated around 10 000 -9 000 YBP (Year Before Present) and sheep a bit earlier, between 11 000 and 10 000 YBP (Zheng et al., 2020; Kamalakkannan et al., 2021). Maternal mitochondrial DNA studies indicate sheep were domesticated in the Fertile Crescent and other places as well (Deng et al., 2020), while for goats, while for goats, a second independent domestication site in either the Indus Valley or China has yet to be verified (Daly et al., 2018).

Initially, both species were hunted by man for meat until the Neolithic revolution and the advent of agriculture, when man became sedentary and started to manage his prey. Goats and their sheep were domesticated for the first thousands of years for meat, and only much later for their milk and wool. Analysis of early Neolithic pottery residuals found in north Africa and Anatolia reveals the first evidence of milk consumption dated at

approximately 5000-6000 YBP (Bleasdale et al., 2021; Rosenstock et al., 2021).

During this period sheep and goats accompanied humans as they migrated to all the continents. As we can see in Figure 2.1, they entered Europe and Africa through three different routes: the Danubian corridor, which follows the Danubio river to the northern Europe, the northern Mediterranean corridor, which goes through Greece and Italy until arriving in Spain, and the southern Mediterranean corridor, which leads to north Africa through the Sinai peninsula(Alberto et al., 2018).

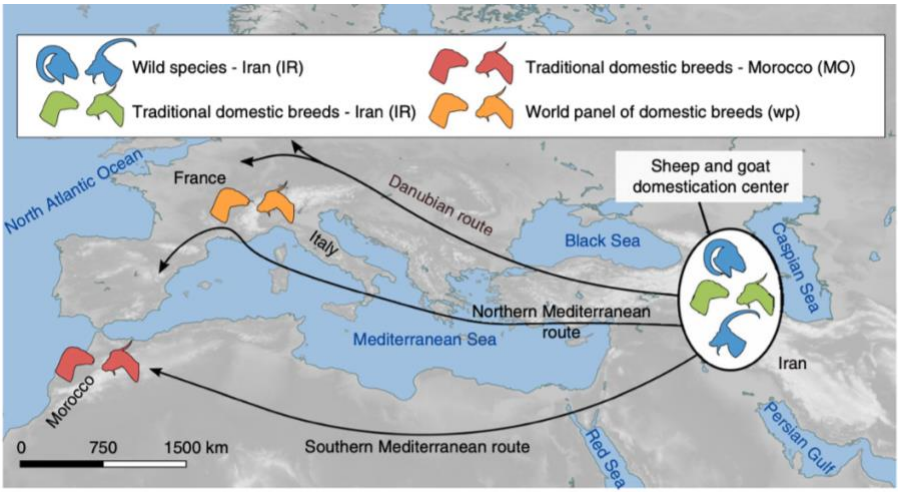


Figure 2.1 - Sampling locations and domestication center of sheep and goat and representation of the main colonization route for Europe and Africa (Alberto et al., 2018)

They spread throughout Asia by means of two different corridors, one leading to the Indian subcontinent through the Khyber pass, and the other from the Eurasian steppes to present day Mongolia and China (Lv et al., 2015; Amills et al., 2017).

As a result of these migrations, goats and sheep colonized and adapted to a wide range of environments, from the subarctic climate of the Alps and the Siberian steppe to the high temperatures of Sahara desert, and from the high altitude of the Tibetan plateaus to the South Pacific islands. This allowed them to develop several specific characteristics. For example, they developed skin protection strategies against solar radiation or extreme below zero temperatures, adaptation to hypoxia and the maintenance of normal circadian rhythms in situations of extreme light/dark oscillations (Chebii et al., 2021). Nevertheless they adapted in response to human selections for different breeding purposes, such as milk, meat and textile productions wherever they settled (Bertolini et al., 2018b).

Today the descendants of these animals number approximately two billion, distributed over five continents, and their number is still increasing due to growing worldwide demand for milk and meat (Mazinani and Rude, 2020). It has been estimated that there are more than 2000 different breeds of goats and sheep worldwide that have had at least one registration event in the last 20 years (https://www.fao.org/home/en). This represents an incredible source of genetic biodiversity that needs to be studied to understand the basis of their adaptation to the environment as well as the basic structure of their different productive traits.

2.2 Effects of Climate change

The latest climate projections, according to the widely used Koppen-Geiger classification represented in Fig. 2.2, indicate that dramatic climate change will characterize most terrestrial regions until the end of the century (Beck et al., 2018). Presented initially by a German climatologist at the end of the 19th century, the Koppen-Geiger classification divides climate into 5 classes and 30 sub-classes based on three main principles: monthly precipitation, air temperature and the observation that the vegetation and the biome of an environment are crystallized representations of its climate. Accordingly, in the next 70 years some areas in Italy that are currently classified as 'temperate-warm' will shift to increasingly arid and warm climate classes (Beck et al., 2018). There are also shorter-term climate change predictions for global warming that predict an average temperature rise of about 1.5 degrees by the end of this decade, longer drought periods and increasing extreme and short-term rainfall events (Tabari, 2020). All these new predictions and updated models regarding the evolution of climate are made possible thanks to the presence of numerous databases of bioclimatic indicators publicly available for the entire scientific community, such as Worldclims (Fick and Hijmans, 2017), CliMond (Kriticos et al., 2012) and CMCC-BioclimInd (Noce et al., 2020).

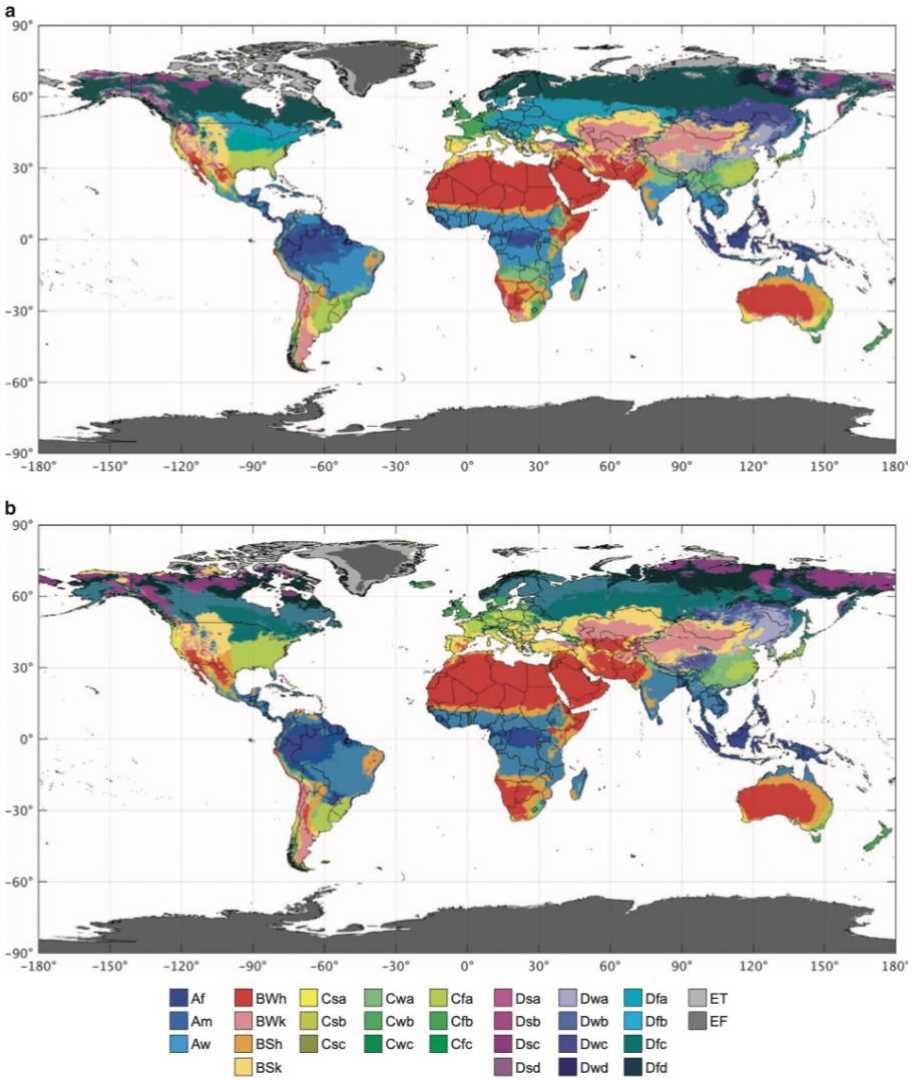


Figure 2.2- representation of the predicted Koppengiger-classification at a global scale for the present day (1980-2016) part a and for the future (2071-2100) part b (Beck et al., 2018)

These predicted changes will greatly influence both animal physiology and adaptation, and consequently, animal farming systems. For example, changes in temperature and humidity can reduce animal feed intake and consumption, because eating and metabolizing nutrients, especially in ruminants, are directly related to body temperature and increased heat stress (Gaughan and Cawdell-Smith, 2015; Sejian et al., 2021).

Inducing animals to eat less not only reduces animal productivity but also their fertility. Without the proper amount of nutrition, body condition scores suffer, signaling a consequent reduction in the rates of conception and pregnancy (Das et al., 2016; van Wettere et al., 2021). Furthermore, a rise in temperature impacts on mating behaviors, resulting in decreasing fertility (Polsky and von Keyserlingk, 2017). Global warming also affects the immune responses of animals at different stages of life (Dahl et al., 2020), while increasing the spread of new pathogens and pathologies in areas where they were not previously present (Baylis and Risley, 2013).

Not only will climate change affect the physiology of an animal but also the way it will be raised. For example, as temperatures rise, the production of forage and crops in terms of quantity and quality will diminish, therefore more effective diets will be needed to keep up the productivity of our animals (D and M, 2017; Calleja-Cabrera et al., 2020). Moreover, more sustainable solutions must be found to reduce the production of greenhouse gases, for example, by extensive farming that exploits local resources more efficiently (Grossi et al., 2019).

From this perspective, small ruminants certainly represent species of domesticated animals that have successfully adapted over time to different climates and food conditions. Their biodiversity can be investigated to

discover possible solutions that will allow our animals to optimally adapt and our farming systems to improve with respect to the new climate scenarios (Joy et al., 2020).

2.3 Impact of Genomics Technologies

Genomics allows us to investigate the signatures left by adaptation to environmental and artificial selection in the genome of small ruminants. After the sequencing the human genome in 2000 (International Human Genome Sequencing Consortium et al., 2001; Venter J. Craig et al., 2001), work started on sequencing the genomes of wild and domesticated animal species. Regarding small ruminants, the sheep genome was first sequenced in 2010 (The International Sheep Genomics Consortium et al., 2010) while recently an improved last version was presented (Davenport et al., 2021), the goat genome was first sequenced later (Dong et al., 2013) followed by newer versions with ARSI being the most accurate to date (Bickhart et al., 2017). For both species the first sequencing was followed by the SNP chip, consisting of specific SNP markers with different densities, making it the perfect, affordable tool for investigating the genetics of different breeds and populations. Presently, there are a couple of SNPChips of medium (50 K) density (Kijas et al., 2009) and one of high (600K) density for sheep (Kijas et al., 2014) while for goats, there is only a mid-density SNP array developed by the International Goat Genome Consortium (Tosser-Klopp et al., 2014), for which there is now a newer 65k version.

These new technologies, however, require increasing amounts of memory space for data storage and computational power to carry out the different analyses, which in turn has led to a parallel increase in costs, beyond the budgets of most small to medium size labs, necessitating the creation of several international projects and consortia.

Increasing computational power of servers and decreasing costs in genotyping and sequencing have resulted in a huge quantity of genomic data pertaining to small ruminants, thanks to the above-mentioned consortia but also to breeders and their associations which have made huge datasets available to the public. Figure 2.3 compares the cost of genotyping one complete human genome in 2020, estimated by the NIH to be less than a thousand dollars to the first complete sequence of the human genome in the early 2000s, when more than 100 million dollars were paid.

Fortunately as regards the field of animal science, and dairy cattle in particular, the cost of genotyping per animal is around 30 dollars, thanks to 50K SNP panels, making the type of data generated with them standard practice in big farms (Lamb et al., 2020). Costs are slightly higher for small ruminants and therefore less applicable for use in routine breeding.

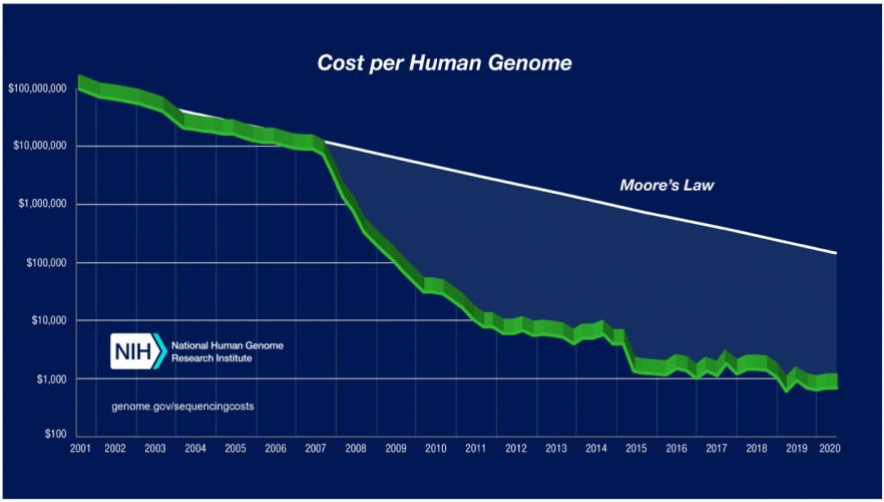


Figure 2.3 - Decreasing cost of a single human genome complete sequencing from 2000 to 2020 (www.genome.gov)

Another factor for the higher cost is the lower value of small ruminants compared to dairy cattle, but thanks to the efforts of the various international consortia, costs are gradually decreasing .Several biodiversity projects have produced a considerable amount of genomic data on small ruminant species (sheep and goats) in Italy; we are currently involved in two such projects: BIOVITA for sheep and the Italian Goat Consortium (IGC2.0) for goats. BIOVITA has genotyped 490 animals from 19 different Italian sheep breeds using medium density SNP array (Ciani et al., 2014a). The Italian Goat Consortium has used the medium 50k genotyping array to create a dataset that now counts more than 1000 animals from more than 30 goat breeds and populations throughout Italy (Talenti et al., 2017b). We are also involved in the CHEESR project (Conservation, Health and Efficiency Empowerment of Small Ruminant) led by the Italian National Association of Goat and Sheep (Assonapa).

Internationally, many projects have been developed to investigate the genomes of small ruminants through genotyping with SNP arrays and through the production of whole genome sequencing data: The international (International Sheep Genomics Consortium), the International goat genome consortium (International Goat Genome Consortium), as well as the ADAPTMAP project (ADAPTmap project), and the VarGOATS project (VarGoats), which are still in progress. A cheaper SNP chip will soon be available for goats. Meanwhile, VarGOATS has recently presented a new dataset of whole genome data from its last resequencing project, containing 1159 goat assembled genomes, of which 1134 sequences from more than 90 local and transboundary domestic populations representing 5 continents, and 35 from 8 wild goat species (Denoyelle et al., 2021).

Genomic information like the above produced by worldwide consortia is becoming fundamental for research in the animal breeding and genetics fields. It can help us, for example, to understand what markers are related to a pathology or a particular phenotypical trait (Talenti et al., 2017a) or the environmental effects on the genome of a breed or the effect of different breeding schemes and management on different populations of farmed animals (Szmatoła et al., 2019). Moreover, genomic information is a fundamental tool to understand the adaptation of local breeds to their specific breeding environment and the relationships between their genotypes and the climatic variables (Mdladla et al., 2018).

2.3.1 Landscape genomics

Landscape genomics consist in the integration of geographic, environmental and ecological data with genomic data to test for spatial patterns of genes and loci under selection and relationship with genotypes and environment (Storfer et al., 2018). This analysis, made possible by the Geographic Information Systems (GIS) (that overlays different kinds of geo-referenced information (Goodchild, 1992)) and genotyping data, help us understand how animals and their genomes evolved and adapted to a specific environment.

Identifying loci and genes responsible for adaptation can be divided into two steps: the first detects outlier loci while the second tries to relate the identified loci to environmental variables. (Li et al., 2017). Landscape genomics analysis was first applied in 2007 by Stefan Joost, developer of the individual-based Spatial Analysis Method (SAM) (Joost et al., 2007) and implemented it in the MatSAM software (Joost et al., 2008). The SAM method is based on computing multiple simultaneous univariate logistic regressions in order to identify possible association between allelic frequencies and climatic and environmental variables It is still successfully used on different species such as cattle, goats and sheep, allowing us to identify possible allelic frequencies a population related to the climatic variables of the region where animals are reared.(Figure 2.4)

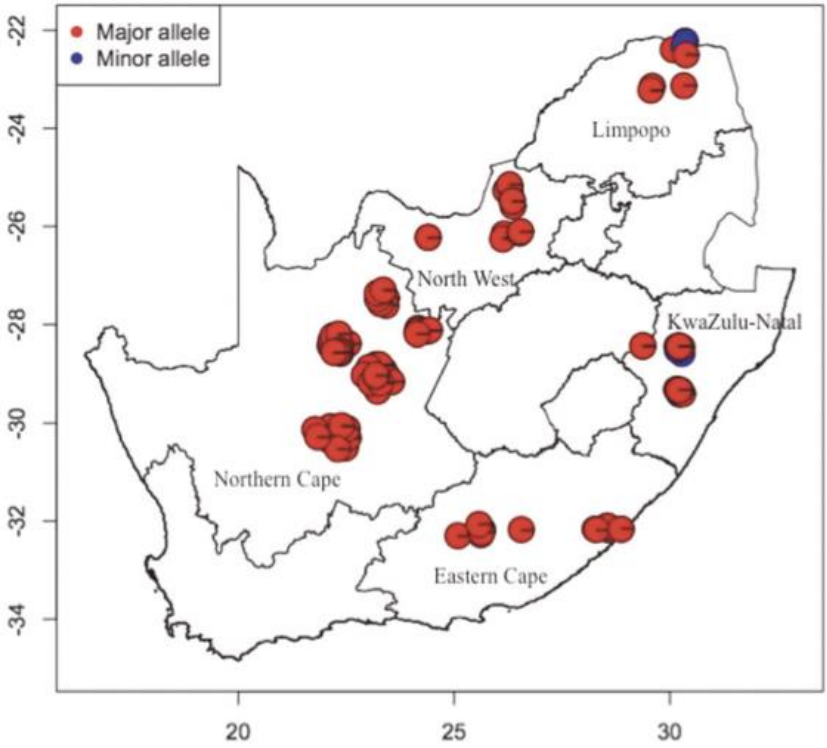


Figure 2.4 - Geographic locations of individuals with the major allele "A" (red) and minor "G" allele (blue) of the outlier SNP on a South African geographic map (Mdladla et al.,2018)

An amplified version of SAM, Samβada (Stucki et al., 2017) adds a neutral genetic structure to the method to improve its performance on huge amounts of data (Rellstab et al., 2015).

Today there are different statistical methods with their relative software to relate loci to environmental variables, methods based on mixed-effects models are among those used most. BAYENV, for example, is based on the Bayesian method (a generalized linear mixed model), and is used compute all the possible correlations between allelic frequencies in a population and the

ecological and climatic variables of the local environment after correcting for population structure and size (Günther and Coop, 2013). Another is Latent factor mixed models (LFMMs) which is built on a fast algorithm using a hierarchical Bayesian mixed model based on a variant of principal component analysis (PCA). Unlike BAYENV in this case the residual population structure in LFMMs is introduced through unobserved or latent factors (Frichot et al., 2013). Finally, there is the Spatial Areas of Genotype Probability (SPAG) that is based on SAM methods and the conditional probability theory, which allows for adaptive genotypes to be combined in different types of intergenic relationships (Rochat and Joost, 2019).

Nowadays when performing landscape genomics analyses, more than one method is used simultaneously because different assumptions on population structure and size can offer complementary evaluations on the same set of variables and genotypes.

2.3.2 Selection signatures

Human needs and environmental adaptation have dictated changes in the morphology, physiology, and behavior of domesticated animals. This process, which has modified the allele frequencies in populations over the years (Gouveia et al., 2014), has left traces on the genome of these species, i.e. signatures that can be detected with data derived from genotyping with SNP array and whole genome sequencing.

Changes in allele frequency, whether due to natural or artificial selection, can be considered as positive, balancing, or negative. Positive selection increases the frequency of a favorable allele inside a population until fixation is reached, while negative selection eliminates a negative allele from the population by progressively reducing its frequency. Balancing selection instead occurs when the heterozygous phenotype is the most valuable and then it is desirable to keep equal frequency of the two different alleles in the populations. This processes not only affect the single specific locus considered by the selection but also, because of the hitch-hiker effect , it consequently affect the frequencies of all the loci that are in linkage disequilibrium with the locus of interest. This kinds of modifications inside the genome are called “selection signatures” (Passamonti et al., 2021).

To identify selection signatures using SNP marker data, several statistical methods have been developed. Some methods detect macroevolution phenomena (which involves traits that are conserved among clades of a phylogeny lineage and which then can be traced back to ancient events), while other methods detect microevolution phenomena (caused by the rapid

prevalence, or the fixation, of a beneficial allele in a population in a more recent period of time) (Vitti et al., 2013).

Livestock science is primarily concerned with microevolution to identify the basis of adaptation and productive trait architecture in different populations and breeds. The principal methods for microevolution can be divided into 5 groups, according to the approach used in (Saravanan et al., 2020) (Figure 2.5):

- Methods, such as Tajima's D (Fu, 1997), Fay and Wu's H statistic (Fay and Wu, 2000), and composite likelihood ratio test (CLR) (Kim and Stephan, 2002) based on site frequency spectrum, which relay on distribution of allele frequency in a population;
- Methods, such as Extended Haplotype Homozygosity (EHH) (Sabeti et al., 2002); the integrated haplotype score (IHS) (Voight et al., 2006) and the linkage disequilibrium decay test (LLD) (Amaral et al., 2008) based on linkage disequilibrium and on the identification of high frequency haplotypes in population;
- Methods, such as Runs of Homozygosity analysis (ROH) (McQuillan et al., 2008) and pooled hetero- zygosity (HP) (Rubin et al., 2010) based on reduced local variability, which identifies genomic regions with very low variability;
- Methods, such as FST index (Wright's fixation index) (Wright, 1949); the FLK (extended Lewontin-Krakauer test) (Bonhomme et al., 2010) based on single site population-differentiation, which compares frequencies at single loci in two populations;
- Methods, such as the cross-population EHH (XP-EHH) (Yang et al., 2016) and the HapFLK (haplotype-based FLK) (Fariello et al., 2013) based

on haplotype differentiation, which compares information on different haplotypes inside to populations;

Another important distinction can be made between methods that are performed intra populations like ROH, iHS and Tajima's D, and methods that are performed inter populations like XPEHH and FST (Saravanan et al., 2020).

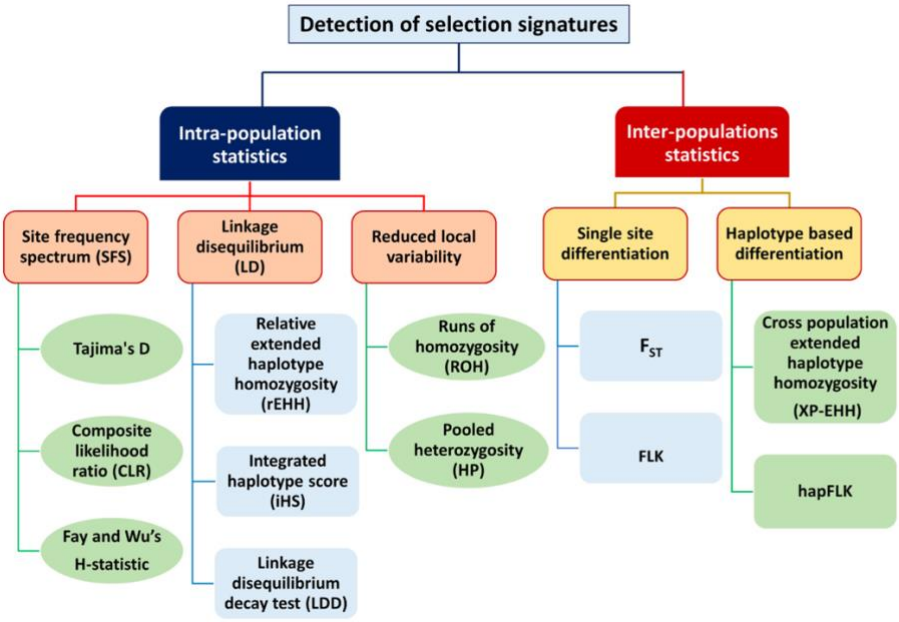


Figure 2.5- Classification of different approaches commonly used for the detection of selection signatures in livestock populations (Saravanan et al., 2020)

Yet another way to classify methods is whether they identify strong changes ('hard sweeps') that can fix a specific allele, such as iHS or events that slightly decreased genome variability ('soft sweeps'), such as HapFLK (Vatsiou et al., 2016). Consequently, when performing selection signature

analysis, combining two or more of these methods enables every signal to be considered, in order to determine if the signal is pertinent or not (Talenti et al., 2017a).

2.3.3 Runs Of Homozygosity

ROH calculations are one of the most frequently used and studied techniques in animal science. First discovered in a study on the human genome in 1999 (Broman and Weber, 1999) and related to possible negative implications in human health ROH is a long and contiguous sequence of homozygous haplotypes in the genome of one individual (Ceballos, 2018). In 2006 the first study on ROH in humans was performed using SNP markers (Gibson et al., 2006), and finally in 2010 it was also adopted for livestock (Maja Ferencakovic et al., 2011)

ROH provide information on individuals but also on the breed or the population to which they belong if the presence of ROH is compared in different individuals. Natural and artificial selection in individuals and populations can be identified due to regions in the genome with high levels of homozygosity and a high presence of long starches of homozygous genotype (also known as "ROH islands"). These indicate interesting adaptative or productive traits as well as specific genes inherited from parents (Peripolli et al., 2016). Unfortunately, as we have seen in human beings studies, not only does ROH contains positive traits and genes, but also the presence of negative homozygous mutations (Nothnagel et al., 2010). Moreover, the high presence of ROH in a single individual is an indicator of the level of inbreeding that has occurred when the parents of that individual are closely related due to artificial selection or for other factors like small population size, bottleneck or genetic drift (McQuillan et al., 2008). ROH not only indicates the level of inbreeding in a population but also the time that inbreeding event occurred.

Short ROH are indicative of earlier inbreeding events while long ROH are related to more recent ones because they are less likely to have been broken or interrupted by mutation or other occurrences during meiosis (Kirin et al., 2010). The generation number at which the ROH-linked inbreeding event occurred can also be estimated based on its exact length in megabases (Onzima et al., 2018).

Detection of ROH today starts from SNP marker data that could have different densities and distances along the genome They could be detected both from genotyping with SNP array with different density (from 50k to 800k) and from whole genome sequencing that could count millions and millions of SNP variants for single individual. Naturally using SNP array data instead of whole genome sequencing data can influence the detection of ROH because a higher SNP number and a higher SNP density in certain regions of the genome increase the possibility of braking a ROH and reducing the presence of the longest one (Ceballos, 2018). This leads to one of the most problematic aspects of ROH detection, which is the absence of standard parameters for defining a ROH. Indeed we need different parameters to define a ROH, such as the minimum length of the ROH, the minimum number of consecutive SNP, the minimum SNP density allowed, the maximum gap between two consecutive ROH, the number of missing SNP per ROH and the number of heterozygous SNP allowed. Unfortunately all these parameters are not universally defined and can vary based on the species we are considering, the structure of the population, the density of the SNP chip we are using or the use of whole genome sequencing (Meyermans et al., 2020).

Table 2.1 - Comparison of pre-set parameters for identification and characterization of ROH in different animal species (Peripolli et al., 2017)

Author	Species	Software	SNP array	Consecutive SNPs/ROH[1]	Density[2] (SNP/kb)	Maximum gap[3] (kb)	Minimum length[4] (kb)	Heterozygous SNP/sliding window	Missing SNP/sliding window
Ferenčaković *et al.* (2011)	Cattle: dual purpose	FORTRAN 90	Illumina Bovine SNP 50 K BeadChip	15	–	–	1000	00	–
Purfield *et al.* (2012)	Cattle: beef, dairy and dual purpose[5]	PLINK v1.07	Illumina BovineHD Genotyping BeadChip assay	58	1/50	100	500	01	02
Purfield *et al.* (2012)	Cattle: beef, dairy and dual purpose[5]	PLINK v1.07	Illumina Bovine SNP 50 K BeadChip	No restriction	1/120	1000	500	01	02
Bjelland *et al.* (2013)	Cattle: dairy	PLINK v1.07	Illumina Bovine SNP 50 K BeadChip	30	–	–	–	00	01
Ferenčaković *et al.* (2013a)	Cattle: dairy and dual purpose[5]	SNP & VARIATION SUITE v7.6.8	Illumina Bovine SNP 50 K BeadChip	15	1/1000	1000	1000	00	05
Karimi (2013)	Cattle: beef, dairy and dual purpose[6]	SNP & VARIATION SUITE v7.6.8 PLINK v1.07 CGATOH	Illumina BovineHD Genotyping BeadChip assay	30	1/50	250	1000	01	04
Biscarini *et al.* (2014a)	Cattle: dairy	PLINK v1.07	Illumina Bovine SNP 50 K Beadchip	No restriction	–	1000	–	01	05
Marras *et al.* (2014)	Cattle: beef, dairy and dual purpose[5]	SAS 9.2 script (SAS Institute 2012)	Illumina bovine SNP 50 K BeadChip	15	–	1000	1000	00	00
Scraggs *et al.* (2014)	Cattle: beef	PLINK v1.07	Illumina bovine SNP 50 K BeadChip	50	–	–	1000	01	01
Mészáros *et al.* (2015)	Cattle: dual purpose	–	Illumina Bovine SNP 50 K BeadChip	30	–	–	1000	00	00
Williams *et al.* (2015)	Cattle: kept for conservation	R Development Core Team (2008)	Illumina BovineHD Genotyping BeadChip assay	–	1/50	100	100	01	02
Zavarez *et al.* (2015)	Cattle: beef	SNP & VARIATION SUITE v7.6.8	Illumina BovineHD Genotyping BeadChip assay	30	1/100	500	4000	02 (ROH ≥ 4 Mb)	05
Zavarez *et al.* (2015)	Cattle: beef	SNP & VARIATION SUITE v7.6.8	Illumina BovineHD Genotyping BeadChip assay	30	1/100	500	500	00 (ROH < 4 Mb)	05
Bosse *et al.* (2012)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	20	1/1000	–	10	01	–
Ai *et al.* (2013)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	–	–	–	500	01	05
Herrero-Medrano *et al.* (2013)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	20	1/1000	1000	10	01	–
Silió *et al.* (2013)	Swine	SNP & VARIATION SUITE v7.6.8	Porcine SNP60 Beadchip	30	1/100	1000	1000	–	02
Saura *et al.* (2015)	Swine	FORTRAN	Porcine SNP60 Beadchip	30	1/100	1000	–	01	02
Zhang *et al.* (2014)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	10	1/500	1000	5000	–	01
Traspov *et al.* (2016)	Swine	PLINK v1.09	Porcine SNP60 Beadchip	–	–	–	500	01	05
Khanshour (2013a)	Horse	PLINK v1.07	Equine SNP50 BeadChip	50	1/50	1000	500	–	–

Author	Species	Software	SNP array	Consecutive SNPs/ROH[1]	Density[2] (SNP/kb)	Maximum gap[3] (kb)	Minimum length[4] (kb)	Heterozygous SNP/sliding window	Missing SNP/sliding window
Al-Mamun *et al.* (2015)	Sheep	PLINK v1.07	Illumina Ovine SNP50 BeadChip	–	–	250	500	01	02
Muchadeyi *et al.* (2015)	Sheep	PLINK v1.07	Illumina Ovine SNP50 Beadchip	20	1/50	500	–	00	02
Guangul (2014)	Goat	CGATOH	47 K SNP bead chip	20	–	1000	1000	01	05

Different studies have tested different sets of parameters in different species, identifying useful and valid parameters but slightly different ones from those reported in Table 2.1 (Peripolli et al., 2016).

Software abounds for estimating ROH starting from SNP chip data, like PLINK, SNP & variation suite and R packages, or from whole genome sequence data like VCFtools /BCFtools. Their internal parameters have to be taken into account as well. Despite varying threshold definitions, the great variety of software available to calculate ROH have made them a sort of gold standard for many genomic analyses in a great number of animal field studies, above all for calculating genomic inbreeding (Gorssen et al., 2021).

2.3.4 Inbreeding and management

Genomic data is also used in animal management and breeding. SNP markers allow breeders and their associations to improve estimates, whether for inbreeding calculations or genomic estimated breeding values (GEBV) (Maltecca et al., 2020).

Parentage information annotated in 'herd' or 'stud' books date back to the Robert Bakwell in the 18th century, to manage the mating between animals. The first herd book ever was the General Studbook, created in 1791 to record thoroughbred horses (Van Vleck and Pollak, 1987).

Traditional inbreeding values or F_{PED} are calculated from pedigree information contained in the herd books of different breeds. Unfortunately, the record of parentage reported by breeders sometimes contained inconsistencies or even mistakes. Parentage assessment for small ruminants was even more difficult (Islam et al., 2020). Genomic data can solve these problems and assure more robust inbreeding values that improve estimates and correct possible mistakes in pedigree. Genomic data can also employ Mendelian sampling for a more precise estimate of inbreeding values, even between close relatives (Leutenegger et al., 2003). There are different methods based on different metrics like maximum likelihood (Milligan, 2003), the diagonal elements of a genomic relationship matrix, (VanRaden, 2008) or the homozygosity-by-descent genomic segments (Druet and Gautier, 2017). One of the most used and easy to calculate relies on the Runs of homozygosity (ROH) and it is called F_{ROH}. It consists in the evaluations of the percentage of the genome of an individual covered by ROH, taking into

account chromosome length and the first and last SNP markers for each of them (McQuillan et al., 2008).

Evaluating inbreeding values of individual animals as correctly as possible still is a fundamental part for animal management, because incorrect values could negatively affect fitness as well as productive and reproductive traits causing "inbreeding depression" (Baes et al., 2019). Indeed, the high level of inbreeding inside a population increases the frequency of possible negative mutations that influence different traits. Deleterious recessive alleles generated by recurrent mutations are usually present in the population at the heterozygous state and selection, genetic drift, and other practices related to animal breeding increase the frequency of homozygotes, resulting in negative effects (Falconer and Mackay, 1996).). ROH are regions with big loads of potentially deleterious mutations inherited from common ancestors, and so them must be carefully considered when evaluating inbreeding depression (Sumreddee et al., 2020). Today many studies analyze inbreeding depression, particularly in cattle, the domesticated species subjected to mating and selection schemes for the longest time (Doekes et al., 2020; Makanjuola et al., 2020).

Lastly, genomic information in animal breeding and management is used in the evaluation of inbreeding values. Genomic information can be employed for canonical evaluation by using the GBLUP method, which is based on the integration of canonical pedigree and phenotypical information with the genotype of individual animals, derived from high density SNP array (Meuwissen et al., 2001). As a result, the field of animal breeding is improved, first of all by reducing the generation interval. An immediate prediction can be made of an animal's breeding value and the consequent

improvement to its population, without waiting to observe the phenotype of its offspring. This is because we link the expression of certain phenotypes to specific genes and regions of the genome and then we assigned a score to that area of the genome based on its influence on the phenotypical or morphological traits we are considering. This method allows us to predict an individual's value for those specific traits only a few days after its birth (Hayes et al., 2009).

CHAPTER 3 My works

3.1 Landscape genomics

3.1.1 Aims

Understanding the phylogeny of different populations and how they have adapted to the environments in which they live today is certainly one of the main goals in the field of modern animal genetics and genomics. Nowadays, it is essential to understand what are the genetic bases of the adaptation to different climates, especially in the face of the climate change expected in the next decades. Small ruminants in particular represent an excellent model for this task, them being found in a wide variety of environments all over the Earth from the arid Sahara Desert to the coldest Tibetan plateaus.

With this work we intended to investigate the panorama of Italian goat breeds using the largest and most complete dataset available to date. The aims were to search for new possible patterns of genetic similarity between the various autochthonous Italian breeds and to find markers and genes related to their adaptation to the great variety of climates that the Italian peninsula offers. We know that in 50 years most of our breeds will be bred in very different climatic conditions from the ones in which they find themselves today. For this reason, a greater knowledge about how they have adapted to different environments will be of help in safeguarding our biodiversity.

Our findings can therefore be considered the starting point for this type of study on the autochthonous breeds of small ruminants in our country. As you will see below, we have detected a genetic pattern in our native breeds along the peninsula that reflects the geopolitical situation of the last few centuries

and highlighted some SNP markers related to environmental variables in the different breeds. We also tried to estimate the future allelic frequencies of these SNPs in the different populations according to the predicted climate change of the regions in which they live.

3.1.2 The climatic and genetic heritage of Italian goat breeds with genomic SNP data

Article published on Scientific Reports 2021, 11 (1), 10986. doi: 10.1038/s41598-021-89900-2

Authors and affiliations

Matteo Cortellari[#1], Mario Barbato[#2], Andrea Talenti[*1,3], Arianna Bionda[1], Antonello Carta[4], Roberta Ciampolini[5], Elena Ciani[6], Alessandra Crisà[7], Stefano Frattini[1], Emiliano Lasagna[8], Donata Marletta[9], Salvatore Mastrangelo[10], Alessio Negro[1], Ettore Randi[11], Francesca M. Sarti[8], Stefano Sartore[12], Dominga Soglia[12], Luigi Liotta[13], Alessandra Stella[14], Paolo Ajmone-Marsan[2], Fabio Pilla[15], Licia Colli[2] and Paola Crepaldi[1]

[1] Dipartimento di Scienze Agrarie e Ambientali – Produzione, Territorio, Agroenergia, Università degli Studi di Milano, Via Celoria 2, Milan, 20133, Italy

[2] Dipartimento di Scienze Animali, della Nutrizione e degli Alimenti and BioDNA Centro di ricerca sulla Biodiversità e sul DNA Antico, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy

[3] The Roslin Institute, University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, United Kingdom

[4] Unità di Ricerca di Genetica e Biotecnologie, Agris Sardegna, 07100, Sassari, Italy

[5] Dipartimento di Scienze Veterinarie, Università di Pisa, Viale delle Piagge 2, 56124 Pisa, Italy

[6] Dipartimento di Bioscienze Biotecnologie e Biofarmaceutica, Università degli Studi di Bari, Via Orabona 4, 70126 Bari, Italy

[7] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) - Research Centre for Animal Production and Acquaculture, 00015 Monterotondo, Roma, Italy

[8] Department of Agricultural, Food and Environmental Sciences, University of Perugia, 06121 Perugia, Italy.

[9] Department of Agriculture, Food and Environment, University of Catania, Via Valdisavoia 5, 95123 Catania IT

[10] Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, 90128 Palermo, Italy

[11] Department of Chemistry and Bioscience, Faculty of Engineering and Science, University of Aalborg, Aalborg, Denmark.

[12] Dipartimento di Scienze Veterinarie, Università degli Studi di Torino, largo Braccini 2, Grugliasco, 10095, Italy

[13] Dipartimento di Scienze Veterinarie, University of Messina, Messina, Italy

[14] Institute of Biology and Biotechnology in Agriculture, National Research Council (CNR), Milan, Italy

[15] Dipartimento Agricoltura, Ambiente e Alimenti Universitá degli Studi del Molise 86100 Campobasso IT

*Corresponding author

[#]equal contributors

Abstract

Local adaptation of animals to the environment can abruptly become a burden when faced with rapid climatic changes such as those foreseen for the Italian peninsula over the next 70 years. Our study investigates the genetic structure of the Italian goat populations and links it with the environment and how genetics might evolve over the next 50 years. We used one of the largest national datasets including >1,000 goats from 33 populations across the Italian peninsula collected by the Italian Goat Consortium and genotyped with over 50k markers. Our results showed that Italian goats can be discriminated in three groups reflective of the Italian geography and its geo-political situation preceding the country unification around two centuries ago. We leveraged the remarkable genetic and geographical diversity of the Italian goat populations and performed landscape genomics analysis to disentangle the relationship between genotype and environment, finding 64 SNPs intercepting genomic regions linked to growth, circadian rhythm, fertility, and inflammatory response. Lastly, we calculated the hypothetical future genotypic frequencies of the most relevant SNPs identified through landscape genomics to evaluate their long-term effect on the genetic structure of the Italian goat populations. Our results provide an insight into the past and the future of the Italian local goat populations, helping the institutions in defining new conservation strategy plans that could preserve their diversity and their link to local realities challenged by climate change.

Introduction

The preservation of animal genetic diversity is fundamental to ensure food security and the development of farming communities (Hoffmann, 2010). Among the key factors shaping genetic and phenotypic diversity there is climatic and environmental heterogeneity (Lv et al., 2014), with indigenous domestic breeds showing better adaptation to local environments than highly productive breeds kept in controlled farming systems in which the effects of climatic challenges are minimized (Pariset et al., 2009; Bruford et al., 2015; Barbato et al., 2020). In this rapidly evolving situation, species that are mostly reared in marginal rural areas of the world such as goats are also more likely to be among those most affected by environmental changes (Nicoloso et al., 2015).

The Italian territory is characterized by a rich environmental diversity, spanning from the polar climate of the Alps to the Mediterranean climate of the south and isles (Fratianni and Acquaotta, 2017). This environmental richness is paired with a reservoir of genetic resources for the caprine species counting over 36 goat breeds registered by the National Goat and Sheep breeds Association (http://www.assonapa.it). Italian goat genetic resources are managed throughout diverse farming systems ranging from intensive and semi-intensive to traditional grazing and transhumance. Importantly, most of them pertain to marginal areas where they play a crucial socio-economic role, contributing to the management of landscapes, biodiversity preservation, and the production of niche traditional products (Marino et al., 2016). The Italian caprine diversity and adaptation to climate has been previously investigated through genotype data analyses (Nicoloso et al., 2015; Bertolini et al., 2018b).

However, no previous work included breeds from the central areas of the country, which hosts several minor, local, and niche populations (Talenti et al., 2017b). Importantly, the latter are specifically adapted to the broad range of Italian eco-climatic scenarios, making them an ideal model to investigate genetic adaptation to climate, which only few studies tried to address in the goat species (Kim et al., 2016; Mdladla et al., 2017; Stella et al., 2018).To investigate the link between territory and genetics, we analyse the second release of the goat genotype dataset assembled by the Italian Goat Consortium, a collaboration across Italian universities that aims to enhance the understanding of the Italian goat genetic variability. This new release of the dataset, called IGC2, expands the previous version, improving the sampling coverage of the central-southern goat populations and completes the whole Italian panorama and its internal connections (Supplementary Figure 3.5). To the best of our knowledge, IGC2 currently represents the largest national-level genotyping effort on goat biodiversity(Talenti et al., 2017b).

The most recent climate predictions by the Koppen-Geiger climate classification foresee hotter and drier climate across the Italian peninsula over the next 70 years (Beck et al., 2018). Such changes will likely affect locally adapted populations by reducing food availability (e.g., pasture and forage crop availability and quality (Nardone et al., 2010)), increase temperature-related health problems (illness, death rates, increased diffusion of vector-borne diseases and parasites), and cause metabolic problems (decreasing productive and reproductive performance or depressing feed intake (Marino et al., 2016; Rojas-Downing et al., 2017)). Locally adapted breeds will be the

hardest hit, mostly relying on grazing (Rischkowsky and Pilling, 2007; Nardone et al., 2010).

We performed genome-wide analyses on the IGC2 dataset to investigate the genetic structure of Italian goat breeds with particular attention to the newly sampled local populations, and link the heritage and genomic structure of current populations with the present and future climatic condition of the rearing areas. Our results will help to understand their environment-driven adaptation and draw effective management plans to face climate change (Mdladla et al., 2018; Rochat and Joost, 2019).

Results and Discussion

Genotyping control and datasets creations

After filtering the initial raw dataset of 1,071 goats for poor quality genotype and related animals we obtained the dataset used for the haplotype sharing analysis, which included 980 animals and 48,396 SNPs. This dataset was further pruned for linkage disequilibrium (LD; $r^2 < 0.2$).

We first balanced the number of animals in the different populations by reducing the size of the nine largest groups, leaving 42,088 SNPs and 802 individuals. We used this dataset to perform population structure analyses (Multi-Dimensional Scaling (MDS), Admixture, and Reynolds distances).

Upon the removal of 2nd-degree related individuals and animals without geographical coordinates, we retained 41,898 SNPs and 489 individuals for Landscape Genomics analyses. See Table 3.1 for the detailing of the different datasets.

Table 3.1 -Composition of the datasets used for the different analysis, with names, codes and number of samples processed and filtered for each population and grouped by the type of analysis.

Breeds ID	Breeds name	Raw dataset (n animals)	Haplotype sharing (n animals)	Population structure (n animals)	Landscape Genomics (n animals)
ALP	Camosciata delle Alpi	143	117	30	43
ARG	Argentata dell'Etna	48	46	30	41
ASP	Capra dell'Aspromonte	24	24	24	18
BEZ	Bezoar (outgroup)	7	7	7	0
BIA	Bianca Monticellana	24	23	23	17
BIO	Bionda dell'Adamello	24	24	24	22
CAP	Capestrina	24	22	22	15
DDS	Derivata di Siria	32	25	25	11
FAC	Facciuta della Valnerina	24	24	24	12
FUL	Fulva del Lazio	22	20	20	14
GAR	Garganica	40	37	30	3
GCI	Grigia Ciociara	43	39	30	18
GIR	Girgentana	59	56	30	9
GRF	Garfagnina	28	25	25	18
JON	Jonica	16	15	15	3
LIV	Capra di Livo	24	22	22	19
MAL	Maltese	16	16	16	10
MES	Messinese	24	23	23	21

MNT_M	Montecristo (mainland)	18	12	12	1
MNT_I	Montecristo (island)	24	23	23	1
MON	Capra di Montefalcone	24	23	23	13
MXS	Incrocio Maltese e Sarda	36	36	30	0
NIC	Nicastrese	24	24	24	14
NVE	Nera di Verzasca	19	19	19	11
ORO	Orobica	23	23	23	5
RCC	Roccaverano	28	28	28	25
RME	Rossa Mediterranea	46	40	30	13
SAA	Saanen	44	44	30	20
SAM	Maltese sampled in Sardinia	15	15	15	5
SAR	Sarda	33	33	30	32
TER	Capra di Teramo	43	30	30	8
VAL	Valdostana	24	24	24	16
VLS	Vallesana	24	17	17	7
VPS	Capra della Val Passiria	24	24	24	24
TOTAL Animals		1071	980	802	489

Population structure

The MDS plot showed a north-south geographic gradient comparable with previous findings on Italian goat population structure (Nicoloso et al., 2015). The first MDS component identified three main groups corresponding to northern Italian, central-southern Italian, and Maltese populations. The

second MDS component discriminated the insular Montecristo goat (MNT_I; Figure 3.1-A) from the other mainland breeds, likely due to the high inbreeding and prolonged geographical isolation (Somenzi et al., in preparation). For this reason, we excluded the two Montecristo populations (MNT_M and MNT_I) from the subsequent population structure and haplotype sharing analysis and to repeat the analyses without them. The new MDS plot without the two MNT populations (Figure 3.1-B) still separated the three main groups on the first component, a structure further supported by the bootstrapped Reynolds' distances phylogenetic tree (Supplementary Figure 3.6).

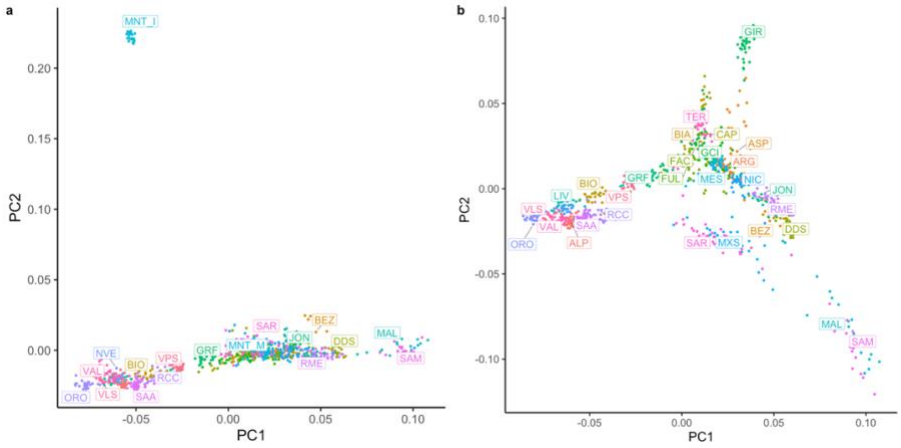


Figure 3.1- A) MDS plot representing all the populations contained in the dataset, B) MDS plot without the two Montecristo populations (MNT_M e MNT_I)

Although overlapping a previous investigation of the Italian caprine population structure (Nicoloso et al., 2015), our improved dataset identified a closer relationship between the central and southern Italian population, more in accordance with the recent known history and geography of the

Country. Until 1860 Italy was divided in many states with tight connections to other European kingdoms (https://www.150anni.it). The north-western part of the country and Sardinia were part of the Sardinian kingdom, tightly connected with the French empire, whereas the north-east part (the Kingdom of Lombardy – Venetia) was under the political influence of the Austrian Empire. Central Italy was ruled by the Papal state, and southern Italy and Sicily were under the Kingdom of the two Sicilies ruled by the Borbone (Figure 3.2) (Riall, 2002).

The ADMIXTURE analysis (Supplementary Figure 3.7) at $K = 2$ separates the Maltese populations (purple component) and the Northern Italy breeds (yellow component), and improves the representation of the North-South gradient over previous studies on Italian goat populations (Nicoloso et al., 2015). At $K = 3$ it resembles the MDS plot distinguishing the central-southern Italian breeds led by the Girgentana (GIR; blue component) and the mean proportion for each breed overlap nicely with the political borders of Italy prior 1860 (Figure 3.2). Each K above 3 distinguishes single or groups of breeds, such as Teramana (TER; $K = 4$) and Valdostana (VAL; $K = 5$). The lowest cross-validation error was recorded at $K = 20$ (Supplementary Figure 3.8) and showed the similar genetic background of those breeds originated from the same geographical regions (north, central, south and Maltese), and some breeds identified by private clusters, once again confirming the uniqueness of GIR, ORO, VAL, TER and SAM, among others (Supplementary Figure 3.7).

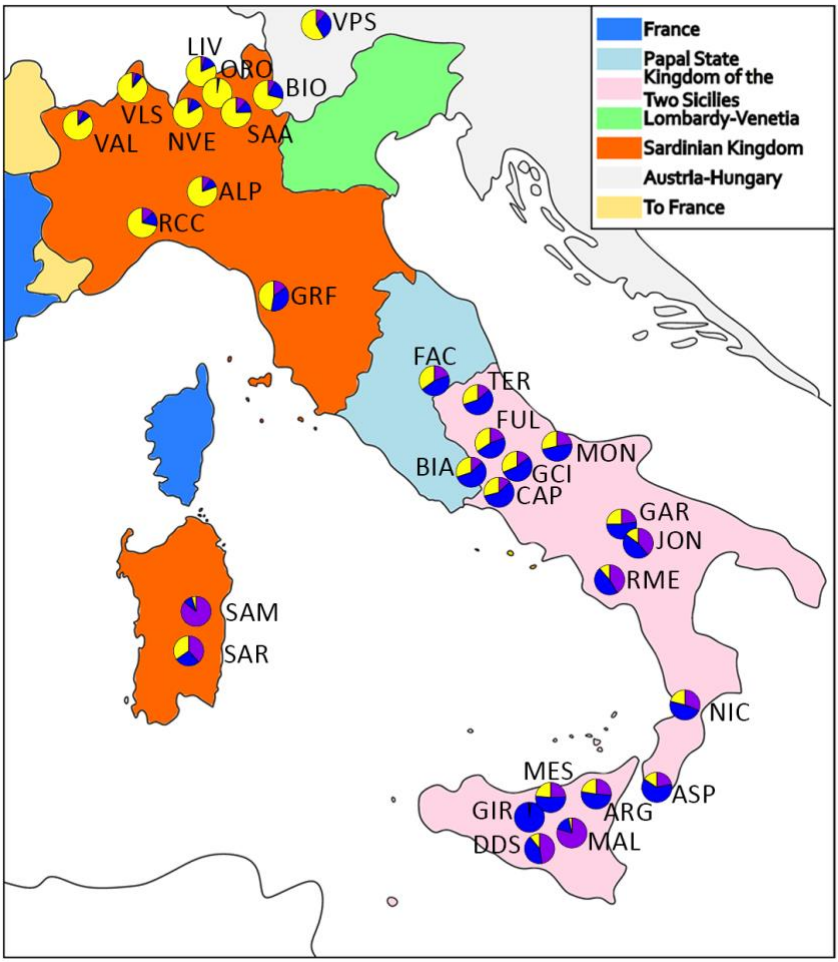


Figure 3.2 - Historical map of Italy pre-unification with pie chart of the ADMIXTURE K = 3 values plotted on the geographical coordinates of the mean sampling location of each breed. The colours reflect the ADMIXTURE component associated with the Maltese (purple), Northern Italy Cluster (yellow), and Central-Southern Cluster (blue). BEZ and MXS were not represented due to the lack of specific geographic coordinates. MNT_I and MNT_M were excluded as extreme outliers. Map was generated in Inkscape v1.0 (https://inkscape.org/); the pie charts were created using R (R Development Core Team, 2011).

The haplotype sharing analysis across populations (Figure 3.3) also highlights the three genetic groups corresponding to admixture K =3 and consistently with the administrative and temporal history of the Italian Peninsula until 1860 (Riall, 2002).

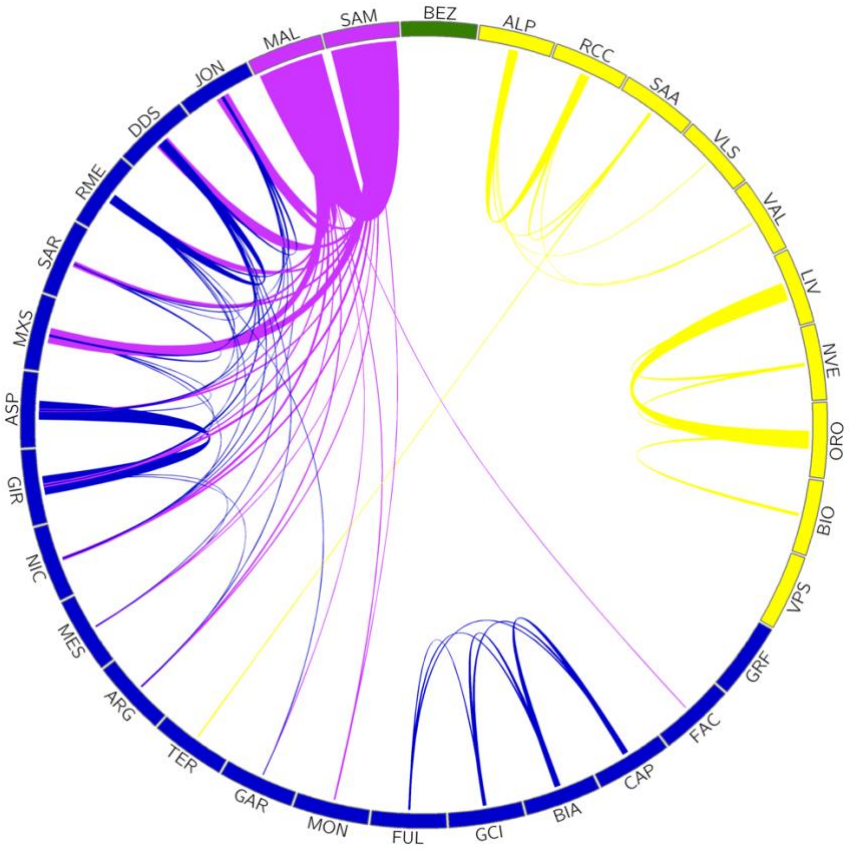


Figure 3.3 - Proportion of the median haplotype shared among the Italian goat populations. The colours reflect the ADMIXTURE components at K = 3, which overlap with the administrative and temporal history of the Italian Peninsula until 1860. The outgroup is highligh in green (extended name reported in Table 1). The figure was generated using Circos v0.69-8 Software (Krzywinski et al., 2009)

We observe that the Northern-Italian populations (yellow cluster) show no haplotype exchange with the other clusters, with the exception of SAA and TER probably due to a recent introgression event. Within the Northern-Italian cluster there is a more pronounced haplotype sharing among the Lombardy breeds (ORO, NVE, LIV and BIO) than among those from the rest of the Alps. The Val Passiria (VPS) together with the Garfagnana (GAR) are the only two populations that do not exchange haplotypes at all, perhaps suggesting a geographical and/or political isolation. Populations from Central-Southern Italy (blue cluster) show large haplotype sharing within and among different clusters, possibly due to breeding and management practices as well as local geographical conditions, such as breeds from the Lazio region (BIA, GCI, CAP, and FUL) have high haplotype sharing among themselves. Lastly, the populations from the isles and in particular the Maltese (MAL and SAM, purple cluster) and Sarda (SAR and MXS) are those that mostly shared haplotypes with all other southern breeds, probably as a consequence of their high productivity and diffusion over the territory. The green color represents the outgroup *Capra aegagrus* that does not exchange haplotypes with any of the other breeds. Importantly, future investigations with dedicated experimental designs aimed to dissect the different effects of selection might aid unfolding the undergoing evolutionary dynamics.

The political subdivision of Italy preceding the unification of the country has probably contributed to maintain the ancient genetic flows from central-north Europe in the north of the country and from Africa and Spain in the south (Stella et al., 2018), with only a minor impact on the population structure of the following 150 years of history of the country.

Landscape genomics

The landscape genomics analyses (LGA) were performed using the climatic variables representing the current climate applying two different approaches: Samβada (Stucki et al., 2017) and LFMM (Caye et al., 2019). We observed no direct overlap between the two methods. However, this is not surprising as simulation studies showed that LFMM is overall more conservative than Samβada, and the two methods tend to have marginal overlap on co-selecting the same signals, with the most significant loci detected by Samβada ignored by LFMM (Caye et al., 2019). Samβada identified 252 genotypes belonging to 216 different SNPs significantly associated (FDR <0.05) with at least one climatic variable (Supplementary Table 1). Among them, 75 SNPs mapped within a gene region annotated in the goat genome (ARS1.2), identifying a total of 62 different genes associated with at least one of the following four representative environmental variables: "Isothermality" (47 genes), "Mean diurnal range" (four genes), "Mean temperature wettest quarter" (three genes) and "Precipitation coldest Quarter" (11 genes) (Supplementary Table 2). Some of these genes had already been identified in other landscape genomics works in relation with different environmental variables, for example ANK3 and BTRC in relation to longitude, and RYR3 with Mean Temperature of the wettest quarter (BIO3) (Mdladla et al., 2018). The DCLK1 gene, in particular, was found in association with the continental goat group compared to the rest of the world (Bertolini et al., 2018b). Details on correlations among representative and excluded variables are shown in Supplementary Table 3.

Initially, we investigated the role in biological pathways of the 62 genes identified by Samβada (Supplementary Table 2), splitting them according to

the associated environmental variable. We identified only one significant pathway ("Circadian rhythm related genes WP3594"; adjusted P-value <0.0045) for those genes associated with "Mean diurnal range" with two genes linked to the circadian clock regulation (*MAPK9* (Yoshitane et al., 2012)) and to hair follicle formation and hair growth in Cashmere goat (*NTKR3* (Liu et al., 2013)).

We also analysed the 62 genes individually to better understand their function. Using the information found, we can clump the most interesting genes into four groups based on the phenotype they affect the most: 1) meat- and growth-related genes, 2) circadian rhythm-related genes, 3) fertility-related genes, and 4) inflammatory response genes.

The first group (meat and growth) is the largest and counts 24 genes, including *HADC9*, which has a role in the feedback inhibition of myogenic differentiation in sheep muscle (Fleming-Waddell et al., 2009), *DLG1*, that is related to adipogenesis and residual feed intake in cows (Seabury et al., 2017), and *KLF12*, which is related to the formation of preadipocytes in goats (Xu et al., 2020). The second group (circadian rhythm) includes 12 genes, such as *MAPK9* and *EYA3*, both related to melanin production and photoperiod regulation (Dupré et al., 2008), and *KCNJ1*, associated with the production of polyunsaturated fatty acid (PUFA) and feed efficiency in cattle (Ibeagha-Awemu et al., 2016; Barbato et al., 2020; Crislip et al., 2020). The third group (fertility-related) includes 15 genes such as *BTRC*, whose mutations can affect spermatogenesis and mammary gland development in mice (Van Den Berg et al., 2014), *PRKD1*, associated to age at puberty in pigs (Nonneman et al., 2016), and *DENND1A*, related to anti-Mullerian hormone and superovulation in dairy cows and to polycystic ovarian syndrome in human

(Nawaz et al., 2018). Finally, the fourth group (inflammatory response) includes eight genes such as *BTLA*, strongly related to rheumatoid arthritis (Lin et al., 2006). This last gene in particular is relevant as a candidate for one of the most relevant infective diseases of goats worldwide, the Caprine Arthritis Encephalitis Virus (CAEV). This virus belongs to the *Retroviriade* virus family, like the human immunodeficiency virus (HIV), and has rheumatoid arthritis among its principal symptoms (Colussi et al., 2019; Schultz et al., 2020). Due to the CAEV importance and the relevance of climatic factors and their change play into pathogens diffusion (Epstein, 2001), this group of genes becomes a potential candidate for studies on livestock resilience to incoming climate challenges.

LFMM identified four SNPs significantly associated (FDR <0.05) with three different climatic variables (Mean Diurnal Range, Mean Temperature Wettest Quarter, and SlopeP) , two of which intercepting NBEA, a gene located within a region involved with wool production in sheep (Wang et al., 2014) and previously associated with continental goat group in the work of (Bertolini et al., 2018b), and the RHOBTB1 gene that is known to be associated to meat quality in cattle (Silva et al., 2020) (Supplementary Table 4).

Future genotypes prediction

The data collected on the current Koppen-Geiger climate classification showed that 21 Italian breeds live in “Temperate” regions, eight in “Cold” regions (BIO, SAA, VLS, TER, MNT_M, LIV, ORO, VPS), two in “Arid” regions (GAR, MNT_I), and one in a “Polar” region (VAL; Figure 3.4-A). BEZ and MXS were not considered for the analysis due to lack of

georeferenced information. If we compare the current Koppen-Geiger classification of their breeding areas with the future predictions (Figure 3.4-B), we observe that, in 70 years from now, only 11 breeds will live in regions that will not change their classification. Such a scenario will likely pose new threats to those populations living in colder climates, whereas those breeds coming from the warmer parts of the country might have a chance to expand their range, with direct repercussions on the genetic diversity and survival of these breeds.

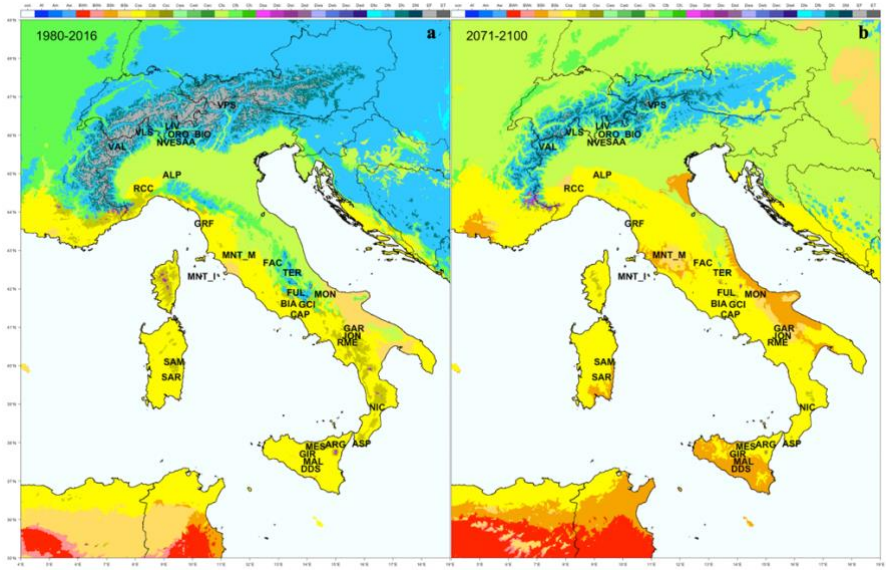


Figure 3.4 - -Koppen-Geiger climate maps with classification for Italy A) present day B) future (extended name reported in Table 1). The Maps were generated using R combining the information available in (Rubel et al., 2017; Beck et al., 2018)

Among them, nine (ASP, BIA, CAP, GRF, MES, NIC, RCC, RME, SAR) populate "Temperate dry hot summer (Csa)" areas, one (GAR) is present in an "Arid step cold (Bsk)" area, and one (NVE) in a "Temperate without dry

season hot summer (Cfa)" region. The remaining 21 breeds populate regions with a warmer and drier climate in the future (Table 3.2).

Table 3.2 - Present and future predicted Koppen-climate class and Anova classification for breed divided in the two groups: HOT/NOTHOT and DRY/NOTDRY (see Materials and Methods). In bold those breeds that will change their Koppen classification in the next 70 years

Breeds	Current class	Future class	Hot-Nothot	Dry-Notdry
GAR	BSk-arid steppe cold	BSk-arid steppe cold	Nothot	Dry
MNT_I	BSk-arid steppe cold	Csa-temperate dry hot summer	Nothot	Dry
MON	Cfa-temperate without dry season hot summer	BSk-arid steppe cold	Hot	Notdry
NVE	Cfa-temperate without dry season hot summer	Cfa-temperate without dry season hot summer	Hot	Notdry
ALP	Cfa-temperate without dry season hot summer	Csa-temperate dry hot summer	Hot	Notdry
FAC	Cfa-temperate without dry season hot summer	Csa-temperate dry hot summer	Hot	Notdry
FUL	Cfb-temperate without dry season warm summer	Csa-temperate dry hot summer	Nothot	Notdry
DDS	Csa-temperate dry hot summer	BSh-arid steppe hot	Hot	Dry
GIR	Csa-temperate dry hot summer	BSh-arid steppe hot	Hot	Dry
MAL	Csa-temperate dry hot summer	BSh-arid steppe hot	Hot	Dry

ASP	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
BIA	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
CAP	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
GRF	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
MES	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
NIC	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
RCC	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
RME	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
SAR	Csa-temperate dry hot summer	Csa-temperate dry hot summer	Hot	Dry
ARG	Csb-temperate dry warm summer	Csa-temperate dry hot summer	Nothot	Dry
GCI	Csb-temperate dry warm summer	Csa-temperate dry hot summer	Nothot	Dry
JON	Csb-temperate dry warm summer	Csa-temperate dry hot summer	Nothot	Dry
SAM	Csb-temperate dry warm summer	Csa-temperate dry hot summer	Nothot	Dry
BIO	Dfb-cold without dry season warm summer	Cfa-temperate without dry season hot summer	Nothot	Notdry

SAA	Dfb-cold without dry season warm summer	Cfa-temperate without dry season hot summer	Nothot	Notdry
VLS	Dfb-cold without dry season warm summer	Cfa-temperate without dry season hot summer	Nothot	Nodry
TER	Dfb-cold without dry season warm summer	Csa-temperate dry hot summer	Nothot	Notdry
MNT_M	Dfb-cold without dry season warm summer	Csb-temperate dry warm summer	Nothot	Notdry
LIV	Dfc-cold without dry season cold summer	Dfb-cold without dry season warm summer	Nothot	Notdry
ORO	Dfc-cold without dry season cold summer	Dfb-cold without dry season warm summer	Nothot	Notdry
VPS	Dfc-cold without dry season cold summer	Dfb-cold without dry season warm summer	Nothot	Notdry
VAL	EF-polar tundra	Dfb-cold without dry season warm summer	Nothot	Notdry

A one-way ANOVA analysis applied on the groups based on the Koppen-Geiger classification identified 27 SNPs that significantly differentiate the groups DRY/NOTDRY (seven within a gene region) and 11 that differentiate the groups HOT/NOTHOT (two within a gene region) (Supplementary Table 5). The linear regression model, applied to verify the variation of the genotype frequencies over time based on the value of their related variables, allowed us to identify five significant SNPs out of nine, intercepting the genes *CHD2*, *ARL13B*, *KLF12*, and *PAK5* for the DRY/NOTDRY group and *RACGAP1* for the HOT/NOTHOT group (Supplementary Table 6). Then, we calculated the expected future variation of allelic and genotypic frequencies of the

significant SNPs in these groups. For instance, the SNP "snp32991-scaffold385-133908" intercepts the *ARL13B* gene and is associated to "Isothermality" with the genotype GG. At present, the frequency of the G allele of this SNP is 0.4296 in the DRY group and 0.6109 in the NOTDRY group and the delta of the variable "Isothermality" for the two groups is respectively -0.1253 for the DRY group and -0.0935 for the NOTDRY group. Using the regressor of the linear regression model (b = 0.3278), we predicted the future G allele frequency for this SNP in both groups (0.3885 and 0.5802 for the DRY and NOTDRY group, respectively) and consequently the expected GG genotype frequency (respectively 0.1509 for the DRY group and 0.3366 in the NOTDRY group). This simplified model suggests a future reduction of the genotype currently associated with the reference variable ("Isothermality") in both groups. Interestingly, the gene intercepted by *ARL13B* interacts with *RABGEF1*, related to the reduction of the circadian cycle in humans according to the GenomeRNAi human phenotypes database (http://www.genomernai.org). In general, the prediction analysis identified SNPs that might go to stabilization of the frequencies or fixation (see "snp44855-scaffold611-263638" and "snp40739-scaffold521-1667886", respectively; Table .33).

Table 3.3 - Predicted genotypic frequencies for the five polymorphisms recorded within genes identified to be significantly different between the groups HOT/NOTHOT or DRY/NOTDRY

SNP	gene	genotype	group	AA current	GG current	AG current	AA future	GG future	AG future
snp32991-scaffold385-133908	ARL13b	GG	DRY	0.33	0.18	0.49	0.37	0.15	0.48

			NOTDRY	0.15	0.37	0.48	0.18	0.34	0.49
snp35938-scaffold431-1169604	KLF12	AA,AG	DRY	0.81	0.01	0.18	0.75	0.02	0.23
			NOTDRY	0.64	0.04	0.32	0.60	0.05	0.35
snp23847-scaffold240-2578654	CHD2	AA	DRY	0.59	0.05	0.36	0.54	0.07	0.39
			NOTDRY	0.36	0.16	0.48	0.33	0.18	0.49
snp44855-scaffold611-263638	PAK5	AG	DRY	0.01	0.84	0.15	0.01	0.80	0.19
			NOTDRY	0.02	0.75	0.23	0.03	0.68	0.29
snp40739-scaffold521-1667886	RACGAP1	GG	HOT	0.06	0.56	0.37	0.05	0.61	0.34
			NOTHOT	0.02	0.73	0.25	0.02	0.77	0.21

Conclusions

This new release of the Italian goat consortium dataset (IGC2) - almost three times the size of the previous iteration - fills in the gaps in terms of completeness and representativeness of the Italian caprine diversity. Our analyses overlap and expand on previous studies providing insight into the

past, present, and future evolution of the populations considered. We confirm the geographic gradient of goat diversity ranging from north to south (Nicoloso et al., 2015), provide fine scale population structure, and highlight the overlap with the geo-political situation in which the breeds evolved. Previous studies have shown how past migrations from Africa and Spain on the one hand, and central Europe and the Alps on the other hand, contributed to shaping the backbone of biodiversity along the peninsula. Nevertheless, the overlap among the three diversity clusters and the political subdivision of Italy up to 160 years ago (Riall, 2002) is an intriguing finding that suggests a role for the past socio-political scenario of the country in the current diversity of Italian goats breeds. By investigating the relationship between genotype and environment, we identified several genes which might play a role in the adaptation to temperature and humidity. Interestingly, we identified a gene that can be a fitting candidate for future studies on the caprine arthritis encephalitis virus (CAEV). Lastly, we predicted the future genotypic frequencies under the light of climate changes and foresee the directionality of changes in genotypes frequencies, an important starting point for future studies aiming at improving these analytical approaches. We infer that improved modelling approaches could deepen and perfect such results and shed light on today's favorable genotypes for specific environmental conditions. These results will likely be instrumental in breeding schemes and genomic selection, assisting locally adapted breeds to cope with the expected climate change toward warmer and drier climates (Beck et al., 2018).

Material and methods

Biological samples

Management and handling of the animals involved in this study were performed following the Italian and European legislation on animal welfare (D.lgs n. 146/2001, Council Directive 98/58/CE) and adhering to the ARRIVE ESSENTIAL 10 guidelines, where applicable. Blood samples were taken by official veterinary surgeons following the recommendations of the European directive 2010/63, without performing any actual experimental research on animals. Experimental protocol was approved by the Ethical committee of the Department of Veterinary Science of the University of Messina (code 046/2020).

Blood sampling collection of the new individuals was performed using Vacutainer tubes with the K-EDTA anticoagulant, then all the samples were stored at −20 °C until genomic DNA was extracted using a commercial kit (NucleoSpin Blood, Macherey-Nagel, GmbH & Co KG, Germany) according to the manufacturer's instructions (Nicoloso et al., 2015). DNA samples were genotyped using the GoatSNP50 BeadChip (Illumina Inc., San Diego, CA) developed by the International Goat Genome Consortium (IGGC) at the Agrotis srl (http://www.lgscr.it, Cremona, Italy), Porto Conte Ricerche s.r.l. (Alghero, Sassari, Italy), and University of Palermo facilities (Italy).

Genotyping control and datasets creations

The IGC2 successfully fills the gaps of the previous dataset (Nicoloso et al., 2015), intercepting the local diversity of several under-represented areas of the country (i.e., the central regions of Italy) and identifying small, indigenous breeds never characterized before. For this work, 19 new Italian goat

populations, for a total amount of 586 individuals, were sampled and added to SNP genotyping data taken from previously studies (Nicoloso et al., 2015; Talenti et al., 2016, 2017a), including seven Iranian Bezoar (*Capra aegagrus*) genotyped by the NEXTGEN project as outgroup for the analyses ("NEXTGEN Project" n.d.). From that, we obtained a final raw dataset consisting of 1,071 goats from 33 Italian breeds and populations and one wild species, *Capra aegagrus* (Table 3.1). Geographical coordinates of the samples at the time of sampling were available for 998 samples (93% of all samples).

The raw dataset was updated to the latest goat genome map (ARS1.2) and pruned to retain SNPs having SNP call rate >98%, individual call rate >95% and minor allele frequency (MAF) >0.05. Then, we pruned loci in linkage disequilibrium (LD), removing one of each couple of SNPs having LD > 0.2 using PLINK v1.90b6 (Chang et al., 2015). Duplicated individuals (identity by state >99%) were removed and for each pair of highly related animals (Mendelian Errors count <100) we excluded the animal occurring in multiple pairs or having the highest missingness. Phasing and imputation of missing genotypes was performed using BEAGLE v4.1 (Browning and Browning, 2007),using sliding windows of ~5 Mb with an overlap of ~2 Mb and allowing two SNPs trimming (~0.15 Mb). The resulting dataset was used for the haplotype sharing analysis. In order to investigate the population structure using comparable population sizes, we created a specific dataset reducing the number of individuals for each population to ≤30 while maintaining the overall within-population diversity, using the 'representative.sample' function implemented in the R package BITE v1.2.0007 (Milanesi et al., 2017).

Lastly, individuals with more than second-degree relatedness were identified using the --genome flag in PLINK and removed to perform the Landscape genomics analyses.

Population structure analysis

Population structure analysis was conducted through MDS of the identity by state (IBS) distances obtained with the flag --cluster in PLINK. Maximum likelihood analysis of population structure was conducted using ADMIXTURE v1.3.0 Software (Alexander et al., 2009). Unsupervised clustering was calculated for *K* values from 2 to 35. We used 5-fold cross-validation (CV) errors for each *K* to evaluate the optimal partitioning, and plots for each *K* were generated using an in-house R script. A phylogeny tree based on Reynolds Genetic distances, with 100 bootstrap replicates, was computed using a custom script. A Neighbour-joining consensus tree was generated using PHYLIP v3.697 (Felsenstein., 1989) and using Bezoar as an outgroup.

The proportion of haplotypes shared among breeds was determined as Identity By Descent (IBD) estimation among individuals, and calculated using RefinedIBD v4.1 (Browning and Browning, 2013) on all the individuals that passed the initial quality check. Sliding windows size was set to of 1Mb, reporting windows of at least 0.2 Mb and allowing 0.05 Mb overlap. We considered the shared haplotypes between two breeds as the median length of shared haplotypes among all the possible pairs of individuals belonging to the breeds considered (individual pairs with no haplotype sharing were assigned length = 0 (Talenti et al., 2018)).

Landscape genomics

For each georeferenced sample, we used the 'extract' function from the R package raster (Hijmans, 2020) to retrieve the values of 19 bioclimatic and elevation variables available from the WorldClim database (Hijmans et al., 2005) as those referring to the time span between 1960 and 1990 as proxy for the current climate, and the estimated future values for 2070 (average for 2061-2080) (Supplementary Table 7). Altitude was used to compute terrain slope through the function raster::terrain. Each variable was retrieved as a raster layer with a spatial resolution of 2.5 arcminutes (~4 km). Pairwise correlations were calculated among the climatic variables using JMP (SAS Institute Inc.).

LGA was performed to assess the genotype/environmental variable association using Samβada v.0.5.3 (Stucki et al., 2017) and LFMM v3.1.2 (Caye et al., 2019) . Analyses were performed using the 'current' bioclimatic variables and a more stringent subset of animals. . To spare computation time, the number of environmental variables was reduced iteratively by randomly removing one of the two most correlated variables until the maximum correlation across all variables was lower than $|r^2| < 0.7$ as implemented in the R function 'caret::findCorrelation' (Kuhn et al., 2020). To reduce the risk of false positive detections, we evaluated the genetic structure of our dataset through principal component analysis, used the scree plot to identify the number of principal components to keep to adequately describe the dataset, and included the selected PCs as population structure predictors for the association analysis (Rellstab et al., 2015; Vajana et al., 2018). A likelihood-ratio test comparing a null and an alternative model was carried out for each genotype. Specifically, null models included the population structure predictors alone, whereas alternative models included population structure

predictors and the focal environmental variable. A genotype was considered significantly associated with the environmental variable if the resulting p-value associated with the likelihood-ratio test statistic was lower than the nominal significance threshold of 0.05 after Benjamini-Hochberg (BH) correction for multiple testing. The R function 'p.adjust' was used to perform corrections for multiple testing (R Development Core Team, 2011).

Gene-level analysis

To further investigate the biology underlying the signals identified, we screened all the SNPs that resulted significantly associated with a WorldClim variable for annotated genes of interest at the exact location of each marker in the ARS1.2 goat genome version (https://www.ensembl.org/biomart). These genes were investigated individually (https://www.genecards.org) and used as input for an enrichment analysis for pathways and ontologies using the online tool Enricher (https://amp.pharm.mssm.edu/Enrichr/).

Future genotypes prediction

Comparing the LGA results with the Koppen-Geiger classification relative to the breeding areas of the different Italian breeds, we tried to predict the future frequencies of the genotypes significantly associated with one or more of the environmental variables considered. First, we estimated the extent of climate change that Italian breeds will face comparing the present and future Koppen-Geiger classification (Beck et al., 2018) (Supplementary Table 8). Then, we grouped the breeds based on temperature according to the current Koppen-Geiger classification, creating the HOT and NOTHOT groups, and humidity, creating the DRY and NOTDRY groups. The HOT group included those breeds that live in Csa and Dfa regions and the NOTHOT group included

breeds that live in Csb, Dfb, Dfc, Bsk, and EF regions. The DRY group included those breeds that live in Csa, Csb, and Bsk regions and the NOTDRY group included those breeds that live in Cfa, Dfb, Dfc, and EF regions (see Table 3.2 and Supplementary Table 8 for the detail of the environmental codes).

We calculated the MAF of all the significant SNPs from LGA in each Italian breed using a custom script. We summarized the MAF in each of the four groups and performed a one-way ANOVA analysis using the R base package(R Development Core Team, 2011) considering the HOT/NOTHOT groups or the DRY/NOTDRY groups as source of variation, identifying the SNPs significantly different in the two couples of groups. We applied a linear regression model (R base Package) only on those SNPs that fall within genomic regions including an annotated gene considering the mean allelic frequencies of the SNP and the mean value of the environmental variable resulted significantly associated with that SNP for each breed in each group. Finally, we calculated the future hypothetical change in allelic and genotypic frequencies only for those SNPs with a statistically significant linear regression model. For each group, we multiplied the delta between the current and the projected future value of the environmental variable associated with the SNP with the regressor of the linear model.

Supplementary Materials

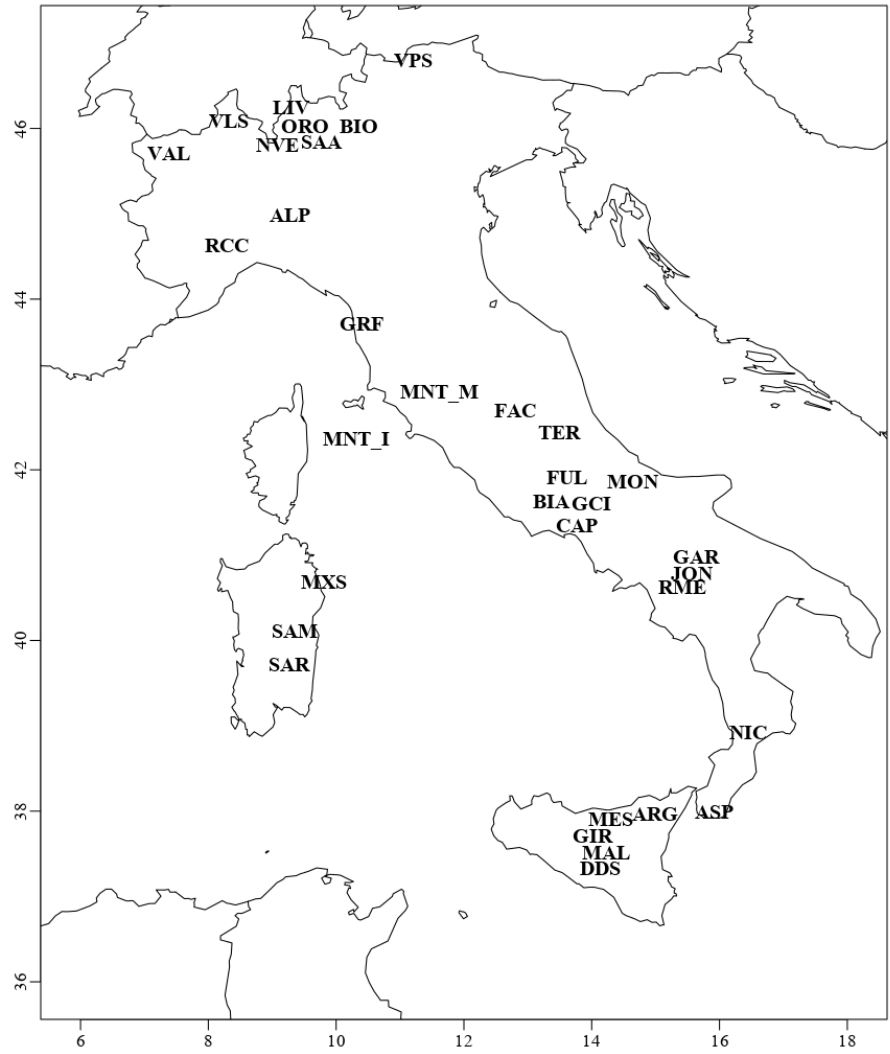


Figure 3.5 - Geographical distribution of the sampling location of the Italian Breeds

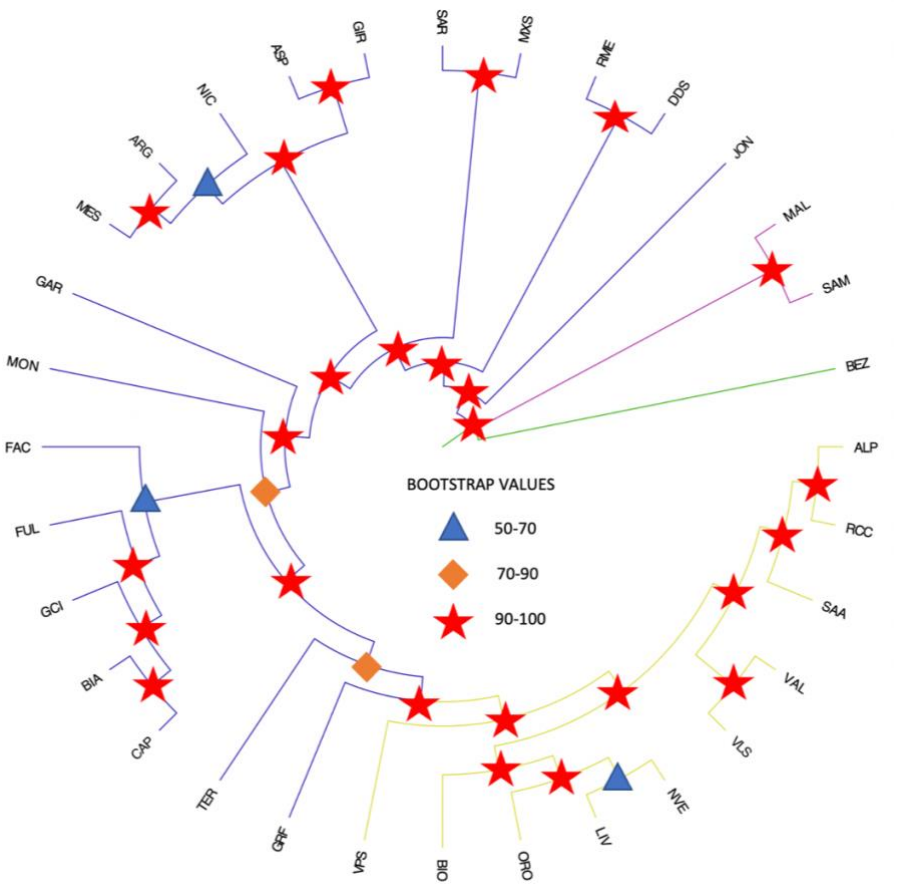


Figure 3.6 - Bootstrapped phylogenetic tree from Reynolds distances. The plot was generated through the FigTree software (http://tree.bio.ed.ac.uk/software/figtree/)

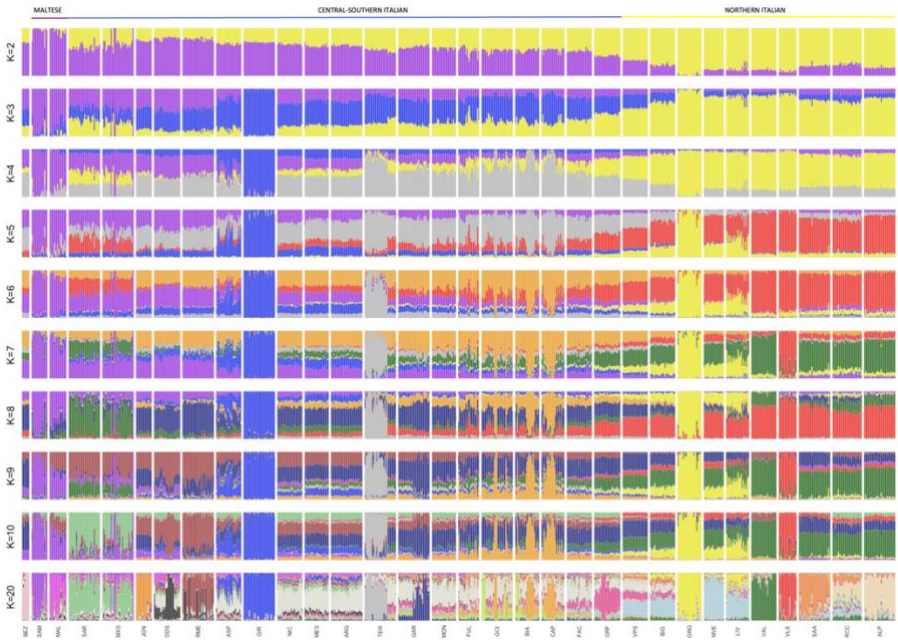


Figure 3.7 - ADMIXTURE plot for all K values 2-10 and 20, produced in R using ggplot2 (https://ggplot2.tidyverse.org/authors.html)

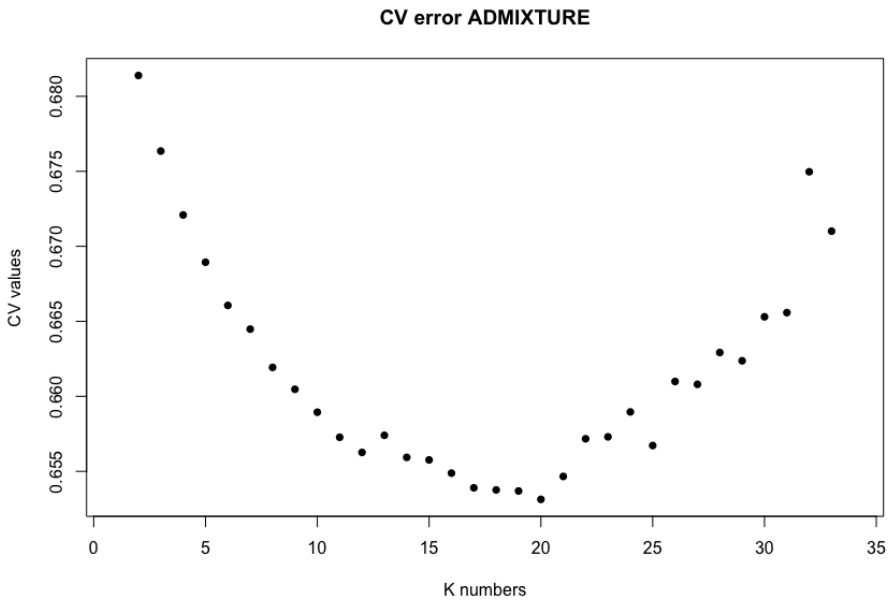


Figure 3.8 - Cross validation error for the different K considered in the ADMIXTURE analyses; the plot was generated through R

All the Supplementary Tables from 1 to 8 can be found at the following link https://doi.org/ 10.1038/s41598-021-89900-2.

3.2 Selection signatures

3.2.1 Aims

Together with the climate, another important factor that influences and differentiates the genomic background of the indigenous Italian goat breeds, especially along the north-south axis of the peninsula, is the presence of different farming managements. In fact, alongside the more modern intensive and semi-intensive breeding systems for the cosmopolitan and highly productive breeds like Saanen, there are the much more extensive and traditional ones characterized by long movements of the flocks through different pastures during the year, the so-called transhumance.

Our aim in this work was to identify what kind of traces in the genome of the different breeds could have been left by the adaptation to different types of breeding systems and what kind of impact did they have on the inbreeding structure of our Italian goat populations.

One of the most suitable and widely used tool for both these purposes is the analysis of the Runs of Homozygosity (ROH) within the different populations: in addition to giving us an idea of which are the most homozygous regions of the genome shared across all the subjects of each Italian goat populations, it can also allow us to calculate the inbreeding value of the single individuals. Moreover, because the longest are the ROH segments the more recent are the inbreeding events they are associated with, we can use them to estimate the period of time in which the inbreeding event occurred.

Lastly, through this study we were able to distinguish the different effects of geographical isolation and transhumance in the populations of southern and northern Italy and, thanks to the comparison between ROH and FST, also to identify about 30 genes differentiating the two different groups of breeds (Northern and Southern) linked to characteristics such as growth, fertility, body size, and climate adaptation.

3.2.2 Runs of homozygosity in the Italian goat breeds: impact of management practices in low input systems

Article published on Genetic Selection Evolution (2021)
(DOI: 10.1186/s12711-021-00685-4)

Authors and affiliations

Matteo Cortellari[#1], Arianna Bionda[#*1], Alessio Negro[1], Stefano Frattini[1], Salvatore Mastrangelo[2], Elisa Somenzi[3], Emiliano Lasagna[4], Francesca M. Sarti[4], Elena Ciani[5], Roberta Ciampolini[6], Donata Marletta[7], Luigi Liotta[8], Paolo Ajmone Marsan[3], Fabio Pilla[9], Licia Colli[3], Andrea Talenti[§10], and Paola Crepaldi[§1]

[1]Dipartimento di Scienze Agrarie e Ambientali – Produzione, Territorio, Agroenergia, Università degli Studi di Milano, Milan, Italy
[2]Dipartimento di Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, Palermo, Italy
[3]Dipartimento di Scienze Animali, della Nutrizione e degli Alimenti and BioDNA Centro di ricerca sulla Biodiversità e sul DNA Antico, Università Cattolica del Sacro Cuore, Piacenza, Italy
[4]Dipartimento di Scienze Agrarie, Alimentari e Ambientali, Università degli Studi di Perugia, Perugia, Italy
[5]Dipartimento di Bioscienze Biotecnologie e Biofarmaceutica, Università degli Studi di Bari, Bari, Italy

[6]Dipartimento di Scienze Veterinarie, Università di Pisa, Pisa, Italy
[7]Dipartimento di Agricoltura, Alimentazione e Ambiente, Università degli Studi di Catania, Catania, Italy
[8]Dipartimento di Scienze Veterinarie, Università degli studi di of Messina, Messina, Italy
[9]Dipartimento Agricoltura, Ambiente e Alimenti, Università degli Studi del Molise, Campobasso, Italy
[10] The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

[#,§]Equal contribution

*Corresponding author

Abstract

Climate and farming systems, several of which are considered as low-input agricultural systems, vary between goat populations from northern and southern Italy and have led to different management practices. These processes have impacted genome shaping in terms of inbreeding and regions under selection and resulted in differences between the Northern and Southern populations. Both inbreeding and signatures of selection can be pinpointed by the analysis of runs of homozygosity (ROH), which provides useful information to assist the management of this species in different rural areas. We analyzed the ROH distribution and inbreeding (F_{ROH}) in 902 goats from the Italian Goat Consortium2 dataset. We evaluated the differences in individual ROH number and length between goat breeds from Northern

(NRD) and Central-southern (CSD) Italy. Then, we identified the signatures of selection that differentiate these two groups using three methods: ROH, ΔROH, and averaged F_{ST}. ROH analyses showed that some Italian goat breeds have a lower inbreeding coefficient, which is attributable to their management and history. ROH are longer in breeds that are undergoing non-optimal management or with small population size. In several small breeds, the ROH length classes are balanced, reflecting more accurate mating planning. The differences in climate and management between the NRD and CSD groups have resulted in different ROH lengths and numbers: the NRD populations bred in isolated valleys present more and shorter ROH segments, while the CSD populations have fewer and longer ROH, which is likely due to the fact that they have undergone more admixture events during the horizontal transhumance practice followed by a more recent standardization. We identified four genes within signatures of selection on chromosome 11 related to fertility in the NRD group, and 23 genes on chromosomes 5 and 6 related to growth in the CSD group. Finally, we identified 17 genes on chromosome 12 related to environmental adaptation and body size with high homozygosity in both groups. These results show how different management practices have impacted the level of genomic inbreeding in two Italian goat groups and could be useful to assist management in a low-input system while safeguarding the diversity of small populations.

Background

Today, in light of the ongoing climate change, the management and conservation of livestock biodiversity are becoming an increasingly

important goal at the global level (Bruford et al., 2015). To face this challenge, it is crucial to draw a precise picture of the genetic structure of the indigenous breeds and populations of farmed animals in different countries. It is necessary to understand the genetic basis of their adaptation, not only to the natural environment, but also to the breeding conditions and management strategies to which they have been subjected (Meuwissen, 2020). From this point of view, Italy provides a good model because it is characterized by a rich biodiversity in all domesticated species thanks to its varied history, environment, climate, and farming traditions (Talenti et al., 2018; Senczuk et al., 2020). Goats, in particular, represent one of the greatest expressions of Italian biodiversity with more than 30 autochthonous breeds and populations reared under very diverse climates and farming conditions, several of which are considered as low-input agricultural systems (Cortellari et al., 2021).In the Northern regions, goats are mainly bred in the Alps, where two diametrically opposed farming systems coexist. On the one hand, in the valleys and hilly regions, modern intensive and semi-intensive farming systems are present, which are specifically suited to milk and cheese production and usually exploit cosmopolitan dairy goat breeds, particularly Saanen and Alpine (Manfredi et al., 2010; Sandrucci et al., 2019). In these systems, medium-to-large flocks are mostly kept indoor with controlled feeding and limited grazing, which is generally conducted in fenced pastures near the farm (Crepaldi et al., 1999). On the other hand, the traditional extensive farms, which can be considered as low-input/low-output systems, are mainly located in the mountainous areas and depend highly on natural grazing. On these farms, small flocks of local breeds are kept indoor during the winter and in pasture for the rest of the year because of extreme variations

in climate and weather conditions, especially during the winter. Some farmers still practice the traditional vertical transhumance (*alpeggio*), which consists in transferring the animals to alpine pastures during the summer only (Crepaldi et al., 1999; Manfredi et al., 2010; Sandrucci et al., 2019). The animals of this farming system are particularly influenced by the climate conditions.Central-southern Italy and the islands count the largest number of goat farms and heads (ISTAT, 2010). In these regions, which are characterized by a hotter and dryer climate (Cortellari et al., 2021) than in Northern Italy, the traditional extensive or semi-extensive farms of autochthonous goat breeds prevail and are generally located in marginal mountainous areas (Di Trana et al., 2015; Paschino et al., 2020). The vertical transhumance was usually combined with a horizontal transhumance; for example, the shepherds transferred all their animals - cattle, sheep, goats, and shepherd dogs - in the mild Apulian plains during the winter and returned to the Abruzzo mountains in the summer (Nannini et al., 2004).These differences between Northern and Southern Italian goat populations in terms of animal nutrition, housing, and mating management, may have contributed to the genetic makeup of the Italian caprine diversity. Among the genomic tools and approaches that are now proposed to characterize animal biodiversity, the analysis of runs of homozygosity (ROH) is certainly one of the most useful (Eusebi et al., 2019). ROH are long stretches of homozygous genotypes in the genome of an individual, which compose a pair of identical haplotypes. They are considered as a standard approach for the calculation of genomic inbreeding values (F_{ROH}) and for the detection of signatures of selection (Gorssen et al., 2021). The length of a ROH can also be a useful indicator of the time of the inbreeding event with which it is associated, i.e.

long ROH are associated with recent events of inbreeding in the history of a breed or of a single individual, whereas short ROH indicate a more ancient event (Kirin et al., 2010). The presence of several ROH in a particular region of the genome of a species or a population, regardless of their length, constitutes a so-called ROH island. The analysis of ROH islands can be a very effective tool to identify the regions of a genome that have been under selective pressure because they can contain variants that are shared between the individuals of a specific population (Nothnagel et al., 2010). For all these reasons, the analysis of ROH from genomic data and the derived inbreeding value (F_{ROH}) are increasingly used as a starting point to develop new management systems of animal populations, together with the more traditional pedigree information (Rodríguez-Ramilo et al., 2019).In this work, our aim was to characterize ROH in 902 goats from the northern and southern Italian groups, estimate their level of inbreeding, and analyze how it has evolved across generations according to management practices.

Methods

Dataset and quality control

In this work, we used the same Italian Goat Consortium2 (IGC2) dataset as described in Cortellari et al. (Cortellari et al., 2021). Among the 34 populations present in that dataset, we decided to exclude the Bezoar, which in the previous work was used as an outgroup, the Maltese×Sarda crosses and the two Montecristo populations due to their unique history of isolation (feral) or farming (mainland). All the animals were genotyped with the Illumina Goat single nucleotide polymorphism (SNP)50 BeadChip. SNPs that had a

missing genotype frequency higher than 0.2, or that were in unplaced scaffolds or on the X chromosome were excluded from the analysis, but we did not apply a threshold for minor allele frequency (MAF) to better identify ROH (Meyermans et al., 2020). Individuals with a call rate lower than 95% were removed. All quality control procedures were carried out with the software PLINK 1.9 (Chang et al., 2015). After the initial quality check on the 986 individuals and the 52,538 SNPs taken from the original IGC2 dataset, 902 animals grouped in 30 breeds and 46,995 SNPs were retained. Population structure of the goats included in the final dataset was investigated by multidimensional scaling analysis (MDS) and by building a phylogeny tree based on Reynolds genetic distances.

Expected heterozygosity (genetic diversity)

For each breed, the PLINK 1.9 software (--hardy option) was used to calculate the expected heterozygosity (H_E), observed heterozygosity (H_O), and Wright's fixation index (F_{IS}), which is defined as the correlation between the homologous alleles within an individual relative to the local population to which that individual belongs (Wright, 1949).

Definition of runs of homozygosity

In order to minimize the discovery of false positives within regions of low marker density, we selected rather stringent criteria (Meyermans et al., 2020). ROH were calculated separately for each individual using the software PLINK 1.9 by applying a sliding window of 20 SNPs. A ROH was called if the following parameters were fulfilled: (i) no heterozygous genotypes, (ii) less than two missing genotypes, (iii) a minimum number of SNPs within a ROH $\geq$ 20, (iv) a minimum ROH length of more than 1 Mb, (v) a minimum

SNP density of two SNPs per Mb, and (vi) a maximum gap of 500 kb between consecutive homozygous SNPs.

ROH distribution and genomic inbreeding

To characterize the ROH distribution, for each breed we estimated: the number of individuals without ROH, the mean number of ROH per individual, the mean total length of ROH per individual, the mean length of a ROH per individual, and the genomic inbreeding coefficient (F_{ROH}) for each individual. The F_{ROH} for each breed was computed following the method proposed by McQuillan (McQuillan et al., 2008):

$$F_{ROH,i} = \frac{L_{ROH}}{L_{AUTO}},$$

where L_{ROH} is the sum of the total length of ROH in individual i and L_{AUTO} is the total length of the autosomes covered by SNPs. In addition, we categorized the ROH for each breed into five length classes (1–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb, and >16 Mb) to compare the distribution of the F_{ROH} across these categories between the considered breeds (Onzima et al., 2018). We focused on these length classes with the intent to investigate the percentage and the impact of ancient and more recent inbreeding events that occurred in the Italian goat breeds.

Identification of the groups of populations

In order to better disentangle the genetic differences between the Italian goat populations analyzed due to climatic conditions and breeding management techniques, we divided them into two large groups according to their geographical distribution: a group of ten populations from the northern Italy breeds (NRD) and a group of 20 populations from the Central-southern Italy

breeds (CSD), which also includes the two Maltese populations (MAL and SAM).

Statistical analysis

We performed two linear mixed models to evaluate the statistical significance of the difference between the ROH parameters identified for each population group (Macciotta et al., 2021) using the statistical software JMP 16 (SAS Institute Inc.). We modelized two variables (Y): (i) a standardized ratio between the length of single ROH in each individual and the length of the corresponding chromosome ($std(\frac{ROH\ length}{CHR\ length})$), and (ii) a standardized ratio between the number of ROH on each chromosome of each individual and the length of the chromosome ($std(\frac{ROH\ number}{CHR\ length})$); both of these models included the same factors:

$$Y = \mu + CHR + POPGROUP * CHR + BREED[POPGROUP] + POPGROUP + id + e,$$

where μ is the mean, CHR is the fixed effect of the autosome (chromosomes 1 to 29), POPGROUP is the fixed effect of the population group (CSD vs NRD), BREED[POPGROUP] is the fixed effect of the breed nested within the population groups (n = 20 in CSD and n = 10 in NRD), $POPGROUP * CHR$ is the interaction between population group and autosome, id is the random effect of the animal, and e is the random residual. The covariance between the animals was assumed to be equal to 0.

Signatures of selection

In this work, we investigated the signatures of selection by using three methods: ROH, ΔROH islands and the Wright's fixation index (F_{ST}).

For the first method, a homozygosity score (H-score) ranging from 0 (0%) to 1 (100%) was obtained for each SNP by counting how many times it appeared in a ROH and dividing the result by the number of the animals. This approach was applied to each population separately, and the top 1% H-scores were considered. SNPs that were within regions of 0.25 Mb were joined together, and regions with more than 15 SNPs were considered as ROH islands. Then, the identified ROH islands were investigated to list the annotated genes they contain based on the reference genome (ARS1) and the associated functions and pathways. The ΔROH score was defined as the difference between the H-scores for the CSD and the NRD groups at a specific SNP. The regions of maximum difference in homozygosity were defined analogously to the ROH islands (top 1% values, SNPs within a region of 0.25 Mb combined together, and regions encompassing >15 SNPs), thus resulting in ΔROH islands.

The Wright's fixation index (F_{ST}) was calculated using the PLINK software, averaging the value of each SNP with the values of five adjacent SNPs in each flanking region to minimize the impact of outlier scores (Onzima et al., 2018). The top 1% averaged F_{ST} values were considered and investigated for annotated genes in the reference genome (ARS1) within a region spanning ± 0.25 Mb from each SNP.Finally, the genes identified by these three methods were analyzed to detect the shared genes.

Results

Dataset composition and quality control

The dataset used for the analyses consisted of 902 goats belonging to 30 populations. The number of analyzed animals per breed are in Table 3.4. The

geographic distributions, the MDS plot, and the phylogeny tree of the studied goat populations are reported in Supplementary Figure 3.12.

ROH description and genetic diversity

In total, 28,383 ROH were identified in the 902 individuals considered; five animals displayed no ROH: one Garganica individual, one Rossa Mediterranea individual, one Grigia Ciociara individual, and two Capra di Teramo individuals. In terms of average number of ROH per animal, the breed with the largest number was Orobica (92.8 ROH per individual), followed by Vallesana (76.6), both of these breeds being raised in the northern regions of Italy. The breeds with the smallest average number of ROH were two Sicilian breeds, Messinese and Argentata dell'Etna with 9.3 and 9.1 ROH per individual, respectively. Argentata dell'Etna and Messinese were also the breeds that, together with the Val di Livo (or Lariana) had the lowest average value of the total ROH length per individual (31.2, 36.8 and 56.8 Mb, respectively), while the two highest values were found in Vallesana (364.6 Mb) and Maltese bred in Sicily (347.5 Mb). When the average length of ROH per individual in each breed was considered, the Capra di Livo-Lariana, Argentata dell'Etna, and Orobica were the breeds with the lowest values, whereas the Capra di Teramo and Roccaverano were those with the highest values (Table 3.4).

Table 3.4 - Dataset composition and mean ROH and genetic parameters for the studied Italian goat breeds

Breed ID	Breed name	Raw dataset	Quality checked	ROH number	ROH total length ± se	ROH length ±se	F_{ROH}	H_E	H_O	F_{IS}
ALP	Camosciata delle Alpi	143	117	31.5	169.1±12.1	5.0±0.2	0.07	0.41	0.4	0.02

ARG	Argentata dell'Etna	48	46	9.1	31.2±5.8	2.9±0.3	0.01	0.41	0.41	0
ASP	Capra dell'Aspromonte	24	24	22.6	113.1±29.1	4.2±0.5	0.05	0.4	0.4	0
BIA	Bianca Monticellana	24	23	30.8	158.7±34.8	4.0±0.6	0.06	0.4	0.39	0.01
BIO	Bionda dell'Adamello	24	24	22.4	93.5±20.3	3.5±0.4	0.04	0.4	0.4	0
CAP	Capestrina	24	22	23.5	117.5±41.2	3.5±0.4	0.05	0.4	0.4	-0.01
DDS	Derivata di Siria	32	25	37.8	232.2±28.8	5.5±0.5	0.09	0.4	0.38	0.03
FAC	Facciuta della Valnerina	24	24	19.5	167.9±48.3	5.4±0.8	0.07	0.41	0.39	0.03
FUL	Fulva del Lazio	22	20	11.8	63.0±14.2	4.1±0.5	0.03	0.41	0.41	-0.02
GAR	Garganica	40	37	21.6	118.8±24.1	4.4±0.3	0.05	0.4	0.4	0
GCI	Grigia Ciociara	43	39	17.5	116.4±21.4	4.8±0.6	0.05	0.41	0.4	0.02
GIR	Girgentana	59	56	74.5	322.7±24.1	4.1±0.2	0.13	0.36	0.36	0.01
GRF	Garfagnana	28	25	33.1	147.5±23.7	4.1±0.4	0.06	0.4	0.4	0.01
JON	Jonica	16	15	24.3	88.4±16.1	3.4±0.4	0.04	0.37	0.41	-0.1
LIV	Capra di Livo-Lariana	24	22	20.6	56.8±7.5	2.6±0.2	0.02	0.4	0.4	-0.02
MAL	Maltese	16	16	70.6	347.5±56.1	4.4±0.4	0.14	0.37	0.36	0.01
MES	Messinese	24	23	9.3	36.8±8.6	3.5±0.6	0.01	0.4	0.41	-0.01
MON	Capra di Montefalcone	24	23	16.7	136.4±45.2	5.2±0.9	0.06	0.4	0.4	0.01
NIC	Nicastrese	24	24	24.2	170.3±40.3	5.3±0.7	0.07	0.4	0.39	0.02
NVE	Nera di Verzasca	19	19	32.6	136.2±27.2	3.7±0.4	0.06	0.38	0.39	-0.02
ORO	Orobica	23	23	92.8	293.6±22.9	3.1±0.1	0.12	0.35	0.36	-0.02
RCC	Roccaverano	28	28	20.7	169.6±37.6	6.1±0.7	0.07	0.41	0.4	0.03
RME	Rossa Mediterranea	46	40	24.2	93.9±12.8	3.4±0.2	0.04	0.39	0.41	-0.05
SAA	Saanen	44	44	25.8	115.9±10.1	4.2±0.2	0.05	0.41	0.41	-0.01

SAM	Maltese sampled in Sardinia	15	15	57.1	294.4±49.0	4.8±0.5	0.12	0.36	0.37	-0.03
SAR	Sarda	33	33	18.1	77.8±16.7	3.4±0.4	0.03	0.41	0.4	0.01
TER	Capra di Teramo	43	30	25.2	234.1±36.1	7.3±0.7	0.1	0.39	0.38	0.01
VAL	Valdostana	24	24	50.9	272.0±49.5	4.7±0.5	0.11	0.37	0.37	0.02
VLS	Vallesana	24	17	76.6	364.6±45.6	4.7±0.4	0.15	0.36	0.35	0.02
VPS	Capra della Val Passiria	24	24	23.1	118.3±24.3	4.4±0.5	0.05	0.4	0.4	0.01

F_{ROH}: ROH-based inbreeding; ROH total length ± se in Mb; ROH length ±se in Mb; H_E: expected heterozygosity; H_O: observed heterozygosity; F_{IS}: Wright's fixation index.

The ROH-based inbreeding values (F_{ROH}) showed that the two breeds with the highest level of inbreeding were Maltese and Vallesana and those with the lowest levels were Messinese, Argentata dell'Etna, and Capra di Livo-Lariana. However, the distribution of the individual F_{ROH} within each population varied among breeds (see the boxplot in Fig. 3.9): some breeds such as Maltese, Capra di Teramo, Vallesana, and Girgentana showed a wide dispersion of the individual inbreeding values, while other breeds such as Saanen, Rossa Mediterranea, Capra di Livo-Lariana, and Messinese showed a more compact distribution. The genomic diversity parameters were similar across all the breeds, with the lowest H_E and H_O respectively in the Orobica (0.35) and Vallesana (0.35), and the highest ones in Roccaverano (0.41) and Saanen (0.41). The F_{IS} values were highest for Roccaverano and Derivata di Siria, and lowest for the Jonica and Rossa Mediterranea breeds (Table 3.4).
The ROH identified in all the populations were classified into the five length classes. The largest numbers of identified ROH belonged to the 1-2 Mb (11,294) and the 2-4 Mb (7525) classes. Nevertheless, the length classes that contributed most to the calculated inbreeding value (F_{ROH}) in the different

populations were 8-16 Mb and > 16 Mb. Analysis of the distribution of the ROH categories among all the populations and of their proportional weight in the definition of the mean F_{ROH} value revealed three categories: (1) one for which the influence of the longest ROH on the total F_{ROH} value appeared to predominate, such as for the Capra di Teramo, Roccaverano, and Montefalcone breeds; (2) one for which the different ROH length classes were well balanced, such as for the Girgentana, Bianca Monticellana , and Nera di Verzasca breeds; and (3) one for which the short ROH were more important, such as for the Orobica, Capra di Livo-Lariana, and Rossa Mediterranea breeds (Fig. 3.9).

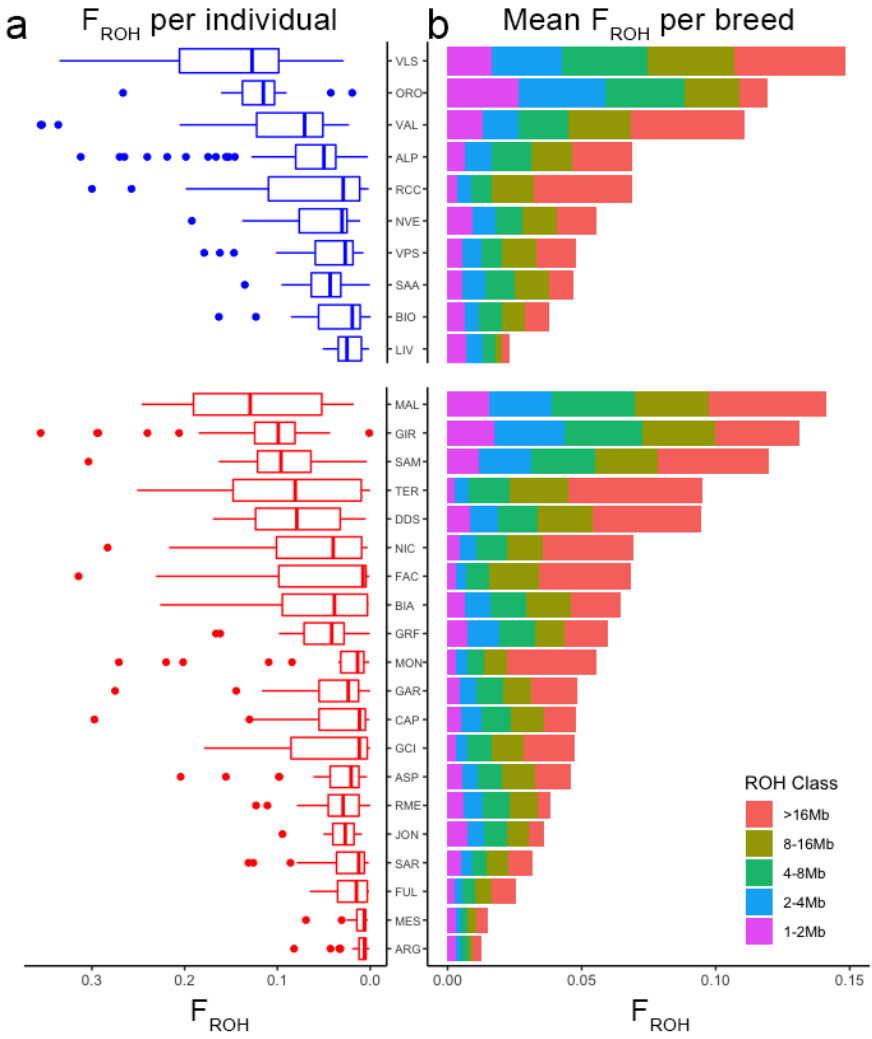


Figure 3.9 - Distribution of FROH values per individual and mean values per breed. Boxplot (a) of the single individual FROH distribution in each population (Northern breeds: blue boxplots, Central-southern breeds: red boxplots) with matching barplot (b) of the mean FROH values (each color representative of the different ROH length classes)

Statistical analysis

The two groups of Italian goat populations included 560 individuals belonging to 20 breeds for the Central-southern (CSD) and 342 individuals belonging to ten breeds for the Northern (NRD) groups. The statistical models showed a significant difference in the number and length of ROH between the two groups. Particularly, the first model (r^2 = 0.18) showed that ROH length was significantly affected by the group (p = 0.003), the breeds within each group (p < 0.0001), and the chromosome (p < 0.0001). ROH were longer for the CSD than the NRD group (LSmean ± SE = 0.06 ± 0.02 *vs* -0.02 ± 0.02). Figure 3.10-a shows the mean standardized ratio between the length of ROH and the length of the corresponding chromosome $std(\frac{ROH\ length}{CHR\ length})$ for each chromosome in the two groups (CSD and NRD): the largest differences were found for chromosomes 3, 13, 25, and 29 and the smallest differences for chromosomes 8, 18, and 20. Interestingly, the smaller chromosomes presented relatively longer ROH in both groups. Previous studies on plants, yeasts, and humans have shown that recombination rates are inversely correlated with chromosome length, which could be due to the lower frequency of multiple crossovers within a chromosome (Kaback, 1996; Copenhaver et al., 1998; Kaback et al., 1999); this has also been reported in goats (Mastrangelo et al., 2017) and other species (He et al., 2020).

When the individual number of ROH per chromosome was modelled (r^2 = 0.40), all the selected factors (group, breed within each group, chromosome, and the interaction between chromosome and group) were significant (p < 0.0001). ROH number was, on average (± SE), larger for the NRD group (0.19 ± 0.03) than the CSD group (-0.07 ± 0.02). Figure 3.10-b shows the

mean standardized ratio between the number of ROH per chromosome and the corresponding chromosome length $std(\frac{ROH\ number}{CHR\ length})$ in the two groups (CSD and NRD): the largest differences were found for chromosome 13 and the smallest for chromosomes 8, 18, 20, and 27. It is worth mentioning that the number of ROH for chromosome 12 was large in both groups, particularly in the CSD group.

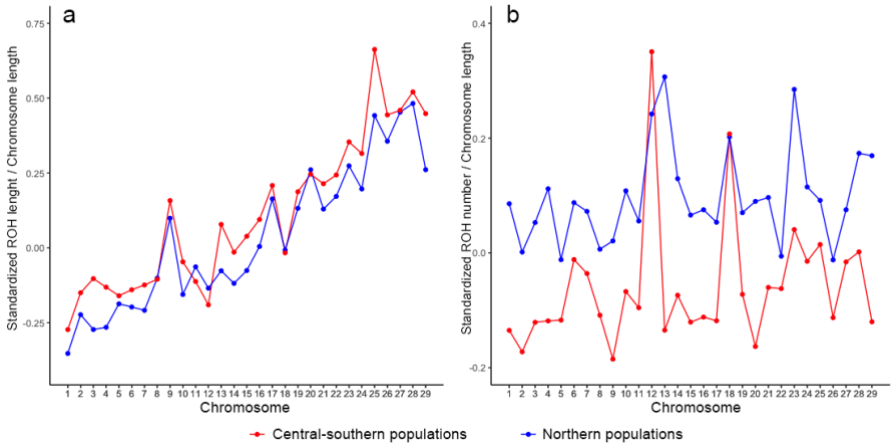


Figure 3.10 - Comparison of mean standardized length and number of ROH in the two groups of Italian goat populations. Graphic representation of the mean standardized length (a) and number (b) of ROH divided by the corresponding chromosome length in the two groups of Italian goat populations

Signatures of selection

For both the CSD and NRD groups, we identified the genomic regions with the highest level of homozygosity, corresponding to the top 1% of SNPs (H-score value > 0.107 for the CSD and > 0.116 for the NRD group) and found 10 regions distributed on seven chromosomes for the CSD group and 15 regions distributed on nine chromosomes for the NRD group. Among these

regions, six were partially or totally shared because they were highly homozygous in both groups, while the remaining 13 were specific to only one of the two groups. Matching the positions of these regions with those of the genes annotated in the goat genome version ARS1 and excluding genes for which a symbol or orthologs were not available (i.e. beginning with "LOC") and transfer RNA gene sequences (TRNA), we identified 133 genes specific to the NRD group, 47 genes specific to the CSD group, and 111 genes common to the two groups.

Then, we identified the regions that showed the largest difference in homozygosity between the two groups (ΔH-score values > 0.06) and found nine regions that were distributed on seven chromosomes and harbored 80 genes of the 291 previously identified genes. These regions were both highly homozygous within a group and capable of differentiating the two NRD and CSD groups. Finally, when the results of the previous analyses were cross-referenced with the top 1% mean F_{ST} values (> 0.09), we identified 44 genes that were shared among all the genes detected by the three methods (Fig. 3.11). In particular, two groups of genes were specific to the CSD group, i.e. one on chromosome 6 and one on chromosome 5, and one group specific to the NRD group on chromosome 11; finally, a gene cluster on chromosome 12 that was revealed by the F_{ST} analysis was highly homozygous in both groups and had a high ΔROH score (Fig. 3.11 and see Supplementary Figure 3.13). The complete list of the identified genes is in Additional file 3 Table S1.

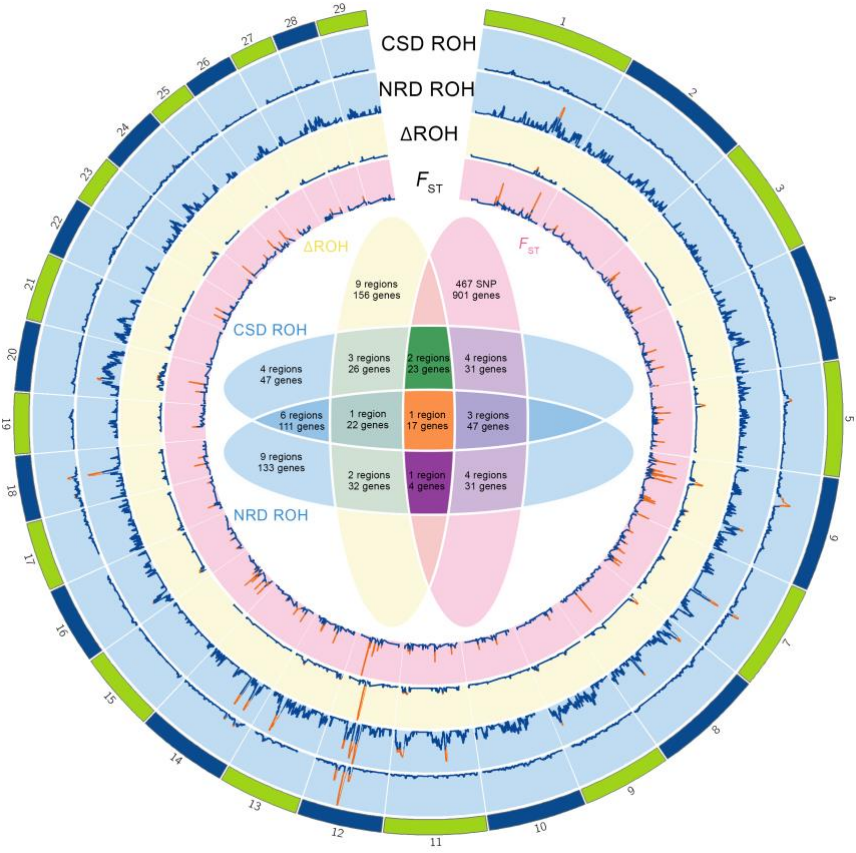


Figure 3.11 - Circos plot of the analysis of signatures of selection with a Venn diagram of the results. Circos plot of the signatures of selection in the two groups of Italian goat populations (external blue tracks), of the ΔROH (middle yellow track) and averaged FST (inner red track). The Venn diagram shows the number of regions and genes shared across methods. CSD Central-southern population group, NRD Northern population group

Discussion

Italy is characterized by a wide variety of breeding environments, managements, and traditions; the effect of this variability is particularly

evident in goats, which have been traditionally bred in low-input systems, which are strongly affected by climatic conditions. In our study, we used the well-established genomic tool of runs of homozygosity to shed light on the impact of different management practices on Italian goat homozygosity.

Our results show that some populations present an extremely small number of ROH per individual and, consequently, a level of genomic inbreeding near zero. Among these, the two Sicilian breeds Argentata and Messinese are known to have been crossbred, mainly due to the typical management of traditional extensive farms that share common pastures (Mastrangelo et al., 2021). Another interesting breed is the Capra di Livo-Lariana, with a very low level of inbreeding that can be explained by historical events, including introgression of many unknown individuals from the surrounding valleys.

Another fundamental aspect that emerges from our work is the possibility of monitoring the inbreeding management in the populations through the evaluation of the relationship among the various ROH length classes. Indeed, regardless of the absolute value of F_{ROH}, the populations that have a large preponderance of long ROH (> 16 Mb) are more likely to have been under a non-optimal management with frequent mating between closely related individuals, a possible consequence of the reduced number of individuals in the population. One example is the Capra di Teramo breed: the earthquakes that hit the regions where this breed is reared had catastrophic consequences on this already endangered population. On the contrary, Orobica has a more balanced ratio between the different ROH length categories and a smaller number of long ROH. In fact, Orobica is one of the first Italian breeds to have been standardized and reflects a long-term efficient management and the particular attention paid by shepherds to the mating plans.

Moreover, the statistical models performed on ROH number, length, and distribution revealed significant differences between the two NRD and CSD groups. In particular, for the NRD group the number of ROH per individual was larger than for the CSD group, whereas for the CSD group the ROH were longer than for the NRD group. The populations from the NRD group have always been bred in isolated valleys, with natural barriers that prevent the exchange of animals; for this reason, a large number of short ROH, indicating ancient inbreeding events, is expected. On the contrary, the breeds from the CSD group might have undergone, in the past, more admixture events due to the horizontal transhumance practice, the sharing of common pastures, and the presence of multi-breed farms; the more recent standardization is represented by longer ROH.

We also identified signatures of selection that characterize Italian goat populations according to their geographic location. For the NRD group, only four genes on chromosome 11 were found across all the analyses. Among these, the most interesting one is *DENND1A*, a fertility-related gene (Jaton, 2018) that is involved in embryogenesis in cattle. Furthermore, other genes worthy of attention belong to highly homozygous regions but were not found by the ΔROH analysis. In particular, two genes on chromosome 11, *HSPA5* and *NR5A1,* are linked to the production of the anti-Müllerian hormone in grazing cows (Gobikrushanth et al., 2019) and are located in a large region on this chromosome that is related to milk production in European, American, and Asian goats (Bertolini et al., 2018a). Another region of interest is located on chromosome 13 and hosts genes that are important for pigmentation such as *ASIP* and *RALY* (Guo et al., 2018).

We found a particularly interesting group of genes for the CSD group in a region of chromosome 6 that distinguishes it from the NRD group. This region harbors different genes related to animal growth and development, such as *LCORL* (Saif et al., 2020), which has been shown to regulate body size in goats and several other mammals, and *HERC6* (Cheng et al., 2020), and *FAM184B* (An et al., 2018). A part of this homozygous region and another region that we also found on chromosome 5 for the CSD group, were previously described by (Mastrangelo et al., 2021), who analyzed only the cluster of Sicilian goats.

Finally, in both NRD and CSD groups, we identified a highly homozygous region on chromosome 12 that was detected by all three analyses and contains 17 genes, which are mostly related to environmental adaptation, for example to hot and arid climates, and body size, including *GJB2*, *GJA3* (Kim et al., 2016), and *PSPC1* (Edea et al., 2018).

Conclusions

Our findings show that the analysis of ROH is a useful tool not only to identify regions under selection in different breeds, but also to evaluate how their management has evolved over generations. However, this is possible only if a representative recent sample of the specific populations is available, with the potential of expanding the study to historical samples to understand the evolution of a breed's inbreeding and signatures of selection. ROH assessment can be adopted as a 'checkpoint' to assess whether selection in a population is leading to an increase in its average homozygosity and inbreeding, therefore indicating whether a fine-tuning of the breeding scheme

is necessary. For these reasons, we recommend the implementation of this tool in the routine evaluation of biodiversity and, consequently, the management of autochthonous populations that are bred in a low-input system as typical of marginal rural areas.

Supplementary materials

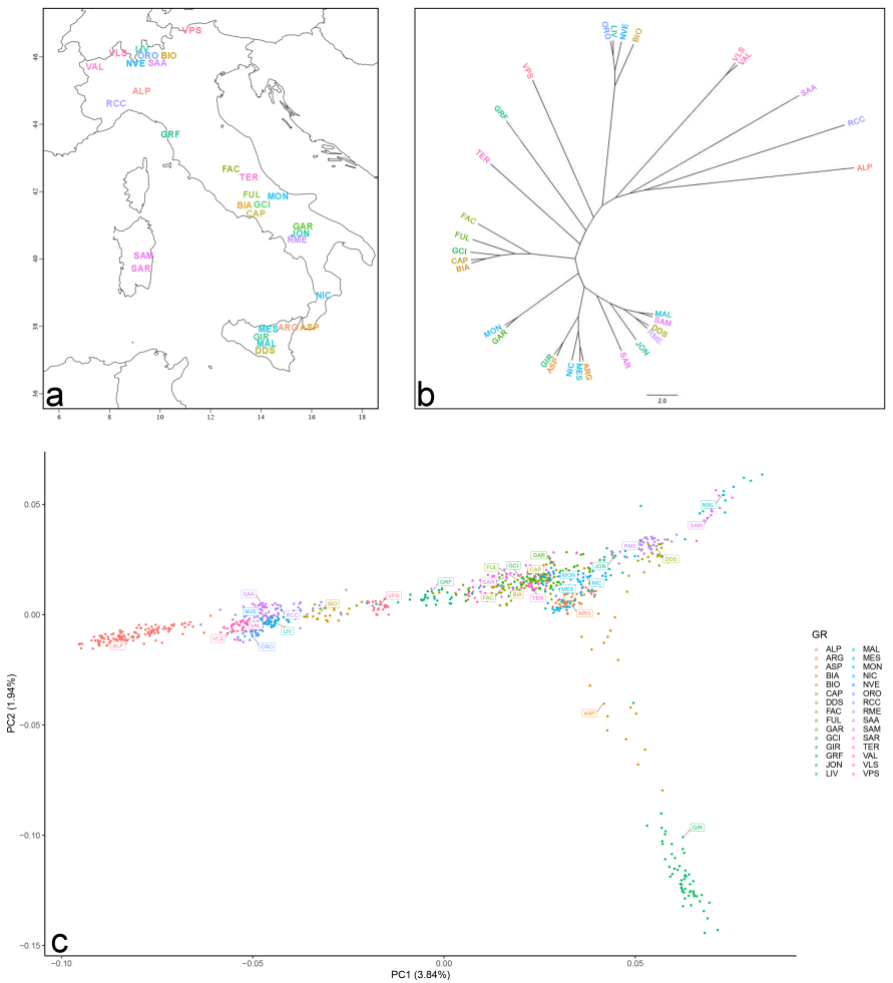


Figure 3.12 - : Geographic distribution (a), phylogeny tree (b), and multidimensional scaling analysis (c) of all the Italian goat breeds included in the study

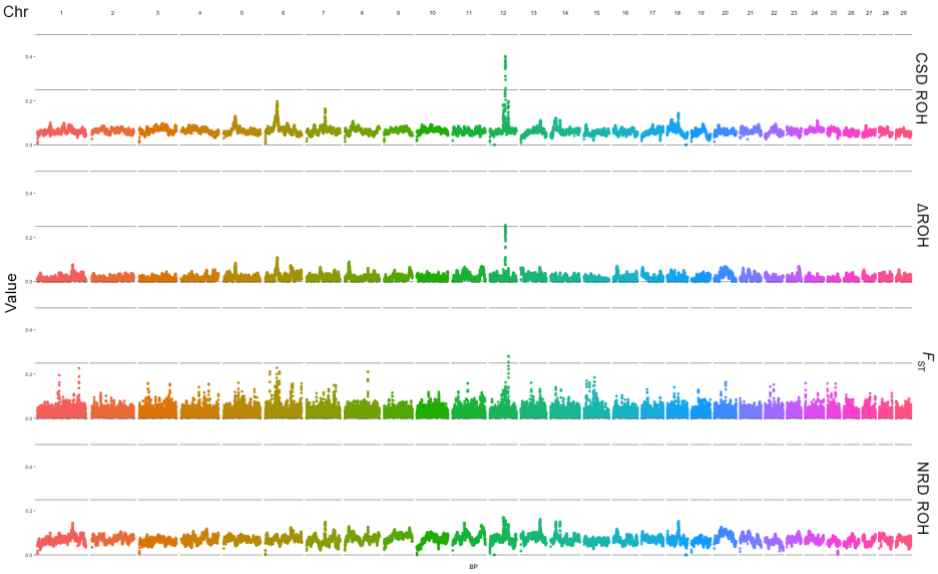


Figure 3.13 - Manhattan plots representing the signals of signatures of selection in the two population groups (CSD and NRD), of the ΔROH, and averaged FST. CSD = Central-southern population group; NRD = Northern population group.

The Additional file 3 Table S1can be found at the following link https://doi.org/10.1186/s12711-021-00685-4.

3.3 Inbreeding and applications

3.3.1 Aims

Monitoring the level of inbreeding within populations of farmed animals has always been one of the main goals of farm management, from both the productive and the animal welfare points of view. The main tool used so far by small ruminant breeders is the information derived from the pedigrees. With the advent of the genomic era, we now have the possibility to combine the canonical pedigree information with the genomic data, making it possible to calculate more accurately the animal relationships and inbreeding values, by considering the so called "mendelian sampling" and by correcting the possible registration errors in the pedigrees. The use of both these types of data is a common practice in the large dairy cattle farms, whereas the use of genomic data has been introduced only recently in the management of small ruminants, whose economic value is lower. For this reason, it has become essential to better understand the relationship between the inbreeding values calculated through pedigree data (F_{PED}) and genomic data (F_{ROH}) within small ruminants populations. This was the aim of my third work, in which we investigated the relationship between F_{ROH} and F_{PED} in 2 goat and 3 sheep breeds representing the different breeding situations that the Italian dairy small ruminant panorama offers.

This study intends to be the first step of the introduction of the genomic data also in Italian sheep and goat populations, which, together with the production of more and more detailed and precise pedigree information, will allow to

implement specific genetic and genomic indexes for our small ruminant populations.

3.3.2 Using pedigree and genomic data toward a better management of inbreeding in Italian dairy sheep and goat breeds.

Article submitted to Animals

Matteo Cortellari[1], Alessio Negro[1,2], Arianna Bionda[1], Silverio Grande[2], Alberto Cesarani[3], Antonello Carta[4], Nicola Macciotta[3], Stefano Biffani[5], Paola Crepaldi[1]

[1]Dipartimento di Scienze Agrarie e Ambientali – Produzione, Territorio, Agroenergia, Università degli Studi di Milano, Via Celoria 2, Milan, 20133, Italy

[2]Ufficio Studi, Associazione Nazionale della Pastorizia (Asso.Na.Pa.), Via XXIV Maggio 44, 00187, Rome, Italy

[3]Dipartimento di Agraria, Università degli Studi di Sassari, Viale Italia 39, Sassari, 07100, Italy

[4]Unità di Ricerca di Genetica e Biotecnologie, Agris Sardegna, 07100 Sassari, Italy

[5]Istituto di Biologia e Biotecnologia Agraria (IBBA), Consiglio Nazionale delle Ricerche, Via Edoardo Bassini 15, 20133, Milano, Italy

Abstract

The inbreeding coefficient is an important parameter for livestock management. Small ruminant breeders and associations mainly rely on pedigree information, but genomic tools are gaining relevance, allowing to overcome possible pedigree inconsistencies. This study aims to investigate the relationship between pedigree-based and genomic inbreeding in a representative number of Italian dairy sheep and goat. Pedigree and genomic data (through a medium-density SNPchip) were obtained from two goat (n. 3,107) and four sheep (n. 2,511) breeds. We estimated pedigree depth (number of fully traced generations, FullGen) and inbreeding (F_{PED}), and two genomic inbreeding indexes, using runs of homozygosity (F_{ROH}) and genomic relationship matrix (F_{GRM}). The correlation between the inbreeding coefficients was assessed by breed and FullGen. A linear regression model (LRM) was fitted for estimating F_{PED} from F_{ROH} for each species.

After quality control, we retained 5,085 animals. Mean inbreeding values were low for all breeds, with higher F_{ROH} than F_{PED} and F_{GRM}. Breed differences can partially depend on different managements (e.g., use of artificial insemination). The correlation between F_{PED} and F_{ROH} was the highest and directly related to pedigree depth. The best LRM to estimate F_{PED} from F_{ROH} was chosen for FullGen $\geq$4 and $\geq$6 for goats and sheep, respectively; after excluding animals with extreme residuals, a new refined model was calculated for each species.

Since massive genotyping is not affordable by small ruminant breeders, it is important to understand the distinction and relationship between differently calculated inbreeding values, also in view of the introduction of genomic

enhanced breeding values. Our study highlights the importance of an accurate pedigree information and, especially if it is not obtainable, of genotyping animals. With correct data, it will be possible to contribute to breed progression mitigating inbreeding depression and loss of genetic variability.

Introduction

The accurate estimate of the inbreeding value of an individual could be of extreme importance for all the domesticated species, both for commercial and conservation purposes (Panetto et al., 2017; Schiavo et al., 2020; Polak et al., 2021). With regards to small ruminants, the pedigree information continues to be the most effective tool for calculating the inbreeding coefficients, even though several authors pointed out that pedigree-based inbreeding values can be severely underestimated due to the high level of inconsistencies in these species (Rodríguez-Ramilo et al., 2019; Wang et al., 2021). Indeed, in goats and sheep the correct parentage relations can be difficult to assess due to several factors, such as the high number of twins, the limited use of artificial insemination, the creation of mating groups with the simultaneous presence of more rams in the same group (Hayes and Goddard, 2008), and the extensive farming system (Paiva et al., 2020). According to (Legarra et al., 2014), the unknown fatherhood rate reaches 50% and 20% in Latxa and Manech/Basco-Béarnaise sheep breeds, respectively.

Today, genomic data obtained by high-throughput SNP genotyping represent a new instrument that should be integrated with the traditional pedigree information to estimate more precisely the relationship between individuals(Villanueva et al., 2021). Moreover, genomic information allows

for a more accurate inbreeding estimation by overcoming pedigree inconsistencies and by accounting for the Mendelian sampling. This will allow not only to capture a better view of the genetic variation between the individuals, which is of great interest primarily in small and endangered populations, but also to eventually estimate unbiased genetic values (Andonov et al., 2017; Cortes-Hernández et al., 2021).

There are different ways to calculate inbreeding using SNP chip data, and those based on Runs of homozygosity (ROH) and Genomic relationship matrix (GRM) are among the most used (Zhang et al., 2015; Caballero et al., 2021). ROH are segments of homozygous genotypes identified for a single individual and composed of a pair of identical haplotypes. ROH give information about a subject's level of auto-zygosity, age of the inbreeding events, and origin, but most importantly, they can be used to assess a reliable inbreeding coefficient (F_{ROH}) (Rodríguez-Ramilo et al., 2019). Indeed, recombination events can disrupt long stretches of DNA over generations (or meiosis) allowing to estimate the age of the inbreeding events: short ROH derive from more distant ancestors, while long ROH suggest that the autozygosity of an individual comes from more recent meiosis events leading to the presence of a recent common ancestor (Ceballos, 2018).

Even if SNP genotyping has been routinely implemented in dairy cattle (Weller et al., 2017), its costs are still too high compared to the value of a single sheep or goat; thus, small ruminant breeders and breeder associations cannot afford to genotype a large number of animals. This supports the need for finding a way to integrate traditional pedigree data with the genomic ones and eventually estimate reliable inbreeding coefficients. Some previous studies have already investigated the genomic variability of Italian sheep and

goat breeds (Ciani et al., 2014b; Nicoloso et al., 2015; Mastrangelo et al., 2018; Cortellari et al., 2021) However, their results were based on a limited number of animals and/or did not elucidate the relationship between traditional pedigree-based and innovative genomic-based parameters. For this reason, our study aims to shed light about the relation between F_{PED} and pedigree depth and the genomic inbreeding F_{ROH} and F_{GRM} in six Italian small ruminant populations.

Materials and methods

Breeds

For this work, we selected a total of six breeds, namely four local dairy sheep breeds - Comisana, Delle Langhe, Massese, and Sarda - and two cosmopolite dairy goat breeds - Camosciata delle Alpi and Saanen. All breeds are mainly used for cheese production, being representative of both different production systems and geographical areas. Ancestry and identification of all animals are recorded in their respective official herd book, managed by the Italian Sheep and Goat Breeders Association (Rome, Italy), hereinafter referred to as Asso.Na.Pa. In all breeds, male selection candidates are DNA tested for parentage assignment. Artificial insemination (AI) is used mainly in Camosciata delle Alpi and Saanen breeds, where from 10 to 15 % of annual births are from semen of foreign AI bucks. All breeds are officially milk recorded and a routine genetic evaluation for milk yields is in place. The selection scheme for Comisana and Massese sheep breeds is based on two different closed nucleus flocks reared at the Genetic Center of Asciano (Siena, Italy); instead, the Sarda sheep is managed using a pyramidal scheme,

with a nucleus flock at the apex (Carta et al., 2009). Finally, selection in Delle Langhe sheep aims to improve milk production through the estimate of the breeding value for milk yield (expressed in kg); a software is available for breeders to optimize genetic breeding values and manage matings to reduce offspring inbreeding.

Datasets and quality control

Pedigree records and genomic data were obtained from the "Conservation, Health and Efficiency Empowerment of Small Ruminant" (CHEESR) repository project, a database created by Asso.Na.Pa. in the framework of the Italian National Rural Development Plan (PSRN) – sub-measure 10.2 project and supported by the European Agricultural Fund for Rural Development (EAFRD). In total, we analyzed 3,107 goats (2,164 Camosciata delle Alpi and 943 Saanen) and 2,511 sheep (1,498 Sarda, 534 Comisana, 375 Massese, and 104 Delle Langhe). All animals were genotyped with the medium density chip, Ovine SNP50 BeadChip for ewes and Goat SNP50 BeadChip for goats. A quality control analysis was performed on both species fitting the same parameters and using PLINK 1.9 software (Chang et al., 2015): a missing genotype frequency ≥ 0.05 and an individual call rate below 0.05. In addition, all the unplaced markers (according to the assembly versions ARS1 for goat and OAR4 for sheep) and those located on the X chromosome were excluded.

Pedigree analysis

All the pedigree analyses were carried out using the Optisel package of the R software (Wellmann, 2019). The function "*summary.Pedig*" was used to calculate both the pedigree-based inbreeding (F_{PED}) and the number of fully traced generations (i.e., whose ancestors are all known, FullGen).

Genomic inbreeding calculation

We used the function --*homozyg* of PLINK 1.9 for the calculation of the ROH applying the following criteria: (i) a sliding window of 20 SNPs, (ii) no heterozygous genotype allowed, (iii) no more than two missing genotypes, (iv) a minimum number of SNPs in a ROH equal to 20, (v) a minimum ROH length of 1 Mb, (vi) a minimum SNP density of 1 SNP per 500 kb, and (vii) a maximum gap of 500 kb between two consecutive homozygous SNPs. In order to minimize false positives discovery within regions of low marker density, rather stringent criteria were selected. ROH distribution was characterized for each breed by estimating the number of individuals without ROH and the mean length, the mean total length, the mean number of ROH and the genomic inbreeding coefficient (F_{ROH}) in each of the individuals considered. We used the following equation in order to calculate the F_{ROH} coefficients:

$$F_{ROH,i} = \frac{L_{ROH}}{L_{AUTO}}$$

where L_{ROH} represent the sum of the total length of ROH per individual *i* and L_{AUTO} represent the total length of the autosomes covered by the SNPs (McQuillan et al., 2008).

The estimation of the GRM and the consequent F_{GRM} values was performed through the use of the --*ibc* parameter implemented in the GCTA v1.93.3 software (Yang et al., 2011). While no minor allele frequency (MAF) threshold was applied to allow a better estimation of ROH (Meyermans et al., 2020), before the estimation of the F_{GRM} we excluded the SNPs with a MAF

< 0.05, being strongly influenced by the frequencies of rare alleles (VanRaden, 2008).

F_{PED}-F_{ROH} estimate

Using the functions "*cor*" and "*lm*" implemented in R base package, we calculated the correlation coefficient between F_{ROH}, F_{GRM}, and F_{PED}, and fitted a linear model to estimate the most probable F_{PED} value from genomic data (F_{ROH}). This calculation was performed on all the individuals of both species together considering increasing FullGen classes (i.e., increasing number of known ancestors).

Then, for each species, we chose the linear model of the FullGen class that maintained at least 600 individuals of the species of interest and, at the same time, had both the highest correlation coefficient between F_{PED} and F_{ROH} and the highest linear model R^2. After that, we applied the coefficients of the chosen FullGen model to estimate the F_{PED} value only on the individuals of the correspondent species with the selected minimum FullGen. We calculated the standardized residuals (Z-score of the differences between the predicted and observed F_{PED} values) and excluded all the animals with a standardized residual falling under the 1^{th} or over the 3^{th} quartile. Lastly, we recalculated the linear model using only the retained subjects and re-estimated their F_{PED} values with the coefficient of this last refined model.

Results

Dataset creation

We applied the same thresholds and filtering parameters to both datasets. After the quality control, we retained 3,086 (2,147 Camosciata delle Alpi and

939 Saanen) individuals and 51,097 markers for goats and 2,484 (1,480 Sarda, 529 Comisana, 371 Massese, and 104 Delle Langhe) individuals and 46,723 markers for sheep. All the 485 individuals for which we could not calculate F_{GRM}, F_{ROH}, or F_{PED} were excluded. Therefore, the final dataset used for estimating the correlation coefficients and the linear model comprised 5,085 animals (3,028 sheep and 2,057 goats) (Table 3.5).

Table 3.5 - Number of subjects in the final dataset, for each species and breed, relative mean value of pedigree, ROH, and GRM-based inbreeding coefficients (FPED, FROH, and FGRM), and their correlation coefficients.

Species/Breed	**N. subjects**	**Mean F_{PED}**	**Mean F_{ROH}**	**Mean F_{GRM}**	**F_{PED}-F_{ROH} correlation coefficient (p-value)**	**F_{PED}-F_{GRM} correlation coefficient (p-value)**	**F_{ROH}-F_{GRM} correlation coefficient (p-value)**
Goats	3028	0.017	0.059	0.012	0.278 (< 0.0001)	0.272 (< 0.0001)	0.604 (< 0.0001)
Camosciata delle Alpi	2093	0.016	0.057	0.013	0.26 (< 0.0001)	0.228 (< 0.0001)	0.671 (< 0.0001)
Saanen	935	0.02	0.062	0.01	0.312 (< 0.0001)	0.351 (< 0.0001)	0.492 (< 0.0001)
Sheep	2057	0.062	0.097	0.006	0.817 (< 0.0001)	0.365 (< 0.0001)	0.477 (< 0.0001)
Sarda	1053	0.093	0.135	0.023	0.804 (< 0.0001)	0.346 (< 0.0001)	0.489 (< 0.0001)
Delle Langhe	104	0.06	0.099	-0.025	0.436 (< 0.0001)	0.225 (0.022)	0.354 (< 0.0001)
Comisana	529	0.018	0.044	-0.008	0.379 (< 0.0001)	-0.130 (0.003)	0.229 (< 0.0001)
Massese	371	0.039	0.067	-0.013	0.318 (< 0.0001)	-0.070 (0.181)	0.224 (< 0.0001)

Inbreeding Correlation and Linear Model

For each sheep and goat breed, we calculated F_{PED}, F_{GRM}, and F_{ROH} for different ROH length classes. As shown in Figure 3.14-A and B, the mean F_{ROH} values are higher than the other inbreeding coefficients in both the species. Instead, F_{PED} values are higher than F_{GRM} in sheep, while the two are similar in goats. It is worth noticing that F_{GRM} shows some negative values in all the breeds and a particularly spread distribution in the Delle Langhe breed. Regarding the F_{ROH} distribution per ROH length class (Figure 3.14-C), the two goat breeds have approximately the same distribution for all the F_{ROH} classes, apart from the one derived from ROH > 16 Mb (suggesting recent inbreeding), which is higher in Saanen than in Camosciata delle Alpi. In the sheep species, Sarda and Delle Langhe breeds present the highest F_{ROH} value and proportion of long ROH classes. In any case, all the inbreeding coefficients are relatively low in both the species, with the highest values in the Sarda sheep.

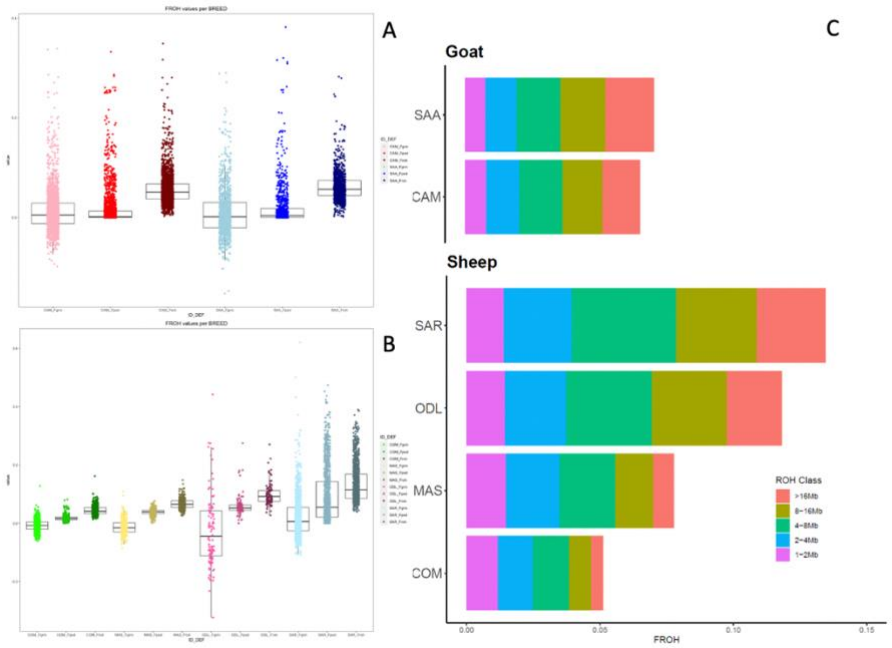


Figure 3.14 - Distribution of F_{GRM} (light colour), F_{PED} (medium colour), and F_{ROH} (dark colour) in goat (A) and sheep (B) breeds, and F_{ROH} per ROH length class in the examined breeds (C).

The correlation coefficient between F_{GRM}, F_{ROH}, and F_{PED} was much higher in sheep than in goats, and, in particular, in the Sarda breed. Also, the correlation between F_{PED} and F_{ROH} was higher than between F_{PED} and F_{GRM}; therefore, we decided to use a linear regression model to estimate F_{PED} from F_{ROH}.

The correlation coefficient between F_{PED} and F_{ROH} and the R^2 of the linear model, reported in Table 3.6, show a progressive increment as the minimum FullGen increases, until we reach FullGen > 6. All the correlations were statistically significant ($p < 0.0001$). Based on these results and the assumptions explained in the material and methods section (highest

correlation value and R^2, and at least 600 subjects) we identified the model calculated for FullGen ≥ 4 as the best one for the goats, and the one calculated on Fullgen > 6 as the best for the sheep.

Table 3.6 - Number of goats and sheep, F_{ROH}-F_{PED} correlation coefficients and p-value, and linear regression model (LRM) coefficients and R^2 per minimum class of fully traced generations in the pedigree (FullGen). Correlation was significant (p-value < 0.0001) for all the classes of FullGen.

Minimum FullGen	N. animals	N. goats	N. sheep	Correlation coefficient	LRM intercept	LRM slope	LRM R^2
0	5085	3028	2057	0.712	-0.028	0.86	0.507
1	4549	2493	2056	0.725	-0.028	0.888	0.526
2	3911	1877	2034	0.735	-0.027	0.91	0.54
3	3311	1358	1953	0.753	-0.026	0.937	0.567
4	2522	717	1805	0.782	-0.027	0.971	0.611
5	1602	167	1435	0.825	-0.028	1.01	0.681
6	927	18	909	0.849	-0.03	1.056	0.72
7	378	2	376	0.847	-0.029	1.085	0.718
8	107	0	107	0.773	0.008	0.997	0.597

Goats

Following the method showed in Figure 3.15, we applied the coefficients of the linear model calculated on all the animals with FullGen ≥ 4 (intercept = -0.03, slope = 0.97) on the 717 goats with FullGen ≥ 4 to estimate the new F_{PED} values (Figure 3.16-A). Then, we excluded the 358 goats with a standardized residual falling under the 1st and over the 3rd quartiles and calculated the new linear model on the goats left: $F_{PED} = -0.03 + 0.91 * F_{ROH}$ ($R^2 = 0.86$, $p < 2.2e\text{-}16$) (Figure 3.16-B).

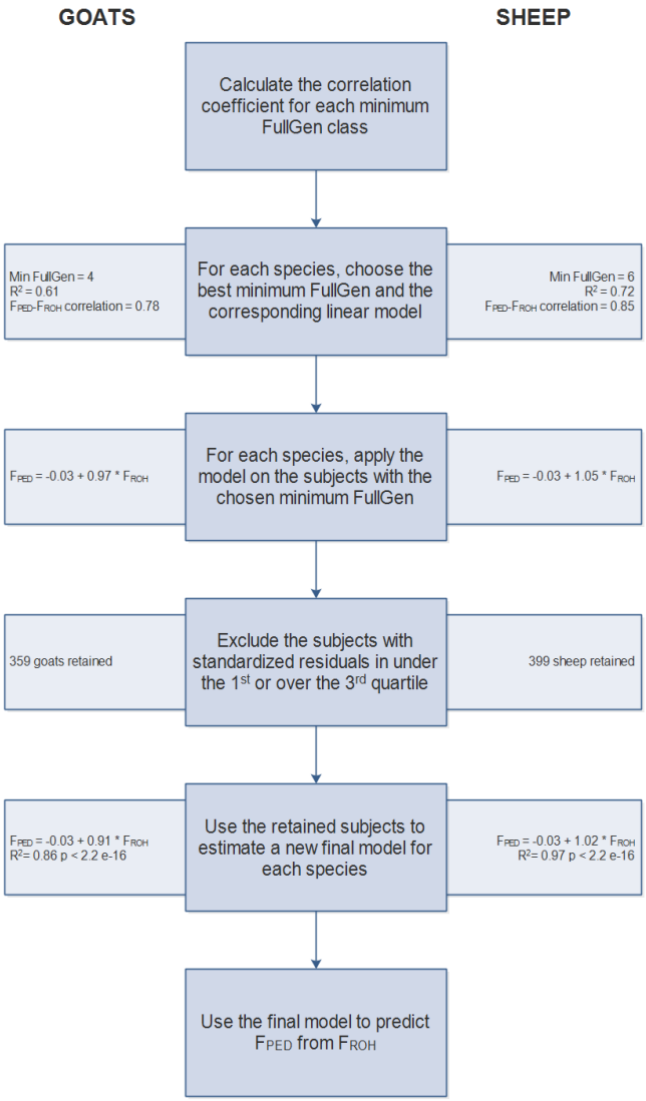


Figure 3.15 - Flowchart of the estimation of a linear regression model estimating F_{PED} from F_{ROH}, for goats and sheep.

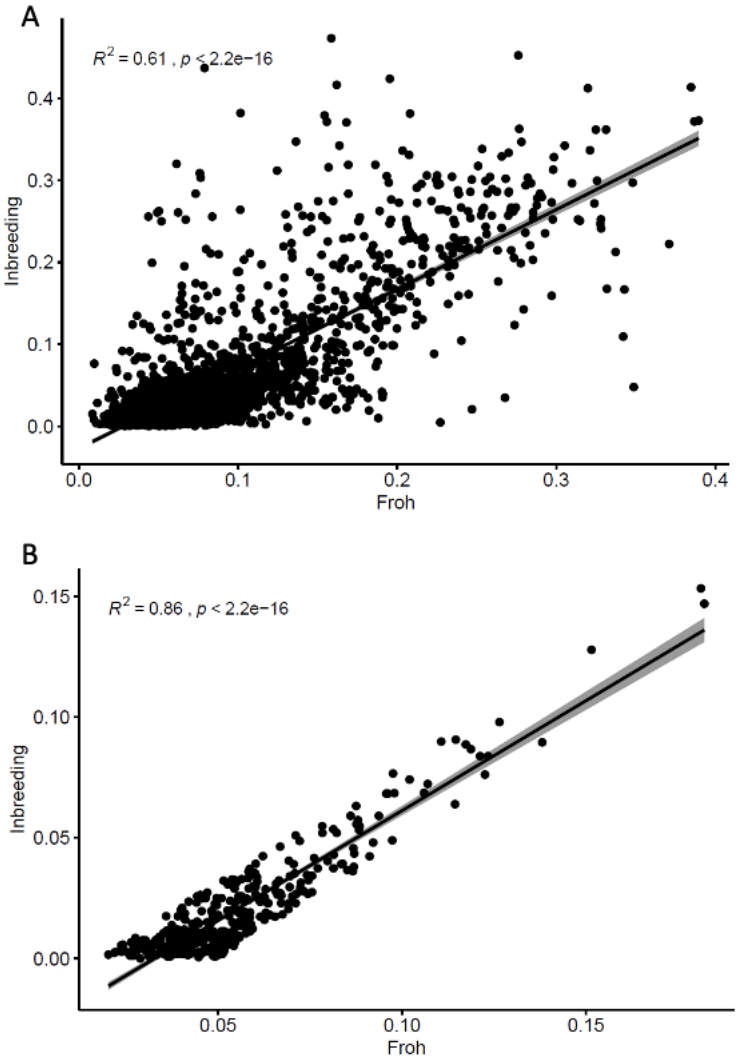


Figure 3.16 - Linear regression models to estimate F_{PED} from F_{ROH}. (A) Linear regression model calculated on all the animals with FullGen ≥ 4. (B) Refined model for goats with FullGen ≥ 4.

Lastly, we used this last regression model to estimate the definitive F_{PED} from F_{ROH} for all the goats with FullGen ≥ 4 and for each 0.05 class of estimated definitive F_{PED} we calculated the corresponding mean and range of F_{ROH} (Table 3.7). It should be noted that, for the class of estimated definitive F_{PED} 0-0.05, the corresponding F_{ROH} values are higher in Saanen than in Camosciata delle Alpi.

Table 3.7 - Number of goats and F_{ROH} mean, standard deviation (SD), 95% confidence interval (CI), and range per class of predicted F_{PED}.

Breed	Predicted F_{PED} class	N. subjects	F_{ROH} mean	F_{ROH} SD	F_{ROH} 95% CI	F_{ROH} range
Goats	0.00-0.05	613	0.054	0.017	0.052 - 0.055	0.008 - 0.088
Camosciata delle Alpi	0.00-0.05	453	0.052	0.017	0.050 - 0.053	0.008 - 0.088
Saanen	0.00-0.05	160	0.058	0.015	0.056 - 0.061	0.024 - 0.088
Goats	0.05-0.10	81	0.103	0.013	0.101 - 0.106	0.088 - 0.138
Camosciata delle Alpi	0.05-0.10	49	0.102	0.011	0.099 - 0.105	0.089 - 0.132
Saanen	0.05-0.10	32	0.106	0.015	0.100 - 0.111	0.088 - 0.138
Goats	0.10-0.15	17	0.162	0.014	0.156 - 0.169	0.143 - 0.182
Camosciata delle Alpi	0.10-0.15	9	0.157	0.014	0.148 - 0.166	0.143 - 0.179
Saanen	0.10-0.15	8	0.168	0.012	0.159 - 0.176	0.153 - 0.182

Sheep

For sheep, we selected the coefficients relative to the regression model calculated for animals with a minimum FullGen equal to 6 (intercept = -0.03, slope = 1.06) and applied them on the 927 sheep with a FullGen ≥ 6 to

calculate the estimated F_{PED} values (Figure 3.17-A). We standardized the residuals and excluded 463 subjects falling under the 1st and over the 3rd quartiles. Finally, we calculated the new model using the 464 retained sheep: $F_{PED} = -0.03 + 1.02 * F_{ROH}$ ($R^2 = 0.97$, $p < 2.2e-16$) (Figure 3.17-B).

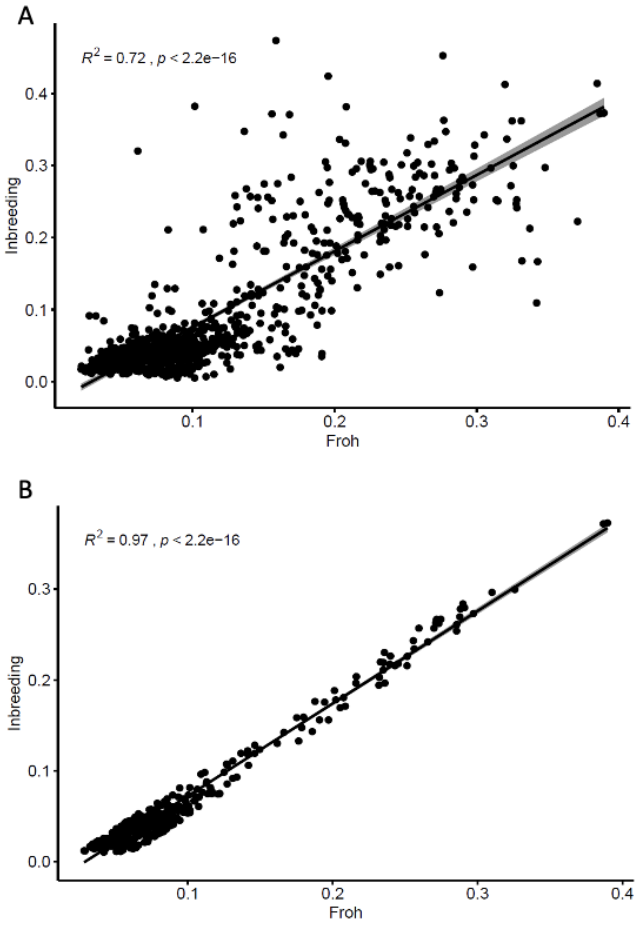


Figure 3.17 - Linear regression models to estimate FPED from FROH. (A) Linear regression model calculated on all the animals with FullGen ≥ 6. (B) Refined model for sheep with FullGen ≥ 6.

The estimated definitive F_{PED} intervals and relative F_{ROH} means and ranges for all the sheep with FullGen ≥ 6 are reported in Table 3.8. There is a significant difference in the breeds' mean F_{ROH} corresponding to an estimated

definitive F_{PED} of 0.00-0.05 and 0.05-0.10: Comisana has the lowest F_{ROH} values, followed by Massese, while Delle Langhe and Sarda present the highest values.

Table 3.8 - Number of sheep and F_{ROH} mean, standard deviation (SD), 95% confidence interval (CI), and range per class of predicted F_{PED}. Only Sarda sheep had predicted F_{PED} values > 0.2.

Breed	**Predicted F_{PED} class**	**N. subjects**	**F_{ROH} mean**	**F_{ROH} SD**	**F_{ROH} 95% CI**	**F_{ROH} range**
Sheep	0.00-0.05	419	0.055	0.014	0.054 - 0.057	0.022 - 0.078
Comisana	0.00-0.05	163	0.045	0.013	0.043 - 0.047	0.022 - 0.078
Massese	0.00-0.05	198	0.061	0.01	0.059 - 0.062	0.029 - 0.078
Sarda	0.00-0.05	33	0.068	0.009	0.065 - 0.071	0.041 - 0.078
Delle Langhe	0.00-0.05	25	0.067	0.012	0.062 - 0.072	0.027 - 0.078
Sheep	0.05-0.10	231	0.097	0.014	0.095 - 0.098	0.078 - 0.127
Comisana	0.05-0.10	2	0.092	0.014	0.072 - 0.111	0.082 - 0.101
Massese	0.05-0.10	59	0.09	0.011	0.087 - 0.093	0.078 - 0.120
Sarda	0.05-0.10	122	0.099	0.015	0.096 - 0.101	0.078 - 0.127
Delle Langhe	0.05-0.10	48	0.1	0.013	0.096 - 0.103	0.079 - 0.124
Sheep	0.10-0.15	89	0.15	0.015	0.146 - 0.153	0.128 - 0.176
Comisana	0.10-0.15	1	0.162			
Sarda	0.10-0.15	79	0.149	0.014	0.146 - 0.152	0.128 - 0.176
Delle Langhe	0.10-0.15	9	0.152	0.018	0.140 - 0.164	0.128 - 0.175
Sheep	0.15-0.20	71	0.199	0.013	0.196 - 0.203	0.177 - 0.225

Sarda	0.15-0.20	68	0.199	0.013	0.196 - 0.203	0.177 - 0.225
Delle Langhe	0.15-0.20	3	0.203	0.022	0.178 - 0.227	0.179 - 0.222
Sheep (Sarda)	0.20-0.25	56	0.25	0.015	0.246 - 0.254	0.225 - 0.274
Sheep (Sarda)	0.25-0.30	27	0.293	0.014	0.288 - 0.298	0.275 - 0.321
Sheep (Sarda)	0.30-0.35	13	0.336	0.013	0.329 - 0.343	0.324 - 0.371

Discussion

The inbreeding coefficient is a fundamental instrument for livestock husbandry: it is necessary for the increase the accuracy of the genomic and genetic breeding value of animals and the consequent improving of the breeds (Aguilar et al., 2020) it is one of the main parameters to monitor the selection programs and identify possible pitfalls in terms of loss of genetic biodiversity; and it helps breeders to choose the best mating schemes to improve animal welfare and to avoid inbreeding depression (which is especially relevant when using the AI) (Maltecca et al., 2020). As previously mentioned, genomic inbreeding coefficients are more reliable than pedigree-based ones, being free from possible registration inaccuracy (Rodríguez-Ramilo et al., 2019; Cortes-Hernández et al., 2021). This is the reason why the implementation of genomic tools at field level has been rising leading to genomic selection programs, not only in cattle, but also in other farm animals such as sheep and goats (Teissier et al., 2018; Cesarani et al., 2019; Zhu et al., 2021).

Nevertheless, the genotyping costs, although consistently reduced in recent years, cannot yet be afforded by the breeders for all their animals. This makes

it necessary to understand how to correctly interpret and compare inbreeding coefficients calculated from different data. Therefore, the present study aimed to explore the pedigree-based (F_{PED}) and genomic inbreeding coefficients (F_{GRM} and F_{ROH}), as well as their relationship, in a representative number of animals belonging to six Italian dairy small ruminant breeds differently managed.

Our results show that there are some differences in the distribution of the inbreeding coefficients, both within and between breeds. For example, F_{GRM} values tend to be more variable because they are highly influenced by the allele frequencies of the rare variants (Zhang et al., 2015); moreover, the particularly spread distribution found in the Delle Langhe breed could depend also on its smaller sample compared to the other breeds. The differences between F_{ROH} and F_{PED} coefficients were marked in goats, probably due to the great number of animals with a zero F_{PED} value. Regarding the percentage of F_{ROH} per ROH length class, the diverse management of the breeds can account for a portion of variability we observed: for example, the higher level of recent inbreeding found in Saanen compared with Camosciata delle Alpi could partially depend on the highest use of AI in the former (33% *vs* 25% in our samples, higher than the values, 11% and 12%, estimated in the whole populations enrolled to the herd books in the same years). The inbreeding estimated on these subjects is, for the most, low; this can depend either on a good management of the population and also from the low level of connection of the Italian small ruminant farms, with minimal exchange of animals except when the breeders make use of the AI or receive animals from a genetic center. The diffusion of AI, in particular in goat, is still limited, but it is rising due to the commercialization of foreign, especially French, semen. Avoiding

an excessive increase of inbreeding is a main goal of the breeders and their associations in order to prevent inbreeding depression, especially in small-sized populations, whose survival would be in danger if a reduction of the number of the breeding animals occurred. If the genomic inbreeding was estimated for all the animals of a farm, or at least a representative number, it would be possible to warn the breeders who are at risk of inbreeding-related issues.

The comparison of inbreeding coefficients with those reported in literature is hampered by the several different parameters that can be defined to estimate them. Average F_{PED} in a population vary based on the considered pedigree depth (i.e., number of generations tracked back), the use of unknown parent groups, and the involved computational algorithm (Colleau, 2002) ,some of which can estimate non-zero inbreeding coefficients also for animals with unknown parents (Aguilar and Misztal, 2008). F_{ROH} can be estimated starting from sliding (as in the present study) or consecutive (not overlapping regions) ROH, both varying according to the parameters used to call the regions. In literature, results about this parameter change considerably also within the same breed. One example is the Sarda dairy sheep: in this study, the average F_{ROH} for this breed was 0.135, using sliding regions, whereas (Cesarani et al., 2019) reported a value of 0.059 using consecutive regions with more strict parameters and (Mastrangelo et al., 2018) estimated an average F_{ROH} of 0.041. In the latter study, the authors reported F_{ROH} values also for Delle Langhe (0.080 *vs* 0.099 in this study), Comisana (0.016 *vs* 0.044 in this study), and Massese (0.055 *vs* 0.067 in this study). The different values between (Mastrangelo et al., 2018) and the present study are mainly due to the minimum number of SNP to call a ROH, that was 30 in that study and 20 in

the present one. Finally, F_{GRM} values change according to the method used to create the genomic relationship matrix (Amin et al., 2007; VanRaden, 2008; Yang et al., 2010) in which different scaling and weighting factors are used, and according to the quality control made on SNPs data, in particular allele frequency. In fact, some studies suggested the use of fixed allele frequency (0.5) when building the GRM. According to (Bjelland et al., 2013), this kind of matrix is a homozygosity measure adjusted to conform to the distribution of the pedigree inbreeding and, therefore, the correlations between F_{PED} and F_{GRM} estimated with this fixed frequency are extremely high.

The analysis of the relationship among inbreeding coefficients showed that the correlation of F_{PED} is generally higher with F_{ROH} (0.30 in goats and 0.82 in sheep) than F_{GRM} (0.27 and 0.35, respectively), especially when the pedigrees are more complete. The results of our analyses are consistent with those previously reported in other studies (Zhang et al., 2015; Alemu et al., 2021). Of interest is the negative correlation observed between F_{PED} and F_{GRM} in Comisana and Massese breeds and already reported in literature for other species (Hidalgo et al., 2021). Pedigree depth was directly related to the correlation between F_{ROH} and F_{PED}; this means that, in order to estimate F_{PED} accurately, at least four to six full generations should be known. It would be important to bring this information to the attention of the breeders, who should genotype especially those animals with a small number of known ancestors. We developed a simple regression model that aims to estimate a more accurate pedigree-based inbreeding coefficient on the basis of F_{ROH}, and can be useful for the breeders to compare the inbreeding of genotyped and non-genotyped animals. We noticed that goats had shallower pedigrees and a higher number of inbreeding coefficients equal to zero. For this reason, since

the two species undergo similar management and often used to be bred together, and all the breeds are used for milk and cheese production, the initial model included both the species. The results of the model also highlighted that different breeds have different mean Fp_{ROH} per estimated F_{PED} class. This aspect should be considered when implementing the genomic inbreeding in the animal evaluation, for example through the estimation of indexes. Indeed, previously published works tried to propose a new model based on ROH for the computation of genomic breeding values in cattle (G_{ROH}), finding excellent levels of accuracy (Luan et al., 2014). This could be a very useful solution in small ruminant populations in which we have less accurate pedigree information and a lower level of connection between farms.

Conclusions

The estimate of the inbreeding is of fundamental importance for livestock husbandry. Alongside the pedigree information, which is still widely used in the population management, genomic data is gaining more and more relevance; nevertheless, massive genotyping is still too expensive for small ruminant breeders. In the light of the above, it is necessary to understand the difference and the relation between inbreeding values estimated from different data. This becomes particularly important when we consider that the breeder associations have recently introduced the index estimation for small ruminants, including both pedigree and genomic data (single-step genomic best linear unbiased prediction – ssGBLUP), and that the latter allow to assess the population situation in terms of genetic variability. Moreover, breeders should be reminded about the importance of an accurate collection of the pedigree information and, especially when it is not possible, of the advantages

of genotyping their sheep and goats. The correctness of the data, indeed, allows to improve small ruminant breeds in different ways: breeders and their associations can choose the best breeding animals and mating schemes, in order to improve welfare, production, and health-related traits; minimize the risk of inbreeding depression; and better understand the effect of their management on inbreeding trends of the sheep and goat populations. On the contrary, inaccurate information might limit the breed genetic progress and lead to the reduction of the genetic variability, which is essential to safeguard livestock biodiversity and to face the future climate challenges.

CHAPTER 4 Conclusions

4.1 Final considerations

In the last five years genomics and tools such as complete sequencing and array SNPs genotyping have become an integral part of animal science studies. These new technologies are being applied to higher-income species such as cattle to provide breeders and their associations with data serving as a concrete basis for their decisions. To this end, my Ph.D. work has focused on lower-income species like small ruminants such as goats and sheep, to disentangle the genetic basis of selection and adaptation in order to provide pertinent data for breeders of these species, which have become increasingly important as a resource in the panorama of world livestock with respect to the challenges of climate change.

The relation between climate change and the genetic background of Italian breeds was the focus of the work presented in the first paragraph of the third Chapter of this thesis. Here the foundations were laid to study the association between the genotypes of Italian breeds and the disparity of climatic environments where they are bred. Important relations were found between SNP markers and variables, such as humidity and temperature, findings deserving further investigation by means of full sequencing data now in production for many of these breeds. The other focal point of this work was to highlight the possible influence of the geopolitical situation of Italy of the past 200 years on the current genetic background of Italian goat breeds, for which several patterns of similarity were found.

Another approach to clarifying the genetic panorama of goat populations and breeds in Italy was from the perspective of traditional breeding typology, and in particular, the signs that selection leaves. Runs Of Homozygosity (ROH)

indicated both the selection signatures left in these populations and the level of genomic inbreeding accumulated over time, based on different types of breeding and animal management. It was also possible to begin to highlight how genomic data could have a strong impact on managing and monitoring inbreeding, especially in small populations.

Our third work took up the last point mentioned, i.e. the relationship between inbreeding data calculated from pedigree data and that calculated from genomic data in populations of small ruminants in Italy. Analysis of these relations revealed that the parameters calculated from the genomic data, in particular one calculated through the ROH (FROH), could have greater reliability and be a fundamental integration to the already used pedigree data. In such way we could arrive to the definition of more reliable genomics index and a much accurate calculation of the effects of inbreeding depression. Moreover, this work also intends to relate the two inbreeding parameters (genomic and pedigree) in a clearer and more understandable way, in order to favor the use and understanding of genomic data by breeders and breeders' associations.

In conclusion, my thesis work underscores the importance of the use and the production of genomic data in the field of small ruminants in order to fully comprehend the great heritage of biodiversity of these species, especially in Italy. In the light of what will be the challenges that the ongoing climate change, investigate small ruminants genetics and genomics will certainly be one of the key points in the animal science field. This will be crucial to understand the basis of adaptation to the different environments and climate (especially warm and arid ones) because, unlike other domesticated species,

sheep and goats have practically colonized and adapted to all the possible climatic situations and ecological environments on our planet.

This is the perfect starting model to determine the genes that lie beyond specific adaptational traits like hot and cold tolerance, circadian rhythm, resistance to specific disease and so on, not only in small ruminants, but also in other domesticate species and even in humans.

The other important aspect my study highlights is the need to implement genomic data as a standard praxis in dealing with small ruminant populations. Evaluating the effects of selection and inbreeding could thereby be significantly improved, both in the most marginal and endangered populations as well as the most productive. With the better understanding of flock genetics that genomic data provides, breeders can improve the management, mating schemes, productivity, and above all, the welfare of their animals.

4.2 Perspectives

Today in Italy we are just starting to use genomic data in the management of sheep and goat populations. In particular, we are working with data deriving from genotyping through medium density 50k SNPchip.

At the international level there many projects are already underway. Their aim is to produce and use of complete sequencing data of various breeds, both sheep and goats, to deepen our knowledge of these species more and more. On the wave of these, in Italy, the next step that we intend to take is starting to produce a complete sequence of our autochthonous breeds and at the same time expand the production of 50k SNPchip, trying to sample and study all the possible biodiversity of small ruminants populations. The hope is that this new data will further our understanding regarding adaptation to climate change, and above all, help breeders and their associations manage their flocks and populations more efficiently and profitably, as well as improve animal welfare and the environment.

CHAPTER 5 References

ADAPTmap project Available at: http://www.goatadaptmap.org/.

Aguilar, I., Fernandez, E. N., Blasco, A., Ravagnolo, O., and Legarra, A. (2020). Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. *J. Anim. Breed. Genet.* 137, 356–364. doi:10.1111/jbg.12470.

Aguilar, I., and Misztal, I. (2008). Technical Note: Recursive Algorithm for Inbreeding Coefficients Assuming Nonzero Inbreeding of Unknown Parents. *J. Dairy Sci.* 91, 1669–1672. doi:10.3168/jds.2007-0575.

Alberto, F. J., Boyer, F., Orozco-Terwengel, P., Streeter, I., Servin, B., De Villemereuil, P., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* 9. doi:10.1038/s41467-018-03206-y.

Alemu, S. W., Kadri, N. K., Harland, C., Faux, P., Charlier, C., Caballero, A., et al. (2021). An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. *Heredity* 126, 410–423. doi:10.1038/s41437-020-00383-9.

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1655. doi:10.1101/GR.094052.109.

Amaral, A. J., Megens, H.-J., Crooijmans, R. P. M. A., Heuven, H. C. M., and Groenen, M. A. M. (2008). Linkage Disequilibrium Decay and Haplotype Block Structure in the Pig. *Genetics* 179, 569–579. doi:10.1534/genetics.107.084277.

Amills, M., Capote, J., and Tosser-Klopp, G. (2017). Goat domestication and breeding: a jigsaw of historical, biological and molecular data with missing pieces. *Anim. Genet.* 48, 631–644. doi:10.1111/age.12598.

Amin, N., van Duijn, C. M., and Aulchenko, Y. S. (2007). A Genomic Background Based Method for Association Analysis in Related Individuals. *PLoS ONE* 2, e1274. doi:10.1371/journal.pone.0001274.

An, B., Xia, J., Chang, T., Wang, X., Miao, J., Xu, L., et al. (2018). Genome-wide association study identifies loci and candidate genes for internal organ

weights in Simmental beef cattle. *Physiol. Genomics* 50, 523–531. doi:10.1152/physiolgenomics.00022.2018.

Andonov, S., Lourenco, D. A. L., Fragomeni, B. O., Masuda, Y., Pocrnic, I., Tsuruta, S., et al. (2017). Accuracy of breeding values in small genotyped populations using different sources of external information—A simulation study. *J. Dairy Sci.* 100, 395–401. doi:10.3168/jds.2016-11335.

Assonapa Available at: http://www.assonapa.it/.

Baes, C. F., Makanjuola, B. O., Miglior, F., Marras, G., Howard, J. T., Fleming, A., et al. (2019). Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. *J. Dairy Sci.* 102, 2807–2817. doi:10.3168/jds.2018-15520.

Barbato, M., Hailer, F., Upadhyay, M., Del Corvo, M., Colli, L., Negrini, R., et al. (2020). Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. *Sci. Rep.* 10, 1279–1279. doi:10.1038/s41598-020-57880-4.

Baylis, M., and Risley, C. (2013). "Infectious Diseases, Climate Change Effects on," in *Infectious Diseases*, eds. P. Kanki and D. J. Grimes (New York, NY: Springer New York), 117–146. doi:10.1007/978-1-4614-5719-0_6.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* 5, 1–12. doi:10.1038/sdata.2018.214.

Bertolini, F., Cardoso, T. F., Marras, G., Nicolazzi, E. L., Rothschild, M. F., Amills, M., et al. (2018a). Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genet. Sel. Evol.* 50, 59. doi:10.1186/s12711-018-0424-8.

Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., et al. (2018b). Signatures of selection and environmental adaptation across the goat genome post-domestication 06 Biological Sciences 0604 Genetics. *Genet. Sel. Evol.* 50, 1–24. doi:10.1186/s12711-018-0421-y.

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., et al. (2017). Single-molecule sequencing and chromatin conformation capture

enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* 49, 643–650. doi:10.1038/ng.3802.

Bjelland, D. W., Weigel, K. A., Vukasinovic, N., and Nkrumah, J. D. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 96, 4697–4706. doi:10.3168/jds.2012-6435.

Bleasdale, M., Richter, K. K., Janzen, A., Brown, S., Scott, A., Zech, J., et al. (2021). Ancient proteins provide evidence of dairy consumption in eastern Africa. *Nat. Commun.* 12, 632. doi:10.1038/s41467-020-20682-3.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., et al. (2010). Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics* 186, 241–262. doi:10.1534/genetics.110.117275.

Broman, K. W., and Weber, J. L. (1999). Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain. *Am. J. Hum. Genet.* 65, 1493–1500. doi:10.1086/302661.

Browning, B. L., and Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459–471. doi:10.1534/genetics.113.150029.

Browning, S. R., and Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi:10.1086/521987.

Bruford, M. W., Ginja, C., Hoffmann, I., Joost, S., Orozco-terWengel, P., Alberto, F. J., et al. (2015). Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. *Front. Genet.* 6, 314–314. doi:10.3389/fgene.2015.00314.

Caballero, A., Villanueva, B., and Druet, T. (2021). On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. *Evol. Appl.* 14, 416–428. doi:10.1111/eva.13126.

Calleja-Cabrera, J., Boter, M., Oñate-Sánchez, L., and Pernas, M. (2020). Root Growth Adaptation to Climate Change in Crops. *Front. Plant Sci.* 11, 544. doi:10.3389/fpls.2020.00544.

Carta, A., Casu, S., and Salaris, S. (2009). Invited review: Current state of genetic improvement in dairy sheep. *J. Dairy Sci.* 92, 5814–5833. doi:10.3168/jds.2009-2479.

Caye, K., Jumentier, B., Lepeule, J., and François, O. (2019). LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Mol. Biol. Evol.* 36, 852–860. doi:10.1093/molbev/msz008.

Ceballos, F. C. (2018). Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. 12.

Cesarani, A., Sechi, T., Gaspa, G., Usai, M. G., Sorbolini, S., Macciotta, N. P. P., et al. (2019). Investigation of genetic diversity and selection signatures between Sarda and Sardinian Ancestral black, two related sheep breeds with evident morphological differences. *Small Rumin. Res.* 177, 68–75. doi:10.1016/j.smallrumres.2019.06.014.

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

Chebii, V. J., Mpolya, E. A., Muchadeyi, F. C., and Domelevo Entfellner, J.-B. (2021). Genomics of Adaptations in Ungulates. *Animals* 11, 1617. doi:10.3390/ani11061617.

Cheng, J., Cao, X., Hanif, Q., Pi, L., Hu, L., Huang, Y., et al. (2020). Integrating Genome-Wide CNVs Into QTLs and High Confidence GWAScore Regions Identified Positional Candidates for Sheep Economic Traits. *Front. Genet.* 11, 569. doi:10.3389/fgene.2020.00569.

Ciani, E., Crepaldi, P., Nicoloso, L., Lasagna, E., Sarti, F. M., Moioli, B., et al. (2014a). Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. *Anim. Genet.* 45, 256–266. doi:10.1111/age.12106.

Ciani, E., Crepaldi, P., Nicoloso, L., Lasagna, E., Sarti, F. M., Moioli, B., et al. (2014b). Genome-wide analysis of Italian sheep diversity reveals a strong

geographic pattern and cryptic relationships between breeds. *Anim. Genet.* 45, 256–266. doi:10.1111/age.12106.

Clutton-Brock, J. (1999). *A natural history of domesticated mammals*. Cambridge University Press.

Colleau, J.-J. (2002). An indirect approach to the extensive calculation of relationship coefficients. *Genet. Sel. Evol.* 34, 409. doi:10.1186/1297-9686-34-4-409.

Colussi, S., Desiato, R., Beltramo, C., Peletto, S., Modesto, P., Maniaci, M. G., et al. (2019). A single nucleotide variant in the promoter region of the CCR5 gene increases susceptibility to arthritis encephalitis virus in goats. *BMC Vet. Res.* 15, 1–6. doi:10.1186/s12917-019-1979-5.

Copenhaver, G. P., Browne, W. E., and Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. *Proc Natl Acad Sci USA*, 6.

Cortellari, M., Barbato, M., Talenti, A., Bionda, A., Carta, A., Ciampolini, R., et al. (2021). The climatic and genetic heritage of Italian goat breeds with genomic SNP data. *Sci. Rep.* 11, 10986. doi:10.1038/s41598-021-89900-2.

Cortes-Hernández, J., García-Ruiz, A., Vásquez-Peláez, C. G., and Ruiz-Lopez, F. de J. (2021). Correlation of Genomic and Pedigree Inbreeding Coefficients in Small Cattle Populations. *Animals* 11, 3234. doi:10.3390/ani11113234.

Crepaldi, P., Corti, M., and Cicogna, M. (1999). Factors affecting milk production and proli®cacy of Alpine goats in Lombardy (Italy). *Small Rumin. Res.*, 6.

Crislip, G. R., Douma, L. G., Masten, S. H., Cheng, K. Y., Lynch, I. J., Johnston, J. G., et al. (2020). Differences in renal BMAL1 contribution to Na+homeostasis and blood pressure control in male and female mice. *Am. J. Physiol. - Ren. Physiol.* 318, F1463–F1477. doi:10.1152/AJPRENAL.00014.2020.

D, H., and M, G. (2017). Climate Change Effects on Livestock Feed Resources: A Review. *J. Fish. Livest. Prod.* 05. doi:10.4172/2332-2608.1000259.

Dahl, G. E., Tao, S., and Laporta, J. (2020). Heat Stress Impacts Immune Status in Cows Across the Life Cycle. *Front. Vet. Sci.* 7, 116. doi:10.3389/fvets.2020.00116.

Daly, K. G., Maisano Delser, P., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., et al. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* 361, 85–88. doi:10.1126/science.aas9411.

Das, R., Sailo, L., Verma, N., Bharti, P., Saikia, J., Imtiwati, et al. (2016). Impact of heat stress on health and performance of dairy animals: A review. *Vet. World* 9, 260–268. doi:10.14202/vetworld.2016.260-268.

Davenport, K. M., Bickhart, D. M., Worley, K., Murali, S. C., Cockett, N., Heaton, M. P., et al. (2021). 42 An Improved, High-quality Ovine Reference Genome Assembly. *J. Anim. Sci.* 99, 23–24. doi:10.1093/jas/skab235.039.

Deng, J., Xie, X.-L., Wang, D.-F., Zhao, C., Lv, F.-H., Li, X., et al. (2020). Paternal Origins and Migratory Episodes of Domestic Sheep. *Curr. Biol.* 30, 4085-4095.e6. doi:10.1016/j.cub.2020.07.077.

Denoyelle, L., Talouarn, E., Bardou, P., Colli, L., Alberti, A., Danchin, C., et al. (2021). VarGoats project: a dataset of 1159 whole-genome sequences to dissect Capra hircus global diversity. *Genet. Sel. Evol.* 53, 86. doi:10.1186/s12711-021-00659-6.

Di Trana, A., Sepe, L., Di Gregorio, P., Di Napoli, M. A., Giorgio, D., Caputo, A. R., et al. (2015). "The Role of Local Sheep and Goat Breeds and Their Products as a Tool for Sustainability and Safeguard of the Mediterranean Environment," in *The Sustainability of Agro-Food and Natural Resource Systems in the Mediterranean Basin*, ed. A. Vastola (Cham: Springer International Publishing), 77–112. doi:10.1007/978-3-319-16357-4_7.

Doekes, H. P., Bijma, P., Veerkamp, R. F., de Jong, G., Wientjes, Y. C. J., and Windig, J. J. (2020). Inbreeding depression across the genome of Dutch Holstein Friesian dairy cattle. *Genet. Sel. Evol.* 52, 64. doi:10.1186/s12711-020-00583-1.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat. Biotechnol.* 31, 135–141. doi:10.1038/nbt.2478.

Druet, T., and Gautier, M. (2017). A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Mol. Ecol.* 26, 5820–5841. doi:10.1111/mec.14324.

Dupré, S. M., Burt, D. W., Talbot, R., Downing, A., Mouzaki, D., Waddington, D., et al. (2008). Identification of melatonin-regulated genes in the ovine pituitary pars tuberalis, a target site for seasonal hormone control. *Endocrinology* 149, 5527–5539. doi:10.1210/en.2008-0834.

Edea, Z., Dadi, H., Dessie, T., Uzzaman, M. R., Rothschild, M. F., Kim, E.-S., et al. (2018). Genome-wide scan reveals divergent selection among taurine and zebu cattle populations from different regions. *Anim. Genet.* 49, 550–563. doi:10.1111/age.12724.

Epstein, P. R. (2001). Climate change and emerging infectious diseases. *Microbes Infect.* 3, 747–754. doi:10.1016/S1286-4579(01)01429-0.

Eusebi, P. G., Martinez, A., and Cortes, O. (2019). Genomic Tools for Effective Conservation of Livestock Breed Diversity. *Diversity* 12, 8. doi:10.3390/d12010008.

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* 193, 929–941. doi:10.1534/genetics.112.147231.

Fay, J. C., and Wu, C.-I. (2000). Hitchhiking Under Positive Darwinian Selection. 9.

Felsenstein., J. (1989). *PHYLIP - Phylogeny Inference Package (Version 3.2)*.

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi:10.1002/joc.5086.

Fleming-Waddell, J. N., Olbricht, G. R., Taxis, T. M., White, J. D., Vuocolo, T., Craig, B. A., et al. (2009). Effect of DLK1 and RTL1 but not MEG3 or MEG8 on muscle gene expression in callipyge lambs. *PLoS ONE* 4. doi:10.1371/journal.pone.0007399.

Fratianni, S., and Acquaotta, F. (2017). “The Climate of Italy,” in (Springer, Cham), 29–38. doi:10.1007/978-3-319-26194-2_4.

Frichot, E., Schoville, S. D., Bouchard, G., and François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* 30, 1687–1699. doi:10.1093/molbev/mst063.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915–925.

Gaughan, J., and Cawdell-Smith, A. J. (2015). “Impact of Climate Change on Livestock Production and Reproduction,” in *Climate Change Impact on Livestock: Adaptation and Mitigation*, eds. V. Sejian, J. Gaughan, L. Baumgard, and C. Prasad (New Delhi: Springer India), 51–60. doi:10.1007/978-81-322-2265-1_4.

Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi:10.1093/hmg/ddi493.

Gobikrushanth, M., Purfield, D. C., Canadas, E. R., Herlihy, M. M., Kenneally, J., Murray, M., et al. (2019). Anti-Müllerian hormone in grazing dairy cows: Identification of factors affecting plasma concentration, relationship with phenotypic fertility, and genome-wide associations. *J. Dairy Sci.* 102, 11622–11635. doi:10.3168/jds.2019-16979.

Goodchild, M. F. (1992). Geographical information science. *Int. J. Geogr. Inf. Syst.* 6, 31–45. doi:10.1080/02693799208901893.

Gorssen, W., Meyermans, R., Janssens, S., and Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genet. Sel. Evol.* 53, 2. doi:10.1186/s12711-020-00599-7.

Gouveia, J. J. de S., Silva, M. V. G. B. da, Paiva, S. R., and Oliveira, S. M. P. de (2014). Identification of selection signatures in livestock species. *Genet. Mol. Biol.* 37, 330–342. doi:10.1590/S1415-47572014000300004.

Grossi, G., Goglio, P., Vitali, A., and Williams, A. G. (2019). Livestock and climate change: impact of livestock on climate and mitigation strategies. *Anim. Front.* 9, 69–76. doi:10.1093/af/vfy034.

Günther, T., and Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics* 195, 205–220. doi:10.1534/genetics.113.152462.

Guo, J., Tao, H., Li, P., Li, L., Zhong, T., Wang, L., et al. (2018). Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Sci. Rep.* 8, 10405. doi:10.1038/s41598-018-28719-w.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92, 433–443. doi:10.3168/jds.2008-1646.

Hayes, B. J., and Goddard, M. E. (2008). Technical note: Prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.* 86, 2089–2092. doi:10.2527/jas.2007-0733.

He, S., Di, J., Han, B., Chen, L., Liu, M., and Li, W. (2020). Genome-Wide Scan for Runs of Homozygosity Identifies Candidate Genes Related to Economically Important Traits in Chinese Merino. *Animals* 10, 524. doi:10.3390/ani10030524.

Hidalgo, J., Cesarani, A., Garcia, A., Sumreddee, P., Larios, N., Mancin, E., et al. (2021). Genetic Background and Inbreeding Depression in Romosinuano Cattle Breed in Mexico. *Animals* 11, 321. doi:10.3390/ani11020321.

Hijmans, R. J. (2020). Geographic Data Analysis and Modeling [R package raster version 3.3-13]. Available at: https://cran.r-project.org/web/packages/raster/index.html.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi:10.1002/joc.1276.

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Anim. Genet.* 41 Suppl 1. doi:10.1111/J.1365-2052.2010.02043.X.

Ibeagha-Awemu, E. M., Peters, S. O., Akwanji, K. A., Imumorin, I. G., and Zhao, X. (2016). High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Sci. Rep.* 6, 1–18. doi:10.1038/srep31109.

International Goat Genome Consortium Available at: http://www.goatgenome.org/.

International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research:, Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi:10.1038/35057062.

International Sheep Genomics Consortium Available at: http://www.sheephapmap.org/.

Islam, R., Liu, Z., Li, Y., Jiang, L., and Ma, Y. (2020). Conservation Assessment of the State Goat Farms by Using SNP Genotyping Data. *Genes* 11, 652. doi:10.3390/genes11060652.

ISTAT (2010). 6° Censimento generale dell'agricoltura. ISTAT Available at: https://www.istat.it/it/archivio/66591.

Jaton, C. (2018). Genome-wide association study and in silico functional analysis of the number of embryos produced by Holstein donors. 101, 10.

Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., et al. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Mol. Ecol.* 16, 3955–3969. doi:10.1111/j.1365-294X.2007.03442.x.

Joost, S., Kalbermatten, M., and Bonin, A. (2008). Spatial analysis method (sam): a software tool combining molecular and environmental data to identify candidate loci for selection. *Mol. Ecol. Resour.* 8, 957–960. doi:10.1111/j.1755-0998.2008.02162.x.

Joy, A., Dunshea, F. R., Leury, B. J., Clarke, I. J., DiGiacomo, K., and Chauhan, S. S. (2020). Resilience of Small Ruminants to Climate Change and Increased Environmental Temperature: A Review. *Animals* 10, 867–867. doi:10.3390/ani10050867.

Kaback, D. B. (1996). Chromosome–size dependent control of meiotic recombination in humans. *Nat. Genet.* 13, 20–21. doi:10.1038/ng0596-20.

Kaback, D. B., Barber, D., Mahon, J., Lamb, J., and You, J. (1999). Chromosome Size-Dependent Control of Meiotic Reciprocal Recombination in

Saccharomyces cerevisiae: The Role of Crossover Interference. *Genetics* 152, 1475–1486. doi:10.1093/genetics/152.4.1475.

Kamalakkannan, R., Kumar, S., Bhavana, K., Prabhu, V. R., Machado, C. B., Singha, H. S., et al. (2021). Evidence for independent domestication of sheep mtDNA lineage A in India and introduction of lineage B through Arabian sea route. *Sci. Rep.* 11, 19733. doi:10.1038/s41598-021-97761-y.

Kijas, J. W., Porto-Neto, L., Dominik, S., Reverter, A., Bunch, R., McCulloch, R., et al. (2014). Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Anim. Genet.* 45, 754–757. doi:10.1111/age.12197.

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., et al. (2009). A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS ONE* 4, e4668. doi:10.1371/journal.pone.0004668.

Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., et al. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* 116, 255–264. doi:10.1038/hdy.2015.94.

Kim, Y., and Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160, 765–777.

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., and McKeigue, P. M. (2010). Genomic Runs of Homozygosity Record Population History and Consanguinity. *PLoS ONE* 5, 7.

Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., et al. (2012). CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling: *CliMond: climate surfaces for bioclimatic modelling*. *Methods Ecol. Evol.* 3, 53–64. doi:10.1111/j.2041-210X.2011.00134.x.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi:10.1101/gr.092759.109.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., et al. (2020). Package ‘ caret .’

Lamb, H. J., Hayes, B. J., Nguyen, L. T., and Ross, E. M. (2020). The Future of Livestock Management: A Review of Real-Time Portable Sequencing Applied to Livestock. *Genes* 11, 1478. doi:10.3390/genes11121478.

Legarra, A., Baloche, G., Barillet, F., Astruc, J. M., Soulas, C., Aguerre, X., et al. (2014). Within- and across-breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech, and Basco-Béarnaise. *J. Dairy Sci.* 97, 3200–3212. doi:10.3168/jds.2013-7745.

Leutenegger, A.-L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F., et al. (2003). Estimation of the Inbreeding Coefficient through Use of Genomic Data. *Am. J. Hum. Genet.* 73, 516–523. doi:10.1086/378207.

Li, Y., Zhang, X.-X., Mao, R.-L., Yang, J., Miao, C.-Y., Li, Z., et al. (2017). Ten Years of Landscape Genomics: Challenges and Opportunities. *Front. Plant Sci.* 8, 2136. doi:10.3389/fpls.2017.02136.

Lin, S. C., Kuo, C. C., and Chan, C. H. (2006). Association of a BTLA gene polymorphism with the risk of rheumatoid arthritis. *J. Biomed. Sci.* 13, 853–860. doi:10.1007/s11373-006-9113-7.

Liu, H., Wang, T., Wang, J., Quan, F., and Zhang, Y. (2013). Characterization of Liaoning Cashmere Goat Transcriptome: Sequencing, De Novo Assembly, Functional Annotation and Comparative Analysis. *PLoS ONE* 8, 1–11. doi:10.1371/journal.pone.0077062.

Luan, T., Yu, X., Dolezal, M., Bagnato, A., and Meuwissen, T. H. (2014). Genomic prediction based on runs of homozygosity. *Genet. Sel. Evol.* 46, 64. doi:10.1186/s12711-014-0064-6.

Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., et al. (2014). Adaptations to Climate-Mediated Selective Pressures in Sheep. *Mol. Biol. Evol.* 31, 3324–3343. doi:10.1093/molbev/msu264.

Lv, F.-H., Peng, W.-F., Yang, J., Zhao, Y.-X., Li, W.-R., Liu, M.-J., et al. (2015). Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep. *Mol. Biol. Evol.* 32, 2515–2533. doi:10.1093/molbev/msv139.

Macciotta, N. P. P., Colli, L., Cesarani, A., Ajmone-Marsan, P., Low, W. Y., Tearle, R., et al. (2021). The distribution of runs of homozygosity in the

genome of river and swamp buffaloes reveals a history of adaptation, migration and crossbred events. *Genet. Sel. Evol.* 53, 20. doi:10.1186/s12711-021-00616-3.

Maja Ferencakovic, Edin HAMZIC, Birgit GREDLER, Ino CURIK, and Johann SÖLKNER (2011). Runs of Homozygosity Reveal Genome- wide Autozygosity in the Austrian Fleckvieh Cattle. 76, 4.

Makanjuola, B. O., Maltecca, C., Miglior, F., Schenkel, F. S., and Baes, C. F. (2020). Effect of recent and ancient inbreeding on production and fertility traits in Canadian Holsteins. *BMC Genomics* 21, 605. doi:10.1186/s12864-020-07031-w.

Maltecca, C., Tiezzi, F., Cole, J. B., and Baes, C. (2020). Symposium review: Exploiting homozygosity in the era of genomics—Selection, inbreeding, and mating programs. *J. Dairy Sci.* 103, 5302–5313. doi:10.3168/jds.2019-17846.

Manfredi, M. T., Di Cerbo, A. R., Zanzani, S., and Stradiotto, K. (2010). Breeding management in goat farms of Lombardy, northern Italy: Risk factors connected to gastrointestinal parasites. *Small Rumin. Res.* 88, 113–118. doi:10.1016/j.smallrumres.2009.12.018.

Marino, R., Atzori, A. S., D'Andrea, M., Iovane, G., Trabalza-Marinucci, M., and Rinaldi, L. (2016). Climate change: Production performance, health issues, greenhouse gas emissions and mitigation strategies in sheep and goat farming. *Small Rumin. Res.* 135, 50–59. doi:10.1016/J.SMALLRUMRES.2015.12.012.

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., and Portolano, B. (2018). Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. 11.

Mastrangelo, S., Di Gerlando, R., Sardina, M. T., Sutera, A. M., Moscarelli, A., Tolone, M., et al. (2021). Genome-Wide Patterns of Homozygosity Reveal the Conservation Status in Five Italian Goat Populations. *Animals* 11, 1510. doi:10.3390/ani11061510.

Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., et al. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in

Valle del Belice sheep. *Genet. Sel. Evol.* 49, 84. doi:10.1186/s12711-017-0360-z.

Mazinani, M., and Rude, B. (2020). Population, World Production and Quality of Sheep and Goat Products. *Am. J. Anim. Vet. Sci.* 15, 291–299. doi:10.3844/ajavsp.2020.291.299.

McHugo, G. P., Dover, M. J., and MacHugh, D. E. (2019). Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. *BMC Biol.* 17, 98. doi:10.1186/s12915-019-0724-7.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–72. doi:10.1016/j.ajhg.2008.08.007.

Mdladla, K., Dzomba, E. F., and Muchadeyi, F. C. (2017). The potential of landscape genomics approach in the characterization of adaptive genetic diversity in indigenous goat genetic resources: A South African perspective. *Small Rumin. Res.* 150, 87–92. doi:10.1016/j.smallrumres.2017.03.015.

Mdladla, K., Dzomba, E. F., and Muchadeyi, F. C. (2018). Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity* 120, 369–378. doi:10.1038/s41437-017-0044-z.

Meuwissen, T. H. E. (2020). Management of Genetic Diversity in the Era of Genomics. *Front. Genet.* 11, 16.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157.

Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. (2020). How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. *BMC Genomics* 21, 1–14. doi:10.1186/s12864-020-6463-x.

Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Garcia, J. F., Ajmone-Marsan, P., et al. (2017). BITE: an R package for biodiversity analyses. *bioRxiv*, 181610–181610. doi:10.1101/181610.

Milligan, B. G. (2003). Maximum-Likelihood Estimation of Relatedness. *Genetics* 163, 1153–1167. doi:10.1093/genetics/163.3.1153.

Nannini, D., Calistri, P., Giovannini, A., Di Ventura, M., Cafiero, M. A., Ferrari, G., et al. (2004). Health management of large transhumant animal populations and risk of bluetongue spread to disease-free areas. *Vet. Ital.* 40, 707–70712.

Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M. S., and Bernabucci, U. (2010). Effects of climate changes on animal production and sustainability of livestock systems. *Livest. Sci.* 130, 57–69. doi:10.1016/j.livsci.2010.02.011.

Nawaz, M. Y., Jimenez-Krassel, F., Steibel, J. P., Lu, Y., Baktula, A., Vukasinovic, N., et al. (2018). Genomic heritability and genome-wide association analysis of anti-Müllerian hormone in Holstein dairy heifers. *J. Dairy Sci.* 101, 8063–8075. doi:10.3168/jds.2018-14798.

Nicoloso, L., Bomba, L., Colli, L., Negrini, R., Milanesi, M., Mazza, R., et al. (2015). Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet. Sel. Evol.* 47, 1–10. doi:10.1186/s12711-015-0140-6.

Noce, S., Caporaso, L., and Santini, M. (2020). A new global dataset of bioclimatic indicators. *Sci. Data* 7, 398. doi:10.1038/s41597-020-00726-5.

Nonneman, D. J., Schneider, J. F., Lents, C. A., Wiedmann, R. T., Vallet, J. L., and Rohrer, G. A. (2016). Genome-wide association and identification of candidate genes for age at puberty in swine. *BMC Genet.* 17, 1–9. doi:10.1186/s12863-016-0352-y.

Nothnagel, M., Lu, T. T., Kayser, M., and Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum. Mol. Genet.* 19, 2927–2935. doi:10.1093/hmg/ddq198.

Onzima, R. B., Upadhyay, M. R., Doekes, H. P., Brito, L. F., Bosse, M., Kanis, E., et al. (2018). Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. *Front. Genet.* 9, 1–13. doi:10.3389/fgene.2018.00318.

Paiva, R. D. M., de Sousa, J. E. R., Ferreira, J., Cunha, E. E., de Paiva, M. P. S. L. M., Gouveia, A. M. G., et al. (2020). Population structure and effect of

inbreeding on milk yield of Saanen goats in Brazilian production systems. *Small Rumin. Res.* 192, 106194. doi:10.1016/j.smallrumres.2020.106194.

Panetto, J. C. do C., Machado, M. A., da Silva, M. V. G. B., Barbosa, R. S., dos Santos, G. G., Leite, R. de M. H., et al. (2017). Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. *Livest. Sci.* 204, 33–38. doi:10.1016/j.livsci.2017.08.008.

Pariset, L., Joost, S., Marsan, P. A., and Valentini, A. (2009). Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genet.* 10, 7–7. doi:10.1186/1471-2156-10-7.

Paschino, P., Stocco, G., Dettori, M. L., Pazzola, M., Marongiu, M. L., Pilo, C. E., et al. (2020). Characterization of milk composition, coagulation properties, and cheese-making ability of goats reared in extensive farms. *J. Dairy Sci.* 103, 5830–5843. doi:10.3168/jds.2019-17805.

Passamonti, M. M., Somenzi, E., Barbato, M., Chillemi, G., Colli, L., Joost, S., et al. (2021). The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock. *Animals* 11, 2833. doi:10.3390/ani11102833.

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., and Baldi, F. (2016). Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.* 48, 255–271. doi:10.1111/age.12526.

Polak, G., Gurgul, A., Jasielczuk, I., Szmatoła, T., Krupiński, J., and Bugno-Poniewierska, M. (2021). Suitability of Pedigree Information and Genomic Methods for Analyzing Inbreeding of Polish Cold-Blooded Horses Covered by Conservation Programs. *Genes* 12, 429. doi:10.3390/genes12030429.

Polsky, L., and von Keyserlingk, M. A. G. (2017). Invited review: Effects of heat stress on dairy cattle welfare. *J. Dairy Sci.* 100, 8645–8657. doi:10.3168/jds.2017-12651.

R Development Core Team, R. (2011). *R: A Language and Environment for Statistical Computing.* doi:10.1007/978-3-540-74686-7.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., and Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* 24, 4348–4370. doi:10.1111/mec.13322.

Riall, L. (2002). *The Italian Risorgimento*. Routledge doi:10.4324/9780203412343.

Rischkowsky, B., and Pilling, D. (2007). *The state of the world's animal genetic resources for food and agriculture*. Rome: Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations.

Rochat, E., and Joost, S. (2019). Spatial Areas of Genotype Probability (SPAG): predicting the spatial distribution of adaptive genetic variants under future climatic conditions. *bioRxiv*, 2019.12.20.884114-2019.12.20.884114. doi:10.1101/2019.12.20.884114.

Rodríguez-Ramilo, S. T., Elsen, J. M., and Legarra, A. (2019). Inbreeding and effective population size in French dairy sheep: Comparison between genomic and pedigree estimates. *J. Dairy Sci.* 102, 4227–4237. doi:10.3168/jds.2018-15405.

Rojas-Downing, M. M., Nejadhashemi, A. P., Harrigan, T., and Woznicki, S. A. (2017). Climate change and livestock: Impacts, adaptation, and mitigation. *Clim. Risk Manag.* 16, 145–163. doi:10.1016/J.CRM.2017.02.001.

Rosenstock, E., Ebert, J., and Scheibner, A. (2021). Cultured Milk: Fermented Dairy Foods along the Southwest Asian–European Neolithic Trajectory. *Curr. Anthropol.*, S000–S000. doi:10.1086/714961.

Rubel, F., Brugger, K., Haslinger, K., and Auer, I. (2017). The climate of the European Alps: Shift of very high resolution Köppen-Geiger climate zones 1800-2100. *Meteorol. Z.* 26, 115–125. doi:10.1127/metz/2016/0816.

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587–591. doi:10.1038/nature08832.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi:10.1038/nature01140.

Saif, R., Henkel, J., Jagannathan, V., Drögemüller, C., Flury, C., and Leeb, T. (2020). The LCORL Locus Is under Selection in Large-Sized Pakistani Goat Breeds. 10.

Sandrucci, A., Bava, L., Tamburini, A., Gislon, G., and Zucali, M. (2019). Management practices and milk quality in dairy goat farms in Northern Italy. *Ital. J. Anim. Sci.* 18, 1–12. doi:10.1080/1828051X.2018.1466664.

Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., and Mishra, B. P. (2020). Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livest. Sci.* 241, 104257. doi:10.1016/j.livsci.2020.104257.

SAS Institute Inc. *JMP*®. Cary, NC Available at: https://www.jmp.com/support/notes/35/282.html.

Schiavo, G., Bovo, S., Bertolini, F., Tinarelli, S., Dall'Olio, S., Nanni Costa, L., et al. (2020). Comparative evaluation of genomic inbreeding parameters in seven commercial and autochthonous pig breeds. *Animal* 14, 910–920. doi:10.1017/S175173111900332X.

Schultz, E. B., Santana, T. E. Z., Silva, F. F., Garcia, A. O., Oliveira, H. R., Rodrigues, M. T., et al. (2020). Short communication : Genetic parameter estimates for caprine arthritis encephalitis in dairy goats. *J. Dairy Sci.* 10. doi:10.3168/jds.2019-17740.

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* 18, 1–25. doi:10.1186/s12864-017-3754-y.

Sejian, V., Silpa, M. V., Reshma Nair, M. R., Devaraj, C., Krishnan, G., Bagath, M., et al. (2021). Heat Stress and Goat Welfare: Adaptation and Production Considerations. *Animals* 11, 1021. doi:10.3390/ani11041021.

Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., et al. (2020). The genetic heritage of Alpine local cattle breeds using genomic SNP data. *Genet. Sel. Evol.* 52, 40. doi:10.1186/s12711-020-00559-1.

Silva, D. B. S., Fonseca, L. F. S., Pinheiro, D. G., Magalhães, A. F. B., Muniz, M. M. M., Ferro, J. A., et al. (2020). Spliced genes in muscle from Nelore

Cattle and their association with carcass and meat quality. *Sci. Rep.* 10, 14701. doi:10.1038/s41598-020-71783-4.

Stella, A., Nicolazzi, E. L., Van Tassell, C. P., Rothschild, M. F., Colli, L., Rosen, B. D., et al. (2018). AdaptMap: exploring goat diversity and adaptation. *Genet. Sel. Evol.* 50, 61–61. doi:10.1186/s12711-018-0427-5.

Storfer, A., Patton, A., and Fraik, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Front. Genet.* 9. doi:10.3389/fgene.2018.00068.

Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., Colli, L., Masembe, C., et al. (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Mol. Ecol. Resour.* 17, 1072–1089. doi:10.1111/1755-0998.12629.

Sumreddee, P., Toghiani, S., Hay, E. H., Roberts, A., Aggrey, S. E., and Rekaya, R. (2020). Runs of homozygosity and analysis of inbreeding depression. *J. Anim. Sci.* 98, skaa361. doi:10.1093/jas/skaa361.

Szmatoła, T., Gurgul, A., Jasielczuk, I., Ząbek, T., Ropka-Molik, K., Litwińczuk, Z., et al. (2019). A comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. *Animals* 9. doi:10.3390/ani9121024.

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Sci. Rep.* 10, 13768. doi:10.1038/s41598-020-70816-2.

Talenti, A., Bertolini, F., Pagnacco, G., Pilla, F., Ajmone-Marsan, P., Rothschild, M. F., et al. (2017a). The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. *Mamm. Genome* 28, 114–128. doi:10.1007/s00335-017-9678-7.

Talenti, A., Dreger, D. L., Frattini, S., Polli, M., Marelli, S., Harris, A. C., et al. (2018). Studies of modern Italian dog populations reveal multiple patterns for domestic breed evolution. *Ecol. Evol.* 8, 2911–2925. doi:10.1002/ece3.3842.

Talenti, A., Frattini, S., Mastrangelo, S., Gerlando, R. Di, Portolano, B., Lasagna, E., et al. (2017b). Italian Goat Consortium: a collaborative project to study

the Italian caprine biodiversity. in *ASPA 22nd Congress* (Perugia: Italian Journal of Animal Science), 75–76.

Talenti, A., Nicolazzi, E. L., Chessa, S., Frattini, S., Moretti, R., Coizet, B., et al. (2016). A method for single nucleotide polymorphism selection for parentage assessment in goats. *J. Dairy Sci.* 99, 3646–3653. doi:10.3168/jds.2015-10077.

Teissier, M., Larroque, H., and Robert-Granié, C. (2018). Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *Genet. Sel. Evol.* 50, 31. doi:10.1186/s12711-018-0400-3.

The International Sheep Genomics Consortium, Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., et al. (2010). The sheep genome reference sequence: a work in progress: The sheep genome reference sequence: a work in progress. *Anim. Genet.* 41, 449–453. doi:10.1111/j.1365-2052.2010.02100.x.

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., et al. (2014). Design and Characterization of a 52K SNP Chip for Goats. *PLoS ONE* 9, e86227–e86227. doi:10.1371/journal.pone.0086227.

Vajana, E., Barbato, M., Colli, L., Milanesi, M., Rochat, E., Fabrizi, E., et al. (2018). Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Front. Genet.* 9, 385–385. doi:10.3389/fgene.2018.00385.

Van Den Berg, I., Fritz, S., Rodriguez, S., Rocha, D., Boussaha, M., Lund, M. S., et al. (2014). Concordance analysis for QTL detection in dairy cattle: A case study of leg morphology. *Genet. Sel. Evol.* 46, 1–14. doi:10.1186/1297-9686-46-31.

Van Vleck, L. D., and Pollak, E. J. (1987). *Genetics for Animal Sciences*. W.H. Freeman Available at: https://books.google.it/books?id=EkR-QgAACAAJ.

van Wettere, W. H. E. J., Kind, K. L., Gatford, K. L., Swinbourne, A. M., Leu, S. T., Hayman, P. T., et al. (2021). Review of the impact of heat stress on reproductive performance of sheep. *J. Anim. Sci. Biotechnol.* 12, 26. doi:10.1186/s40104-020-00537-z.

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980.

VarGoats Available at: http://www.goatgenome.org/vargoats.html.

Vatsiou, A. I., Bazin, E., and Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: A comparison of recent methods. *Mol. Ecol.* 25, 89–103. doi:10.1111/mec.13360.

Venter J. Craig, Adams Mark D., Myers Eugene W., Li Peter W., Mural Richard J., Sutton Granger G., et al. (2001). The Sequence of the Human Genome. *Science* 291, 1304–1351. doi:10.1126/science.1058040.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E., et al. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genet. Sel. Evol.* 53, 42. doi:10.1186/s12711-021-00635-0.

Vitti, J. J., Grossman, S. R., and Sabeti, P. C. (2013). Detecting Natural Selection in Genomic Data. *Annu Rev Genet* 47, 97–120. doi:10.1146/annurev-genet-111212-133526.

Voight, B. F., Kudaravalli, S., Wen, X., Pritchard, J. K., and Gudjonsson, S. (2006). A Map of Recent Positive Selection in the Human Genome. *PLoS Biol.* 4, e72–e72. doi:10.1371/journal.pbio.0040072.

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, e107101. doi:10.1371/journal.pone.0107101.

Wang, Z., Zhou, B., Zhang, T., Yan, X., Yu, Y., Li, J., et al. (2021). Assessing Genetic Diversity and Estimating the Inbreeding Effect on Economic Traits of Inner Mongolia White Cashmere Goats Through Pedigree Analysis. *Front. Vet. Sci.* 8, 665872. doi:10.3389/fvets.2021.665872.

Weller, J. I., Ezra, E., and Ron, M. (2017). Invited review: A perspective on the future of genomic selection in dairy cattle. *J. Dairy Sci.* 100, 8633–8644. doi:10.3168/jds.2017-12879.

Wellmann, R. (2019). Optimum contribution selection for animal breeding and conservation: the R package optiSel. *BMC Bioinformatics* 20, 25. doi:10.1186/s12859-018-2450-5.

Wright, S. (1949). THE GENETICAL STRUCTURE OF POPULATIONS. *Ann. Eugen.* 15, 323–354. doi:10.1111/j.1469-1809.1949.tb02451.x.

Xu, Q., Lin, Y., Wang, Y., Bai, W., and Zhu, J. (2020). Knockdown of KLF9 promotes the differentiation of both intramuscular and subcutaneous preadipocytes in goat. *Biosci. Biotechnol. Biochem.* 84, 1594–1602. doi:10.1080/09168451.2020.1767497.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569. doi:10.1038/ng.608.

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* 88, 76–82. doi:10.1016/j.ajhg.2010.11.011.

Yang, J., Li, W. R., Lv, F. H., He, S. G., Tian, S. L., Peng, W. F., et al. (2016). Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol. Biol. Evol.* 33, 2576–2592. doi:10.1093/molbev/msw129.

Yoshitane, H., Honma, S., Imamura, K., Nakajima, H., Nishide, S. Y., Ono, D., et al. (2012). JNK regulates the photic response of the mammalian circadian clock. *EMBO Rep.* 13, 455–461. doi:10.1038/embor.2012.37.

Zhang, Q., Calus, M. P., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16, 88. doi:10.1186/s12863-015-0227-7.

Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., et al. (2020). The origin of domestication genes in goats. *Sci. Adv.* 6, eaaz5216–eaaz5216. doi:10.1126/sciadv.aaz5216.

Zhu, S., Guo, T., Yuan, C., Liu, J., Li, J., Han, M., et al. (2021). Evaluation of Bayesian alphabet and GBLUP based on different marker density for genomic prediction in Alpine Merino sheep. *G3 GenesGenomesGenetics* 11, jkab206. doi:10.1093/g3journal/jkab206.

CHAPTER 6 Other works

6.1 Published

6.1.1 Genome-Wide Patterns of Homozygosity Reveal the Conservation Status in Five Italian Goat Populations.

Published on Animals 2021, 11 (6), 1510. DOI: 10.3390/ani11061510

Authors

Mastrangelo, S.; Di Gerlando, R.; Sardina, M. T.; Sutera, A. M.; Moscarelli, A.; Tolone, M.; **Cortellari, M**.; Marletta, D.; Crepaldi, P.; Portolano, B.

Abstract

The application of genomic technologies has facilitated the assessment of genomic inbreeding based on single nucleotide polymorphisms (SNPs). In this study, we computed several runs of homozygosity (ROH) parameters to investigate the patterns of homozygosity using Illumina Goat SNP50 in five Italian local populations: Argentata dell'Etna (N = 48), Derivata di Siria (N = 32), Girgentana (N = 59), Maltese (N = 16) and Messinese (N = 22). The ROH results showed well-defined differences among the populations. A total of 3687 ROH segments >2 Mb were detected in the whole sample. The Argentata dell'Etna and Messinese were the populations with the lowest mean number of ROH and inbreeding coefficient values, which reflect admixture and gene flow. In the Girgentana, we identified an ROH pattern related with recent inbreeding that can endanger the viability of the breed due

to reduced population size. The genomes of Derivata di Siria and Maltese breeds showed the presence of long ROH (>16 Mb) that could seriously impact the overall biological fitness of these breeds. Moreover, the results confirmed that ROH parameters are in agreement with the known demography of these populations and highlighted the different selection histories and breeding schemes of these goat populations. In the analysis of ROH islands, we detected harbored genes involved with important traits, such as for milk yield, reproduction, and immune response, and are consistent with the phenotypic traits of the studied goat populations. Finally, the results of this study can be used for implementing conservation programs for these local populations in order to avoid further loss of genetic diversity and to preserve the production and fitness traits. In view of this, the availability of genomic data is a fundamental resource.

6.1.2 Genomic Variability of Cirneco Dell' Etna and the Genetic Distance with Other Dog Breeds.

Published on Italian Journal of Animal Science 2021, 20 (1), 304–314. DOI: 10.1080/1828051X.2021.1873076

Authors

Cortellari, M.; Bionda, A.; Talenti, A.; Ceccobelli, S.; Attard, G.; Lasagna, E.; Crepaldi, P.; Liotta, L.

Abstract

Cirneco dell'Etna is an old Italian breed of scent hunting dogs. Commonly used genomic measures such as heterozygosity, fixation indexes, and runs of homozygosity can help to improve knowledge about its genetic diversity. This study aimed to: (i) investigate Cirneco's genomic background, (ii) quantify its genomic inbreeding, and (iii) detect genomic regions differentiating the Cirneco's two allowed coat colours, self-coloured fawn and tan and white. Canine 230 K SNP BeadChips was used to investigate 24 Cirneco (19 self-coloured fawn, and 5 tan and white) and other 106 dogs from eight phylogenetically and historically related breeds. The genetic distance, ancestry, and relationship among breeds were explored by multidimensional scaling, Reynolds distances, phylogenetic tree, and admixture analysis. The genomic inbreeding (FROH) was calculated for each breed. Averaged Wright's fixation index FST was used to identify the genes that most

differentiated the two groups of Cirneco. All analyses highlighted that Segugio Italiano and Kelb tal Fenek are the closest breeds to Cirneco. Within the breed, tan and white subjects showed a more heterogeneous genetic background and a lower inbreeding in comparison with self-coloured fawn ones, even though more than half of the latter presented a superimposable admixture. The gene that most differentiated these two groups is Microphthalmia-Associated Transcription Factor (MITF), previously associated with white spotting in other breeds. Given the small size of the Cirneco population and its open registry, its management should carefully com- bine morphological and genealogical evaluations with genetic tools to identify the best breeders while maintaining an acceptable genetic pool.

6.1.3 From phenotypical to genomic characterization of the Mannara dog: an Italian shepherd canine resource.

Published on Italian Journal of Animal Science 2021, 1431–1443. doi:10.1080/1828051X.2021.1972852.

Authors

Liotta, L., Bionda, A., **Cortellari, M.**, Negro, A., and Crepaldi, P.

Abstract

Mannara dogs have long been bred in Sicily (Italy) to work alongside shepherds as flock guardians. This study provides a morphologic, genealogic, and genomic characterization of the Mannara dog, useful in light of its recognition process and to improve the breed standard. Morphologic measurements of body, head, and chest were taken on 111 adult Mannara dogs. The whole population pedigree was used to calculate the inbreeding coefficient (F) and five effective population size (Ne) parameters. Twelve Mannara dogs were genotyped using the Canine 230 K SNP BeadChips and compared with Maremma sheepdog, Caucasian shepherd dog, Cane Corso Italiano, and Neapolitan mastiff for population structure, heterozygosity, and runs of homozygosity. The morphometric evaluation showed that Mannara dogs generally accords with the provisional standard and can be classified as a large/giant, meso-dolicomorphic, and meso- cephalic breed. The population

consists of 375 individuals, one third of which are founders and the remaining belong to 58 litters; presenting low inbreeding (F 1⁄4 0.7%) and balanced sires and dams. The Ne estimates range widely: two (NeN1⁄4159 and NeFi1⁄450) exceed the FCI limit for breed recognition and one (NeCi1⁄425) did not. Genetically, all the included populations are well distinct, with the Maremma sheepdog being the nearest to the Mannara dog. Five Mannara have a single ancestral component, while the others show higher admixed proportions. The genomic inbreeding and heterozygosity confirm the good management of the breed. Our analyses suggest that the Mannara breed should continue the recognition process, pivotal to preserving an invaluable canine resource for the Sicilian agriculture.

6.1.4 Genetic trend of the junctional epidermolysis bullosa in the German shorthaired pointer in Italy

Published on Veterinary Record Open 2021,8:e15. DOI:10.1002/vro2.15

Authors

Frattini, S., Polli, M., **Cortellari, M.**, Negro, A., Bionda, A., Riva, J., Rizzi, R.; Marelli, S.; Crepaldi, P.

Abstract

Background: Epidermolysis bullosa (EB) is a hereditary heterogeneous group of mechanobullous disorders caused by mutations in several structural skin proteins observed in both humans and animals. In this work, we report the incidence and the genetic trend of the junctional epidermolysis bullosa (JEB), a major type of EB, in the Italian German Shorthaired Pointer (GSPs) population in a 10 years span.

Methods: In this study, we monitored the genetic trend of JEB in the Italian population of the GSPs from 2009 to 2018 in 750 animals. The studied mutation was the insertion (4818+207 ins 6.5 kb) of repetitive satellite DNA within intron 35 of the LAMA□ gene. Results: Allele frequencies showed a reduction of the mutated (C) allele during the years, with the only exception of 2017, when 13 dogs were diagnosed as carrier for the genetic pathology. A regression logistic analysis was performed, including sex, coat color and

their interaction. Our results showed that there was a statistically significant association with coat color.

Conclusions: The simplicity and the low cost of the analysis for the detection of this pathology suggests that a deeper identification of carrier dogs will allow better breeding strategies and management, leading to a rapid JEB eradication.

6.1.5 Echocardiographic evaluation of the mitral valve in Cavalier King Charles Spaniels

Published on Animals 2020 ,10, 1454. DOI: 10.3390/ani10091454

Authors

Bagardi, M.; Bionda, A.; Locatelli, M.; **Cortellari**, **M**.; Frattini, S.; Negro, .A; Crepaldi, P.; Brambilla P. G.

Abstract

This prospective cross-sectional study aimed to: (1) characterize echocardiographic features of mitral valve in MMVD affected Cavalier King Charles Spaniels (CKCS), focusing on dogs classified as American College of Veterinary Internal Medicine (ACVIM) class B1; (2) compare echocardiographic data in ACVIM B1 dogs divided on the basis of age at time of MMVD diagnosis, in order to understand if different aged subjects had different echocardiographic patterns. Length (AMVL), width (AMVW) and area (AMVA) of the anterior mitral valve leaflet, mitral valve prolapse, diameters of the mitral valve annulus in diastole (MVAd) and systole (MVAs) of 90 CKCS in different ACVIM classes, 64 of which in class B1, were measured. Valvular measurements were indexed to body weight using Wesselowski's scaling exponents. The presence of heart murmur did not discriminate between A and B1 classes ($p = 0.128$). Heart enlargement was more frequent in males ($r2 = 0.07$, $p = 0.013$). Within class B1, older subjects

showed significantly higher values of AMVA, AMVW, MVAd, MVAs and lower sphericity index (SI). Since many CKCS with MMVD have no murmur and their mitral valve has peculiarities, a specifically designed echocardiographic screening should be realized. In addition, different aged B1 dogs have different echocardiographic patterns that may imply different genetic and prognostic profiles.

6.1.6 A genomic study of Myxomatous Mitral Valve Disease in Cavalier King", Animals

Published on Animals 2020 ,10, 1895. DOI: 1 0.3390/ani10101895

Authors

Bionda, A.; **Cortellari**, **M**.; Bagardi, M.; Frattini, S.; Negro, A.; Locatelli , C.; Brambilla, P.G.; Crepaldi, P

Abstract

Cavalier King Charles spaniels (CKCSs) show the earliest onset and the highest incidence of myxomatous mitral valve disease (MMVD). Previous studies have suggested a polygenic inheritance of the disease in this breed and revealed an association with regions on canine chromosomes 13 and 14. Following clinical and echocardiographic examinations, 33 not-directly-related CKCSs were selected and classified as cases (n = 16) if MMVD was present before 5 years of age or as controls (n = 17) if no or very mild MMVD was present after 5 years of age. DNA was extracted from whole blood and genotyped with a Canine 230K SNP BeadChip instrument. Cases and controls were compared with three complementary genomic analyses (Wright's fixation index—FST, cross-population extended haplotype homozygosity—XP-EHH, and runs of homozygosity—ROH) to identify differences in terms of heterozygosity and regions of homozygosity. The top 1% single-nucleotide polymorphisms (SNPs) were selected and mapped, and

the genes were thoroughly investigated. Ten consensus genes were found localized on chromosomes 3-11-14-19, partially confirming previous studies. The HEPACAM2, CDK6, and FAH genes, related to the transforming growth factor β (TGF-β) pathway and heart development, also emerged in the ROH analysis. In conclusion, this work expands the knowledge of the genetic basis of MMVD by identifying genes involved in the early onset of MMVD in CKCSs.

6.1.7 Genotypic and allelic frequencies of MDR1 gene in dogs in Italy

Published on Veterinary Record Open 2020; 7:e000375. DOI:10.1136/ vetreco-2019-000375

Authors

Marelli, S.; Polli, M.; Frattini, S.; **Cortellari, M**.; Rizzi, M.; Crepaldi, P.

Abstract

Background A mutation in the canine multidrug resistance MDR1 gene (also referred as ABCB1), encoding for the multidrug resistance (MDR) P-glycoprotein (P-gp) transponder, causes a pathological condition known as 'ivermectin toxicosis'. The causative mutation, known since 2001, has been described to affects sheep herding breeds related to collie lineage. The present study is

a retrospective investigation of the presence of MDR1 mutated allele in Italian dog populations in a 5 years'

time lapse. The aim of the research is to offer a deep knowledge in MDR1 allelic and genotypic frequencies in canine breeds and populations raised in Italy.

Methods Genotype data for the 4-bp deletion (c296_299del4) in MDR1 gene from 811 dogs belonging to 32 breeds/populations were collected.

Results The mutated allele has been found in 9 out of 31 breeds: Rough Collie, Smooth Collie, Border Collie, Bearded Collie, Shetland Sheepdog,

Australian Shepherd, White Swiss Shepherd, Old English Sheepdog, Whippet and also in crossbreed. The breeds with the highest allelic mutation frequency are Smooth and Rough Collies with 75 per cent and 66 per cent of mutant MDR1 allele, respectively. Conclusions The results support the usefulness of this genetic analysis to optimize medical care in dogs at risk of multidrug resistance and to create an objective basis in breeding program definition and in the risk evaluation in different breeds.

6.1.8 The SNP-based profiling of Montecristo feral goat populations reveals a history of isolation, bottlenecks and the effects of management choices.

Published on Genes 2022; 13(2), 213. DOI: 10.3390/genes13020213

Authors

Somenzi, E.; Senczuk, G.; Ciampolini, R.; **Cortellari**, M.; Vajana, E.; Tosser-Klopp, G.; Pilla, F.; Ajmone-Marsan, P.; Crepaldi P ;. Colli L.

Abstract

The Montecristo wild goat is an endangered feral population that has been on the homony- mous island in the Tuscan Archipelago since ancient times. The origins of Montecristo goats are still debated, with authors dating their introduction either back to Neolithic times or between the 6th and 13th century of the Common Era. To investigate the evolutionary history and relationships of this population we assembled a 50K SNP dataset including 55 Mediterranean breeds and two nuclei of Montecristo goats sampled on the island and from an ex situ conservation project. Diversity levels, gene flow, population structure, and genetic relationships were assessed through multiple approaches. The insular population scored the lowest values of both observed and expected heterozygosity, high- lighting reduced genetic variation, while the ex situ nucleus highlighted a less severe reduction.

Multivariate statistics, network, and population structure analyses clearly separated the insular nucleus from all other breeds, including the population of Montecristo goats from the mainland. Moreover, admixture and gene flow analyses pinpointed possible genetic inputs received by the two Montecristo goat nuclei from different sources, while Runs of Homozygosity (ROHs) indicated an ancient bottleneck/founder effect in the insular population and recent extensive inbreeding in the ex situ one. Overall, our results suggest that Montecristo goats experienced several demographic fluctuations combined with admixture events over time and highlighted a noticeable differentiation between the two *nuclei.*

CHAPTER 7 Phd Summary

7.1 Attended courses

- All mandatory “Transferable skills” courses organized for PhD students of the XXXIV cycle
- PhD course “Genomics for ecological and evolutionary from Dna sequencing to data analysis”
- PhD course “Bioinformatics and functional genomics”
- PhD course “Statistics for veterinary and animal science-1”
- PhD course “Statistics for veterinary and animal science-2”
- “Corso di formazione e aggiornamento esperti per la valutazione delle razze caprine del libro genealogico e del registro anagrafico”, Brescia, May 6-7 , 2019
- “Settimana formativa sul calcolo degli indici genetici e genomici: dal single trait genetico al single step (ssGBLUP)”, Milano, November 5-8, 2019
- Post-graduate course on “ Characterization, management and exploitation of genomic diversity in animals “, 9-13 December 2019, Wageningen University & Research
- On line Course on “GENOMIC APPROACHES with RANDOM REGRESSION (RR) MODELS”,14-16 September, 2020
- “Corso di Analisi statistica di base per le scienze zootecniche”, ASPA, online, 8 June-15 July,2021
- Corso “Principles of data science applied to livestock”, ASPA ,Agripolis campus, Padova 13-17 September, 2021

7.2 Attended Congress

- 23rd Congress of the Association for Animal Science and Production - ASPA, June 11th to 14th 2019, Sorrento (NA), Italy.
- European Federation of Animal Science – EAAP, Virtual Meeting, 1st - 4th December 2020
- 24rd Congress of the Association for Animal Science and Production - ASPA, September 21st to 24th 2021, Padova (PD), Italy.

7.3 Teaching activities

- Lab activities of the course of "Miglioramento genetico e biotecnologie applicate alla zootecnia" under the supervision of the Professor Paola Crepaldi. During this course I also held the seminar "Utilizzo delle ROH per valutare le zone del genoma sotto selezione"
- Lab activities of the course of "Metodologie sperimentali per l'Agricoltura" under the supervision of the Professor Paola Crepaldi.
- Online seminar titled "Runs of Homizygosity" held during the course of "Miglioramento genetico e biotecnologie applicate alla zootecnia" of the Professor Paola Crepaldi

7.4 Congress proceedings

7.4.1 Oral presentation

- P. Crepaldi, S. Frattini, A. Negro, **M. Cortellari**, A. Talenti. Metigree: a genomic tool to enhance the well-being of Italian mongrel dogs (ASPA 23rd Congress, Sorrento, June 11th-14th, 2019)
- S. Frattini, **M. Cortellari**, A. Talenti, A. Negro, C. Biagini, M. Polli, P. Crepaldi. An haplotype view of Cystinuria in dog (ISAG 37th Congress ,Leida, July 7-12, 2019)
- Bionda, **M. Cortellari**, A. Negro, S. Frattini, A. Talenti, L- Liotta and P. Crepaldi. Genomic evaluation of the Italian shepherd dogs (ASPA 24rd Congress, Padova, September 21-24, 2021)
- M. Cortellari, A. Negro, A. Bionda, S. Frattini, S. Grande and P. Crepaldi. Inbreeding depression in small ruminants: from pedigree to genomic estimation (ASPA 24rd Congress, Padova, September 21-24, 2021)

7.4.2 Posters

- **M. Cortellari**, F. Galluzzo, A. Negro, A. Talenti, S. Frattini, S. Mastrangelo, G. Minozzi, F. Pilla, G. Pagnacco, P. Crepaldi. Cabannina: genomic characterization of a local Italian Breed (ASPA 23rd Congress, Sorrento, June 11th-14th, 2019)
- **M. Cortellari**, A. Negro, A. Talenti, S. Frattini, S. Mastrangelo, F. Pilla, P. Ajmone-Marsan, P. Crepaldi. Genomic meat traceability: from breeders to consumers (ASPA 23rd Congress, Sorrento, June 11th-14th, 2019).
- S. Frattini, **M. Cortellari**, A. Talenti, A.Negro, M. Caprioglio, and P. Crepaldi. Investigation of genomic variation of coat color genes in Italian goat breeds (ISAG 37th Congress ,Leida, July 7-12, 2019)
- **M. Cortellari**, A. Negro, A. Bionda, A. Cesarani2, N. Macciotta, S. Grande, S. Biffani and P. Crepaldi. Genomic inbreeding of Nicastrese: conservation

of an autochthonous Italian goat breed (EAAP online meeting 1-4 December 2020)

- **M. Cortellari**, A Negro, A. Bionda, A. Cesarani, N. Macciotta, S. Grande, S. Biffani and P. Crepaldi. Pedigree and genomic inbreeding comparison in the Italian Delle Langhe dairy sheep breed (EAAP online meeting 1-4 December 2020)
- A. Cesarani, S. Biffani, A. Negro, **M. Cortellari**, S. Grande, P. Crepaldi and N. Macciotta. Application of single-step GBLUP in Italian Comisana sheep (EAAP online meeting 1-4 December 2020)
- A. Negro, **M. Cortellari**, A. Bionda, S. Biffani, S. Grande and P. Crepaldi. Exploring pedigrees: an overall picture of biodiversity in Italian small ruminants (ASPA 24rd Congress, Padova, September 21-24, 2021)
- **M. Cortellari**, M. Pampuri, R. Pasquariello and P. Crepaldi. The effect of an extender inducing sperm capacitation on fertility of Italian dairy cattle: a field experience (ASPA 24rd Congress, Padova, September 21-24, 2021)
- **M. Cortellari**, A. Bionda, S. Frattini, A. Talenti, A. Negro, L. Liotta and P. Crepaldi. Runs of Homozygosity in 21 Italian dog populations (ASPA 24rd Congress, Padova, September 21-24, 2021)
- A. Bionda, **M. Cortellari**, A. Negro, S. Frattini, A. Talenti, S. Sechi, M. Zedda, L. Liotta, R. Cocco and P. Crepaldi. Fonni's dog: genetic variability and relationship with other breeds (ASPA 24rd Congress, Padova, September 21-24, 2021)
- S. Biffani, A. Cesarani, A. Negro, **M. Cortellari**, S. Grande, P. Crepaldi and N. Macciotta. Application of single-step GBLUP in Italian Comisana sheep (ASPA 24rd Congress, Padova, September 21-24, 2021)
- A. Negro, **M. Cortellari**, A. Bionda, S. Frattini, S. Biffani, S. Grande and P. Crepaldi. Genomic tools to support breed assignment in small ruminants (ASPA 24rd Congress, Padova, September 21-24, 2021)

7.5 Papers

7.5.1 Published

- Somenzi, E.; Senczuk, G.; Ciampolini, R.; **Cortellari**, M.; Vajana, E.; Tosser-Klopp, G.; Pilla, F.; Ajmone-Marsan, P.; Crepaldi P ;. Colli L. The SNP-based profiling of Montecristo feral goat populations reveals a history of isolation, bottlenecks and the effects of management choices. Genes 2022; 13(2), 213. DOI: 10.3390/genes13020213
- **Cortellari, M**.; Bionda, A.; Talenti, A.; Ceccobelli, S.; Attard, G.; Lasagna, E.; Crepaldi, P.; Liotta, L.. Genomic Variability of Cirneco Dell'Etna and the Genetic Distance with Other Dog Breeds. Italian Journal of Animal Science 2021, 20 (1), 304–314. doi: 10.1080/1828051X.2021.1873076
- Mastrangelo, S.; Di Gerlando, R.; Sardina, M. T.; Sutera, A. M.; Moscarelli, A.; Tolone, M.; **Cortellari, M**.; Marletta, D.; Crepaldi, P.; Portolano, B.. Genome-Wide Patterns of Homozygosity Reveal the Conservation Status in Five Italian Goat Populations. Animals 2021, 11 (6), 1510. doi: 10.3390/ani11061510
- **Cortellari, M**.; Barbato, M.; Talenti, A.; Bionda, A.; Carta, A.; Ciampolini, R.; Ciani, E.; Crisà, A.; Frattini, S.; Lasagna, E.; Marletta, D.; Mastrangelo, S.; Negro, A.; Randi, E.; Sarti, F. M.; Sartore, S.; Soglia, D.; Liotta, L.; Stella, A.; Ajmone-Marsan, P.; Pilla, F.; Colli, L.; Crepaldi, P.. The Climatic and Genetic Heritage of Italian Goat Breeds with Genomic SNP Data. Scientific Reports 2021, 11 (1), 10986. doi: 10.1038/s41598-021-89900-2
- Liotta, L., Bionda, A., **Cortellari, M.**, Negro, A., and Crepaldi, P.. From phenotypical to genomic characterization of the Mannara dog: an Italian

shepherd canine resource. Italian Journal of Animal Science 2021, 1431–1443. doi:10.1080/1828051X.2021.1972852.

- Frattini, S., Polli, M., **Cortellari, M**., Negro, A., Bionda, A., Riva, J., Rizzi, R.; Marelli, S.; Crepaldi, P.. Genetic trend of the junctional epidermolysis bullosa in the German shorthaired pointer in Italy. Veterinary Record Open 2021, 8. doi:10.1002/vro2.15.
- **Cortellari**, **M**.; Bionda, A.; Negro, A.; Frattini, S.; Mastrangelo, S.; Somenzi, E.; Lasagna, E.; Sarti, F.M.; Ciani, E.; Ciampolini, R.; Marletta, D.; Liotta, L.; Ajmone Marsan, P.; Pilla, F.; Colli, L.; Talenti A.; Crepaldi P.. Runs of homozygosity in the Italian goat breeds: impact of management practices in low-input systems. Genetic Selection Evolution 2021,53:92., doi:10.1186/s12711-021-00685-4
- Marelli, S.; Polli, M.; Frattini, S.; **Cortellari, M**.; Rizzi, M.; Crepaldi, P.. Genotypic and allelic frequencies of MDR1 gene in dogs in Italy. Veterinary Record Open 2020; 7:e000375. doi:10.1136/ vetreco-2019-000375
- Bagardi, M.; Bionda, A.; Locatelli, M.; **Cortellari**, **M**.; Frattini, S.; Negro, .A; Crepaldi, P.; Brambilla P. G.. Echocardiographic evaluation of the mitral valve in Cavalier King Charles Spaniels. Animals 2020 ,10, 1454. doi: 10.3390/ani10091454
- Bionda, A.; **Cortellari**, **M**.; Bagardi, M.; Frattini, S.; Negro, A.; Locatelli , C.; Brambilla, P.G.; Crepaldi, P.. A genomic study of Myxomatous Mitral Valve Disease in Cavalier King. Animals 2020 ,10, 1895. doi: 1 0.3390/ani10101895