UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Sede di Piacenza

Dottorato di ricerca per il Sistema Agro-alimentare

Ph.D. in Agro-Food System

Cycle XXXIV

S.S.D. AGR/17

Investigation on small ruminants biodiversity and adaptation

Candidate:

Elisa Somenzi

Matriculation n: 4816135

Academic Year 2020/2021

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Sede di Piacenza

Dottorato di ricerca per il Sistema Agro-alimentare

Ph.D. in Agro-Food System

Cycle XXXIV

S.S.D. AGR/17

Investigation on small ruminants biodiversity and adaptation

Coordinator:

Ch.mo Prof. Paolo Ajmone Marsan

Tutor: Prof. Paolo Ajmone Marsan

Candidate:

Elisa Somenzi

Matriculation n: 4816135

Academic Year 2020/2021

Contents

List of publication

This thesis is based on the work contained in the following papers:

- **Chapter 1:** Somenzi, E., Ajmone-Marsan, P., & Barbato, M. (2020). Identification of ancestry informative marker (AIM) panels to assess hybridisation between feral and domestic sheep. Animals, 10(4). https://doi.org/10.3390/ani10040582

- **Chapter 2**: Somenzi, E., Senczuk, G., Ciampolini, R., Cortellari, M., Vajana, E., Tosser-klopp, G., … Colli, L. (2022). The SNP-Based Profiling of Montecristo Feral Goat Populations Reveals a History of Isolation, Bottlenecks, and the Effects of Management. Animals, 13(213), 1–14. https://doi.org/10.3390/genes13020213

- **Chapter 3**: Cortellari, M., Bionda, A., Negro, A., Frattini, S., Mastrangelo, S., Somenzi, E., … Crepaldi, P. (2021). Runs of homozygosity in the Italian goat breeds: impact of management practices in low-input systems. *Genetics Selection Evolution*, *53*(1). https://doi.org/10.1186/s12711-021-00685-4

- **Chapter 4**: Passamonti, M. M., Somenzi, E., Barbato, M., Chillemi, G., Colli, L., Joost, S., … Ajmone-Marsan, P. (2021). The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock. *Animals*, *11*(10), 2833. https://doi.org/10.3390/ani11102833

- **Chapter 5**: Barbato M, Vajana E, Selmoni O, Somenzi E, Del Corvo M, Tixier-Boichard, Joost S, Lazzari B, Stella A, Ajmone Marsan P. (2022). Deciphering climate-mediated adaptation in European sheep. (Manuscript)

Introduction

Biodiversity

The international Convention on Biological Diversity has defined biodiversity as "the variability among living organisms from all sources including, *inter alia*, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species and of ecosystems" (United Nations, 1992). All this variability is expressed, and can be described, in a multiplicity of different ways. According to Gaston (Gaston & Spicer, 2013) biological diversity can be divided into three basic building blocks: i) ecological diversity, ii) organismal diversity, and iii) genetic diversity. Ecological diversity concerns variations among biomes, that is, differences occurring between niches and habitats. Organismal diversity is related to the biological variability occurring on a taxonomic level. Lastly, genetic diversity encompasses the variation in genetic make-up at individual and population levels and characterizing the differences among species and breeds. Despite being described as separate blocks, these three elements are strictly interconnected and should be considered together for a better characterization of biodiversity.

Since the international Convention on Biological Diversity took place in 1992, awareness of the need for biodiversity preservation has considerably increased. In October 2010 the tenth meeting of the Conference of the Parties, involving more than 150 government leaders, adopted a revised and updated Strategic Plan for Biodiversity, including the Aichi Biodiversity Targets, which are a series of 20 targets for biodiversity preservation to be achieved in the 2011-2020 period (Convention on Biological Diversity, n.d.). The Aichi Biodiversity Target 13 stated that: "*By 2020, the genetic diversity of cultivated plants and farmed and domesticated animals and of wild relatives, including other socio-economically as well as culturally valuable species, is maintained, and strategies have been developed and implemented for minimizing genetic erosion and safeguarding their genetic diversity"* (Convention on Biological Diversity, n.d.). Nevertheless, the Global Assessment Report on Ecosystem and Biodiversity Services (Brondízio, Settele, Díaz, & Ngo, 2019) identified that the current rate of global species extinction is tens to hundreds of times higher compared to the average over the last 10 million years. This negative trend not only relates to wildlife, but also domesticated breeds of mammals of which by 2016 560 were considered extinct, with another approximately 1,000 considered to be at risk (Brondízio et al., 2019).

Livestock Biodiversity

Livestock species constitute a key component of agricultural biodiversity. Livestock are all those domesticated or semi-domesticated mammalian species that are used for the production of food and agriculture (Gregoire Leroy et al., 2018).

Since the domestication of livestock occurred, around 10,000 YBP, farmers have managed their animals in a wide variety of environments, resulting in thousands of different populations adapted to local conditions. Livestock keepers influenced for centuries the characteristic of their animals through selective breeding, but it was during the 18th Century, in the United Kingdom, that farmers started to select and cross animals based on their phenotype and pedigree, leading to the development of the modern concept of breed (Wykes, 2004). From the second half of the twentieth century, artificial insemination became a widely used practice that allowed to obtain a large progeny from a single highly performant sire. Consequently, the use of a limited number of popular sires rapidly increased the productivity of several breeds but also drastically reduced their effective population size. In the last few decades, with the introduction of quantitative genetics methods in breeding, farm animal selection pressure was further intensified, increasing the average inbreeding level of several breeds (Ajmone-Marsan, 2010).

However, high selection pressure is not the only threat to farm animal genetic diversity. In industrialized and developing countries the demand for animal products has rapidly increased. As a consequence, farmers have been replacing traditional and locally adapted breeds with a reduced number of “improved”, highly selected breeds. The livestock sector in developing countries is facing challenges which are causing a progressive abandonment of rural systems along with traditionally farmed local breeds. The factors that are reducing the income from livestock include, among others, degradation and unavailability of natural resources, loss of workforce caused by outmigration, social instability, wars and sanitary problems (FAO, 2015). In this context, climate change is an additional burden affecting livestock production systems. The increase in temperature and the higher frequency of extreme climatic events impact local ecosystems, jeopardizing feed availability and increasing the spread of diseases.

These negative factors have resulted in an increasing number of livestock breeds classified by FAO as being at risk of extinction, which is defined as the total number of breeding males less than or equal to 20. The percentage of livestock breeds classified *At risk* raised from 15 percent to 17 percent between 2005 and 2014, with a further 58 percent of breeds classified with *Unknown risk status* due to the lack of recent population data (FAO, 2015). This uncontrolled loss of genetic diversity will negatively impact on ability of livestock systems to respond to future challenges such as changes in environmental conditions, market demands, husbandry practices and emerging diseases. Many

different strategies have been developed and tried with the aim of minimizing genetic erosion. In this context, phenotypic and genetic characterization of breeds are the key first steps. Genomics is currently contributing to the latter, allowing fine-grained genetic knowledge to develop breeding strategies for conservation programs.

Small ruminant biodiversity

Small ruminants possess the highest genetic variability among livestock species (Kijas et al., 2012)(Colli et al., 2015). Sheep and goats are widely distributed all over the world, with a head count of 1.2 billion and 1 billion respectively, mostly localized in developing countries (FAO, 2015). The Domestic Animal Diversity Information System (DAD-IS) ("Domestic Animal Diversity Information System (DAD-IS)," n.d.) reported that the current number of local and transboundary goat breeds is 662 with more than twice this number for sheep, with 1382 breeds. A total of 414 goat breeds have *Unknown* risk status and 91 are classified *At Risk*, while788 sheep breeds have *Unknown* risk status and 191 are classified *At Risk* (FAO, 2015). These classifications suggest that several small ruminants populations are likely to go extinct, the majority of which are local breeds with unique adaptation features, whose loss would represent an irreversible loss of genetic diversity.

Small ruminant domestication and history

Domestication has been defined as "distinctive coevolutionary and mutualistic relationship between domesticator and domesticate" (Zeder, 2015). This process, involving biological and social aspects, acted as a strong evolutionary force that has shaped livestock genomes. Indeed, all domesticated species differ in morphological, physiological and behavioral traits compared to wild relatives, and have developed specific characteristics compatible with human necessities. Domestication gradually resulted in morphological, physiological and genetic differences between domestic and wild populations. In the past few decades, the remarkable technological advances in genomic technologies have produced information that sheds light on the domestication processes and subsequent dispersal of the major livestock species throughout the world (Pitt, Sevane, et al., 2019)(Zeder & Hesse, 2000)(Chessa et al., 2009).

Sheep (Ovis aries)

Among livestock species sheep was one of the first to be domesticated, approximately between 11,000 and 9,000 YBP, coinciding with the Neolithic agricultural revolution that occurred in ancient Mesopotamia. The origin and taxonomy of the genus *Ovis* is still an unsolved problem (Rezaei et al.,

2010). several wild species have been proposed as potential progenitors of domestic sheep, but none of these have been conclusively confirmed as being the living ancestor of sheep (Barbato, Hailer, Orozco-Terwengel, et al., 2017). It is known that several domestication events, as inferred by multiple mitochondrial lineages (Chessa et al., 2009) in modern sheep breeds, gave rise to domestic sheep. Sheep that first colonized Europe had coarse wool and were reared mainly for meat, but during the fifth millennium B.P. in Southwest Asia and the fourth millennium B.P. in Europe, specialization for "secondary" products such as wool became apparent. Sheep selected for fine wool appear to have replaced more primitive domestic populations during a second migration wave (Chessa et al., 2009). The primitive breeds survived the second migrations of improved breeds from Southwest Asia by returning to a feral or semi-feral state on islands without predators or by occupying inaccessible areas, as occurred for the Soay sheep on St Kilda Archipelago and the European Mouflon on Sardinian Island. (Barbato, Hailer, Orozco-Terwengel, et al., 2017). The majority of current breeds are descendant of the later migratory episodes of sheep improved for secondary production traits (Chessa et al., 2009; Dýrmundsson & Ninikowski, 2010).

Hence, domesticated sheep were first translocated from the Fertile Crescent to northern Europe following human migrations ~6,000 YBP, when agro-pastoralism became the main system of food production (Zeder, 2008). The first evidence of breeding domestic sheep in Europe was found in Greece, at the time of the ancient Greek civilization, and in southern France around the Neolithic Period (Zeder, 2008). Importantly, the development of woolen clothes might have helped humans in colonizing colder areas. The Roman Empire, , spread domestic sheep along trade routes throughout Europe and at the time of the Middle Ages, sheep became a primary economic source for many countries in Europe. An important contribution to the development of modern sheep breeds came from the British agriculturalist Robert Bakewell in 18th century. Bakewell promoted the scientifically driven approach of selective breeding for trait selection, a substantial improvement with respect to the subjective breeding that had previously been used (Wykes, 2004). Since then, sheep breeds have been selected for different phenotypic characteristics and productive traits, leading to them being reproductively isolated and a fragmentation of the founder populations. The development of modern selection methods quickly improved the productivity breeds, albeit at the considerable cost of loss in genetic diversity (Taberlet et al., 2008).

Goat (Capra Hircus)

Genus *Capra* is thought to have emerged in the Pliocene period (Ropiquet & Hassanin, 2006) and largely expanded during late Pleistocene (250 000 YBP) (Nomura et al., 2012). During the late Eurasian Saalian (130 000 – 160 000 YPB) and the following Würm glacial periods (12 000 – 71 000 YBP) wild goat population experienced a drastic population bottleneck (Colli et al., 2015) but some lineages

survived in Near East refuge areas (Pala et al., 2012) where they were domesticated. Domestication of the wild ancestor of goat, the bezoar (*Capra Aegagrus*), most likely occurred through the prey pathway, in which wild goats were initially hunted and only later directly managed by humans for ensuring their availability (Zeder, 2008). Based on zooarchaeological findings and genomic studies, goat domestication seems to have occurred between 10 000 and 11 000 YBP in an area spanning from Southeastern Anatolia to the Zagros Mountains in Central Iran (Colli et al., 2015; Zeder & Hesse, 2000). Domesticated goats subsequently spread in Europe following human migration routes. According to Zeder and colleagues (Zeder, 200) the goats arrived on the island of Cyprus around 9,000 -10,500 YBP and southern-east Italy around 8,000 YBP then in the following 700 years, goats appeared in all the coastal areas of southern France and the Iberian Peninsula colonization (Zeder, 2008). Goats dispersion in North Africa occurred around 6000-7000 YBP while southern Africa was goats were only found 2000 YBP (Pereira et al., 2009). In the same period goats also spread all over Asia, reaching China around 4000 YBP (Cai et al., 2020). Following human settlement all over the world, goats adapted over the course of centauries to a large variety of environments, developing a wide range of different morphological and physiological characteristics. Environmental and human-driven selection resulted in the development of around 600 local goat breeds, mostly distributed in Asia (56% of the total) and in Africa (30% of the total) (FAO, 2015). Goats are mostly farmed for milk and meat production, and are a key source of income especially in rural and marginal areas as well as in developing countries (Nicoloso et al., 2015).

Adaptation

Animal husbandry has been practiced for many centuries in the diverse and sometimes extreme environments. A combination of natural and anthropogenic selection has resulted in the development of breeds that are highly productive and other that are adapted to survive under unfavourable conditions (FAO, 2015). Natural selection has been defined as the preferential retention (or elimination) of certain genetic variants in a population as a response to an environmental selective pressure. Three main modes of selection have been described: positive, balancing, and negative (Vitti, Grossman, & Sabeti, 2013). Positive selection occurs when a fitness-enhancing variant rapidly increases its frequency in a population. Conversely, negative selection removes deleterious mutations (Zeng et al., 2018). Lastly, balancing selection promotes the increase in frequency of the heterozygotes (Charlesworth, 2006).

In the current changing climatic scenario, livestock adaptation to adverse conditions, such as feed scarcity, drought, diseases and heat stress, is a crucial aspect. The identification of adaptation-related genes and genetic variants will facilitate future selection in livestock breeding. Recently molecular approaches have been used to identify genes under selection in response to environmental stressors. These include: i) genome wide association studies (GWAS), which use phenotypes related to adaptation, ii) Landscape Genomics, based on the use of environmental variables as proxies for phenotypes and iii) the analysis of patterns of genomic diversity within and between populations to identify selection signatures. Molecular approaches have detected genomic signatures of adaptation resulting from centuries of natural and artificial selection pressure on genomes.

Since domestication, small ruminants have developed morphological and physiological characters that increase their ability to adapt. Nearly 25% of sheep in the world have a fat tail which acts as storage, enabling animals to survive harsh semi-arid desert conditions when food is in short supply (Ahbara et al., 2019; S. Mastrangelo et al., 2019; Moradi, Nejati-Javaremi, Moradi-Shahrbabak, Dodds, & McEwan, 2012; Passamonti et al., 2021; Z. Yuan et al., 2017). Several recent investigations of the adaptation of sheep to harsh environments have identified adaptation-related genes (Álvarez et al., 2020b; Eydivandi, Roudbar, Ardestani, Momen, & Sahana, 2021; Zhang et al., 2021). Signatures of adaptation to desert and arid environment were found in different native sheep breeds (Ji Yang et al., 2016) as well as genes associated with heat stress adaptation (DNAJC28, GNRH1 and MREG), high altitude adaptation and hypoxia (MITF, FGF5, MTOR, TRHDE and TUBB3) (Ji Yang et al., 2016). Modification in metabolism (Lv et al., 2014a) and body size (Kominakis et al., 2017) have been identified as adaptation-related traits, enhancing fitness in harsh environments. In unfavourable climatic condition characterized by higher frequency of disease (e.g. high humidity level) also adaptation via disease-resistance can enhance population fitness. Thus, signature of selection related to immune system functions (HERC2, CYFIP1, ZBP1, PRDX1, MAST2 and LURAP) have been found in

sheep (Abied, Xu, et al., 2020; Mwacharo et al., 2017) as well as several types of parasite and pathogen resistance, including resistance to gastrointestinal nematode infection, found on OAR2 (Al Kalaldeh, Gibson, Lee, Gondro, & Van Der Werf, 2019; Atlija, Arranz, Martinez-Valladares, & Gutiérrez-Gil, 2016; Estrada-Reyes et al., 2019; Keane et al., 2006) and pneumonia (PADI2) (Y. H. Cao et al., 2021). In their work, Bertolini and colleagues (Francesca Bertolini, Servin, et al., 2018) analysed more than three thousands goats from all over the world and identified selection signatures in genes associated with several important pathways. Heat stress can affect the organism response to insulin stimulus and oxidative stress: genes related to glucose pathway (IGF2) and protection against oxidative stress (GPR37L1, INS) were found under selection. Adaptation to different latitude was suggested by the identification of genes under selection related to circadian clock rhythm (SOX14, NOCT, RAI1) (Francesca Bertolini, Servin, et al., 2018). Genes associated with inflammatory response and resistance to diseases (BTLA, NOS2, TGFB2 and TLR4) have also been found in goats (Estrada-Reyes et al., 2019)(Cortellari, Barbato, Talenti, Bionda, Randi, et al., 2021). Genes under selection associated with development and function of the nervous and endocrine systems in goat breeds adapted to hot environment were also found (E. S. Kim et al., 2016).

Investigation on biodiversity and adaptation

Genomic tools

The ever-reducing cost of DNA sequencing and genotyping has allowed larger and larger amounts of genomic data to be produced. Whole genome sequencing (WGS), made possible because of reducing sequencing costs, has identified several different types of genetic variants including single nucleotide polymorphisms (SNPs) insertions and deletions (indels), structural variation (SV), and copy-number variation (CNV) that have been used as genetic markers. Until recently, two classes of genetic markers were mainly used in genetic studies on livestock species: mitochondrial DNA (mtDNA) sequence variants and microsatellite loci or Simple Sequence Repeats (Hiendleder, Kaupe, Wassmuth, & Janke, 2002)(Lenstra, 2005). However, whole genome sequencing has also established reference genomes for major livestock species (Davenport et al., 2021)(Bickhart et al., 2017), which were used as initial step for the development of multi-locus bead arrays of single nucleotide polymorphisms (SNP). A single nucleotide polymorphism is defined as the variation in a single nucleotide occurring at a specific position in the genome. SNP are considered the most common polymorphism in the genome, with an estimated occurrence of one every 1 kb on average. Current technological advances allow the simultaneous genotyping of 10,000 to 1,000,000 SNPs in the genome using DNA array technologies (Nicolazzi et al., 2015). SNP microarrays are composed of a collection of oligonucleotide probes designed to match a specific genomic region and attached to a solid surface, such as glass slides or silicon quartz wafers. Currently two commercial are widely for SNP array platforms are widely used: Illumina's Infinium iSelect Microarray or BeadChip (Illumina Inc, 2016) and Affymetrix's GeneChip or AxiomArray (Affymetrix Inc., 2020). SNP arrays used in livestock species are available with different marker densities: i) low-density, below 20,000 SNPs, ii) medium density, around 50,000 SNPs, iii) medium to high density, ranging from 150,000 to 200,000 SNPs and iv) high, with more than 500,000 SNP. Low density panels of markers have been developed for specific purposes such as estimation of biogeographical ancestry in human populations (Halder, Shriver, Thomas, Fernandez, & Frudakis, 2008; Kosoy et al., 2008; Paschou et al., 2008, 2007), discrimination of breeds (F Bertolini et al., 2018; Dimauro et al., 2015; F. Bertolini, G. Galimberti, 2015; Wilkinson et al., 2011), identification of the origin of animal products origin (Dimauro et al., 2013)(Orrù et al., 2009), and for detection of hybrids (Ettore Randi et al., 2014; Russell et al., 2019; Somenzi, Ajmone-Marsan, & Barbato, 2020).

Methods to analyse biodiversity and adaptation

Along with the production of a huge number of datasets also the number of approaches and software for data analysis has significantly increased. Methods for investigating biodiversity and adaptation on

livestock have been widely reviewed (Groeneveld et al., 2010; G. Leroy, 2014; Passamonti et al., 2021; Saravanan et al., 2020; Vitti et al., 2013; Worku & Tadesse, 2017). The rationale behind the main methods used in this thesis is described below.

Molecular diversity

One of the basic parameters for quantifying genetic diversity of a breed is the Observed heterozygosity (H_o). The value of H_o ranges between 0 and 1 where a low H_o indicates a low genetic variability and may suggest historic population bottleneck events, a small population size, or a deliberate inbreeding such as the line-breeding systems. Conversely, high values indicate high genetic variability and outbreeding. Expected heterozygosity (H_e) is the level of heterozygosity expected under Hardy-Weinberg equilibrium (HWE) which states that in a large panmictic population subjected to no evolutionary force, including mutation, genetic drift, selection and migration, allelic and genotypic frequencies remain constant across generations (M. Nei, 1973)(Masatoshi Nei, 1978).

The difference between H_e and H_o can be used to infer the inbreeding level in a population. A significantly lower observed heterozygosity compared to expected is a signal of inbreeding (Curik, Ferenčaković, & Sölkner, 2014). Another way to evaluate genomic inbreeding in individuals is via the calculation of the F_{ROH}. F_{ROH} is the proportion of the autosomal genome covered by runs of homozygosity (ROH) (McQuillan et al., 2008) which are contiguous homozygous segments of the genome (Gibson, Morton, & Collins, 2006). The analysis of ROHs allows the population history and demography, bottleneck events, founder effect etc to be inferred and signatures of selection to be detect (Ceballos, Joshi, Clark, Ramsay, & Wilson, 2018; Meyermans, Gorssen, Buys, & Janssens, 2020; Peripolli et al., 2017). Long ROH segments have been associated with recent inbreeding, while a large number of short ROHs is related to remote events including population bottlenecks and founder effects (Peripolli et al., 2017). The analysis of ROHs has also been applied to the detection of signatures of selection: a ROH segment, in a specific position of the genome, shared by several individuals in a population can be interpreted as a sign of positive selection (Saravanan et al., 2020).

The Effective population size (*Ne*) is related to the level of variability existing in a population. *Ne* can be described as the number of individuals in a population who contribute offspring to the next generation (Barbato, Orozco-terWengel, Tapio, & Bruford, 2015). It is important to distinguish *Ne* from the census size (N), the latter being the number of individuals in a population, which is usually larger than the effective population size. The difference in the demographic structure between the idealized population and the real one can be evaluated from the ratio between *Ne* and N. Large values of *Ne* are associated with genetically stable populations with a large genetic pool. Instead, a small *Ne* results in unpredictable allele frequencies, loss or fixation of some alleles, and an increased risk of extinction.

In domesticated species *Ne* values tend to be lower than in wild species, due to non-random mating (Kristensen, Hoffmann, Pertoldi, & Stronen, 2015)(Peripolli et al., 2017).

Population structure

The Principal Component Analysis (PCA) is a statistical procedure widely used to represent the genetic variation within and between populations. The procedure applies orthogonal linear transformation to a set of observations described by several correlated variables (e.g. genotype data of one or more populations or individuals). The most relevant relationships underlined in the data are expressed as a set of variables called principal components. Principal components are ordered on the basis of the share of total variance explained. The first principal component explains the largest variance while the succeeding components show in a decreasing order the remaining variance. As in Figure 1 of Chapter 1 values of the first two principal components for each sample are visualized in a Cartesian coordinate system, were individuals more genetically related to each other appear closer in the graph.

Another approach to analyse genetic relationship among individuals or populations is the use of phylogenetic network. Phylogenetic networks are graphs used for representing complex evolutionary history with parallel, divergent and convergent evolution between different breeds (Bandelt & Dress, 1992)(Dopazo, Dress, & Von Haeseler, 1993). This kind of representation, rather than simple branching trees, better fits the non-treelike history of breeds, with overlapping clustering solutions, which are very common patterns in domesticated species. In this thesis the Neighbour-Net algorithm has been used for constructing phylogenetic networks using a Reynolds' genetic distance matrix as input. Reynolds' distances are the most used genetic distances when studying livestock species because, unlike other genetic divergence measures, assumes that genetic differentiation occurs only by genetic drift, without mutations (Reynolds, Weir, & Cockerham, 1983). This approach is more appropriate for short-term evolution such as the divergence between livestock breeds (Barker, 1990). Reynold's distances are computed as:

$$\Theta_w = \sqrt{\frac{\sum_l \ \sum_u . (X_u - Y_u)^2}{2 \ \sum_l (1 - \sum_u X_u \ \ Y_u)}}$$

The formula estimates the coancestry coefficient (Θ) using the information of two different populations (X and Y) with L loci analysed (where X_u and Y_u are the frequency of the u^{th} allele in the two populations). As shown in Figure 4 of Chapter 2, the Neighbour-Net reconstruction based on Reynolds' distances allowed us to identify genetic relationship among breeds and the length of the branches highlighted evidence of prolonged isolation and genetic drift occurring in some breeds.

To investigate patterns of splits and migrations occurring between populations in Chapter 2 (figure 6) we applied the statistical model developed by Pickrell and Pritchard (Pickrell & Pritchard, 2012) and

implemented in the software *Treemix*. The program constructs a maximum-likelihood tree based on the covariances of the allele frequencies and relates populations to a common ancestor. Furthermore, the statistical model allows vectors between branches to be added to represent migration events that are thought to have occurred between populations. The best fitting number of edges (*m*) to include can be inferred from the second-order rate of change in likelihood (Δ*m*) across incremental values of *m*, as proposed by Fitak (Fitak, 2021).

Model-based clustering is a common approach to identify ancestry proportions in a set of individuals. The model-based clustering method implemented by Pritchard *et al.* (Pritchard, Stephens, & Donnelly, 2000) and subsequently improved by Alexander and colleagues (Alexander, Novembre, & Lange, 2009), assumes a number *K* of putative ancestral populations chosen by the investigator and models each sample as a mixture of these *K* clusters. Admixture proportions of each individual are estimated using a Bayesian approach based on a Markov Chain Monte Carlo (MCMC) algorithm, to sample the posterior distribution using Linkage Disequilibrium and Hardy-Weinberg information. This approach has been used in Chapter 1 (Figure 4) and 2 (Figure 5) for investigating ancestry proportions of the samples used in the two studies.

Population history reconstruction

Recent studies have sought to shade light on breed histories and demography. Currently, several approaches are available for testing the likelihood of alternative breed historical and demographical scenarios based on the comparison of genomic data. Simulation-based methods such as Approximate Bayesian Computation (ABC) have been specifically developed for the analysis of complex population genetic histories, including changes in population size as well as divergence, migrations and admixture events. In Chapter 2 of this thesis an Approximate Bayesian Computation-random Forest approach (ABC-RF) (Collin et al., 2021) was used for investigating recent demographic history of the population under study. The implementation of supervised machine learning (SML) approach in the Approximate Bayesian Computation method increases the efficacy of the analysis by using the provided data samples as a training set to simulate new data points. The ability of SML methods to use simulation to extend observed data is especially crucial for population genetics applications, where adequately sized data sets are not always available. The use of ABC-RF approach allowed us to: (i) simulate multiple historical models; (ii) evaluate the accuracy of parameters estimated during the analysis; and (iii) rank models based on the approximate posterior probabilities.

Landscape genomics

In Chapter 5 a landscape genomics approach has been applied for identifying adaptation-related genes (Manel, Schwartz, Luikart, & Taberlet, 2003; Pariset, Joost, Gargani, & Valentini, 2012).

Landscape genomics combines population genomics and environmental data to quantify the effects of the environment on genetic variation through the detection of significant associations between population genetic makeup and habitat characteristics (Pariset et al., 2012). Landscape genomics relies on two main technological advances, firstly, the development of Geographic Information Systems (GIS) (Goodchild, 1992), which facilitated the overlay of different types of environmental variables, secondly, the availability of genotypes from geo-referenced samples.

Aims of the thesis

The aim of this thesis was to study population history, structure, genetic diversity and environmental adaptation of sheep and goats local breeds, for contributing to the process of characterization and conservation of small ruminants genetic resources. In this context, this thesis addresses five specific goals:

1) To develop an algorithm for the selection of ancestry informative markers for the identification of feral x domestic sheep hybrids.
2) To assess molecular diversity, population structure, history and relationship of the feral goat of Montecristo.
3) To investigate the impact of different management practices on genomic inbreeding on Italian goat populations.
4) To review methods and approaches for detecting adaptation in ruminants.
5) To detect genes associated to environmental adaptation in European local sheep breeds.

Chapter 1: Identification of Ancestry Informative Marker (AIM) Panels to Assess Hybridisation between Feral and Domestic Sheep

Elisa Somenzi, Paolo Ajmone-Marsan and Mario Barbato *

Department of Animal Science, Food and Nutrition – DIANA, and Nutrigenomics and Proteomics Research Center – PRONUTRIGEN, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; elisa.somenzi@unicatt.it (E.S.); paolo.ajmone@unicatt.it (P.A.-M.)

Simple Summary: Once present in the entirety of Europe, mouflon (wild sheep) became extinct due to intense hunting, but remnant populations survived and became feral on the Mediterranean islands of Corsica and Sardinia. Although now protected by regional laws, Sardinian mouflon is threatened by crossbreeding with domestic sheep causing genetic hybridisation. The spread of domestic genes can be detrimental for wild populations as it dilutes the genetic features that characterise them. This work aimed to identify diagnostic tools that could be applied to monitor the level of hybridisation between mouflon and domestic sheep. Tens of thousands of genetic markers known as single nucleotide polymorphisms (SNPs) were screened and we identified the smallest number of SNPs necessary to discriminate between pure mouflon and sheep. We produced four SNP panels of different sizes which were able to assess the hybridisation level of a mouflon and we verified that the SNP panels efficacy is independent of the domestic sheep breed involved in the hybrid. The implementation of these results into actual diagnostic tools will help the conservation of this unique and irreplaceable mouflon population, and the methodology applied can easily be transferred to other case studies of interest.

Abstract: Hybridisation of wild populations with their domestic counterparts can lead to the loss of wildtype genetic integrity, outbreeding depression, and loss of adaptive features. The Mediterranean island of Sardinia hosts one of the last extant autochthonous European mouflon (*Ovis aries musimon*) populations. Although conservation policies, including reintroduction plans, have been enforced to preserve Sardinian mouflon, crossbreeding with domestic sheep has been documented. We identified panels of single nucleotide polymorphisms (SNPs) that could act as ancestry informative markers able to assess admixture in feral *x* domestic sheep hybrids. The medium-density SNP array genotyping data of Sardinian mouflon and domestic sheep (*O. aries aries*) showing pure ancestry were used as references. We applied a two-step selection algorithm to this data consisting of preselection via Principal Component Analysis followed by a supervised machine learning classification method based on random forest to develop SNP panels of various sizes. We generated ancestry informative marker (AIM) panels and tested their ability to assess admixture in mouflon *x* domestic sheep hybrids both in simulated and real populations of known ancestry proportions. All the AIM panels recorded high correlations with the ancestry proportion computed using the full medium-density SNP array. The AIM

panels proposed here may be used by conservation practitioners as diagnostic tools to exclude hybrids from reintroduction plans and improve conservation strategies for mouflon populations.

1. Introduction

Genetic hybridisation among related species is increasingly studied due to its role in the adaptation, evolution, and diversification of species (Barbato et al., 2020; Bruford, Ginja, Hoffmann, Joost, Wengel, et al., 2015; Dowling & Secor, 1997; Hedrick, 2013; X. J. Hu et al., 2019). Hybridisation due to crossbreeding between interfertile species can occur due to range overlap of the populations (Iacolina, 2018), or can be promoted by anthropogenic activities (Miller, 2012; Olden, Poff, Douglas, Douglas, & Fausch, 2004; R. Oliveira et al., 2015). Habitat degradation and livestock translocation have recently increased the rate of hybridisation events worldwide, contributing to the erosion of genetic diversity, and, in some cases, to the extinction of locally adapted wild animals (Allendorf, Leary, Spruell, & Wenburg, 2001; Mallet, 2005). In Europe, the genetic diversity represented by many locally adapted wild species is threatened due to hybridisation with their domestic counterpart (Iacolina, 2018). Among these species is the European mouflon (*Ovis aries musimon*) present on the Mediterranean islands of Corsica and Sardinia (Barbato, Hailer, Orozco-Terwengel, et al., 2017; Sanna et al., 2015). The European mouflon is considered a remnant of the first wave of sheep domestication that occurred in the Fertile Crescent ~11,000 years ago (YA), whereas the current domestic sheep (*Ovis aries*) have been associated to a second wave of domestication which occurred ~5,000 years later (Chessa et al., 2009). The European mouflon was introduced inCorsica and Sardinia ~6-7,000 YA following human migrations and established feral populations (i.e., domesticates that have returned to the wild state) which survived until today in the harshest mountainous area of the islands (Chessa et al., 2009). Sardinia hosts the largest extant autochthonous European mouflon population (Barbato, Hailer, Orozco-Terwengel, et al., 2017)(Sanna et al., 2015)(Mereu et al., 2019). Throughout the last century the Sardinian mouflon was reduced to the brink of extinction due to intense hunting and habitat erosion, until it was declared endangered and protected under local Governmental laws (Legge Regionale n23 del 1998). Since then, conservation efforts have been pursued to increase the mouflon presence in the island through translocation and genetic rescue. Since the arrival of the second wave of domestication in Europe, the sheep population brought to Sardinia has lived in sympatry with the already established mouflon population residing in the islands. Crosses between mouflon and domesticated sheep are documented since Roman times (Barbato, Hailer, Orozco-Terwengel, et al., 2017) and evidence of introgression from domestic sheep to mouflon has been occasionally reported by microsatellites and single nucleotide polymorphism (SNP) genotyping (Barbato, Hailer, Orozco-Terwengel, et al., 2017; Lorenzini, Cabras, Fanelli, & Carboni, 2011). Importantly, strong signals of recent introgression were detected in an enclosed mouflon population used for restocking across the

island (Barbato, Hailer, Orozco-Terwengel, et al., 2017). Such an occurrence is not surprising, as the identification of hybrids based on morphological features is unreliable, especially when backcrosses occur and individuals no longer show a distinguishable intermediate phenotype between parental taxa (Allendorf et al., 2001). Hence, establishing new populations from crossbred founders represents a threat for the Sardinian autochthonous genetic diversity. The use of genome-wide analysis using DNA arrays with tens or hundreds of thousands of SNPs allows the accurate detection of hybrid individuals despite confounding morphological evidence (Rhymer & Simberloff, 1996)[17]. However, the large-scale use of DNA arrays can be challenging for the average financial availability of conservation projects (Ruane, J., & Sonnino, 2006).

Small sets, or panels, of ancestry informative markers (AIMs) have been developed and used to infer population genetic parameters in several species. AIM panels have been used to estimate biogeographical ancestry and structure in human populations (Halder et al., 2008; Kosoy et al., 2008; Paschou et al., 2008, 2007), to discriminate among breeds and geographical origin of Italian sheep breeds (Dimauro et al., 2015), to identify cattle breeds (F Bertolini et al., 2018; F. Bertolini, G. Galimberti, 2015; Wilkinson et al., 2011), to trace the origin of animal products (Dimauro et al., 2013; Orrù et al., 2009), and for breed assignment and analysis of individual ancestry in cattle (Frkonja, Gredler, Schnyder, Curik, & So, 2012; Gorbach et al., 2010; Kumar et al., 2019; Lewis, Abas, Dadousis, Lykidis, & Paschou, 2011) and pig populations (Liang et al., 2019). Moreover, AIM panels have been developed to identify hybrids in wildlife conservation projects on wolf (Ettore Randi et al., 2014), wild cat, and mule deer (Russell et al., 2019).

In this work, we applied supervised machine learning approaches on mid-density DNA array data and identified AIMs able to correctly discriminate mouflon *x* domestic sheep crosses when tested both on real and simulated data. Our results provide a fundamental conservation tool for the affordable identification of hybrid mouflon *x* domestic sheep and for the quantification of their admixture level.

2. Materials and Methods

We collected publicly available genotype data (~50k SNPs) from 23 non-admixed feral Sardinian mouflon (MSar) and 23 Sarda sheep (SAR) (Figure S1, Supplementary Materials), and from 28 Sardinian mouflon hybrids (MxS) showing extensive levels of admixture according to previous analyses (Barbato, Hailer, Orozco-Terwengel, et al., 2017; E. Ciani et al., 2013).

The Sarda sheep is an autochthonous Sardinian sheep breed counting almost four million heads in Sardinia. It is reared for its high milk production and accounts for almost all of the sheep presence in the island. Additionally, we collected genotypic data from 26 Lacaune (LAC), 28 Australian Poll Merino (MER), and 23 New Zealand Texel (TEX) made available by the Sheep HapMap project (Kijas et al.,

2012). Lacaune is a milk-producing breed which has been recently imported into Sardinia, whereas Merino and Texel are two cosmopolitan breeds reared for wool and meat production, respectively (Kijas et al., 2012).

Non-autosomal variants, markers with missing data (> 0), minor allele frequency (MAF) ≤ 0.1, and markers significantly out of Hardy–Weinberg equilibrium (HWE ≤0.001) were excluded from the subsequent analyses. Pruning was performed using PLINK v1.9 (Christopher C Chang et al., 2015).

2.1. Aims Identification and Panel Development

To identify the markers from the medium-density SNP array genotyping data for MSar and SAR able to detect admixture between mouflon and domestic sheep, we applied a two-fold approach which included (1) a preselection step aimed at removing the least significant markers, and (2) supervised machine learning classification approaches to identify the most informative markers among those left after preselection.

2.2. Preselection

The dataset was first reduced based on Principal Component Analysis (PCA) results. PCA is a widely used dimensionality reduction technique that generates new uncorrelated variables (principal components; PCs) sorted according to the variance they explain. The contribution of each SNP to every PC is expressed as a loading score. A PCA analysis of the reference breeds was performed using R v3.4.3 (R Core Team, 2021). The first PC discriminated mouflon against domestic sheep, whereas the second PC described within-mouflon structure (see Results). Similarly, the following PCs identified within-population substructure. Consequently, we only considered the first PC. The loading scores of the first PC were squared (see (F. Bertolini, G. Galimberti, 2015; Paschou et al., 2007)), and those SNPs associated with loading values exceeding 1.5x the interquartile range of the loading distribution were retained for further analyses (hereon: GW1; Figure S2, Supplementary Materials).

2.3. Supervised Classification

SNPs which passed preselection were submitted to supervised machine learning classification, as implemented in Boruta v6.0.0 (Miron B. Kursa, 2010), in order to select the most discriminant markers. This algorithm is based on a random forest learning process that allows, after a training step, to derive a set of rules for data classification. The Boruta algorithm measures the importance of an attribute (here, a SNP) through the loss of classification accuracy due to random permutation of the attribute values between objects (Miron B. Kursa, 2010). Eventually, Boruta assigns the binary classification “confirmed important” and “confirmed unimportant” to each of the features tested. A single run of the Boruta algorithm was performed genome-wide to identify a first subset of significant SNPs (GW1) and those SNPs listed as “confirmed important” were assembled as a GW2 panel. To obtain an even smaller AIM panel, we performed 20 independent iterations of the classification algorithm on the GW1

panel; those SNPs listed as “confirmed important” in at least 19 scans out of 20 were pooled in a third panel (GW3) (Figure S3, Supplementary Materials).

With the aim to identify AIM panels which forcibly intercepted all autosomes, we applied the iteration algorithm to each autosome separately (Chromosome Wide, CW). Feature selection was performed for each autosome and SNPs listed as important were assembled in a panel named CH1. Lastly, to further reduce the number of significant SNPs, a last panel (CH2) was generated by selecting from CH1 the three most discriminant SNPs per each autosome.

2.4. AIM Panel Validation

The accuracy of each of the five AIM panels (three genome-wide and two chromosome-wide) to assess ancestry levels, or levels of admixture, was tested using genotype data from both simulated and real hybrid populations (MxS). To test the AIM panels with simulated hybrids, we developed Hybridiser v0.1, an R script able to simulate F1 hybrid individuals (available at https://github.com/barbatom/Hybridiser). Given two parental populations, Hybridiser computes the allele frequency at each locus, and then generates hybrid genotypes at each locus by selecting an allele from each of the parental populations with probability equal to the parental allele frequency. A dataset of 270 simulated hybrids (HYBs) was generated pooling 90 MSar *x* SAR first-generation crosses (F1) and the reciprocal backcrosses F1 *x* MSar (BC1M), and F1 *x* SAR (BC1S), each comprising 90 individuals.

Admixture v1.3.0 (Alexander et al., 2009) was used to perform supervised clustering tests to evaluate the ancestry proportions of both simulated and real hybrids using the pure mouflon and domestic populations as ancestry sources. To measure how well the AIM panels estimated the admixture level compared to that determined by the full set of markers, we compared the admixture results from each of the AIM panels (selected markers) with those from the full medium-density genotype array (full set of markers) using the coefficient of determination (r^2).

To test if the AIM panels performed better than an equally sized set of SNPs chosen at random, we generated 5,000 random AIM sets, and for each random set we performed supervised admixture analysis. Finally, we computed coefficients of determination values between the ancestry assignment of the full set and the reduced random panel. Being computationally challenging, this test was performed on the smallest AIM panel set we generated (GW3). The coefficient of determination values obtained using the 5,000 random SNP sets were standardised by z-scores. An empirical *p*-value for the correlation obtained using the GW3 panel was calculated according to Davison and Hinkley (Davison & Hinkley, 1997) as $p = (1 + x)/(1 + n)$, where x is the number of random sets that produced an r^2 score greater than or equal to that calculated using GW3 and n is the total number of random set tested (Figure S4, Supplementary Materials). Significance was set for *p*-values lower than 0.01.

Finally, we evaluated the ability of the five AIM panels to detect mouflon hybrids with domestic sheep breeds other than Sarda. For each mouflon *x* domestic sheep population, we generated a dataset of simulated hybrids composed of: i) 90 F1 offspring between MSar and the test domestic breed, ii) 90 backcrosses between F1 and the test domestic breed and iii) 90 individuals obtained as backcross between F1 and MSar. The AIM panels were tested on simulated hybrid populations of MSar with MER, LAC, and TEX, respectively. Coefficients of determination were computed between the supervised admixture results obtained using AIM panels (selected markers) and the full medium-density genotype array (full set of markers).

3. Results

Pruning for missing data, MAF, and HWE left 33,481 non-rare, neutral SNPs. PCA was performed on the two reference populations (MSar and SAR) to evaluate how many PCs contributed in discriminating mouflon and domestic sheep. As expected, PC1 (18.7% of the total variance), split mouflon and domestic sheep as two distinct clusters (Figure 1, Figure S1, Supplementary Materials). PC2 (7.07% of the total variance) identified subpopulation structure in mouflon exclusively (Figure 1). Consequently, PC1 was considered the only relevant component for the preselection step.

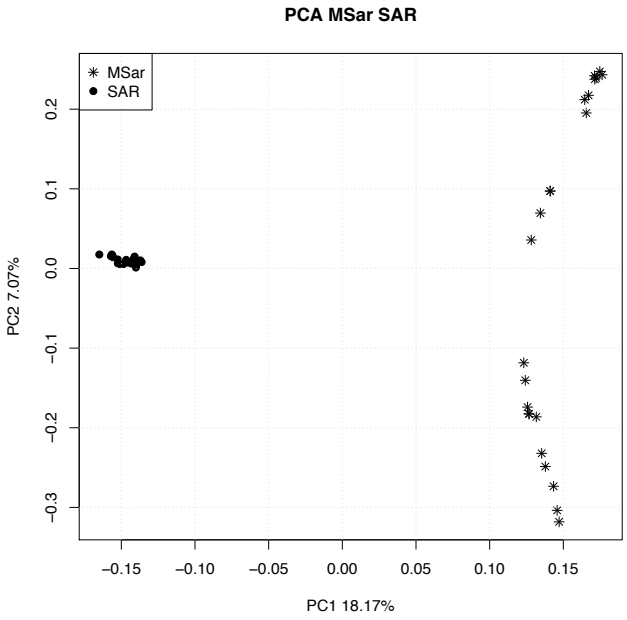


Figure 1. Principal Components Analysis (PC1 vs. PC2) of the two reference populations (MSar and SAR) analysed using the full single nucleotide polymorphism (SNP) set. In brackets are the percentage of variance explained by each component.

3.1. AIM Identification and Panel Development

The PCA-based preselection step identified 1,279 SNPs that contributed the most to discriminate between mouflon and domestic sheep reference populations (GW1; Table 1, Figure S2, Supplementary Materials). This first panel was submitted to both a single run and 20 iterations of

random forest selection that identified 131 (GW2), and 51 SNPs (GW3), respectively (Table 1). A single run of random forest was applied chromosome-wide and identified 933 SNPs (CH1), with a range of 10 to 73 SNPs per chromosome (Table S1; Figure S5, Supplementary Materials). The selection of the three most significant SNPs per chromosome from CH1 selected 78 SNPs (CH2) (Table 1).

Table 1. Characteristics of the ancestry informative marker panels. The data processes used for marker selection were: preselection (Pre), Random Forest (RF), iterated Random Forest (iRF), and top-markers choice (tc). N is the number of SNPs in each panel. The SNP distribution per chromosome can be found in Supplementary Table S1.

Panel Name	Scope	Method	N
GW1	Genome-wide	Pre	1279
GW2	Genome-wide	Pre + RF	131
GW3	Genome-wide	Pre + RF + iRF	51
CH1	Chromosome-wide	Pre + RF	933
CH2	Chromosome-wide	Pre + RF + tc	78

3.2. Panels Validation

PCA analyses were used to visually compare the performance of the full set of SNPs and the five AIM panels applied to simulated (HYB) and real hybrid (MxS) populations.

The PCA of HYB using the full set of SNPs discriminated the parental populations (SAR and MSar) at opposite sides of the graph and positioned the hybrid populations according to their ancestry proportions, with F1 at the centre of the plot and the two backcrosses BC1S and BC1M closer to SAR and MSar, respectively (Figure 2; Figure S6, Supplementary Materials). PCA of MxS using the full set of markers identified four individuals overlapping with the pure ancestry mouflon cluster, while the others were distributed along a gradient between MSar and SAR (Figure 3; Figure S7, Supplementary Materials).

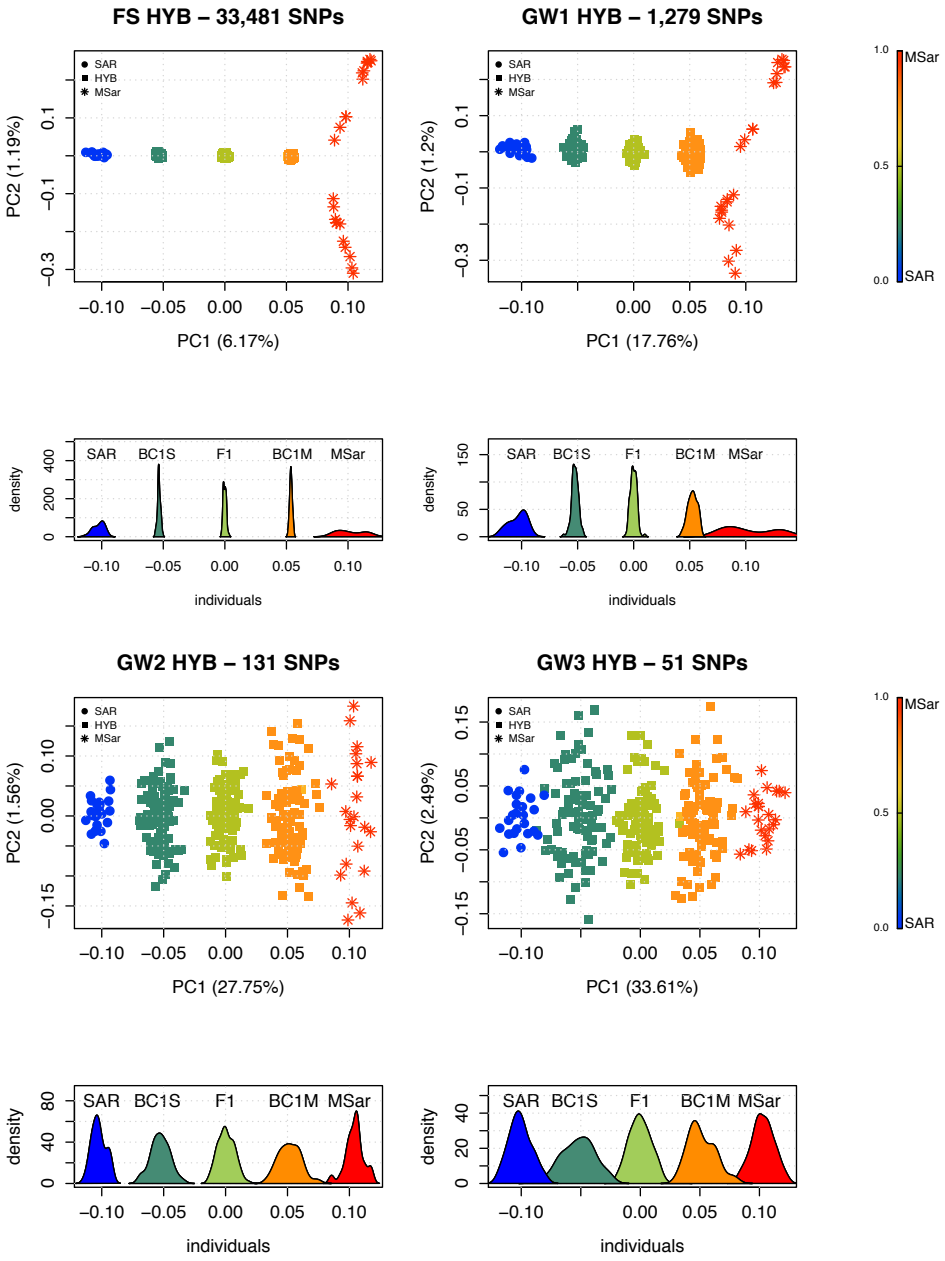


Figure 2. PCA and density distribution of the PC1 obtained using the full SNP set (top-left panel) and three AIMs on reference populations and simulated hybrids (HYB) using the genome-wide discovery approach. BC1S and BC1M are the simulated F1 backcrossed with SAR and MSar, respectively. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components.

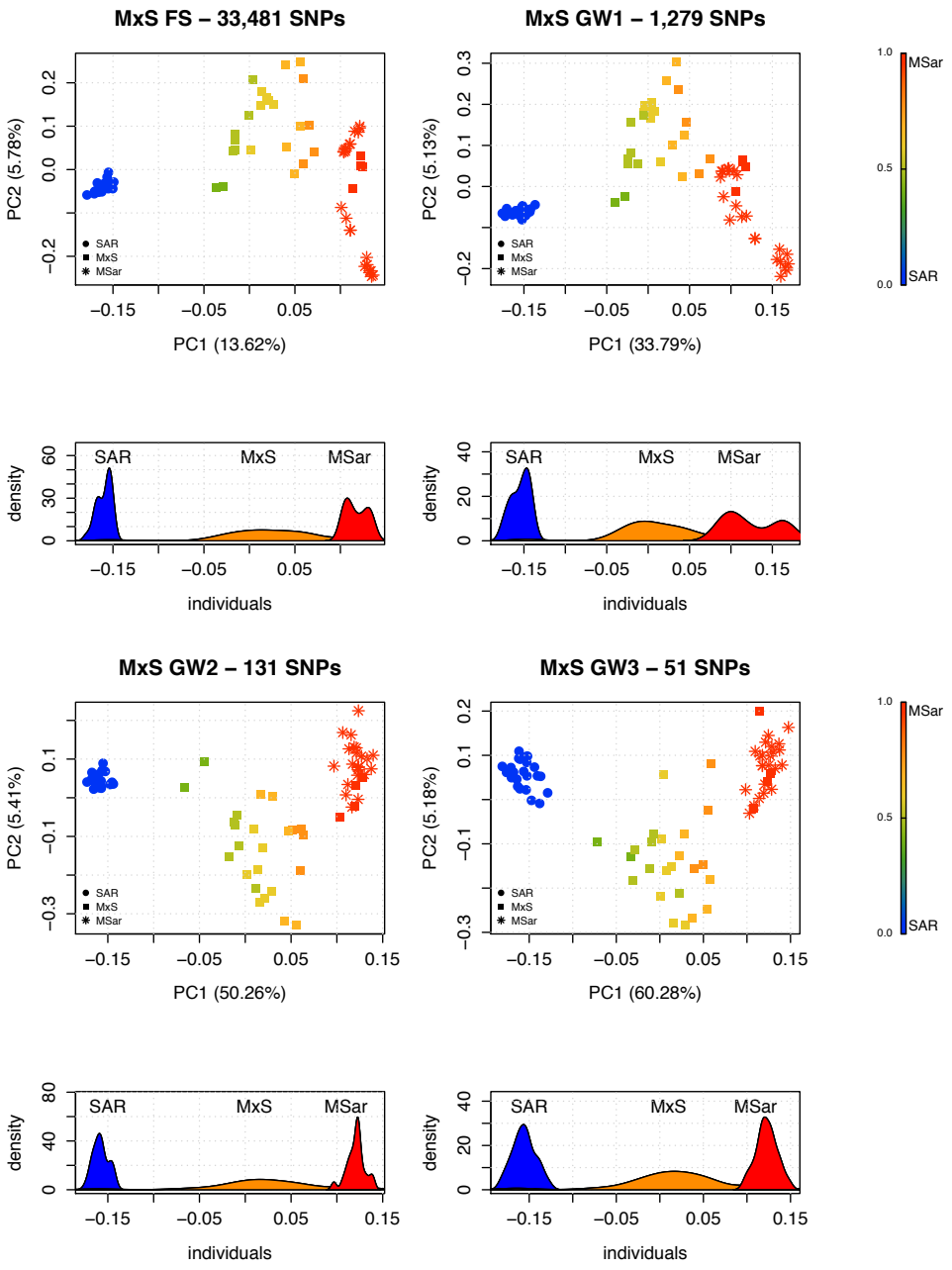


Figure 3. PCA and density distribution of the PC1 obtained using the full SNP set (top-left panel) and three AIMs on reference populations and real mouflon x domestic hybrids (MxS) using the genome-wide discovery approach. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components. The analysis was performed using 33,481 SNPs and the three GW panels.

PCA performed on HYB using the AIMs showed clustering comparable with the PCA results obtained using the full set of SNPs (Figure 2; Figure S6). Each AIM panel was able to discriminate the simulated ancestry proportions and showed distinct clusters for each simulated hybrid population, with only minimal cluster overlap for three individuals when using the smallest SNP set (GW3; Figure 2). When applied to the real population (MxS), GW1, GW2, GW3 and CH1 discriminated the pure and admixed individuals (Figure 3; Figure S7), whereas CH2 showed some overlap occurring between the pure mouflon and the admixed individuals having the smallest amount of domestic ancestry (Figure S7).

A quantitative assessment of the ancestry proportions in HYB was obtained through supervised Admixture using MSar and SAR as reference populations. HYB showed ancestry proportions coherent with the expected values of each offspring group (Figure S8, Supplementary materials); mean and standard deviation were 0.497 ± 0.0059 for F1, 0.234 ± 0.0059 for BC1S, and 0.76 ± 0.005 for BC1M. Supervised Admixture on MxS using the full set highlighted the heterogeneity of ancestry proportions in this population, and the presence of four individuals showing pure ancestry (Figure 4).

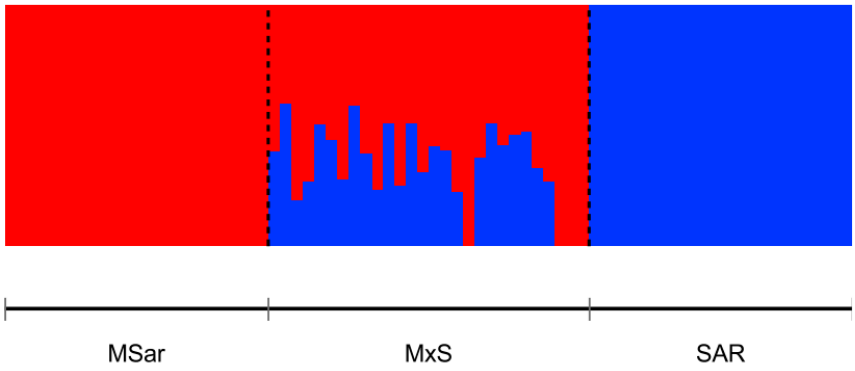


Figure 4. Supervised Admixture plot of MxS dataset obtained using the full set of SNPs. MSar and SAR were used as prior populations.

Coefficients of determination were then calculated between the ancestry proportions obtained using the full set of SNPs and AIM panels. The coefficients of determination were high overall across all panels ($r^2 \geq 0.960$), with GW1 and CH2 scoring the highest and lowest, respectively (Table 2). As expected, values were proportional to the number of AIMs in the panels. In simulated individuals, GW1 scored $r^2 = 0.990$ and CH1 $r^2 = 0.989$, while MxS had $r^2 = 0.997$ for both the panels. Coefficients of determination were slightly higher for AIMs identified using a genome-wide approach instead of chromosome-wide. The genome-wide GW3 panel performed better than the chromosome-wide CH2 panel despite the lower number of SNPs. AIM panels tested on HYB recorded higher correlation values compared to MxS.

Table 2. Coefficient of determination values (r^2) calculated between the ancestry percentages using the full set of SNPs and the AIM panels in the simulated (HYB) and case study (MxS) populations. N is the number of SNPs in each panel.

AIMs	N	HYB	MxS
GW1	1279	0.997	0.99
GW2	131	0.985	0.971
GW3	51	0.966	0.966
CH1	933	0.997	0.989
CH2	78	0.961	0.946

We further tested the ancestry assignments obtained using GW3 (the smallest panel among the AIMs generated) against the null hypothesis that the same results could be obtained using any equal sized set of SNPs chosen at random. The null hypothesis was rejected with high significance (*p*-value <0.001; Figure S4, Supplementary Materials).

AIM panels were further tested to detect admixture between mouflon (MSar) and three commercial sheep breeds (TEX, APM, and LAC). Coefficients of determination calculated between the ancestry percentage obtained using the AIM panels and full set of SNP ancestry resulted in $r^2 > 0.92$ (Table 3).

Table 3. Coefficient of determination values calculated between the ancestry percentages obtained using the full SNP set and the AIMs in three commercial sheep breeds.

Breed	Acronym	GW1	GW2	GW3	CH1	CH2
New Zealand Texel	TEX	0.995	0.971	0.945	0.993	0.920
Australian Poll Merino	APM	0.995	0.968	0.938	0.993	0.923
Lacaune	LAC	0.994	0.962	0.928	0.992	0.923

4. Discussion

Hybridisation between wild species and their domestic counterpart may lead to genetic homogenization and loss of local adaptation (Allendorf et al., 2001). Sardinian mouflon is the largest extant autochthonous European mouflon population and represents a unique genetic heritage threatened by hybridisation with domestic sheep (Barbato, Hailer, Orozco-Terwengel, et al., 2017)(Mereu et al., 2019)(Lorenzini et al., 2011).

In this work we implemented a supervised machine-learning-based classification approach to identify highly discriminant markers which can be included in rapid and low-cost diagnostic assays with the aim to support mouflon conservation (VonHoldt et al., 2013). The accuracy of AIM panels to detect introgression depends on the quality and sample size of the reference populations, with increasing probability of capturing most of the existing within-population variability when purer/non-admixed and larger reference datasets are used (Hulsegge et al., 2013). To the best of our knowledge, previous work on AIM panel discovery performed on humans, domestic and wild animals always accounted >100 reference samples (Dimauro et al., 2013, 2015; F. Bertolini, G. Galimberti, 2015; Halder et al., 2008; Paschou et al., 2008; Ettore Randi et al., 2014; Russell et al., 2019; Wilkinson et al., 2011). Due to the large genetic distance occurring between sheep and mouflon, we achieved reliable feature selection using a sample size of 23 individuals for each reference population. Our results showed that the AIMs we identified can accurately discriminate Sardinian mouflon ancestry from other cosmopolitan sheep breeds as well. As expected, panels counting a higher number of SNPs, such as GW1 and CH1, resulted in better performances in hybrid identification than other panels. However, previous research showed that reliable results in individual assignment can be achieved also with a number of SNPs lower than 100 (Dimauro et al., 2013; Frkonja et al., 2012; Kersbergen et al., 2009; Lewis et al., 2011; Paschou et al., 2007; VonHoldt et al., 2013). Accordingly, we obtained positive results also with less than 100 AIMs, as shown by tests performed with GW3 and CH2 panels. We tested AIM identification using both chromosome-wide and genome-wide approaches as done in previous research (see (F Bertolini et al., 2018). Our results confirmed the genome-wide approach to

perform better than chromosome-wide. Chromosome-wide selection may introduce feature redundancy by selecting several SNPs on different chromosomes but carry identical discriminant power. Conversely, the genome-wide approach might be better at selecting highly discriminant SNPs which also capture a wider range of the focal population diversity. Noticeably, the distribution of AIMs in the genome-wide approach appears to be not homogeneous along the genome. Indeed, two phylogenetically close references, as in this case, are likely to share most of the genome, with few genomic regions responsible for much of the genomic distance, further supporting the genome- over the chromosome-wide approach as best suited to intercept such regions.

We assessed the performance of the panels in identifying crosses between domestic sheep and feral mouflon using both simulated and real data. As expected, using the AIMs on simulated data performed moderately better than on real admixed mouflon samples. Real admixed populations present a more complex genetic make-up, influenced by demography, selection, inbreeding and introgression events occurred during centuries. Conversely, the simulated individuals were generated from the same reference populations used to select the best AIMs. In addition, the mating system applied in simulations generates simplified admixture patterns with respect to those occurring in real populations. However, the use of simulated hybrid genotypes may aid in AIM panels testing to overcome the lack of real admixed samples. When available, real hybrids samples can guarantee the accuracy of the AIMs panels.

5. Conclusions

The detection of hybrids is a fundamental task in biodiversity conservation, and genetics provides an unmatched tool to identify even cryptic levels of hybridization. Applying a machine learning-based approach, we identified reduced SNP panels which proved effective in identifying domestic sheep introgression in Sardinian mouflon. The AIM panels we propose can be applied for large-scale assessment of the Sardinian mouflon population, and aid in the conservation of this unique genetic resource. Lastly, the method we present can be easily applied to other wildlife and domestic species conservation to develop accurate and affordable tools for hybrid identification.

Supplementary information

Table S1 – distribution of SNPs per chromosome: for each AIMs panel is reported the exact number of SNPs per chromosome and the percentage related to FS.

	FS	GW1	%	GW2	%	GW3	%	CH1	%	CH2	%
1	3760	126	3,35	10	0,01	5	0,13	66	1,75	3	0,07
2	3585	148	4,12	16	0,44	7	0,19	73	2,03	3	0,08
3	3259	129	3,95	15	0,46	9	0,27	64	1,96	3	0,09
4	1765	77	4,36	7	0,39	1	0,05	65	3,68	3	0,16
5	1537	57	3,70	4	0,26	-	-	46	2,99	3	0,19
6	1723	65	3,77	11	0,63	3	0,17	54	3,13	3	0,17
7	1504	49	3,25	4	0,26	3	0,19	47	3,12	3	0,19
8	1411	44	3,11	1	0,07	1	0,07	41	2,90	3	0,21
9	1398	72	5,15	6	0,42	2	0,14	46	3,29	3	0,21
10	1151	43	3,73	10	0,86	4	0,34	33	2,86	3	0,26
11	686	27	3,93	5	0,72	3	0,43	23	3,35	3	0,43
12	1065	29	2,72	3	0,28	-	-	31	2,91	3	0,28
13	1059	50	4,72	6	0,56	1	0,09	27	2,54	3	0,28
14	724	23	3,17	2	0,27	1	0,13	27	3,72	3	0,41
15	1057	50	4,73	1	0,09	-	-	43	4,06	3	0,28
16	996	38	3,81	5	0,50	2	0,20	35	3,51	3	0,41
17	885	34	3,84	1	0,11	1	0,11	32	3,61	3	0,28
18	891	32	3,59	3	0,33	1	0,11	31	3,47	3	0,30
19	739	39	5,27	3	0,40	2	0,27	23	3,11	3	0,33
20	715	31	4,33	4	0,55	1	0,13	15	2,09	3	0,33
21	546	11	2,01	-	-	-	-	15	2,74	3	0,40
22	678	29	4,27	5	0,73	2	0,29	23	3,39	3	0,41
23	679	20	2,94	1	0,14	1	0,14	21	3,09	3	0,54
24	443	20	4,51	3	0,67	2	0,45	15	3,38	3	0,44
25	639	18	2,81	5	0,78	-	-	10	1,56	3	0,44
26	586	18	3,07	-	-	-	-	27	4,60	3	0,67

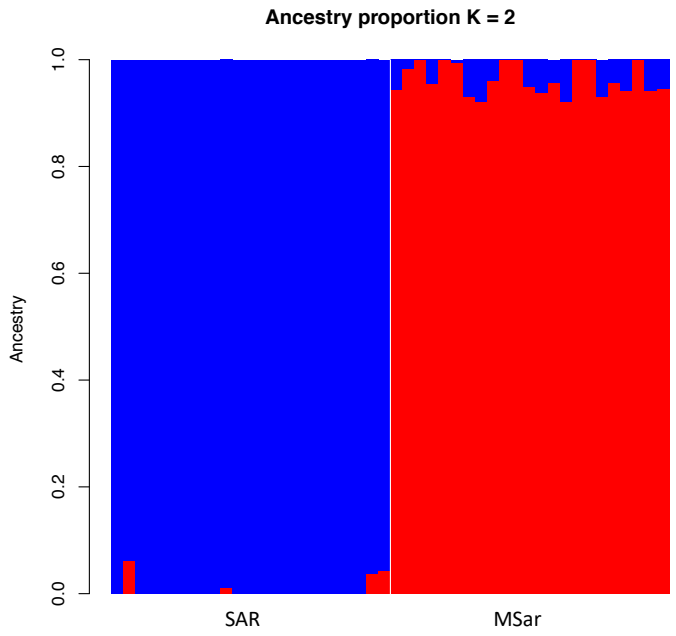


Figure S1 – Unsupervised Admixture analysis for K=2 evaluated on the full set of reference populations (SAR and MSar)

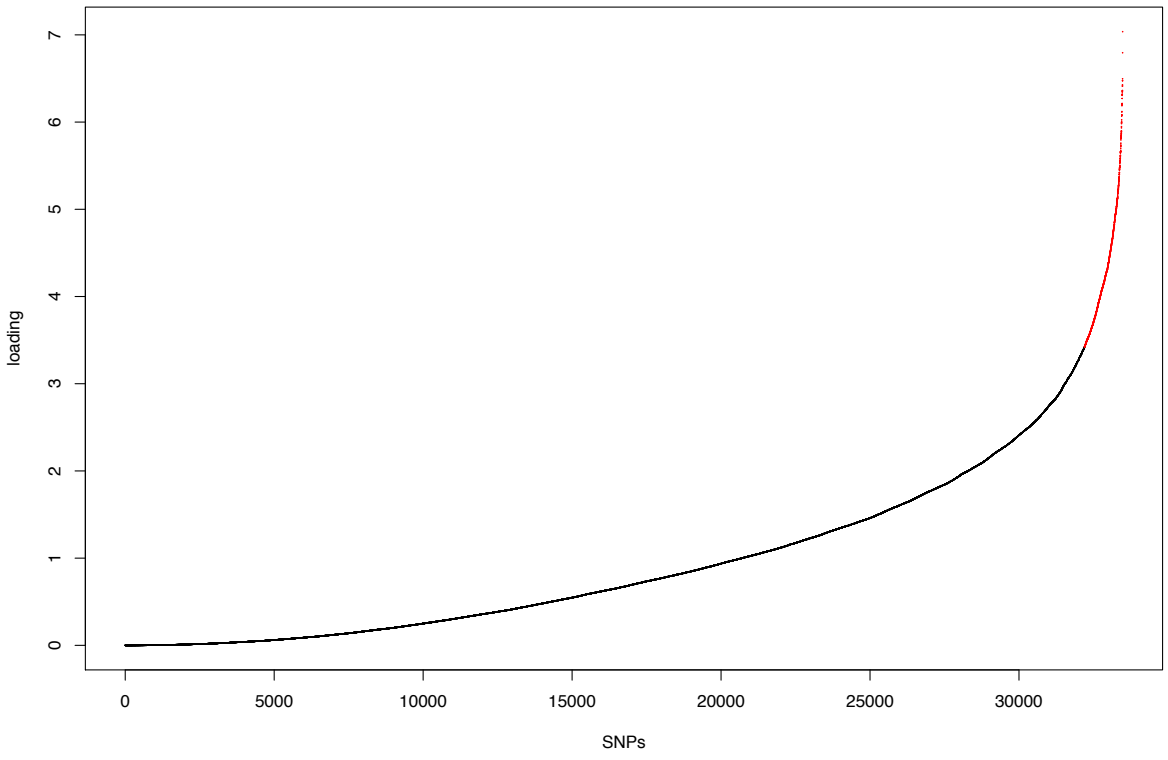


Figure S2 – Plot of PC1 loadings, squared and ordered. Red dots represent SNPs selected.

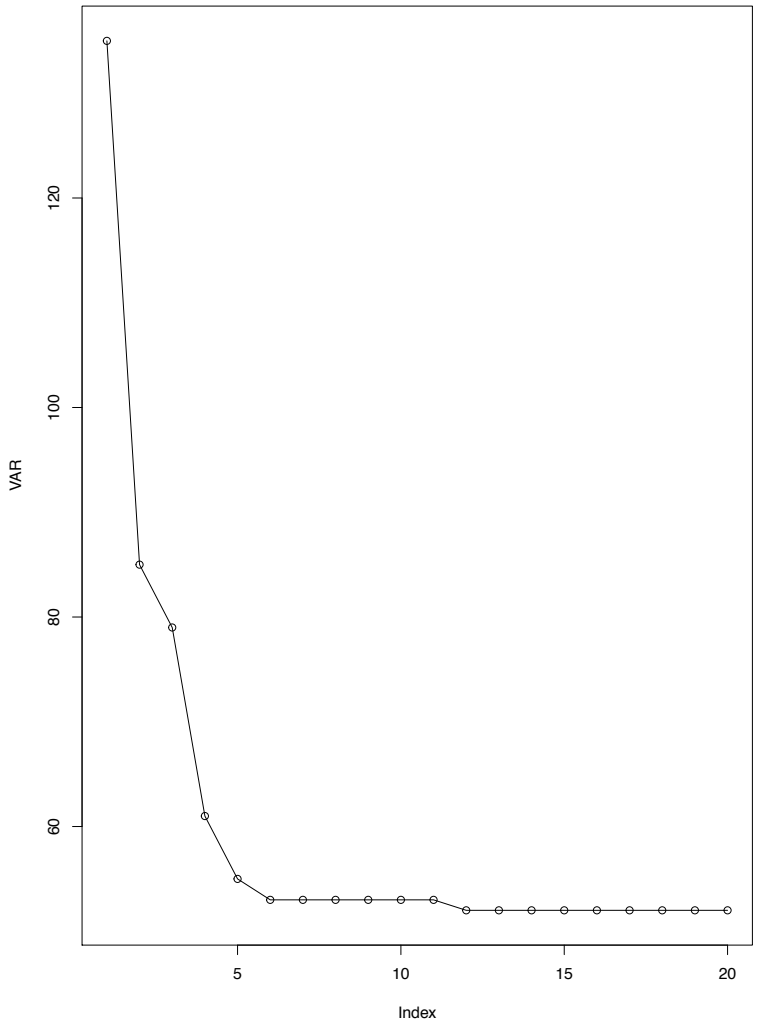


Figure S3 – For each one of the 20 subsequent iterations is shown the consensus number of SNPs “confirmed important” in all the performed iterations.

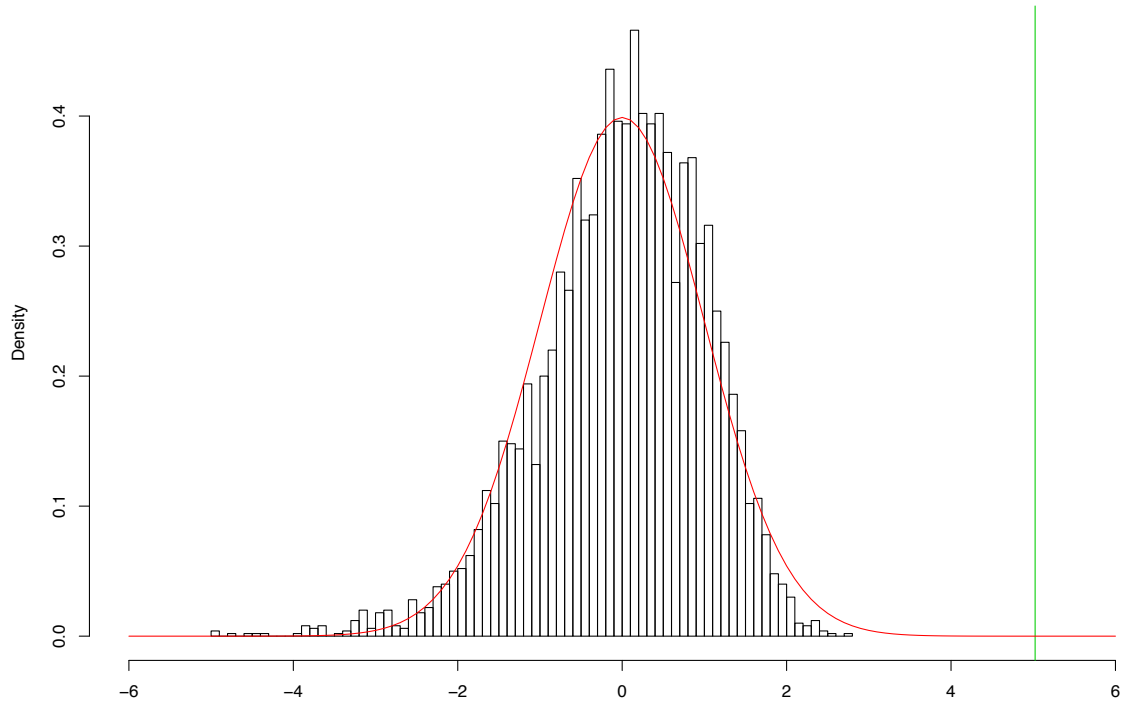


Figure S4 – plot of the Pearson correlation values density distribution of random panels test. The green line represents the correlation value obtained for the GW3 panel.

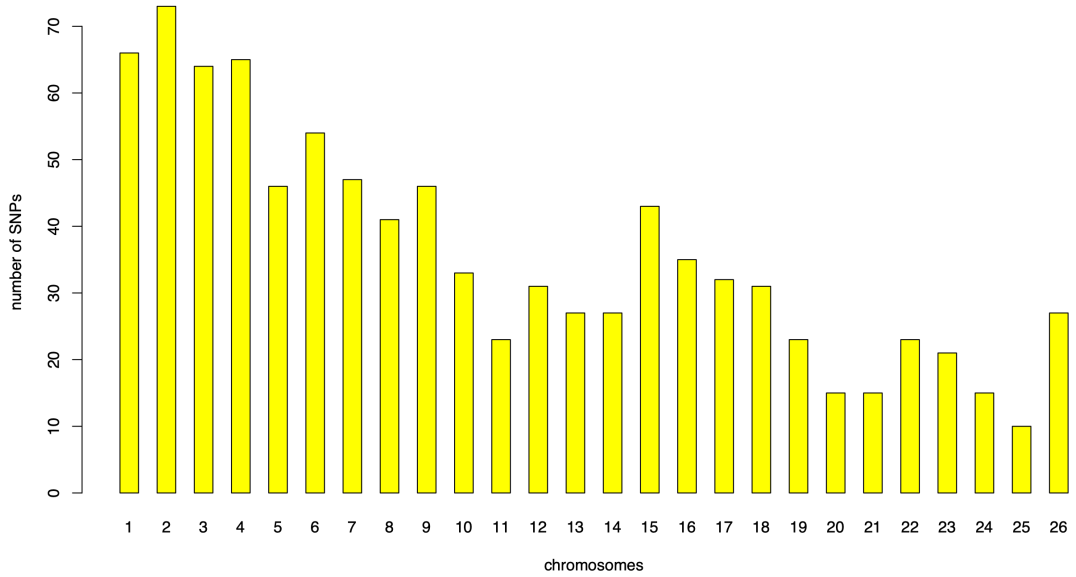


Figure S5 – distribution of SNPs per chromosome in panel CH1

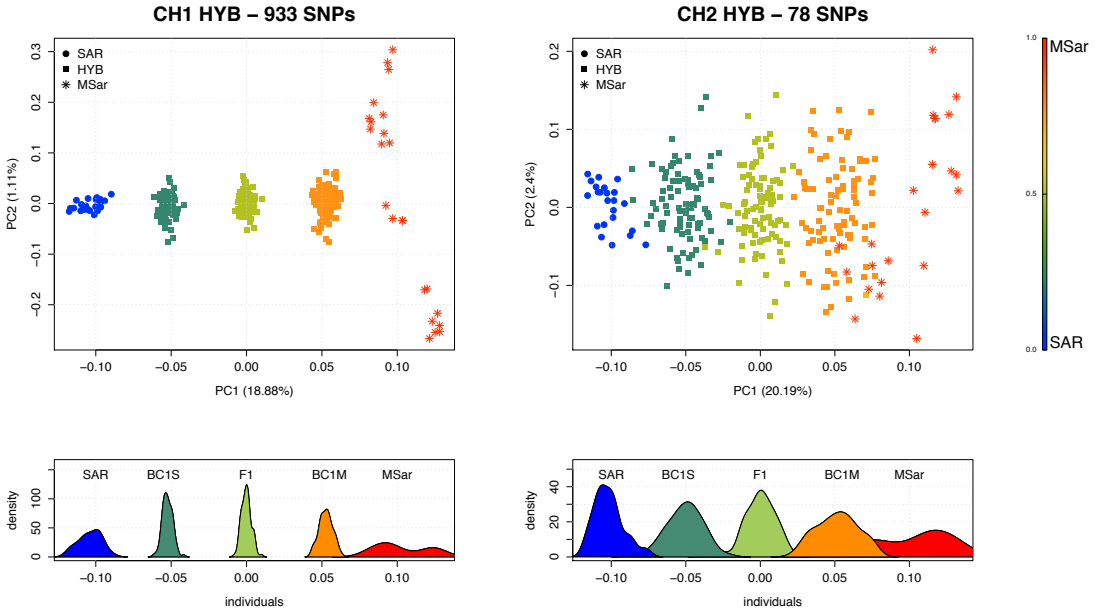


Figure S6 –– Principal Components Analysis (PC1 vs PC2) analysis and density distribution of mouflon (MSar), hybrid (HYB), and domestic sheep (SAR) populations. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components. The analysis was performed using panel CH1 and CH2.

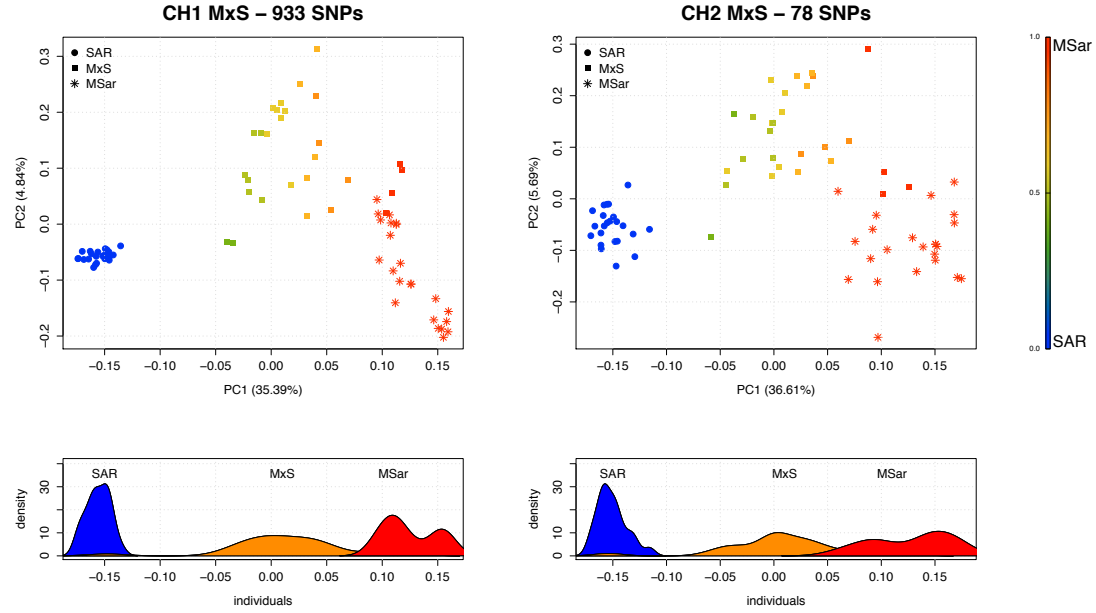


Figure S7 –Principal Components Analysis (PC1 vs PC2) analysis and density distribution of mouflon (MSar), mouflon x domestic hybrid (MxS), and domestic sheep (SAR) populations. The gradient legend on the right side of the plot shows the transition gradient from sheep to mouflon genetic components. The analysis was with CH1 and CH2.

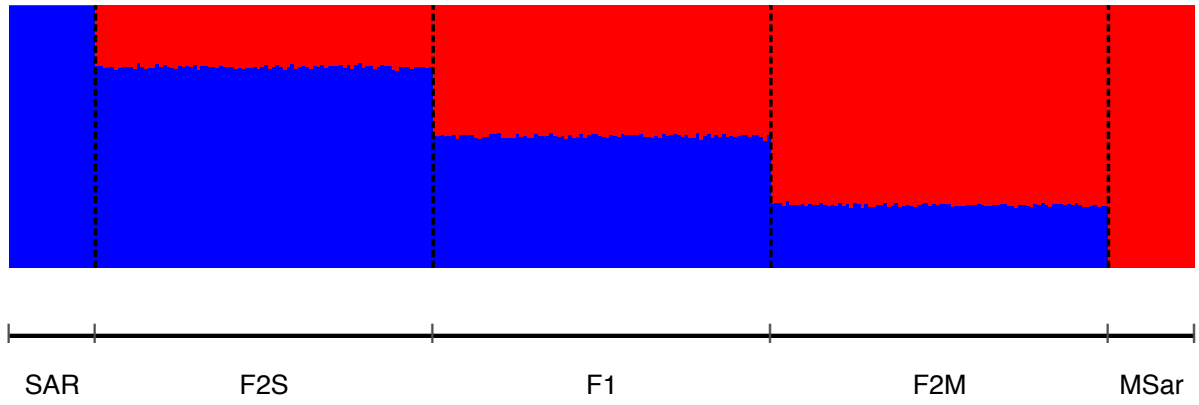


Figure S8 – supervised admixture K=2 plot evaluated with full set on the HYB dataset

Chapter 2: The SNP-Based Profiling of Montecristo Feral Goat Populations Reveals a History of Isolation, Bottlenecks, and the Effects of Management

Elisa Somenzi [1,*], Gabriele Senczuk [2], Roberta Ciampolini [3], Matteo Cortellari [4], Elia Vajana [1,5], Gwenola Tosser-Klopp [6], Fabio Pilla [2], Paolo Ajmone-Marsan [1,7], Paola Crepaldi [4] and Licia Colli [1,8]

[1]DIANA Dipartimento di Scienze Animali, della Nutrizione e degli Alimenti, Università Cattolica del S. Cuore, Piacenza, Italy; elia.vajana@epfl.ch (E.V.); paolo.ajmone@unicatt.it (P.A.-M.); licia.colli@unicatt.it (L.C.)
[2]Dipartimento di Agricoltura, Ambiente e Alimenti, University of Molise, 86100 Campobasso, Italy; g.senczuk@unimol.it (G.S.); pilla@unimol.it (F.P.)
[3]Dipartimento di Scienze Veterinarie, University of Pisa, 56124, Pisa, Italy; roberta.ciampolini@unipi.it
[4]Dipartimento di Scienze Agrarie ed Ambientali—Produzione, Territorio, Agroenergia, University of Milano, 20133 Milan, Italy; matteo.cortellari@unimi.it (M.C.); paola.crepaldi@unimi.it (P.C.)
[5]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), 1015, Lausanne, Switzerland
[6]GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326 Castanet-Tolosan, France; gwenola.tosser@inrae.fr
[7]Centro di Ricerca Nutrigenomica e Proteomica-PRONUTRIGEN, Università Cattolica del Sacro Cuore, 29122, Piacenza, Italy
[8]Centro di Ricerca sulla Biodiversità e sul DNA Antico-BioDNA, Università Cattolica del Sacro Cuore, 29122, Piacenza, Italy

Abstract: The Montecristo wild goat is an endangered feral population that has been on the homonymous 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, highlighting 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.*

1. Introduction

For centuries, Mediterranean islands have hosted several feral goat populations, strongly adapted to survive in arid environments and feed scarcity (Horwitz & Bar-Gal, 2006). Among these is

the Montecristo feral goat, a free-ranging population inhabiting the homonymous Italian island since ancient times. The Montecristo Island is a relatively small island (ca. 1000 hectares) in the Mediterranean basin. Located in the Thyrrenian sea 60 kms from the coasts of Tuscany, this island has been a Nature Reserve since 1971 (European Committee for the Conservation of Nature and Natural Resources, 1986; Pavan, 1971), and hosts several endangered and endemic species. The Montecristo feral goats are characterised by phenotypic traits shared both with the domestic goat *Capra hircus* (i.e., the small size and the wide colour variability) and with the semi-wild goat populations of the Mediterranean basin (i.e., scimitar shape horns, present in both sexes but more prominent in males) (F. Ciani & Masseti, 1991; Doro et al., 2016). Due to the occurrence of mixed phenotypic traits, the taxonomic status of Montecristo feral goats has long been debated, with some authors referring to this population as *Capra aegagrus,* to point to a closer relationship with the Mediterranean feral goats (Masseti, 2016).

Despite several hypotheses having been proposed, the origins of the Montecristo goats are unknown; indeed, any precise inference has been hampered so far by the lack of clear archaeological data owing to the island soil that prevents the formation of fossils (Spagnesi, Toso, & Masseti, 2003). Some authors (Masseti, 2009) have suggested that Montecristo goats were introduced onto the island during Neolithic times, while others (Angelici, Laurenti, & Nappi, 2009; Bruno & Sauli, 1976; Toschi, 1953) date the first occurrence much more recently between the 6th and 13th centuries, when the goats were exploited as a food resource by the monk community settled on the island. In the mid-19th century the presence of feral goats on the island was reported by Alexandre Dumas in his famous novel "The Count of Montecristo": *"At every step that Edmond took he disturbed the lizards glittering with the hues of the emerald; afar off he saw the wild goats bounding from crag to crag*", and was further documented by the naturalists exploring the island in the same period (Masseti, 2016). From the late-19th century until the mid-20th Montecristo Island was exploited as a Savoia royal family's game reserve. In the same period, the introduction of goats from the mainland for restocking purposes was reported (Pelliccioni Raganella, Lazzaro, Gotti, & Baccetti, 2015). Between the 1950s and the 1960s the island was owned by a private company, which established a game reserve (Masseti, 2016). Due to excessive hunting the original nucleus of goats was possibly reduced to less than 10 individuals (Angelici et al., 2009) and subsequently restocked with goats from the mainland (Masseti, 2016). When Montecristo Island was declared a Nature Reserve in 1971 to preserve the autochthonous species and the local goats (Pavan, 1971), hunting was finally forbidden, leading to a steady increase in the number of goats, which soon became a threat to the island ecosystem. Since then, several selective culling campaigns have been performed to reduce population size (Pelliccioni Raganella et al., 2015).

Currently, a small population of goats are farmed in Tuscany in the province of Grosseto, which are described as the descendants of animals allegedly moved from Montecristo Island to the Italian mainland in the last decades of the 20th century for conservation and research purposes. The information regarding the origin, recent history, and management of these nuclei is scanty, with their establishment being dated back either between the end of the 1970s and 1980s, or even before the Nature Reserve was set up in 1971 ("Preserving the Biodiversity of the 'Il Felcetone' Farm in Tuscany," n.d.). These small nuclei of animals, i.e., each one including one male and four females, were hosted at different breeding facilities on the mainland (Ciampolini R., pers. comm.) and likely underwent strong demographic fluctuations in the following decades.

Starting from the last decade of the 20th century, a few studies have been carried out to characterize Montecristo goats by means of molecular markers. In 1990, Randi and colleagues analysed allozyme loci of 20 samples from the Montecristo goat population, and underlined the occurrence of several introductions of domestic goats from mainland that contributed to the original gene pool. More recently, Doro and colleagues (Doro et al., 2016) analysed the complete mitochondrial DNA sequence of a single male specimen of Montecristo feral goat, highlighting a similarity with Western European domestic lineages. Analyses of Montecristo goats microsatellites and mtDNA were part of a LIFE project focussed on Montecristo island (Zanichelli, Giannini, De Pietro, & Puppo, 2014), which highlighted the absence of bottleneck events in the insular population as well as the presence of two unique mitochondrial haplotypes. Furthermore, a clear differentiation between insular and ex situ stocks was assessed (Pelliccioni Raganella et al., 2015). The animals sampled on the mainland, in fact, were not assigned to the gene pool typical of insular Montecristo goats, since they displayed mitochondrial haplotypes and microsatellite alleles not found in the insular population (Zanichelli et al., 2014).

Several authors have discussed the origin and the conservation status of the Montecristo goats during the last decades (Gippoliti, 2016; Masseti, 2009). However, the combined lack of pre-1980s demographical information and extensive genomic sampling has not allowed light to be shed on the genetic heritage of this feral stock, so far.

To tackle this gap, the present work assembled and analysed a 50K SNP dataset including 55 domestic goat breeds and 50 Montecristo individuals, from both the island and the mainland, to investigate the demographical history of the Montecristo population in the context of the Mediterranean basin. To our knowledge, this is the most comprehensive study on the Montecristo feral goat molecular diversity to date, which can represent a first step towards marker-assisted conservation of this peculiar goat population.

2. Materials and Methods

2.1. Dataset Construction and Filtering

The Montecristo goat samples included 32 individuals (MNT_I) from the free-ranging population inhabiting Montecristo Island sampled between 1995 and 2012, and 18 goats (MNT_M) reared ex situ at a farm in Seggiano on the Mount Amiata (Grosseto province, Tuscany, Italy). The latter population was sampled in 2016 and comprises the descendants of two of the four starting nuclei of individuals allegedly moved from the island of Montecristo to the Italian mainland between the end of the 1970s and the beginning of the 1980s.

The animals were genotyped with the Illumina GoatSNP50 BeadChip (Tosser-Klopp et al., 2014) as described in Cortellari et al.(Cortellari, Barbato, Talenti, Bionda, Randi, et al., 2021). Genotypes were merged with publicly available data representing 55 goat breeds from the Mediterranean basin and south-western Asia (Figure 1; Supplementary Table S1) (Cortellari, Barbato, Talenti, Bionda, Randi, et al., 2021; Nicoloso et al., 2015; Stella et al., 2018).

Figure 1. Geographical distribution of the 57 breeds included in the dataset. The labels indicate the centroids of sampling locations. The inset shows a male Montecristo goat with light-coloured phenotype (Photo from https://www.ruminantia.it/vi-raccontiamo-le-razze-la-capra-di-montecristo/ accessed on 11 12 November 2021). For the correspondence between labels and breed names refer to Supplementary Table S1.

Data from different sources were re-mapped on the goat reference genome ARS1, merged, and quality-controlled using PLINK 1.9 (Christopher C Chang et al., 2015). Individuals and markers exceeding the following thresholds were removed: (i) SNP call rate < 0.98; (ii) individual genotype call rate < 0.96; (iii) minor allele frequency (MAF) < 0.1. SNPs with unknown map position or located on sex chromosomes were removed. Pruning for linkage disequilibrium (LD) was performed using the '--indep-pairwise' function in PLINK (Christopher C Chang et al., 2015), where SNPs with r2 > 0.2 were removed from sliding windows of 50 SNPs and a step size of five SNPs. Breeds with sample size larger

than 30 individuals were thinned to a subset of 30 representative samples using the function representative.sample implemented in R package BITE (Milanesi et al., 2017).

To mitigate the bias possibly deriving from the reduced number of polymorphic loci in MNT_I population, we assembled a second dataset including only the markers scored as polymorphic in MNT_I (poly-MNT_I dataset). The two datasets were subjected to the same analyses and the results were compared.

2.2. Estimation of Genetic Diversity, Population Structure, and Migration Events

Observed and Expected heterozygosities (H_O and H_E) and the inbreeding coefficient (F_{IS}) values were calculated using the software Arelquin v3.5.2.2, which allows analysing genomic data (Excoffier & Lischer, 2010). To minimize the effects of the different numbers of polymorphic loci on the estimates of diversity, heterozygosity values were corrected over the number of usable SNPs with the formula proposed by Colli and colleagues (Colli et al., 2018).

Principal Component Analysis (PCA) was performed with the '*--pca-clusters*' flag in PLINK v1.9 (Christopher C Chang et al., 2015) in both and unsupervised (i.e., standard) and a supervised fashion to account for outlier's behaviour, using the '*--pca-clusters*' function. In this approach, outlier populations are first identified through unsupervised PCA, then principal components are calculated leaving the outlier populations out in the supervised analysis; finally, the individuals belonging to the outlier populations are assigned coordinates and projected onto the supervised PC axes. Results were visualized in R v3.6.1 (R Core Team, 2021).

Reynolds' unweighted distances between breeds were calculated with Arlequin (Excoffier & Lischer, 2010) and used to build a Neighbour-net graph with SplitsTree v4.14.6 software (Huson & Bryant, 2006). Admixture v 1.3.0 (Alexander et al., 2009) was used to evaluate population structure through a maximum-likelihood-based approach. Analyses were performed for K values ranging from 2 to 45. The best fitting K was identified as the one scoring the lowest cross-validation error value.

The occurrence of migration events was investigated with the software Treemix v. 1.13 (Pickrell & Pritchard, 2012) by setting windows of 500 consecutive SNPs and testing migration events (*m*) from 0 to 11, with 5 iteration each. An ad hoc statistic implemented in the R package OptM was used to identify the most likely number of migration edges (Fitak, 2021). The robustness of the nodes of Treemix underlying graph was estimated through 100 bootstrap replicates run for the best *m* value. A consensus tree was produced using the *consense.exe* executable in PHYLIP v3.695 (Felsenstein, 2004) and plotted with the BITE function *treemix.bootstrap* (Milanesi et al., 2017).

Lastly, we used the LD-based method implemented in the SNeP v1.1 software (Barbato et al., 2015) to evaluate the changes in Effective Population size (*Ne*) during the last 1000 generations. This analysis

was performed on the two Montecristo feral goats' populations MNT_I and MNT_M, and on the three geographically closest breeds, namely, Garfagnana (GRF), Sarda (SAR) and Corse (CRS).

2.3. Runs of Homozygosity and Heterozygosity-Rich Regions

Continuous stretches of homozygous sequences, i.e., Runs of homozygosity (ROHs), were scored on the two Montecristo *nuclei* to assess population demographic history. ROHs were computed with PLINK v1.9 (Christopher C Chang et al., 2015) as described in (Cortellari, Bionda, Negro, Frattini, Mastrangelo, et al., 2021). Number of ROHs per animal, average ROH length and ROH distribution across length classes (0–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb, and >16 Mb; see Supplementary Table S2) were calculated. Scored ROHs were visualized chromosome-wise in R v3.6.1 (R Core Team, 2021) and the ROH coverage-derived genomic inbreeding was computed with the dedicated function in R package detectRUNS (Biscarini, Cozzi, Gaspa, & Marras, 2018). The per-proportion of times each SNP falls inside a run in a given population was calculated using the R package detectRUNS (Biscarini et al., 2018). SNPs with a value in the top 0.1% of the percentile distribution were considered as statistically significant. Each significant SNP was annotated with the R package GALLO (Fonseca, Suárez-Vega, Marras, & Cánovas, 2020), considering an interval of 1000 bp upstream and downstream the examined marker. Genomic inbreeding derived from ROH coverage was computed for the two Montecristo populations using the dedicated formula implemented in the package detectRUNS (Biscarini et al., 2018). Heterozygosity-rich regions (HHRs) were evaluated with the R package detectRUNS through the "Sliding Windows" method with the following criteria: (i) the window size was set to 10 SNPs; (ii) the window threshold was kept with the default 0.05 value; (iii) the minimum number of SNPs in a HHR was 5; and (iv) the minimum length of a HHR was 500 bps. For identifying shared regions of interest, the same approach used for ROHs was applied. The top 0.1% SNPs in the percentile distribution of the number of times each SNP falls inside a run were considered as to be significant. Significant markers were annotated with the R package GALLO (Fonseca et al., 2020).

2.4. Approximate Bayesian Computation

The Approximate Bayesian Computation–Random Forest approach (ABC-RF) implemented in the DiyABC—rf1.0 software (Collin et al., 2021) was used to test alternative scenarios describing the recent demographic history of Montecristo goats. This approach allows to: (i) simulate multiple historical models; (ii) evaluate the parameter estimated during the analysis; and (iii) rank the best fitting model based on the approximate posterior probabilities.

A dedicated dataset was assembled to carry out ABC-RF analysis, including the two Montecristo populations (MNT_M, MNT_I) together with the geographically and genetically closest breeds as identified in previous analyses: Garfagnana (GRF), Sarda (SAR) and Corse (CRS). To reduce computational burden the dataset was further pruned for LD, reducing the number of SNPs to 9757.

To estimate the extent of a possible loss of information in the reduced dataset, we computed the Pearson's correlation coefficient between eigenvector values obtained with the main and the reduced datasets.

To test our hypotheses on recent demographic history, we modelled four different scenarios based on the results of population structure analyses and available historical records on the Montecristo goat population. In Scenario 1 (Supplementary Figure S1, panel a) we tested the hypothesis of the continental population (ancestors of GRF) diverging as first from the insular ones. The Montecristo goat population was modelled as deriving from an admixture event involving the Sarda and Corsa domestic breeds. Under Scenario 2 (Supplementary Figure S1, panel b) we assumed a more ancient origin of the Montecristo goat population, which diverged before the split between the Sarda and Corse breeds. Scenario 3 (Supplementary Figure S1, panel c) was designed to test a different origin of the two Montecristo populations with the insular one originating from the Corse and the mainland one from the Sarda breed. In Scenario 4 (Supplementary Figure S1, panel d) a possible derivation of the Montecristo population from the Corse breed was tested.

3. Results

3.1. Dataset Construction and Filtering

After quality control routines, a working dataset including 1251 animals and 43,252 SNPs was retained for subsequent analyses; while the number of markers was further reduced to 33,123 SNPs in the poly-MNT_I dataset containing only the markers that were polymorphic in the MNT_I population. The comparison between the results obtained from the main dataset and the poly-MNT_I datasets showed no substantial differences. Therefore, only the results obtained from the main working dataset were shown here.

3.2. Estimation of Genetic Diversity, Population Structure, and Migration Events

The Observed heterozygosity (H_O) corrected value calculated for the insular Montecristo goats (MNT_I) was the lowest recorded in the dataset (i.e., $H_O = 0.271$), and lower than that of the bezoar ($H_O = 0.281$) (Supplementary Table S1). The continental population MNT_M, instead, scored a higher value ($H_O = 0.324$), which is not far from the lower end of the range of H_O values scored for the domestic goat breeds (0.348 (VLS) $< H_O < 0.417$ (MLG). The corrected values of Expected Heterozygosity (H_E) were higher than H_O for both Montecristo populations (MNT_M: $H_E = 0.377$ vs. $H_O = 0.324$ and MNT_I: $H_E = 0.347$ vs. $H_O = 0.281$), with the insular population scoring the lowest value in the dataset. For the whole dataset the *p*-values associated to the inbreeding coefficient F_{IS} estimates were not statistically significant, except for the bezoar $F_{IS} = 0.233$ (significant at $p < 0.005$. Supplementary Table S1).

The first and second Principal Components of the unsupervised PCA together accounted for 5.57% of the total variance (Figure 2). On PC1 (3.99% of variance) the breeds were distributed following a

seamless north–south geographical pattern, with the northern Italian populations on the left side of the plot and the northern African ones on the right side (Figure 2). The MNT_M population was positioned close to two Spanish breeds, Murciano Granadina (MUG) and Malagueña (MLG).

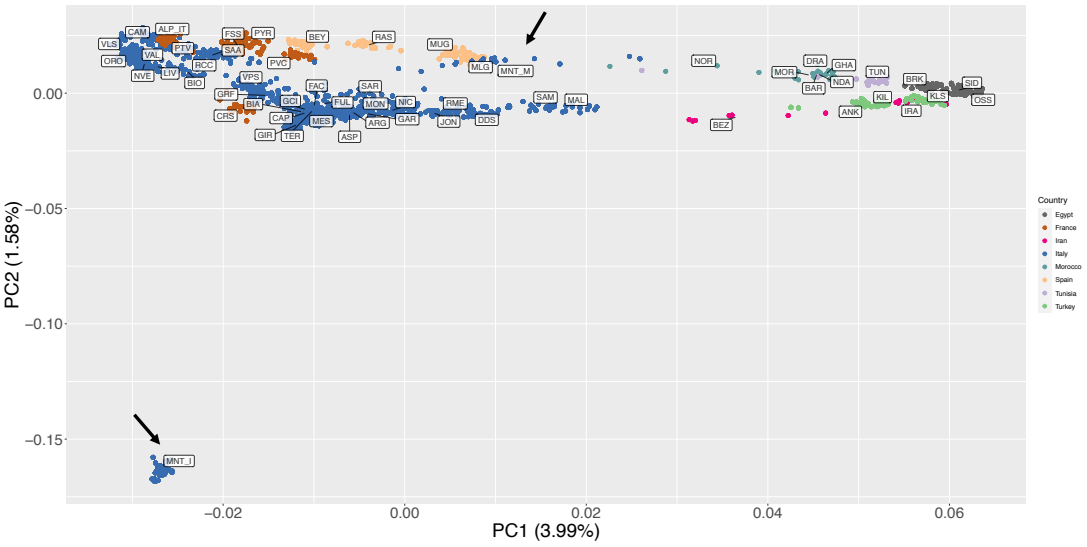


Figure 2. Unsupervised Principal Components Analysis (PC1 vs. PC2). The percentages of variance explained by each component are given into brackets. Arrows indicate position of Montecristo populations in the figure.

The second PC (1.58% of variance) highlighted an extreme outlier behaviour of the insular Montecristo goats MNT_I, which lies at the bottom left corner of the plot at a great distance from all other populations (Figure 2).

As detailed previously, the supervised Principal Component Analysis (Figure 3; Supplementary Figure S2) was then carried out to minimize the impact of outlier populations. In this analysis, SPC1 (4.1% of variance) clearly separated the European breeds from the African and south-western Asian ones. Together, SPC1 and SPC2 (5.5% of variance explained overall) showed a clear clustering of the breeds based on their country of origin and confirmed a north-to-south genetic pattern within Italy, as already highlighted in previous studies (Nicoloso et al., 2015; Stella et al., 2018)**.** The insular Montecristo goat population MNT_I clustered with breeds from central Italy (GRF, FAC, FUL), Sardinia (SAR), and Corse (CRS), while the scatter of points belonging to the ex situ MNT_M individuals was placed close to the origin of the axes and stretched between the SAR and Moroccan populations (Figure 3).

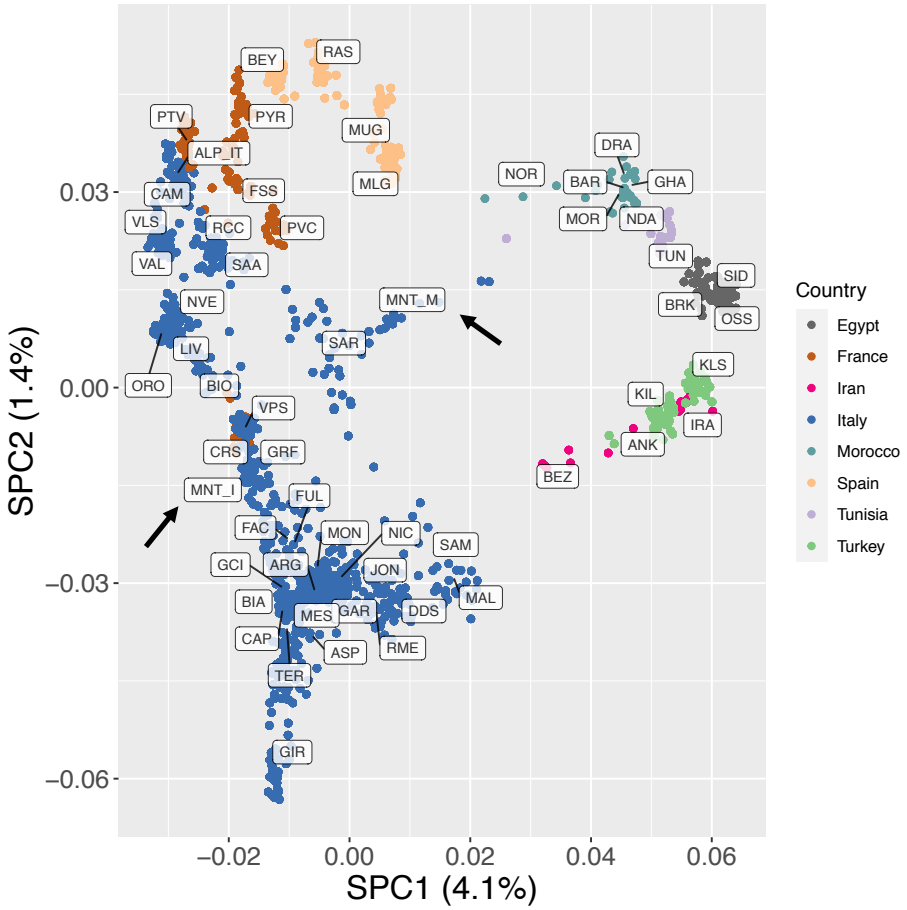


Figure 3. Supervised Principal Component Analysis (SPC1 vs. SPC2). The percentages of variance explained by each component are given into brackets.

The Neighbour-network based on Reynolds genetic distances (Figure 4) confirmed the geographical structuring of diversity already pointed out by PCA analyses. Both Montecristo goat populations were positioned on long branches, usually interpreted as evidence of prolonged isolation likely combined with genetic drift. Similar to SPCA results, the MNT_I population clustered with the Sarda, Corse, and Garfagnana breeds, while the MNT_M population was with some central Italian (FAC) and Spanish breeds (MLG, MUG, RAS, and BEY).

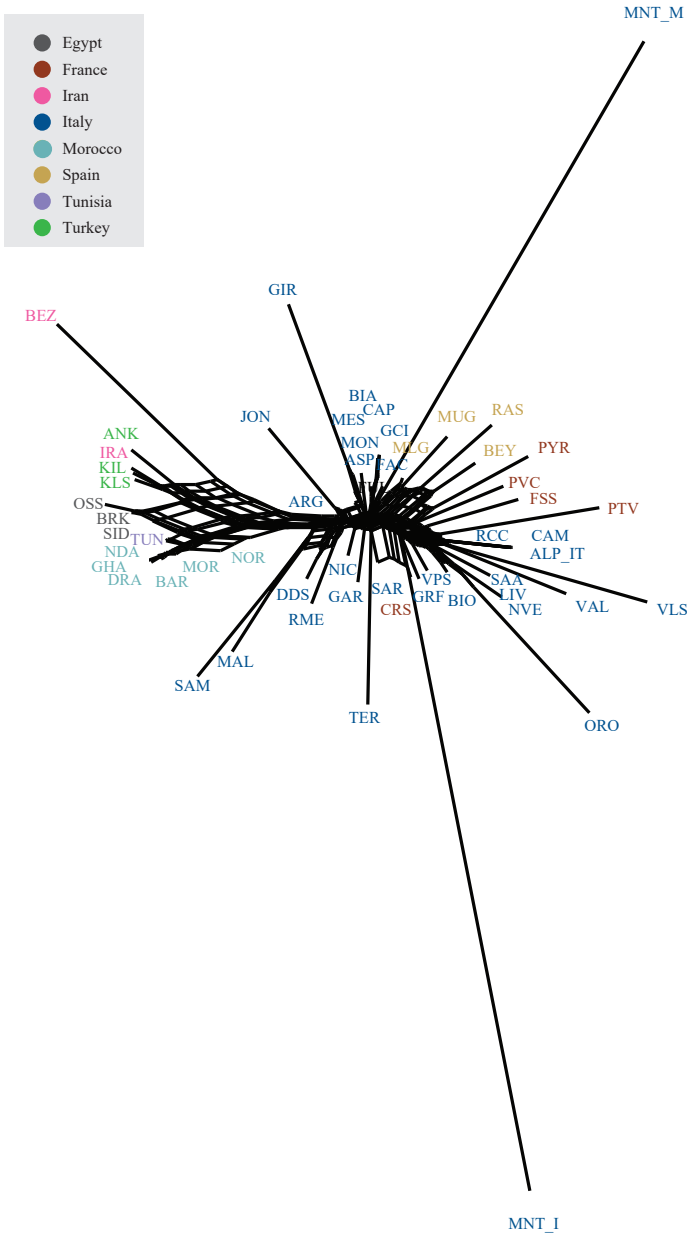


Figure 4. Neighbour-network reconstruction based on Reynolds' genetic distances between breeds.

Among the tested admixture scenarios (K = 2 to K = 45; Figure 5), the best fitting resolution was identified at K = 31 according to the CV error decay (Supplementary Figure S3).

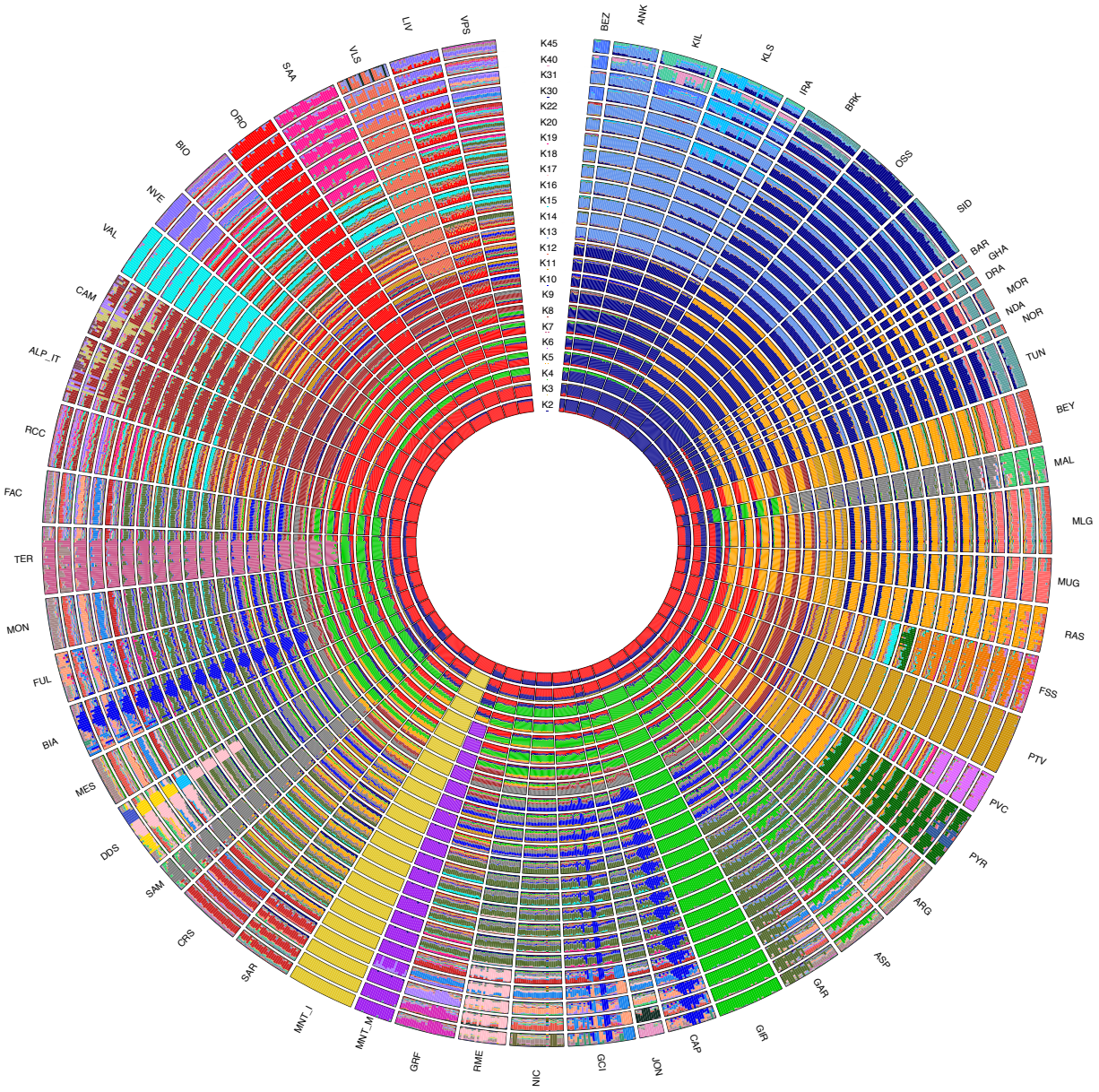


Figure 5. ADMIXTURE software analysis with putative ancestral population (K) computed from 2 to 45. The reconstruction at K = 31 had the smallest cross-validation error.

The model assuming two ancestral populations (K = 2) firstly separated the populations on a geographical basis, with the breeds from European assigned to a cluster different from that of African and south-western Asian ones. Already at this low K, the two Montecristo populations showed different behaviours: MNT_I was clearly assigned to the European cluster, while MNT_M displayed an admixed genomic background including both ancestral components. At K = 3 the insular MNT_I population clearly clustered apart from all other breeds. Conversely MNT_M (i) showed the

occurrence of a genomic component shared with the breeds from northern Africa, southern France and Spain at K = 5; (ii) was assigned to a separated cluster at K = 6; and (iii) showed a sub-structure at K = 30 (Figure 5).

Regarding the Treemix analysis, the Evanno statistic calculated over five iterations for *m* from 0 to 10 indicated *m*6 as the most likely number of gene flow events (Supplementary Figure S4). In the corresponding tree-based graph, most of the nodes were supported by high bootstrap values (Figure 6). Several migration edges connected the northern African breeds with each other and with those from Spain and southern France. MNT_I was positioned on the same branch as CRS, the latter being also connected to the MNT_M basal node by a migration edge.

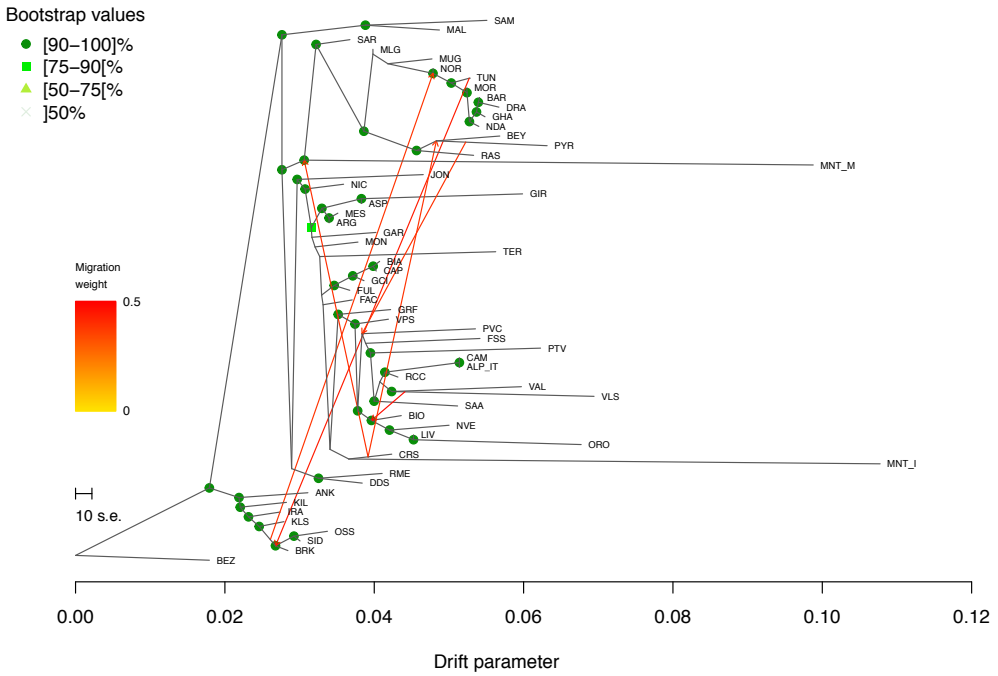


Figure 6. Treemix graph corresponding to m = 6. Robustness of the nodes was computed over 100 bootstrap replicates.

SNeP analysis pinpointed a marked but gradual reduction in effective population size over time for both MNT_I and MNT_M. The decline in *Ne* was consistent among the two Montecristo populations and less steep compared to the behaviour of the domestic SAR, CRS, and GRF (Supplementary Figure S5).

3.3. Runs of Homozygosity and Heterozygosity-Rich Regions

The average number of ROHs per animal was 239.7 and 66.35 for MNT_I and MNT_M, respectively, and the total number of ROHs identified was 1128 for the MNT_M and 7191 for the MNT_I population. ROHs of the two populations were classified into five length classes (0–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb, and >16 Mb. Supplementary Table S2): in the MNT_I population, the highest number of ROHs was

scored in the shortest length class (0–2 Mb), with the number of ROHs in the remaining classes gradually decreasing to 71 ROHs in the >16 Mb class. Conversely, the MNT_M population showed a more homogeneous distribution of ROHs across the five length classes, with the highest frequency in the 4–8 Mb class (Supplementary Table S2) and with an occurrence of ROHs >16 Mb class almost as frequent as in the other classes. The values of genomic inbreeding derived from the ROH coverage were 0.270 for the MNT_M and 0.312 for the MNT_I.

The scored ROHs were plotted chromosome-wise to obtain further insight on the contrasting behaviour of the two Montecristo goat populations, which showed a clear difference in the genomic distribution of the homozygous stretches (Supplementary File S1). The MNT_I individuals, in fact, showed the occurrence of generally short ROHs uniformly distributed along the chromosomes and interspersed with heterozygous stretches, and without major differences between individuals. The MNT_M animals conversely displayed a highly variable behaviour in terms of ROH occurrence, length, and position. Some individuals showed uninterrupted ROHs spanning over long chromosome tracts, sometimes as long as >80% of the chromosome (e.g., see chr4) and accompanied by very extended and completely heterozygous regions (e.g., chr6 and chr19) (Supplementary File S1). Moreover, the single-chromosome plots also pointed at the occurrence of regions consistently homozygous/heterozygous across individuals. This evidence was further investigated by identifying ROH regions including the top 0.1% of the most represented SNPs, which were found on chromosomes 1, 2, 3, and 11 for MNT_I, and on chromosome 16 for MNT_M (See Supplementary Table S3). In MNT_I shared ROH islands harboured the genes TNP1, SMARCAL1, MARCHF4, and ST6GALNAC5, while for the MNT_M population the highly shared ROH region on chromosome 16 included the genes 5S_rRNA, RRP15, KCTD3, BRINP3, GPATCH2, ESRRG, TGFB2, and SPATA17. Several unannotated sequences were also found in the ROHs of both populations.

For MNT_I common HRRs were found on chromosomes 6 and 13 (Supplementary Table S3). The HRR on chromosome 6 included the genes PPP3CA, EMCN, CNOT6L, SHLD1, and GPCPD1. HHRs shared by MNT_M individuals were located on chromosomes 1, 2, 8, 13, 14, 16, 18, 22, and 24 and spanned the genes OTOL1, LYPD6B, NFIB, BMP1, OPTN, BEND7, MCM10, PHYH, ASAP1, ADCY8, TATDN3, RPS6KC1, ANGEL2, NSL1, LOXHD1, KATNAL2, ST8SIA5, PIAS2, SKOR2, and SMAD4.

3.4. Approximate Bayesian Computation

According to the value of the Pearson correlation coefficient calculated between the main vs. reduced dataset PCA loadings (0.971), loss of information due to the reduction in the number of SNPs was excluded. Scenario 2 (divergence of the Montecristo goat population before the split between the Sarda and Corse breed) resulted in the most supported one (64.8% of the votes in model choice prediction). However, the simulations never fit the observed data (Supplementary Figure S7), which

suggests that none of the four models devised could thoroughly account for the complex demographic history of the Montecristo goats.

4. Discussion

In this study, we performed the first genome wide assessment of the genetic variation in Montecristo feral goats to shed light on their levels of polymorphism, population structure, and demographic history, and to evaluate their historic relationships with goat breeds across the Mediterranean area.

Previous investigations on the origin of Montecristo feral goats were based either on phenotypic traits, or on a limited number of allozymic/microsatellite loci or on partial fragments of the mtDNA control region (Doro et al., 2016; E Randi, Tosi, Toso, Lorenzini, & Fusco, 1990; Zanichelli et al., 2014). Here, the availability of new genotype data from two different *nuclei* of Montecristo goats, i.e., the free-ranging insular population and the captive-bred nucleus, also enabled us to compare their genomic make-up and to evaluate the effectiveness of the ex situ conservation project that has been carried out on the Italian mainland since the last decades of the 20th century. Overall, the results consistently highlighted marked molecular differences between the in situ and ex situ nuclei. The in situ population showed the lowest recorded value of observed heterozygosity in the dataset (Supplementary Table S1), a strong outlier behaviour in the unsupervised PCA (Figure 2), a clear independent clustering in admixture analysis (Figure 5), and a high value of genomic inbreeding ($F_{ROH} = 0.312$). These findings are in line with expectations, considering the geographical isolation and the repeated bottleneck events that have characterised the demographic history of the insular population, at least from the 18th century. They also agree with previous studies on insular populations, i.e., goats from the Mediterranean basin (Cardoso et al., 2018) and Soay sheep from the Scottish island of Hirta (McHugo et al., 2019; Selli et al., 2021) In all cases, insular populations had increased levels of inbreeding and a reduced variability compared to the nearby mainland populations. ROH analysis identified a high frequency of short ROHs that can be traced back to ancient bottlenecks and founder effects (Supplementary Table S2). Moreover, in the supervised PCA (Figure 3), Neighbour-network (Figure 4), and Treemix graph (Figure 6), the insular MNT_I population showed a genetic proximity to local breeds from Central Italy and the nearby islands of Corse and Sardinia, as expected based on the geographical location of Montecristo Island in the Tuscan archipelago, and the recorded inputs of domestic stocks during the 20th century.

Taken together, the results obtained from the insular population account for an ancestral genomic background shared with domestic breeds of the Thyrrenian sea area, subsequently moulded by a history of prolonged isolation, with several ancient bottlenecks likely accompanied by gene flow events from domestic stocks mainly from neighbouring areas.

Conversely, the population from the mainland displayed higher observed heterozygosity (MNT_M H_O = 0.325) and lower genomic inbreeding (F_{ROH} = 0.270) compared to the insular population (MNT_I H_O = 0.272; F_{ROH} = 0.312). The inbreeding level of the mainland population is remarkably higher than in domestic goat breeds from Italy (Salvatore Mastrangelo et al., 2021) and other areas of the world (Burren et al., 2016; Nandolo et al., 2019), while observed heterozygosity is only slightly lower than those observed in other breeds.

This is the result of the peculiar demographic history of the ex situ population. Four original ex situ nuclei were established, each one including only one male and four females, which inevitably contributed to a reduced amount of starting genetic variation. The nuclei were hosted at different farms and bred separately, without exchange of males and females for several generations (Ciampolini R., per. comm.). This led to a fast increase in inbreeding due to prolonged reproductive isolation, as testified by the extensive occurrence of long Runs of Homozygosity.

A few years ago, the only two remaining nuclei were finally merged into a single population (Ciampolini R., per. comm.). This recent merging is the likely cause of the higher-than-expected H_O in the mainland population, as an increase in observed heterozygosity is expected to occur following the removal of a reproductive barrier between two formerly separated populations, particularly when they are highly inbred and heavily affected by genetic drift due to small number of founders and small population size. This is further confirmed by the occurrence of long stretches of completely heterozygous regions (Supplementary File S1).

The peculiar genomic makeup of the ex situ population, characterised by a mosaic of alternated long stretches of completely homozygous/heterozygous regions in most of the individuals (Supplementary File S1), likely affected the behaviour of MNT_M in several analyses, making the identification of the actual relationship with the insular population and the other breeds difficult. The PCA/sPCA, population structure and neighbour-network results (Figures 2, 4, and 5), in fact, pointed to an affinity between the ex situ nucleus and the Italian Sarda, the Spanish and northern African breeds. This evidence is not consistent with the behaviour of MNT_I in the same analyses, nor with the known history of MNT_M after the establishment of the ex situ nuclei, even if the possible occurrence of unrecorded crossbreeding events with other domestic goats (e.g., Sarda or central Italian breeds) cannot be ruled out.

The more supported scenario obtained from the ABC analyses accounted for a first separation of the central Italian Garfagnana (GRF), followed by that of the Montecristo populations, and later by that of the Sarda (SAR) and Corse (CRS) breeds. This scenario would suggest a shared ancestry between the two MNT populations and their common, but more ancient, origin with the Sardo–Corse domestic stocks. However, these results are to be taken with caution, as simulated data did not fit the observed

ones (Supplementary Figure S6), possibly because of the MNT_M peculiar genetic makeup. Overall, current results are not conclusive on the relationships between the two Montecristo goat populations and further analyses are needed, possibly based on whole genome sequence and the analysis of haplotype blocks.

Conversely, some interesting genes were mapped in the most shared ROH/HRR regions. In MNT_I, the most shared ROHs harboured, among others, the genes TNP1 and SMARCAL1, respectively associated to spermiogenesis (Miyagawa et al., 2005) and genome integrity (Bansbach, Bétous, Lovejoy, Glick, & Cortez, 2009) (Supplementary Table S3). The highly shared HRR on chr6, instead, harboured the gene PPP3CA, previously associated to fecundity traits and litter size in small ruminants (E, Zhao, & Huang, 2019; Islam et al., 2020), while the gene GPCPD1 found in chr13 HRR was identified as playing a potential role in the metabolism of lipids and the lipoprotein pathway in sheep (Ji Yang et al., 2016). These results may suggest that both the fixation of specific variants in genes related to reproduction and integrity of the genome on the one hand, and the maintenance of consistent heterozygosity of genes involved in fecundity and energy metabolism on the other hand, may have played a role in the adaptation to the harsh environment of the island despite recurrent demographic fluctuations.

Conversely, molecular evidence points to a fixation of variants in genes involved in disease resistance, which may be alarming, and to the retention of heterozygosity in genes related to the development of internal organs and fertility in the case of the Montecristo goat population from the mainland.

5. Conclusions

In this study, we present the first comprehensive analysis of the demographic history and genome-wide molecular variation in Montecristo feral goats. According to our results, the insular population faced several demographic fluctuations over the centuries, which partially eroded the original genomic make-up in combination with gene flow events from other goat breeds. This population shares its ancestry with breeds from the surrounding areas of Sardinia, Corse, and Tuscany and does not carry signatures of recent inbreeding. Conversely, the molecular diversity of the mainland Montecristo goat population seems to have been severely impacted by the dynamics of the ex situ breeding that confounds the reconstruction of past population history. Overall, our findings represent a starting point for the implementation of marker-assisted monitoring and conservation plans to preserve the genomic heritage of the feral goats from Montecristo Island.

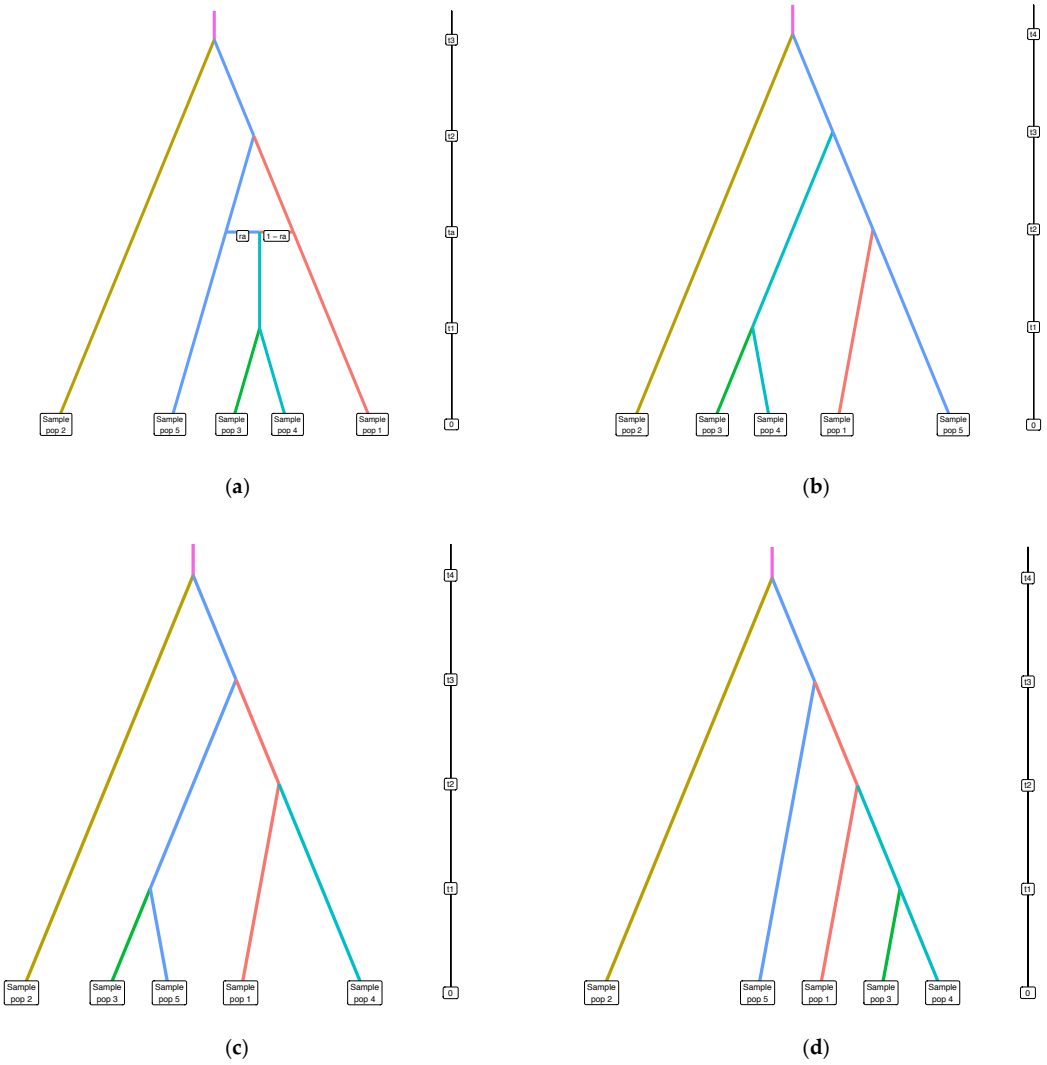
Supplementary information

Supplementary Table 1: Breed code, breed name, country of origin, source dataset (AdaptMap (Stella et al., 2018); IGC2 (Cortellari, Barbato, Talenti, Bionda, Randi, et al., 2021)), number of individuals pre-QC (Raw dataset)), number of individuals in the working dataset (WD), observed heterozygosity (H_O), expected heterozygosity (H_E) values corrected over the number of usable loci, inbreeding coefficient (F_{IS}). Statistical significance as follows: ***= P< 0.005.

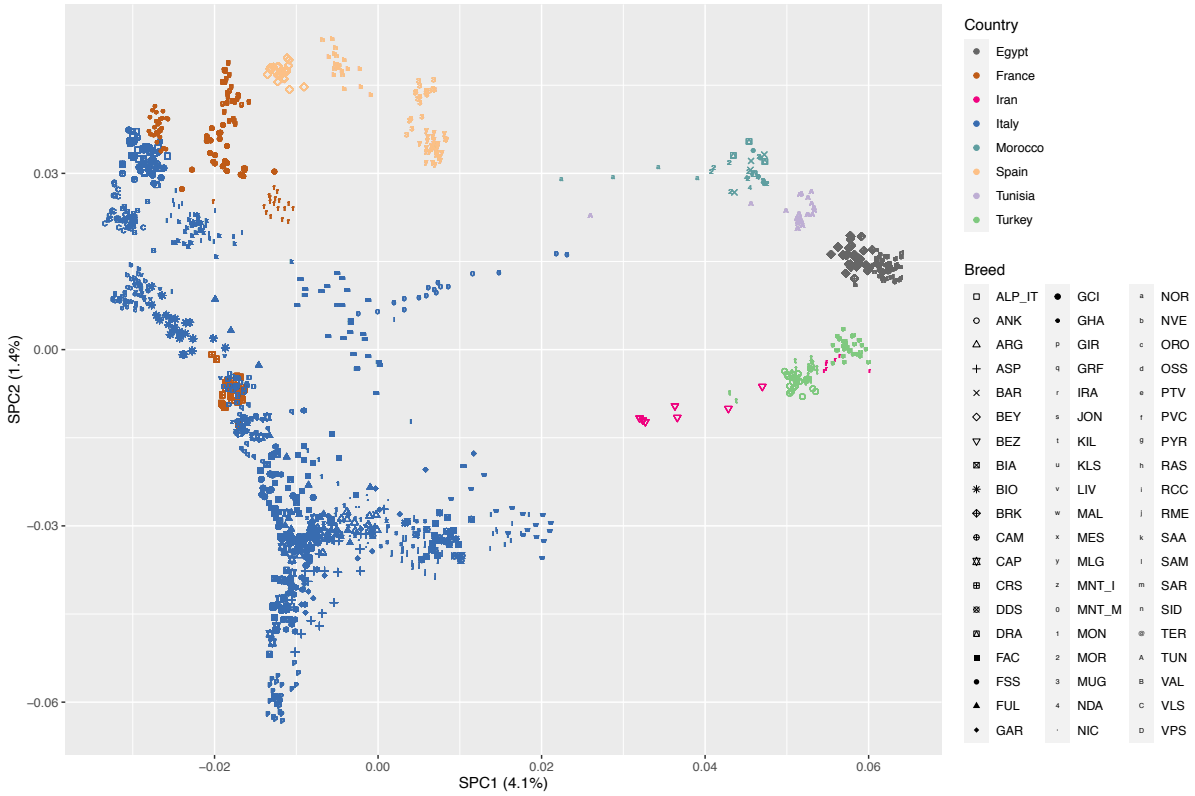
Breed acronm	Breed name	County of Origin	Source Dataset	N of animas Raw dataset	N of animals Workig Dataset	Ho	He	Fis
ALP_IT	Alpine	Italy	AdaptMap	160	30	0.399	0.405	0.025
ANK	Ankara	Turkey	AdaptMap	20	20	0.398	0.397	0.006
ARG	Argentata dell'Etna	Italy	IGC2	48	29	0.416	0.413	0.008
ASP	Capra dell'Aspromonte	Italy	IGC2	24	24	0.402	0.406	0.026
BAR	Barcha	Morocco	AdapMap	4	4	0.391	0.395	0.038
BEY	Bermeya	Spain	AdaptMap	24	24	0.404	0.397	-0.006
BEZ	Bezoar	Iran	AdaptMap	7	7	0.281	0.376	0.233***
BIA	Bianca Monticellana	Italy	IGC2	24	24	0.393	0.403	0.041
BIO	Bionda dell'Adamello	Italy	IGC2	24	24	0.404	0.405	0.017
BRK	Barki	Egypt	AdaptMap	153	28	0.397	0.403	0.020
CAM	Camosciata	Italy	IGC2	143	29	0.406	0.404	0.005
CAP	Capestrina	Italy	IGC2	24	22	0.404	0.403	0.016
CRS	Corsa	France	AdaptMap	30	29	0.401	0.404	0.020
DDS	Derivata di Siria	Italy	IGGC2	32	30	0.382	0.402	0.060
DRA	Draa	Morocco	AdaptMap	4	4	0.387	0.394	0.035
FAC	Facciuta della Valnerina	Italy	IGC2	24	23	0.395	0.410	0.055
FSS	Fosses	France	AdaptMap	26	26	0.394	0.394	0.013
FUL	Fulva del Lazio	Italy	IGC2	22	22	0.415	0.411	0.009
GAR	Garganica	Italy	IGC2	40	27	0.403	0.403	0.011
GCI	Grigia Ciociara	Italy	IGC2	43	30	0.405	0.409	0.023
GHA	Ghazalia	Morocco	AdaptMap	4	4	0.386	0.395	0.044
GIR	Girgentana	Italy	IGC2	59	30	0.357	0.371	0.025

GRF	Garfagnana	Italy	IGC2	28	27	0.399	0.405	0.028
IRA	Iranian goat	Iran	AdaptMap	9	9	0.381	0.399	0.075
JON	Jonica	Italy	IGC2	16	11	0.417	0.388	-0.083
KIL	KIl	Turkey	AdaptMap	25	25	0.404	0.401	0.003
KLS	Kilis	Turkey	AdaptMap	40	30	0.401	0.400	0.006
LIV	Capra di Livo	Italy	IGC2	24	22	0.407	0.402	0.003
MAL	Maltese	Italy	AdaptMap	16	16	0.365	0.379	0.047
MES	Messinese	Italy	IGC2	24	22	0.413	0.409	0.008
MLG	Malaguena	Spain	AdaptMap	42	30	0.417	0.412	0.000
MNT_M	Montecristo	Italy	IGC2	18	17	0.325	0.377	-0.086
MNT_I	Montecristo	Italy	IGC2	32	30	0.272	0.347	-0.007
MON	Capra di Montefalcone	Italy	IGC2	24	23	0.403	0.408	0.029
MOR	Moroccan goat	Morocco	AdaptMap	10	10	0.380	0.396	0.068
MUG	Murciano-granadina	Spain	AdaptMap	20	20	0.402	0.397	0.001
NDA	Noire de l'Atlas	Morocco	AdaptMap	4	4	0.384	0.396	0.055
NIC	Nicastrse	Italy	IGC2	24	24	0.396	0.408	0.045
NOR	Nord	Morocco	AdaptMap	4	4	0.394	0.398	0.059
NVE	Nera di Verzasca	Italy	IGC2	19	18	0.395	0.394	0.003
ORO	Orobica	Italy	IGC2	23	23	0.360	0.364	0.004
OSS	Oasis	Egypt	AdaptMap	72	30	0.373	0.395	0.056
PTV	Poitevine	France	AdaptMap	29	27	0.375	0.381	0.014
PVC	Provencale	France	AdaptMap	18	18	0.409	0.397	-0.015
PYR	Pyrenean	France	AdaptMap	27	27	0.374	0.388	0.041
RAS	Blanca de Rasquera	Spain	AdaptMap	20	20	0.378	0.388	0.036
RCC	Roccaverano	Italy	IGC2	28	28	0.401	0.416	0.052
RME	Rossa Mediterranea	Italy	IGC2	46	21	0.410	0.396	-0.038
SAA	Saanen	Italy	IGC2	44	30	0.416	0.410	-0.001
SAM	Maltese Sampled in Sardina	Italy	IGC2	15	14	0.375	0.371	-0.010
SAR	Sarda	Italy	IGC2	33	27	0.403	0.409	0.026

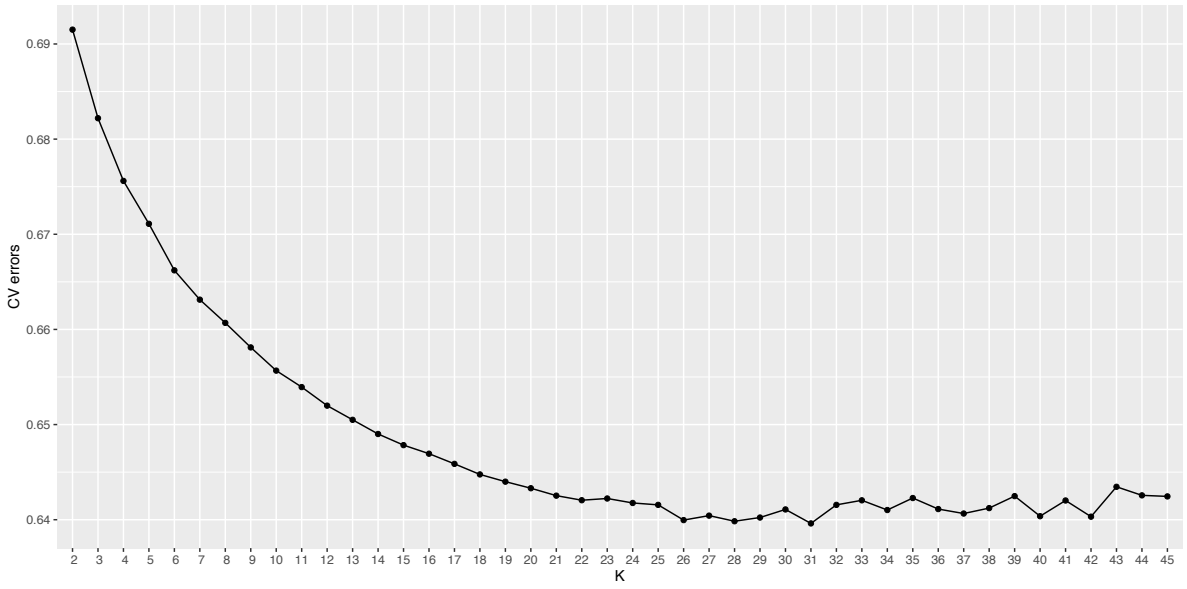
SID	Saidi	Egypt	AdaptMap	60	30	0.384	0.400	0.044
TER	Capra di Teramo	Italy	IGC2	43	30	0.385	0.377	-0.016
TUN	Tunisian	Tunisia	AdapMap	23	23	0.401	0.401	0.015
VAL	Valdostana	Italy	IGC2	24	24	0.371	0.384	0.039
VLS	Vallesana	Italy	IGC2	24	23	0.348	0.370	0.051
VPS	Capra della Valpassiria	Italy	IGC2	24	24	0.402	0.407	0.030
TOT.				1871	1251	0.399	0.396	0.025



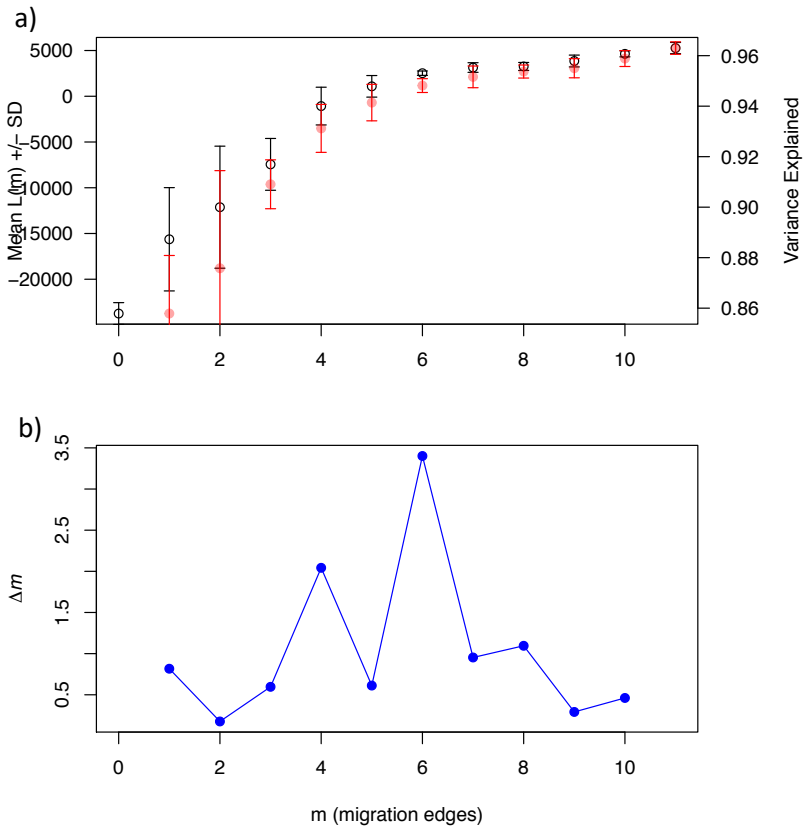
Supplementary Figure 1: Schematic representation of: (a) Scenario 1; (b) Scenario 2; (c) Scenario 3; and (d) Scenario 4 of ABC-RF anaysis. Populations are coded as follow: pop 1: CRS; pop 2: GRF; pop 3: MNT_M; pop 4: MNT_I; pop 5: SAR.



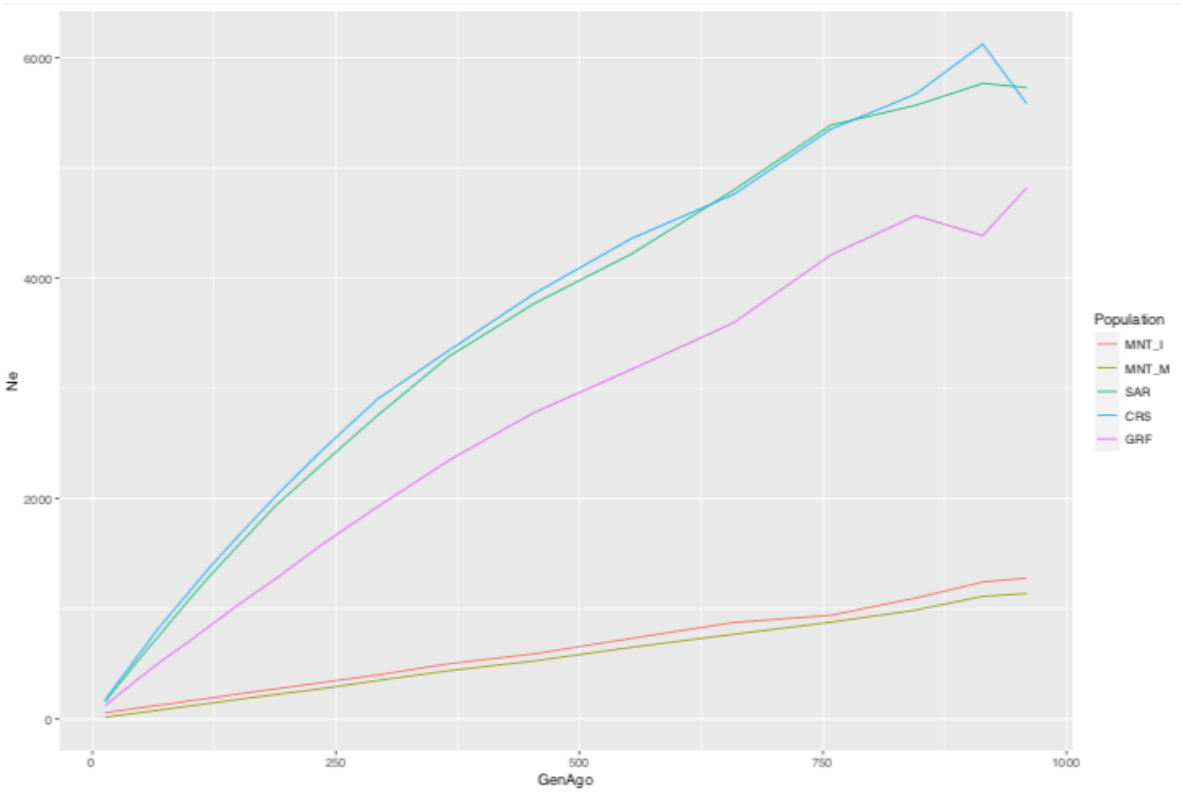
Supplementary Figure 2: Supervised Principal Component Analysis (SPC1 vs. SPC2). The percentages of variance explained by each component are given into brackets.



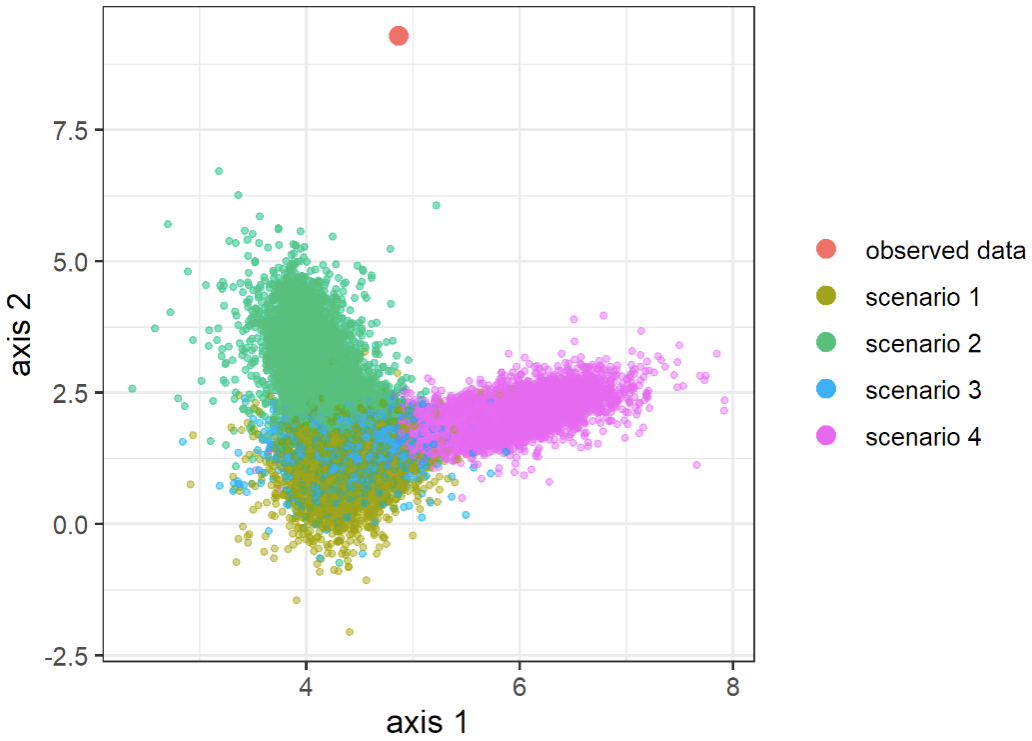
Supplementary Figure 3: Cross-Validation CV error values calculated for Admixture K values 2-45.



Supplementary Figure 4: (a) The mean and standard deviation (SD) calculated across 5 iterations for the composite likelihood L(m) (left axis, black circles) and proportion of variance explained (right axis, red circles). (b) The second-order rate of change (Dm) across values of m.



Supplementary Figure 5: Changes in estimated effective population size (Ne) along the last 1,000 generations.



Supplementary Figure 6: Projection of the datasets from the training set on the first two linear discriminant analysis (LDA) axes. The location of the observed dataset in the LDA projection is indicated by the red dot.

Supplementary Table 2: Summary of the number of ROH segments for different length categories scored in the two Montecristo goat nuclei.

Length category (Mb)	MNT_M	MNT_I
0-2	188	3209
2-4	252	2408
4-8	258	1155
8-16	217	348
>16	213	71
Tot.	1128	7191

Supplementary table 3: *Table summarising the 0.1% most shared regions of Runs of Homozygosity (ROH) and Heterozygosity Rich Regions (HRR) identified in the two Montecristo populations with the genes included.*

Run	Population	Chr	Start (bp)	End (bp)	genes
ROH	MNT_I	1	38552149	39474884	-
ROH	MNT_I	2	30528787	31463338	ENSCHIG00000006284, ENSCHIG00000006293, ENSCHIG00000006309, TNP1, SMARCAL1, MARCHF4
ROH	MNT_I	3	38397951	38432723	-
ROH	MNT_I	3	52713763	55922240	ST6GALNAC5
ROH	MNT_I	11	96495649	96534269	ENSCHIG00000004988
ROH	MNT_M	16	13996242	20982479	ENSCHIG00000000862, ENSCHIG00000000869, ENSCHIG00000000876,

					ENSCHIG00000000880, ENSCHIG00000000882, 5S_rRNA, RRP15, KCTD3, BRINP3, GPATCH2, ESRRG, ENSCHIG00000021102, TGFB2, SPATA17
HRR	MNT_I	6	23704450	24651938	PPP3CA, EMCN
HRR	MNT_I	6	92891918	93001162	CNOT6L
HRR	MNT_I	13	47213471	47526560	SHLD1, GPCPD1
HRR	MNT_M	1	105425339	105469270	OTOL1
HRR	MNT_M	2	89186937	89310361	LYPD6B
HRR	MNT_M	2	89822439	89891526	ENSCHIG00000000666
HRR	MNT_M	8	29737003	29990995	NFIB
HRR	MNT_M	8	69147795	69192956	BMP1
HRR	MNT_M	13	27167011	27654015	ENSCHIG00000000499, OPTN, BEND7, MCM10, PHYH
HRR	MNT_M	14	71387238	72223053	ASAP1, ADCY8
HRR	MNT_M	16	69226610	69455422	TATDN3, RPS6KC1, ANGEL2, NSL1
HRR	MNT_M	18	43777759	43932769	ENSCHIG00000004436, ENSCHIG00000004439
HRR	MNT_M	22	44176757	44243832	ENSCHIG00000015055
HRR	MNT_M	24	46431005	46854646	LOXHD1, KATNAL2, ST8SIA5, PIAS2
HRR	MNT_M	24	47100548	47587014	ENSCHIG00000002761, ENSCHIG00000002755, SKOR2
HRR	MNT_M	24	50751818	50783212	SMAD4

Supplementary file 1: ROHs of the two Montecristo populations plotted chromosome-wise.

Available at: **https://doi.org/10.3390/genes13020213**

Chapter 3: Runs of homozygosity in the Italian goat breeds: impact of management practices in low-input systems

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, Agro- energia, 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 Bio- DNA 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 Messina, Messina, Italy.

9 Dipartimento Agricoltura, Ambiente e Alimenti, Università degli Studi del Molise, Campobasso, Italy.

10 The Roslin Institute, University of Edinburgh, Edinburgh, UK

Abstract

Background: 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.

Results: 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.

Conclusions: 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, Ginja, Hoffmann, Joost, Wengel, 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 (T. H. E. Meuwissen, Sonesson, Gebregiwergis, & Woolliams, 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 (Senczuk et al., 2020; Talenti et al., 2018). 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, Barbato, Talenti, Bionda, Randi, 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, Di Cerbo, Zanzani, & Stradiotto, 2010; Sandrucci, Bava, Tamburini, Gislon, & Zucali, 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, Corti, & Cicogna, 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, Barbato, Talenti, Bionda, Randi, 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 (Paulina G. Eusebi, Martinez, & Cortes, 2020). 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, Meyermans, Janssens, & Buys, 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, Lu, Kayser, & Krawczak, 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, Elsen, & Legarra, 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, Barbato, Talenti, Bionda, Randi, 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 (Christopher C 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 ≥ 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. 1, n.d.). 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 ($std\left(\frac{ROH\ lengh}{CHR\ lenght}\right)$) , and (ii) a standardized ratio between the number of ROH on each chromosome of each individual and the length of the chromosome ($std\left(\frac{ROH\ number}{CHR\ lenght}\right)$); 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 1. The geographic distributions, the MDS plot, and the phylogeny tree of the studied goat populations are reported in Additional file 1: Fig. S1.

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 1).

Table 1: Composition of the dataset and mean ROH and genetic parameters for each of the studied Italian goat breeds.

Breed ID	Breed name	Raw dataset	Quality checked	ROH number	ROH total length	ROH length	FROH	He	Ho	Fis
ALP	Camosciata delle Alpi	143	117	31.5	169.1	5.0	0.07	0.41	0.40	0.02
ARG	Argentata dell'Etna	48	46	9.1	31.2	2.9	0.01	0.41	0.41	0.00
ASP	Capra dell'Aspromonte	24	24	22.6	113.1	4.2	0.05	0.40	0.40	0.00
BIA	Bianca Monticellana	24	23	30.8	158.7	4.0	0.06	0.40	0.39	0.01
BIO	Bionda dell'Adamello	24	24	22.4	93.5	3.5	0.04	0.40	0.40	0.00
CAP	Capestrina	24	22	23.5	117.5	3.5	0.05	0.40	0.40	-0.01
DDS	Derivata di Siria	32	25	37.8	232.2	5.5	0.09	0.40	0.38	0.03
FAC	Facciuta della Valnerina	24	24	19.5	167.9	5.4	0.07	0.41	0.39	0.03
FUL	Fulva del Lazio	22	20	11.8	63.0	4.1	0.03	0.41	0.41	-0.02
GAR	Garganica	40	37	21.6	118.8	4.4	0.05	0.40	0.40	0.00
GCI	Grigia Ciociara	43	39	17.5	116.4	4.8	0.05	0.41	0.40	0.02
GIR	Girgentana	59	56	74.5	322.7	4.1	0.13	0.36	0.36	0.01
GRF	Garfagnana	28	25	33.1	147.5	4.1	0.06	0.40	0.40	0.01
JON	Jonica	16	15	24.3	88.4	3.4	0.04	0.37	0.41	-0.10
LIV	Capra di Livo-lariana	24	22	20.6	56.8	2.6	0.02	0.40	0.40	-0.02
MAL	Maltese	16	16	70.6	347.5	4.4	0.14	0.37	0.36	0.01
MES	Messinese	24	23	9.3	36.8	3.5	0.01	0.40	0.41	-0.01
MON	Capra di Montefalcone	24	23	16.7	136.4	5.2	0.06	0.40	0.40	0.01
NIC	Nicastrese	24	24	24.2	170.3	5.3	0.07	0.40	0.39	0.02
NVE	Nera di Verzasca	19	19	32.6	136.2	3.7	0.06	0.38	0.39	-0.02
ORO	Orobica	23	23	92.8	293.6	3.1	0.12	0.35	0.36	-0.02
RCC	Roccaverano	28	28	20.7	169.6	6.1	0.07	0.41	0.40	0.03
RME	Rossa Mediterranea	46	40	24.2	93.9	3.4	0.04	0.39	0.41	-0.05
SAA	Saanen	44	44	25.8	115.9	4.2	0.05	0.41	0.41	-0.01
SAM	Maltese sampled in Sardinia	15	15	57.1	294.4	4.8	0.12	0.36	0.37	-0.03
SAR	Sarda	33	33	18.1	77.8	3.4	0.03	0.41	0.40	0.01
TER	Capra di Teramo	43	30	25.2	234.1	7.3	0.10	0.39	0.38	0.01
VAL	Valdostana	24	24	50.9	272.0	4.7	0.11	0.37	0.37	0.02
VLS	Vallesana	24	17	76.6	364.6	4.7	0.15	0.36	0.35	0.02
VPS	Capra della Val Passiria	24	24	23.1	118.3	4.4	0.05	0.40	0.40	0.01

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 LivoLariana. However, the distribution of the individual F_{ROH} within each population varied among breeds (see the boxplot in Fig. 1): 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} highest values were found for Roccaverano and Derivata di Siria, and the lowest for the Jonica and Rossa Mediterranea breeds (Table 1).

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.

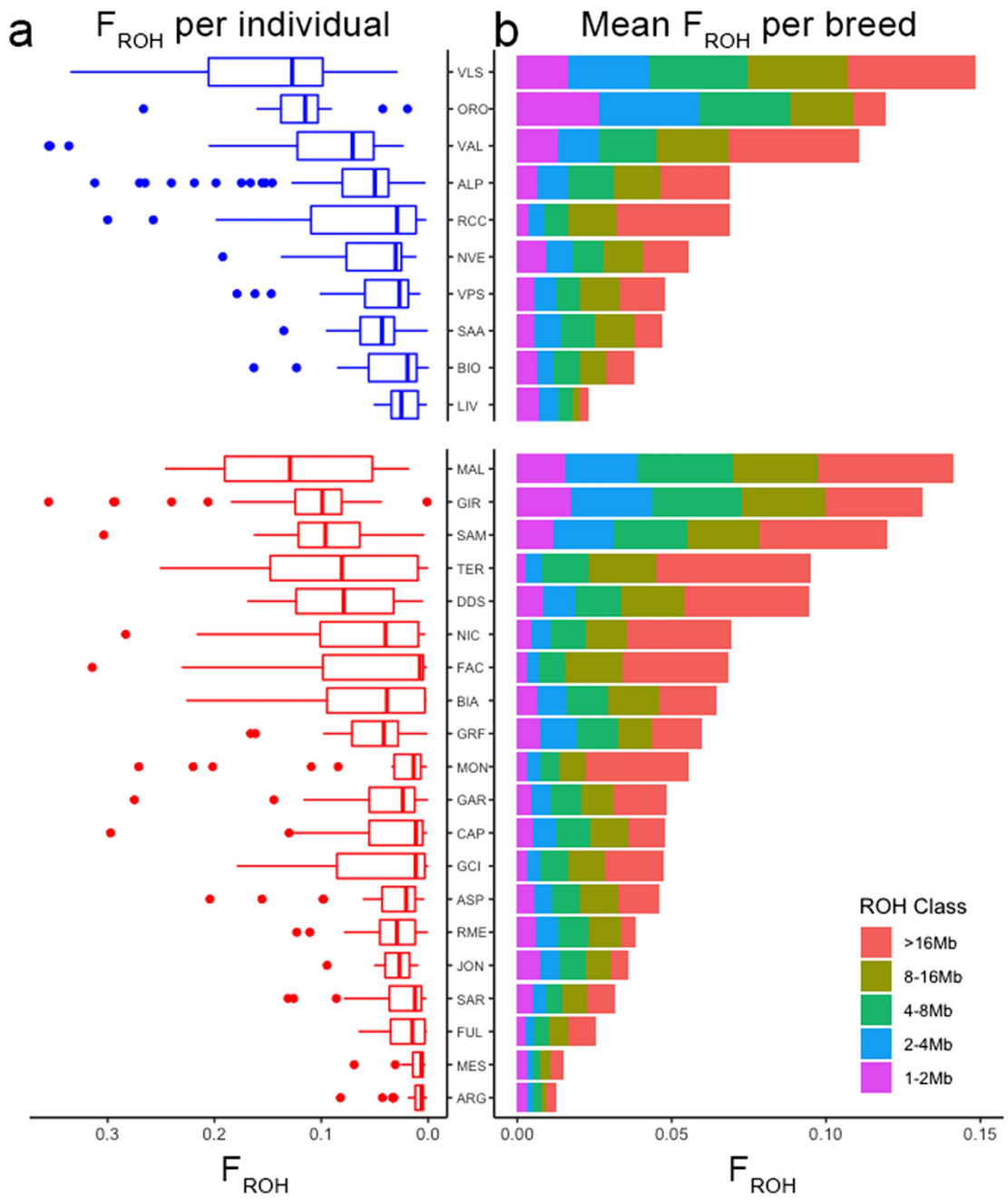


Figure 1: Distribution of F_{ROH} values per individual and mean values per breed. Boxplot (**a**) of the single individual F_{ROH} distribution in each population (Northern breeds: blue boxplots, Central-southern breeds: red boxplots) with matching barplot (**b**) of the mean F_{ROH} values (each color representative of the different ROH length 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. 1).

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 < 0.000). ROH were longer for the CSD than the NRD group (LSmean ± SE = 0.06 ± 0.02 vs − 0.02 ± 0.02). Figure 2a shows the mean standardized ratio between the length of ROH and the length of the corresponding chromosome $std\left(\frac{ROH\ lengh}{CHR\ lenght}\right)$ for each chromosome in the two (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 (Copenhaver, Browne, & Preuss, 1998; D. Kaback, 1996; D. B. Kaback, Barber, Mahon, Lamb, & You, 1999); this has also been reported in goats (Salvatore 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 2b shows the mean standardized ratio between the number of ROH per chromosome and the corresponding chromosome length $std\left(\frac{ROH\ number}{CHR\ length}\right)$ 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.

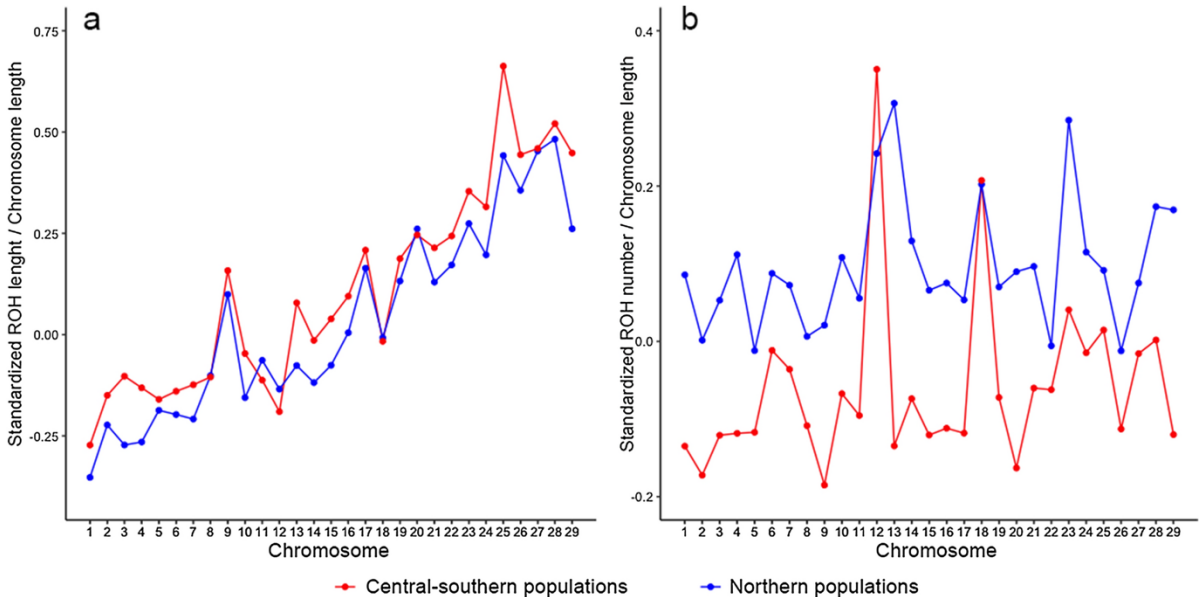


Figure 2: Comparison of mean standardized length and number of ROH in the two groups of Italian goat populations. Graphic representation of the mean standardized length (**a**) and number (**b**) of ROH divided by the corresponding chromosome length in the two groups of Italian goat populations

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 ten 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.

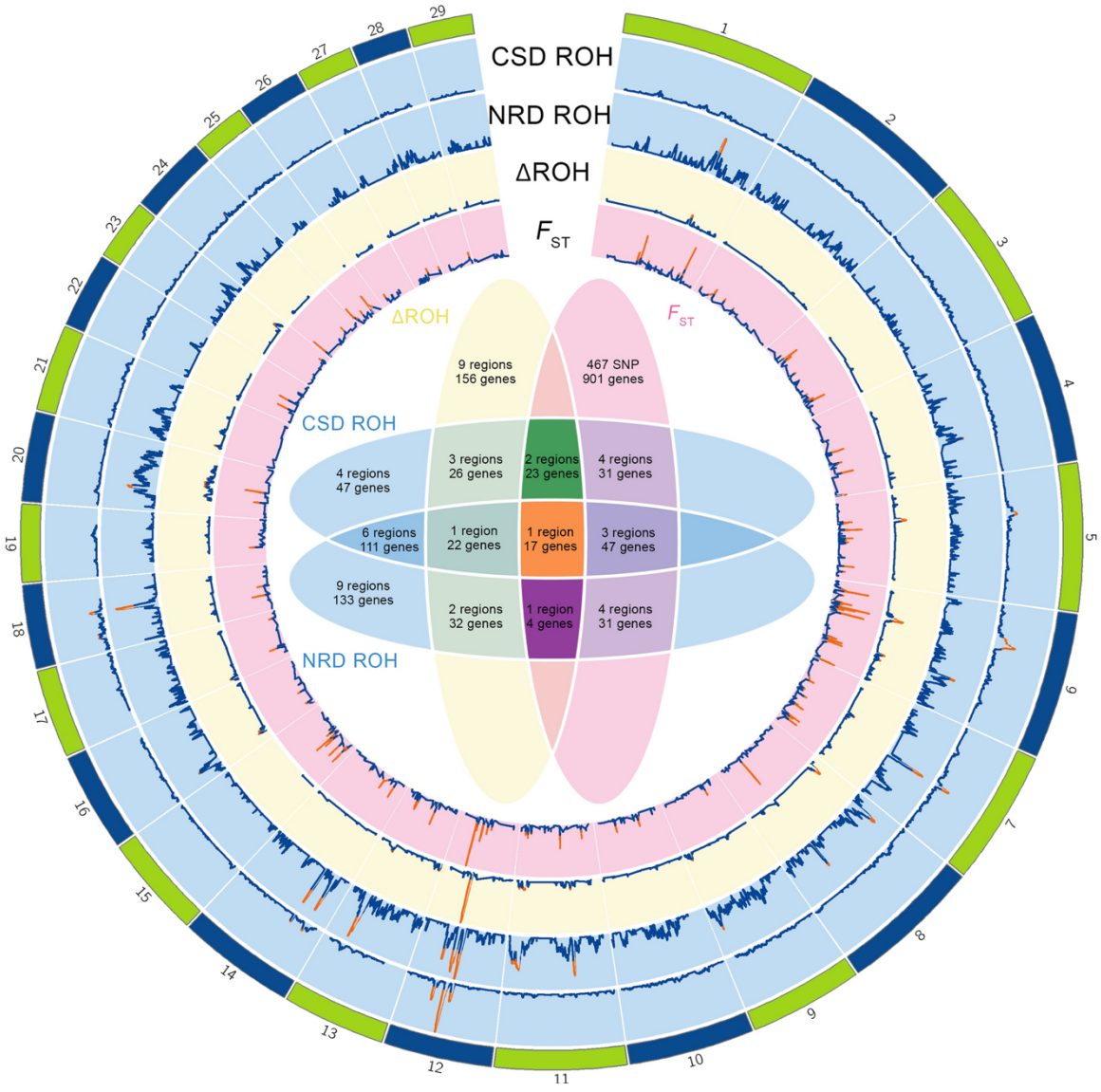


Figure 3: Circos plot of the analysis of signatures of selection with a Venn diagram of the results. Circos plot of the signatures of selection in the two groups of Italian goat populations (external blue tracks), of the ΔROH (middle yellow track) and averaged F_{ST} (inner red track). The Venn diagram shows the number of regions and genes shared across methods. *CSD* Central-southern population group, *NRD* Northern population group

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). 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 and see Additional file 2: Fig. S2). The complete list of the identified genes is in Additional file 3: Table S1.

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 (Salvatore 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 et al., 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 (Francesca Bertolini, Cardoso, et al., 2018). 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, *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 (Salvatore 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* (E. S. 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 information

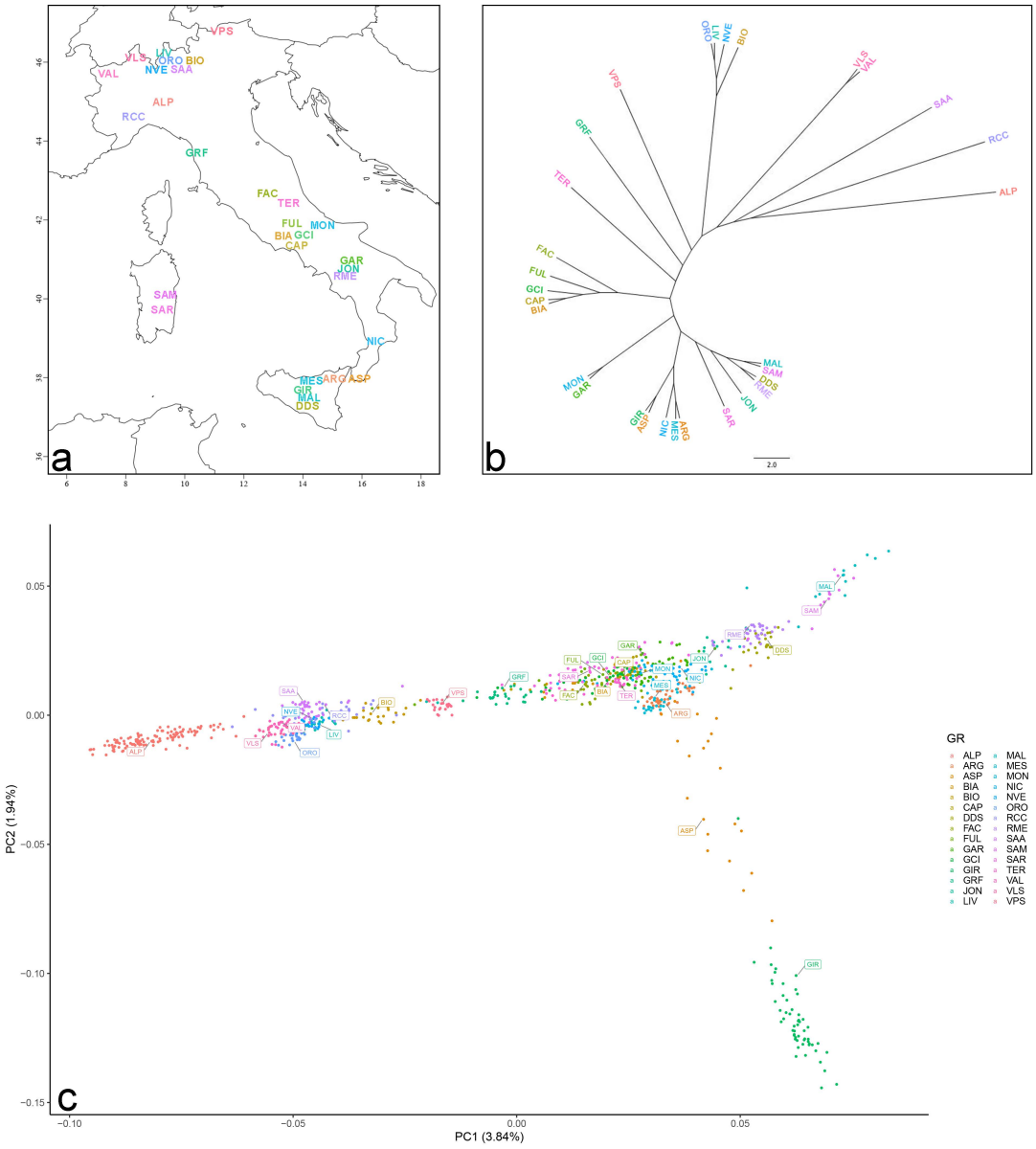


Figure S1: Geographic distribution (a), phylogeny tree (b), and multidimensional scaling analysis (c) of all the Italian goat breeds included in the study.

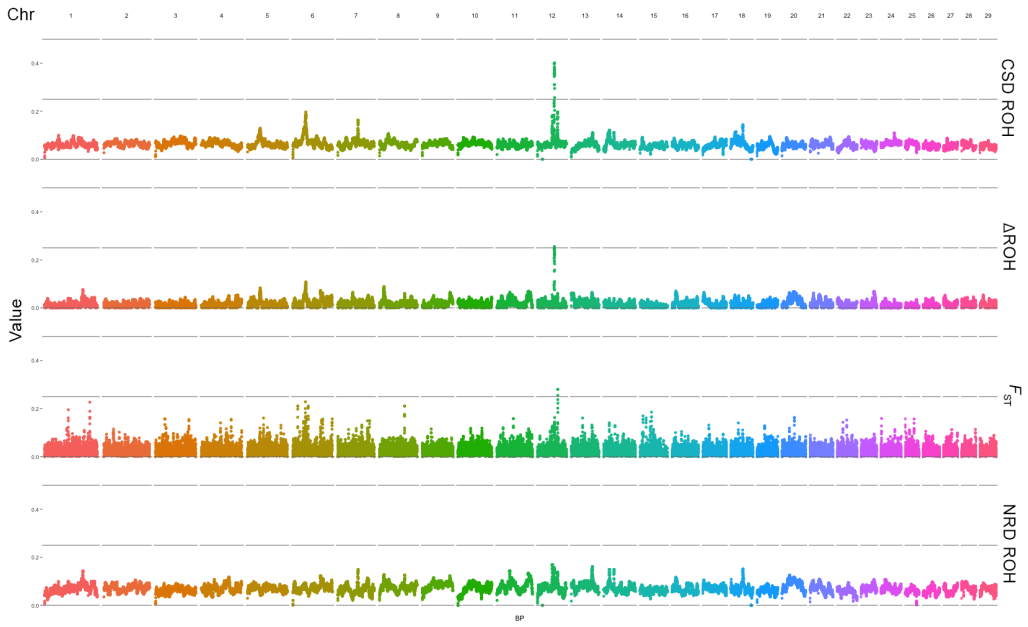


Figure S2: Manhattan plots representing the signals of signatures of selection in the two population groups (CSD and NRD), of the ΔROH, and averaged F_{ST}. CSD = Central-southern population group; NRD = Northern population group.

Table S1: List of the genes identified by the analyses of the signatures of selection: top 1% homozygosity score in CSD and NRD groups, ΔROH, and averaged F_{ST}. CSD = Central-southern population group; NRD = Northern population group.

Available at: https://doi.org/10.1186/s12711-021-00685-4

Chapter 4: The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock

Matilde Maria Passamonti [1,†], Elisa Somenzi [1,†], Mario Barbato [1], Giovanni Chillemi [2], Licia Colli [1,3], Stéphane Joost [4], Marco Milanesi [2], Riccardo Negrini [1], Monia Santini [5], Elia Vajana [4], John Lewis Williams [1] and Paolo Ajmone-Marsan [1,6,*]

[1] Department of Animal Science, Food and Nutrition—DIANA, Università Cattolica del Sacro Cuore, Via Emilia Parmense, 84, 29122 Piacenza, Italy; matildemaria.passamonti1@unicatt.it (M.M.P.); elisa.somenzi@unicatt.it (E.S.); mario.barbato@unicatt.it (M.B.); licia.colli@unicatt.it (L.C.); riccardo.negrini@unicatt.it (R.N.); johnlewis.williams@unicatt.it (J.L.W.)

[2] Department for Innovation in Biological, Agro-Food and Forest Systems—DIBAF Università Della Tuscia Via S. Camillo de Lellis snc, 01100 Viterbo, Italy; gchillemi@unitus.it (G.C.); marco.milanesi@unitus.it (M.M.)

[3] Research Center on Biodiversity and Ancient DNA–BioDNA, Università Cattolica del Sacro Cuore, Via Emilia Parmense, 84, 29122 Piacenza, Italy

[4] Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland; stephane.joost@epfl.ch (S.J.); elia.vajana@epfl.ch (E.V.)

[5] Impacts on Agriculture, Forests and Ecosystem Services (IAFES) Division, Fondazione Centro Euro-Mediterraneo Sui Cambiamenti Climatici (CMCC), Viale Trieste 127, 01100 Viterbo, Italy; monia.santini@cmcc.it

[6] Nutrigenomics and Proteomics Research Center–PRONUTRIGEN, Università Cattolica del Sacro Cuore, Via Emilia Parmense, 84, 29122 Piacenza, Italy

***** Correspondence: paolo.ajmone@unicatt.it

† Equal contribution.

Simple Summary: After domestication in specific regions, livestock followed human migrations and colonized the whole world. During this population expansion, human and natural selection, together with demographic events, molded the livestock genome leading to local breeds and populations able to produce milk, meat, wool and tractive power in many different agro-climatic conditions. The climate is changing, with temperatures and the frequency of extreme climatic events increasing, which affects livestock welfare and production efficiency, particularly of the highly productive breeds. Genomics is now able to explore the DNA of local breeds adapted to extreme environments in search of genes carrying signatures of selection for adaptation. This review summarizes methods used to accomplish this task, giving examples of results achieved and perspectives for future breeding.

Abstract**:** Livestock radiated out from domestication centres to most regions of the world, gradually adapting to diverse environments, from very hot to sub-zero temperatures and from wet and humid conditions to deserts. The climate is changing; generally global temperature is increasing, although there are also more extreme cold periods, storms, and higher solar radiation. These changes impact livestock welfare and productivity. This review describes advances in the methodology for studying livestock genomes and the impact of the environment on animal production, giving examples of discoveries made. Sequencing livestock genomes has facilitated genome-wide association studies to localize genes controlling many traits, and population genetics has identified genomic regions under

selection or introgressed from one breed into another to improve production or facilitate adaptation. Landscape genomics, which combines global positioning and genomics, has identified genomic features that enable animals to adapt to local environments. Combining the advances in genomics and methods for predicting changes in climate is generating an explosion of data which calls for innovations in the way big data sets are treated. Artificial intelligence and machine learning are now being used to study the interactions between the genome and the environment to identify historic effects on the genome and to model future scenarios.

1. Introduction

Climate change is generally causing an increase in global temperatures (see Box 1). The most recent estimates (Masson-Delmotte et al., 2018) suggest that a 1.5 °C warming compared to the 1850–1900 baseline will be reached in the second half of the current decade, but, in addition, there are longer cold periods and increased levels of solar radiation (Ames & Insley, 1975; Belasco, Cheng, & Schroeder, 2015; Lees et al., 2019; Pezzopane et al., 2020; Toghiani, Hay, Roberts, & Rekaya, 2020) (Figure 1).

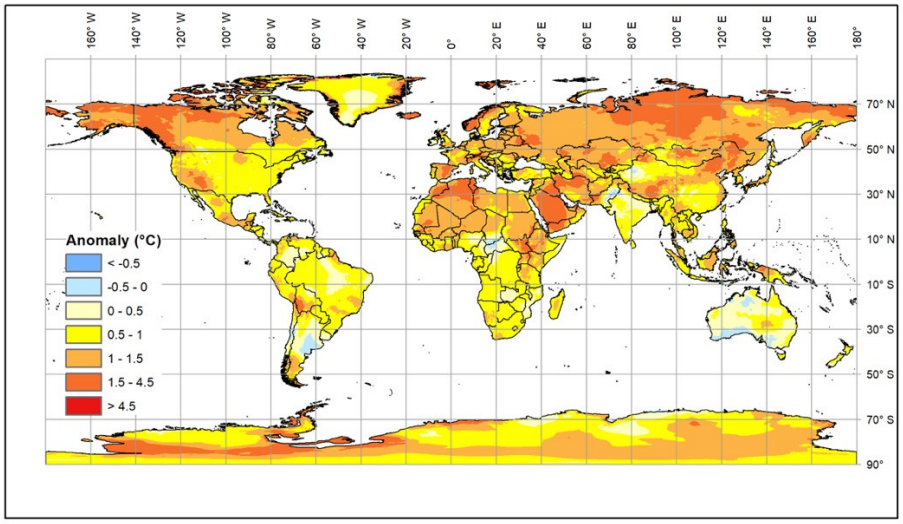


Figure 1. Map of annual temperature changes in the period 1992–2020 compared to 1950–1978, created using ERA5 climate Reanalysis tools. The areas showing warming are in yellow-red and those showing cooling are in blue.

These changes affect both extensive and intensive farming systems (Rust, Holman, Bloomfield, Cuthbert, & Corstanje, 2019; Thornton, van de Steeg, Notenbaert, & Herrero, 2009). The impact of environmental changes on animals affects their health, growth, and fertility as well as the diseases to which they are exposed. In addition, availability and types of feed may change because of the impact of the climate on the production and quality of grains, pasture and forage crops (Rotter & Van de Geijn, 1999; Wheeler & Reynolds, 2013), which will affect nutrition as well as animal health and metabolism (Ortiz-Bobea, Knippenberg, & Chambers, 2018). Livestock can adapt to gradual changes in environmental temperature. However, rapid changes or extended periods of extreme conditions

reduce their welfare and productivity and are potentially life threatening. Therefore, the current rapid rise in global temperature is increasingly exposing livestock to stress in many countries. Some local breeds that have been kept in areas with adverse conditions, such as high temperature and humidity or drought, have become adapted over many generations; these breeds are an invaluable resource for research and breeding. It is urgent to understand the biological mechanisms underlying their adaptability, and, in particular, to identify genomic regions and genes that control such mechanisms in order to facilitate the rapid selection of livestock resilient to climate change. This review focuses on ruminants and on the current state of knowledge on genetics controlling adaptation.

2. Impacts of Climate Change on Livestock

With increasing global temperatures, more productive livestock are at greater risk (see Box 2), because they have higher feed intake and feed consumption, which is directly related to animal heat production (J. B. Gaughan & Cawdell-Smith, 2017). Animals eat less to counteract high temperatures, and nutrients are prioritized to support maintenance rather than production and reproduction. In the central U.S., for example, severe losses of beef cattle kept in feedlots have been reported because of heat waves in summer and extreme snowstorms and wind in winter (Hahn et al., 2001). Climate related economic losses as a result of animal death and reduced performance have been seen (T. Mader, Davis, Gaughan, & Brown-brandl, 2003). Cattle, sheep, pigs and chickens reduce their feed intake by 3–5% for each unit increase in temperature above 30 °C (National Research Council, 1981). Reproduction is particularly affected. Hahn (Hahn, 1996) reported that conception rates in dairy cows are reduced by 4.6% per unit change above 70 in the temperature humidity index (THI) (Thom, 1959). For beef cattle kept in range or pasture management systems, a decrease in pregnancy rates of 3.2% and 3.5% was observed for each unit increase in average THI above 70 and an increase in average temperature above 23.4 °C, respectively. Among environmental variables, temperature has the greatest effect on cow pregnancy rates (Amundson, Mader, Rasby, & Hu, 2006).

Climate change further includes altered rainfall patterns that, combined with geographical factors such as soil type, affect crop production (Chapman, Chakraborty, Dreccer, & Howden, 2012; Lean & Rind, 2009; Shukla et al., 2019). Drought reduces biomass (Calleja-Cabrera, Boter, Oñate-Sánchez, & Pernas, 2020), increases lignin accumulation in plant tissues, and reduces proteins, resulting in less digestible forages (Tubiello, Soussana, & Howden, 2007) and insufficient energy to meet livestock requirements (Hidosa & Guyo, 2017; Morton, 2007). Increased occurrence of prolonged drought is therefore of great concern to pasture-based livestock systems (Tubiello et al., 2007), especially those in environments which cannot support arable production (O'Mara, 2012).

Climate change influences the distribution of animal pathogen vectors and parasite range (Baylis & Risley, 2013) which, together with the decreased immune response of animals under stress (triggered by cortisol), exposes livestock to higher risks of disease. Early springs, warmer winters and changes in rainfall distribution affect the seasons in which pathogens, parasites and vectors are present, potentially increasing proliferation and survival of these organisms. Bluetongue recently spread northward from Africa to Europe (Baylis & Githeko, 2006) as a consequence of climate-driven ecosystem changes and the associated expansion of the geographic range of the insect *Culicoides imicola*, the vector of the virus (Wilson & Mellor, 2009). Other vectors such as the tick *Rhipicephalus appendiculatus,* which is the host for the protozoan pathogen *Theileria parva*, are predicted to shift their geographic range due to climate change, moving southward from central sub-Saharan Africa towards southern Africa (Olwoch, Reyers, Engelbrecht, & Erasmus, 2008). Higher temperatures in Europe have increased parasite burdens such as helminths, with a shift from species traditionally found in temperate zones such as *Ostertagia ostertagi* to tropically adapted species, particularly *Haemonchus contortus* (Fox, Glenn, Davidson, White, & Hutchings, 2015; Kenyon, Sargison, Skuce, & Jackson, 2009). In addition to temperature, increased rainfall and humidity have affected the distribution of parasites. Leptospirosis in humans has been linked to transmission from livestock, with many outbreaks reported following extreme weather events around the world (Lau, Smythe, Craig, & Weinstein, 2010).

Box 1. Climate Data and Tools.

High resolution meteorological data are used to evaluate climate trends and variability and to predict the frequency of extreme events. Where meteorological data are not available, advanced climate modelling produces "Climate Reanalysis" datasets for a comprehensive description of the climate in three-dimensional grids. "Climate Reanalysis" has become an essential tool for modelling meteorological data to provide services to sectors dependent on climate assessments, forecasts and projections, including ecosystem management, agriculture, and livestock farming (Buontempo et al., 2020; Escarcha, Lassa, & Zander, 2018)*. Climate modelling is also able to produce short- to long-term climate predictions (months to a few decades ahead), and projections extending over many decades at the global level. Bioclimatic indicators allow the ever–increasing climate datasets to be combined and condensed and are valuable for both expert and non-expert users. Bioclimatic indicators from several global datasets are available from WorldClim* (Fick & Hijmans, 2017)*, CHELSA* (Karger et al., 2017)*, CliMond* (Kriticos et al., 2012)*, ecoClimate* (Lima-ribeiro, Varela, & Oliveira, 2015)*, ENVIREM* (P.O. Title & Bemmels, 2017)*, MERRAclim* (Vega, Pertierra, & Olalla-tárraga, 2017)*, CMCC-BioclimInd* (Noce, Caporaso, & Santini, 2020) *and the latest, KGClim* (Cui, Liang, Wang, & Liu, 2021)*. The FAO (Food and Agriculture Organization of the United Nations), provides Global Agro-Ecological Zoning*

(GAEZ) indicators of the likely variation in agricultural resources over time. Agrometeorological indicators from 1979 to the present and agro-climatic indicators from 1951 to 2099 derived from Climate Reanalysis and projections are available from Climate Change Service (C3S) of the Copernicus programme (Thépaut, Dee, Engelen, & Pinty, 2018). *Frequency, duration, timing and severity of extreme weather events can be calculated using indicators and indices for climate extremes such as those defined by the Expert Team on Climate Change Detection Monitoring and Indices (ETCCDI)* (Sillmann, Kharin, Zhang, Zwiers, & Bronaugh, 2013; Sillmann, Kharin, Zwiers, Zhang, & Bronaugh, 2013)

3. Becoming Adapted

Archaeological evidence and molecular analysis of present-day DNA variation suggest that livestock were domesticated in specific regions of different continents. The Fertile Crescent region in Southwest Asia is one of these. Here the wild progenitors of cattle, sheep, goats and pigs progressively adapted to a closer relationship with humans and finally became dependent on human care. Archaeozoological and mitochondrial DNA diversity data have confirmed that domestication of these species occurred in a climatically homogeneous area around the Fertile Crescent, comprising South-eastern Anatolia and the Iranian Zagros Mountains (Naderi et al., 2007; Zeder, 2008, 2015). After domestication, livestock followed human migrations and, with agricultural expansion, colonized the whole world (Legge, 1996; J. Vigne, 2011). Technological advances have facilitated the study of ancient DNA (aDNA) from well–preserved archaeological remains, which is shedding light on the spatiotemporal dynamics of domestication and on the physiological and neurobiological changes that livestock species underwent during the transition from the wild to a domestic existence, as well as on the subsequent adaptation to different environments and selection for functional traits (Mchugo, Dover, & Machugh, 2019). For example, these studies, have shown that cattle and goat domestication took place over relatively large geographical areas and extended time frames (Daly et al., 2018; Park et al., 2015; Verdugo et al., 2019), with frequent events of admixture and introgression, sometimes from several wild relative species (Park et al., 2015; Zheng et al., 2020). Over millennia livestock species have adapted to thrive in a range of environments, with different temperature, humidity, water and fodder availability and quality, pathogen and parasite challenges, and also to satisfy human needs for food, wool, fibre and tractive power.

At the genetic level, signals of adaptive changes driven by domestication have been found in genes related to nervous system development (Librado et al., 2017)(Pendleton et al., 2018) including kit ligand (*KITLG*), the treacle ribosome biogenesis factor *1* (*TCOF1*), and fibroblast growth factor receptor 1 (*FGFR1*) (Librado et al., 2017). Other signatures of selection, or of adaptive introgression from wild relatives, have been found in genes implicated in adaptation to feed and farming regimes. A variant

in the cytochrome P450 2C19 gene (*CYP2C19*) has been under positive selection in goats. *CYP2C19* is a member of the *CYP2C* subfamily of the cytochrome P450 superfamily of genes (Hannemann, Bichet, Ewen, & Bernhardt, 2007) which confers protection against a mycotoxin produced by Fusarium spp. fungi in cereals (Daly et al., 2018). Therefore, the increased frequency of the variant is most likely a response to an increasingly cereal-based diet contained in waste by-products. Alleles that may have been introgressed into domesticated goats from *Capra caucasica* (Weinberg, 2002), a West Caucasian tur–like species, have been found in a genomic region harbouring genes that affect immune function and parasite resistance, including *SERPINB3, SERPINB4, CD1B, COL4A4, BPI, MAN2A1*, and *CD2AP*. In particular, the mucin 6 oligomeric mucus/gel–forming gene (*MUC6*), which encodes a gastro-intestinally secreted mucin, is nearly fixed in goats for the Tur–derived haplotype, which confers enhanced immune resistance to gastrointestinal pathogens (Zheng et al., 2020). The fixation of this introgressed variant may be the consequence of the adaptive advantage it provided in farm environments, where there is increased exposure to parasites and disease (Zheng et al., 2020).

Recently, the characterization of the paleo-epigenome and paleo-microbiomes of domestic species have facilitated the exploration of their role in the adaptation of mammalian livestock to their environment (Liu, Weyrich, & Llamas, 2020). Data on the epigenomic profiles or microbiota composition in ancient livestock may provide information on diet, lifestyle, health status and exposure to stressors, and thus help us to explore the mechanisms of adaptation and interaction with the environment on a micro-evolutionary scale.

Animals adapt to the environments in which they live and to external stress by acclimation to a particular stressor or to a range of stressors (Collier, Baumgard, Zimbelman, & Xiao, 2019; John B. Gaughan, Sejian, Mader, & Dunshea, 2019). Adaptation can be crucial for survival, but often negatively affects the productivity and profitability of livestock systems. The ability to adapt depends in part on the flexibility of behavioral traits (Mignon-grasteau et al., 2005) and in part on morphological and physiological changes that better adapt animals for survival. For example, about 25% of sheep in the world are fat tail or fat rump breeds that are adapted to harsh semi-arid desert conditions where food availability is sporadic. The fat tail or rump acts as a store, to enable the animals to survive long periods when food is in short supply (Galal, Rasoul, Annous, & Shoat, 2005).

Cattle adapted to prolonged heat stress have increased hemoglobin and red cell numbers (Mazzullo et al., 2014), which may also protect them against blood borne parasites such as theileriosis. *Bos taurus taurus* cattle that have been raised over many generations in cool and temperate climates have long hair, subcutaneous fat, and often a dark coat colour. In contrast, *Bos taurus indicus* cattle that were originally from hotter tropical climates have short hair, little subcutaneous fat, low metabolism, and a body conformation to aid heat dispersion, with high surface to volume ratio, large ears and loose

skin, especially around the dewlap (Hansen, 2004; Utsunomiya, Milanesi, & Fortes, 2019). To increase performance while maintaining environmental resilience, crosses between taurine and indicine cattle have been developed (Trail, Gregory, Marples, & Kakonge, 1985). The crossbred animals show better adaptation to high temperature and humidity, and to parasites, e.g., resistance to *Boophilus microplus* ticks increases in proportion to *Bos taurus indicus* ancestry in the cross (Madalena, 2002).

Box 2. Heat Stress Indicators.

The level of heat stress experienced by an animal is the result of a combination of air temperature, relative humidity (Hill & Wall, 2015) *and other climate factors including wind speed and solar radiation* (V. P. R. da Silva et al., 2010)*. Depending on the management system, these parameters may make different contributions to the risk of thermal stress* (Herbut, Angrecka, & Walczak, 2018)*. Environmental parameters can be measured and used to construct indices and set thresholds to define risk situations.*

Most of the indices defining thermal stress risk have been developed for cattle, especially for dairy cows that are particularly susceptible to high temperatures. The Temperature Humidity Index (THI) (Thom, 1959) *takes into account the effect of air temperature and humidity. THI was originally developed as a general indicator of heat stress for humans, but today is also applied to livestock. Over the years, the model and threshold values used to define heat stress conditions have been modified* (Herbut et al., 2018)*, and corrections are now applied if cooling systems are used in the housing* (St-Pierre, Cobanov, & Schnitkey, 2003)*. THI does not take into account the cumulative effect of high temperature* (Herbut et al., 2018) *or the impact of wind speed and solar radiation, which are important when estimating the level of heat stress experienced by an animal. The Equivalent Temperature Index (ETI) includes air speed in the formula* (Baêta, Meador, Shanklin, & Johnson, 1987)*, although solar radiation is not considered* (Herbut et al., 2018)*. The THI adjusted (THIadj) index considers both the wind speed and the solar radiation, as well as breed and coat colour* (T. L. Mader, Davis, & Brown-Brandl, 2006)*. The Respiration Rate index (RR) is an extension of THIadj that also takes into account whether animals are in a shaded area or under the sun* (Eigenberg, Brown-Brandl, Nienaber, & Hahn, 2005)*.*

Other prediction models that have been developed to overcome the limitations of THI include the heat load index (HLI), which incorporates "black globe" temperature measurements substituting air temperature, animal factors (genotype, coat colour and health status) and management strategies (shade availability, days on feed, manure management and temperature of drinking water). These factors are used to modify the threshold to define the heat stress, and combined with factors to account for location-specific variables in different geographic areas (Rashamol et al., 2019)*. HLI is considered a better predictor than THI as it includes the interaction between climatic variables and*

animal thermal exchange mechanisms (Rashamol et al., 2019). *The Accumulate Heat Load Unit (AHLU) index, based on HLI, is a measure of the animal's heat load balance* (J B Gaughan, Mader, Holt, & Lisle, 2008). *The AHLU may increase or decrease over time depending on HLI values. A zero AHLU value indicates that the animal is in thermal balance* (J B Gaughan et al., 2008). *The HLI has also been extended to create a Comprehensive Climate Index (CCI) that can also be used under cold conditions* (T. L. Mader et al., 2010).
A comprehensive review of models for predicting heat stress response in livestock is given in Rashamol et al (Rashamol et al., 2019).

Senepol cattle were developed on the island of St Croix to create a breed that was polled, easily managed and tolerant of the tropical environment by crossing red polled taurine cattle with African Zebu cattle (Flori et al., 2012). Some of these cattle have very short hair and reduced follicle density, giving the phenotype referred to as "SLICK". SLICK is controlled by a single genetic locus and carriers of the Slick variant have lower core temperature than non-SLICK contemporaries (Olson, Lucena, Chase, & Hammond, 2003). Interestingly, the effect of SLICK is most likely through increased sweat production rather than the decrease in hair length and density (Dikmen et al., 2008). The SLICK variant in Senepol cattle was initially mapped to chromosome 20 (Mariasegaram et al., 2007), and later the causative variation was identified in the prolactin receptor gene (*PRLR*). A single base deletion in exon 10 causes a frameshift that introduces a stop codon and results in the truncation of the protein (Littlejohn et al., 2014). Other criollo cattle breeds, such as Carora and Limonero, that were brought to the Americas from Spain 500 years ago (Martinez et al., 2012) display a similar SLICK phenotype. However, these breeds do not carry the same prolactin variant that was identified in the Senepol cattle, although a genome-wide association analysis located the causative variant in or near to *PLRL*. DNA sequencing of SLICK Limonero cattle revealed three variants within the prolactin receptor gene that create premature stop codons in exon 11, one of which is also found in SLICK Carora cattle (Porto-neto et al., 2018). Recently, three novel variants were discovered in the *PLRL* gene in six Caribbean Basin cattle breeds. All create premature stop codons and increase heat tolerance. The occurrence of mutations in the prolactin receptor in several cattle breeds that are adapted to tropical climates and that have distinct evolutionary histories is unlikely to be by chance. Indeed, prolactin levels have been shown to be involved in thermoregulation in humans (Mundel, Bunn, Hooper, & Jones, 2007), showing that certain physiological processes and specific genes can be targeted by environmental pressure. The SLICK variant has now been introgressed into other breeds, including the highly productive Holstein dairy breed, creating more heat tolerant animals (Dikmen et al., 2014).

Nevertheless, adaptation generally requires changes in the combination of alleles of many genes; for example, the genomic analysis of admixture between *Bos taurus taurus* and Zebu (*Bos taurus indicus* cattle) in Africa showed that more than 150 loci were under selection for local adaptation (Bahbahani, Tijjani, Mukasa, & Wragg, 2017). The ability of livestock to successfully adapt to extreme climatic conditions and to tolerate a wide range of parasites has resulted in local populations with specific characteristics. These populations are valuable resources that, if well characterized, could be exploited to create breeds suited to new conditions arising from climate change.

Box 3. The Genome and Genomics.

The publication of the human genome sequence in 2001 (International Human Genome Consortium, 2001) *was a landmark that opened new opportunities in molecular genetics. The same approach that was used to sequence the human genome was used to produce draft sequences for the major livestock species; the first was the chicken in 2004* (International Chicken Genome Sequencing Consortium, 2004)*, followed by the cow in 2009* (The Bovine genome Sequencing and Analysis Consortium, 2009)*, then the pig* (Groenen et al., 2012)*, sheep* (Faraut et al., 2014) *and goat* (Dong et al., 2013) *in 2012. These genomes became references against which DNA and RNA sequences from these species were aligned and compared. With the rapidly advancing sequencing technologies, which progressed from automated Sanger sequencing to next-generation high throughput short read sequencing* (Metzker, 2010)*, large numbers of individuals were sequenced at low resolution. Alignment of these sequences with the reference genomes revealed huge numbers of variations among individuals, in particular, Single Nucleotide Polymorphisms (SNP). This SNP data led to the development of genome-wide genotyping panels. A range of low (few thousand) to high (many hundred thousand) density SNP panels is commercially available, including some targeted to specific traits, and others that include SNP for several species to reduce costs of genotyping. Knowledge of the genome sequence from large numbers of individuals in a population enables low density SNP genotype data to be used to estimate higher density genotypes by "imputation"* (Berry & Kearney, 2011)*.*

The analysis of phenotype and genotype in genome-wide association studies enables genetic loci with a major effect on the phenotype to be identified (e.g., (Igoshin et al., 2019; Porto-neto, Reverter, & Prayaga, 2014; Vanvanhossou et al., 2020)*). In some cases the genes and causative polymorphisms controlling variations in target traits have been identified (e.g.,* (Raven, Cocks, & Hayes, 2014)*). Perhaps the most important advance coming from the availability of genome-wide SNP panels is that the idea of genome-based selection envisioned by Meuwissen and colleagues more than a decade ago has now been realized* (T. H. Meuwissen, Hayes, & Goddard, 2001)*. Other applications of the SNP panels include the analysis of population structure, history and diversity (e.g.,* (Kijas et al., 2009; Michailidou et al., 2019; B. Yang et al., 2017) *to guide conservation strategies* (Bruford, Ginja,

Hoffmann, Joost, Orozco-terWengel, et al., 2015) *and the identification of regions of the genome that are under selection (e.g.,* (L. Xu et al., 2015)*).*

Next generation sequencing (NGS) has also facilitated the study of gene expression by enabling the analysis of the whole transcriptome (I. Yang & Kim, 2015)*. Depending on how samples are processed and analysed, this approach can examine the expression of genes (e.g.,* (Marino et al., 2016; L. Wang et al., 2017)*), variations in splice sites* (Zappaterra, Gioiosa, Chillemi, Zambonelli, & Davoli, 2020)*, and non-coding RNAs* (Kern, Wang, & Chitwood, 2018; X. Miao, Luo, Zhao, & Al., 2016) *as well as short, micro-RNAs* (Pasquariello et al., 2017) *that have a regulatory role.*

Further advances in sequencing technology are opening new opportunities. Long read, single molecule sequencing has enabled haplotype resolved genome sequences to be produced by separating the sequence reads originating from the maternally and paternally inherited chromosome (Koren et al., 2018; Low et al., 2020)*. Long read technologies such as Pacific Biosciences and Oxford Nanopore can produce full length sequences of transcripts to reveal isoforms present in different tissues or diverse physiological states. These technologies are also able to distinguish modified bases in the DNA, specifically methylation, in order to examine epigenetic patterns directly and explore the regulation of gene expression* (Jin et al., 2018)*. The Functional Annotation of Animal Genomes Consortium* (Giuffra, Tuggle, & FAANG Consortium, 2018) *is assembling data on genome structure, expression, and regulation using a range of new technologies. For an extensive review of the state of livestock genomics see Georges et al* (Georges, Charlier, & Hayes, 2019)*.*

4. Seeking Adaptive Genes

Several molecular genetic approaches have been used to identify adaptation-related genes. Genome wide association studies (GWAS) use phenotypes related to adaptation recorded directly on the animals. Landscape Genomics approaches use environmental variables as proxies for phenotypes. Other methods analyse the patterns of genomic diversity within and between populations and the level of admixture in specific genomic regions to identify selection signatures of adaptation. These approaches use genomic tools that may focus on individual loci through to whole genomic sequence analyses (see Box 3) and dedicated software (Table 1).

4.1. Genome-Wide Association Studies

Genome-wide association studies (GWAS) identify the association between variations in the genome, the genotype, with variations in phenotype displayed by individual animals belonging to a same breed or population. GWAS therefore requires both genotype and phenotype data on each individual (Cantor, Lange, & Sinsheimer, 2010; Tam et al., 2019). Fulfilling such conditions is difficult for complex phenotypes, and not always feasible when the target population is small or isolated (Hatzikotoulas,

Gilly, & Zeggini, 2014), which is often the case in adaptation studies. Moreover, costs for genotyping and trait recording represents a further hurdle in reaching an adequate sample size. For these reasons, GWAS carried out in livestock to understand the genetic control of complex traits, are invariably low powered and results between studies on the same traits are often inconsistent. In addition, the genetic associations identified are likely to differ depending on the way that a trait is measured, the genetic background and the environment. Livestock GWAS have primarily been used to identify genetic variants associated with specific production traits or disease responses (Sharma, Seop, Dang, Sudrajad, & Kim, 2015). GWAS that identify the genes controlling climate adaptation traits (e.g., efficient thermoregulation, feed utilization, and immunity) would accelerate selection for animals more resilient to climatic challenges (B. J. Hayes, Lewin, & Goddard, 2013).

Several statistical tests have been applied to identify marker–trait associations in GWAS, from single marker regression, to mixed model and Bayesian approaches that use different marker effect distributions as prior information, to haplotype based GWAS (B. Hayes, 2013). In all cases, corrections have to be applied for multiple testing and for population structure in order to avoid a high number of false positives. As most traits involved in adaptation are highly complex and have a low to moderate heritability, a large cohort of animals has to be investigated to reach a sufficient statistical power in GWAS. (Bouwman et al., 2018; Goddard & Hayes, 2009).

A GWAS of cattle indigenous to Benin (Vanvanhossou et al., 2020) identified several potential candidate genes associated with stress and immune response (*PTAFR, PBMR1, ADAM, TS12*), feed efficiency (*MEGF11, SLC16A4, CCDC117*), and conformation and growth (*VEPH1, CNTNAP5, GYPC*). The study of cold stress in Siberian cattle breeds identified two candidate genes (*MSANTD4* and *GRIA4*) on chromosome 15, putatively involved in cold shock response and body thermoregulation (Igoshin et al., 2019). GWAS in taurine, indicine and cross-bred cattle identified *PLAG1* (BTA14), *PLRL* (BTA20) and *MSRB3* (BTA5) as candidate genes for several traits important for adaptation to extensive tropical environments (Porto-neto et al., 2014). A GWAS of the Frizarta dairy sheep breed, which is adapted to a high relative humidity environment, identified 39 candidate genes associated with body size traits including *TP53, BMPR1A, PIK3R5, RPL26,* and *PRKDC* (Kominakis et al., 2017). An association analysis of genotype-by-environment (GxE) interactions with growth traits in Simmental cattle showed that birth weight was affected by temperature, while altitude affected weaning and yearling weight. Genes implicated in these traits included neurotransmitters (*GABRA4* and *GABRB1*), hypoxia-induced processes (*PLA2G4B, PLA2G4E, GRIN2D,* and *GRIK2)* and keratinization (*KRT15, KRT31, KRT32, KRT33A, KRT34,* and *KRT3)*, all processes that play a role in physiological responses associated with adaptation to the environment (Braz, Rowan, Schnabel, & Decker, 2021).

Enhancing efficiency would reduce the impact of changes in feed availability on livestock systems and potentially reduce methane production, which contributes to climate change. Residual feed intake (RFI), that is, the difference between actual feed intake and the theoretical energy requirements of an animal (Koch, Swiger, Chambers, & Gregory, 1963), has been used to select for increased feed efficiency (FE) (Herd & Bishop, 2000; Tortereau, Weisbecker, Marcon, Bouvier, & François, 2020). A GWAS of RFI in Nellore cattle identified QTL on chromosomes 8 and 21 affecting the trait. Putative candidate genes on BTA 8 are *CCDC171* and *CLCN3* (Santana et al., 2014), while candidates on BTA11 are *DEPP1,* expression of which is induced by fasting, *TUBB3* and *PTSG1* (P. S. N. De Oliveira et al., 2014).

A GWAS for temperament scores carried out on crossbred steers in a feedlot identified five SNP on BTA 1, 24, and 29 and 13 SNP on BTA11 (Riley et al., 2016). Functional candidate genes close to these loci had roles in neural function included synaptotagmin 4 (BTA 24), FAT atypical cadherin 3 (BTA 29*)*, tubulin tyrosine ligase-like 1 (BTA 5), spermatogenesis associated 17 *(*BTA 16), stanniocalcin 2 (BTA 20), and GABAA receptor γ 3 (BTA 21). A GWAS of 3,274 Charolais beef cows detected four significant and 12 suggestive chromosomal regions associated with several functional and behavioral traits including aggressiveness (Vallée, Daures, Arendonk, & Bovenhuis, 2016). A recent GWAS analysis of 1,370 Brahman cattle clustered in two groups of temperament identified nine SNP located in intergenic regions near candidate genes *ACER3*, *VRK2*, *FANCL* (Paredes-Sanchez et al., 2020).

4.2. Selection Signatures

Natural or artificial selective pressure causes an increase or decrease in the frequency of genetic variants in a population. Selection can be positive, balancing, or negative (Vitti et al., 2013). Positive selection increases the frequency of fitness-enhancing variants in a population whereas negative selection removes unfavourable mutations to restore DNA functional integrity (Zeng et al., 2018). Balancing selection retains more than one allele of a gene where heterozygotes have higher fitness (Charlesworth, 2006). The genes in the genomic region in linkage disequilibrium with the genes under selection will also increase or decrease in frequency through the hitch-hiker effect (Barton, 2000), changing the expected patterns of molecular variation and giving a “selection signature”.

Tajima’s D statistic (See Box 4) has been used to analyse wild and domestic sheep data to identify a genomic region involved in the resistance to pneumonia (Y. H. Cao et al., 2021). A scan of Russian cattle genomes using Tajima’s D statistic detected signatures of selection most likely resulting from adaptation to cold environments (Yurchenko et al., 2018). Fay and Wu’s H statistic has been used with cattle data to detect signals of recent positive selection involving genes associated with innate immune response (Y. Chen et al., 2020).

Signatures of recent selection associated with aggressiveness have been identified on chromosome X by comparing the Lidia cattle breed, which has been selected for aggressive responses, with two Spanish breeds showing docile behaviour. The most significant selection signature included the monoamine oxidase A gene (*MAOA*) (P.G. Eusebi, Cortés, Carleos, Dunner, & J., 2018). A further refinement of the analysis identified a variable number of tandem repeats in the gene, with the Lidia breed having fewer repeats compared with the docile breeds (P.G. Eusebi et al., 2020). Favourable genetic and phenotypic relationships between docility and meat quality, feedlot performance, ease of transport and reproductive traits have been reported (Hamlyn-Hill, 2012). Temperamental animals generally are not as well adapted to stress and have slow growth rates, poor carcass conformation and poor immune function (Burdick et al., 2011; Café et al., 2011). Differences in docility have also been found between *Bos taurus taurus* and *Bos taurus indicus* cattle (e.g., (Burrow, 2001) and between beef and dairy breeds (Hoppe, Brandt, Nig, Erhardt, & Gauly, 2010).

Signatures of selection related to feed adaptation have been found in sheep using an F_{ST} approach (Lv et al., 2014b). Of the seventeen genes under climatic selection, nine were related to energy metabolism. The strongest selection signal was around *TBC1D12,* on OAR22, which plays a role in GTPase regulation. The F_{ST} approach was also applied to Siberian cattle populations in order to understand the genetic basis of adaptation to cold environments (Igoshin, Yudin, Aitnazarov, Yurchenko, & Larkin, 2021). Results identified several genes that have been implicated in thermal adaptation in cattle, such as *GRIA4*, *COX17*, *MAATS1*, *UPK1B*, *IFNGR1*, *DDX23*, *PPT1*, *THBS1*, *CCL5*, *ATF1*, *PLA1A*, *PRKAG1,* and *NR1I2.*

With regard to hot environments, Li and colleagues (R. Li et al., 2020) investigated selection signatures of bovine heat tolerance in Dehong cattle, a Chinese indigenous zebu breed, using an F_{ST} approach. Results indicated that genes involved in heat shock (*HSF1*), oxidative stress response (*PLCB1*, *PLCB4*), coat color (*RAB31*), feed intake (*ATP8A1*, *SHC3*) and reproduction (*TP63, MAP3K13, PTPN4, PPP3CC, ADAMTSL1, SS18L1, OSBPL2, TOX, RREB1,* and *GRK2*) may play a role in heat adaptation.

Pairwise comparison of genetic differentiation of sheep breeds adapted to different environments identified selection signatures in the genes *MITF, FGF5, MTOR, TRHDE* and *TUBB3* that have been associated with high-altitude adaptation (Edea, Dadi, Dessie, & Kim, 2019). An F_{ST} statistic approach applied to cattle breeds reared in different environments identified several genes under positive selection for thermal tolerance (Freitas et al., 2021). HapFLK detected the Nebulin Related Anchoring Protein gene (*NRAP)* to be under selection for adaptation to cold environments (Buggiotti et al., 2021), *ACSS2, ALDOC, EPAS1, EGLN1* and *NUCB2* to be under selection for high-altitude adaptation in cattle (X. Wang et al., 2021), and *DNAJC28, GNRH1* and *MREG* to be associated with heat stress adaptation in sheep (Molotsi, Cloete, Taylor, & Whitacre, 2018).

iHS methods have been used to detect signatures of adaptation to environments in French Charolais cattle, sheep and goats (Álvarez et al., 2020a; Francesca Bertolini, Servin, et al., 2018; E. S. Kim et al., 2016; Mwacharo et al., 2017; Saravanan et al., 2021). Cross-population EHH-based tests have been used to detect hot climate adaptation in cattle (Dutta et al., 2020; Pitt, Bruford, et al., 2019; Singh et al., 2020) and sheep (Álvarez et al., 2020b; Eydivandi et al., 2021; Zhang et al., 2021), and hypoxia adaptation in new world camelids (R. Fan et al., 2020). Detecting runs of homozygosity (ROHs) to find regions containing genes associated with adaptation has been demonstrated in several domestic species (Abied, Xu, et al., 2020; Álvarez et al., 2020a; Freitas et al., 2021; Macciotta et al., 2021; Saravanan et al., 2021).

Box 4. Approaches for Selection Signature Detection.

Selection on a locus, whether artificial for production or natural for adaptation, is associated with the reduction of genetic diversity in the region, creating a "selection signature". Tajima's test (Tajima, 1989) *is able to detect positive selection sweeps that occurred recently, as it identifies regions with high numbers of rare, low-frequency variants that are the result of recent mutation* (Simonsen, Churchill, & Aquadro, 1995)*. Fay and Wu statistics* (Fay & Wu, 2000)*, in contrast, assess the relationship between ancestral and derived alleles, which enables both positive and negative recent selection occurring in medium- to high-frequency alleles to be detected. However, knowledge of ancestral alleles is necessary to apply the method* (Saravanan et al., 2020)*.*

Various approaches have been used to assess positive and negative selection in populations. Wright's fixation index (F_{ST}) measures differences in allele frequencies between populations based on individual loci. F_{ST} has been used in many studies of livestock to explore differences among populations. A more recent approach to analyse population differentiation is the hapFLK metric (Fariello, Boitard, Naya, SanCristobal, & Servin, 2013)*, which improves on single locus statistics by testing haplotype differentiation. hapFLK corrects frequency estimates, accounting for the genetic relationship between populations using Reynolds genetic distances.*

Selection for a favourable allele of a gene increases the levels of linkage disequilibrium (LD) around the locus under selection, until recombination occurs to reduce the extent of LD (Y. Kim & Neilsen, 2004)*. Selection signatures can therefore be found by detecting regions of strong LD relative to their prevalence within a population* (Sabeti et al., 2002, 2007)*. Alleles at linked loci are referred to as haplotypes. Extended haplotype homozygosity (EHH) methods measure the decay of haplotype homozygosity as a function of genetic distance. The integrated Haplotype Score (iHS)* (Voight, Kudaravalli, Wen, & Pritchard, 2006) *is calculated from the integrals of the observed decay of EHH for the ancestral and derived alleles surrounding the locus under selection. Divergence between values from the genomic average is indicative of selection. This approach requires phased data and*

knowledge of the ancestral state for each allele, and it has low power when one allele is at high frequency or fixed. Cross-population methods such as XP-EHH (Sabeti et al., 2007) *and Rsb* (Tang et al., 2007) *calculate EHH profiles between two populations, removing the need to know the ancestral state. These methods have high power for detecting selective sweeps that have reached fixation. Selective sweeps generate runs of homozygosity (ROH) when both parents pass on the same haplotypes that are inherited from one generation to the next* (Gibson et al., 2006).

4.3. Local Ancestry Inference

Local ancestry inference (LAI) identifies the ancestors of each genomic region at the chromosome level. LAI is also described as local ancestry deconvolution or chromosome painting. Local ancestry information can help to understand fine scale admixture and the population genetic history, identify recent targets of selection, guide the selection of reference panels for genotype imputation, and improve the detection power of genetic association studies of admixed populations (Atkinson et al., 2021; Brisbin et al., 2012; Pasaniuc, Sankararaman, Kimmel, & Halperin, 2009; Schubert, Andaleon, & Wheeler, 2020; Tang et al., 2007). Identifying the ancestry of chromosomal segments in admixed individuals facilitates the accurate identification of the history of genetic variants under selection (Pasaniuc et al., 2009), particularly where adaptive introgression has fixed or nearly fixed regions of the genome with specific population ancestry (J. Wu, Liu, & Zhao, 2021).

Most approaches to profile local ancestry divide the genome into windows and assign ancestry to each window by comparing it against a reference panel (Brisbin et al., 2012; Dias-Alves, Mairal, & Blum, n.d.; Guan, 2014; Maples, Gravel, Kenny, & Bustamante, 2013; Pasaniuc et al., 2009; Salter-Townshend & Myers, 2019; Tang, Coram, Wang, Zhu, & Risch, 2006). New methods do not require the explicit definition of a reference population (Popescu & Huber, 2015; Utsunomiya et al., 2020). The most popular algorithms for LAI rely on hidden Markov models (HMM), an extension of a Markov chain, to identify the transformation of a genomic region from the reference, which is often not obvious (W. Zhao, Ma, Chen, Fu, & Zhang, 2019). These methods provide the posterior probabilities for each possible ancestry state at each ancestry-informative site along the chromosome (Schubert et al., 2020; J. Wu et al., 2021). The estimates obtained depend largely on reference populations; therefore, approaches to identify convergent signals of ancestry across multiple tests using different references have been developed (Barbato, Hailer, Orozco-Terwengel, et al., 2017).

LAI has been widely applied to identify adaptive introgression related to climatic stressors in livestock. Adaptive introgression from wild to domestic sheep of loci affecting climatic adaptation and resistance to pneumonia has been identified using LAI (Barbato, Hailer, Orozco-Terwengel, et al., 2017; Y. H. Cao et al., 2021). Using LAI and multiple-reference adjustments, ancestry components of indicine origin were found in cattle breeds from Central Italy that are associated with resilience to harsh

environments and climatic conditions (Barbato et al., 2020). A region of indicine introgression into Italian local taurine breeds has been identified on BTA18 containing *KLHL36, USP10, KIAA0513* and *FAM92B*, all of which are related with residual feed intake (Barbato et al., 2020). This introgression could provide an adaptive advantage enabling animals to use low quality feed efficiently.

Introgression of genes regulating the response to hypoxia from yak into Tibetan cattle that facilitated the adaptation of the latter to high altitude was also identified by LAI (D.-D. Wu et al., 2018). Similarly, adaptive introgression of genes related to oxygen transportation from Argali sheep to Tibetan domestic sheep may be a key factor conferring high-altitude resilience (X.-J. Hu et al., 2019). Local ancestry signals in African cattle have identified the genomic components of indicine cattle related to heat tolerance and water reabsorption, along with innate-immune resistance to tick and tick-borne diseases (K. Kim et al., 2020). LAI tests have provided evidence of adaptive introgression between llama and alpaca for coat colour, fibre characteristics, and adaptation to high altitude and harsh environment (R. Fan et al., 2020).

4.4. Landscape Genomics

Landscape genomics explores the interaction between the genome and the environment to better understand evolution by combining landscape ecology and population genetics (Jelinski, 1997; Manel et al., 2003). Two advances enabled landscape genomics to be realized. The first was the development of Geographic Information Systems (GIS) (Goodchild, 1992), which facilitated the overlay of diverse geo-referenced information, in this case genetic and environmental data. The second was the availability of large numbers of genetic markers, specifically single nucleotide polymorphisms, that are easily assayed. The development of the software MatSAM to compare a large number of allele frequencies with eco-climatic variables brought these two advances together as landscape genomics (Joost et al., 2007). The MatSAM software (Joost, Kalbermatten, & Bonin, 2008) has been successfully used for landscape genomics analyses of plant and animal species, including sheep (Joost et al., 2007), goats (Pariset, Joost, Ajmone Marsan, Valentini, & Ec, 2009) and fish (Tonteri, Vasemagi, Lumme, & Primmer, 2010). These studies used GIS to store both genetic and environmental variables retrieved from open access databases to create gene–environment matrices that are processed by logistic regressions. Several software programs using different models have been developed for land-scape genomic analysis; improvements of these have an ever-increasing capability to efficiently analyse big data sets of genomic and environmental variables (see Box 5).

Landscape genomics approaches were used to understand the genetic adaptation of South African goats, finding that climatic variables explained 17% of their overall diversity. Using SAM software (see Box 5 and (Joost et al., 2007)), 843 SNPs were identified that were associated with longitude, while LFMM software (Frichot, Schoville, Bouchard, & François, 2013) found that 714 SNPs were associated

with temperature and precipitation (Mdladla & Dzomba, 2018), with only one locus in common that included *DGKB*. These SNPs were close to genes involved in 205 biological pathways, all of which are potentially related to adaptation. Among the genes identified, several have been associated with thermoregulation in hot environments (e.g., *PLCB1*). In the analysis of a goat database of more than 1000 animals covering 33 Italian populations using landscape genomics methods and LFMM (Stucki et al., 2017), identified many loci putatively associated with environmental variables, although there was no overlap in loci identified by each of the methods. Samβada identified 62 genes associated with temperature or precipitation; among these, *RYR3* has been associated with mean temperature and *ANK3* and *BTRC* with longitude (Cortellari, Barbato, Talenti, Bionda, Randi, et al., 2021). The LFMM analysis identified four SNPs associated with Mean Diurnal Range and Mean Temperature. These SNP were near *NBEA*, located within a region involved with wool production in sheep (Z. Wang et al., 2014), and *RHOBTB1*, which is known to be associated with meat quality in cattle (D. B. S. Silva et al., 2020). As observed before, methods implemented in Samβada and LFMM produce non-overlapping results. The two software are suited to the analysis of population having specific genetic structure (see Box 5) and their use is suggesed as complementary rather than alternative tools. Colli et al. (Colli, Negrini, Nicoloso, & Crepaldi, 2014) applied landscape genomics software based on the SAM approach to analyse 43 European and West Asian goat breeds. Using AFLP markers, four loci were identified that were significantly associated with diurnal temperature range, frequency of precipitation, relative humidity and solar radiation.

A landscape genomic analysis of 57 sheep breeds using the SAM approach found that the *DYMS1* microsatellite locus was associated with the number of wet days, which largely affects parasite load (Joost et al., 2007). In an earlier study this locus was shown to be associated with parasite resistance (Buitkamp, Filmether, Stear, & Epplen, 1996).

Box 5. Landscape Genomics Software.

With the availability of increasing numbers of measures of environmental variables and an increasing number of genetic markers, the MatSAM software (Joost et al., 2008) *was developed to process many simultaneous univariate association models. Samβada* (Stucki et al., 2017) *is able to compute univariate and multivariate logistic regressions, integrate and make an intelligent selection of significant models, calculate pseudo R2, Moran's I, and Geographically Weighted Regressions. This software has High Performance Computing (HPC) capacities to handle the large datasets created when several million SNPs, produced by high-throughput sequencing, are combined with hundreds of environmental variables. Samβada is also supported by R-SamBada* (Duruz et al., 2019)*, an R software package that provides a complete pipeline for landscape genomic analyses, from the retrieval of environmental variables at sampling locations to gene annotation using the Ensembl genome browser.*

Other landscape genomics software include BAYENV (Gunther & Coop, 2013)*, which uses the Bayesian method to compute correlations between allele frequencies and ecological variables, taking into account differences in sample size and population structure; LFMM* (Frichot & Francois, 2015; Frichot et al., 2013)*, which identifies gene-environment associations and SNPs with allele frequencies that correlate with clines of environmental variables; and SGLMM* (Guillot & Vitalis, 2014)*, which extends the BAYENV approach* (Coop, Witonsky, Rienzo, & Pritchard, 2010) *by using a spatially explicit model and calculating inferences with an Integrated Nested Laplace Approximation and Stochastic Partial Differential Equation (SPDE). BayPass* (M. Gautier, 2015) *builds on BAYENV to capture linkage disequilibrium information. BAYESCENV* (Villemereuil & Gaggiotti, 2015) *produces an* F_{ST}*-based genome scan, taking into account environmental differences between populations. The latest version of LFMM* (Caye, Jumentier, Lepeule, & François, 2019) *improves on both scalability and speed with respect to other GEA methods using a least-squares approach to estimate cofounders. Moreover, LFMM uses several categories of genomic data which are not restricted to genotypes.*

Landscape genomics studies often use population genomics software (e.g., LOSITAN based on the FDist model (Antao, Lopes, Lopes, Beja-pereira, & Luikart, 2008; Beaumont & Nichols, 1996)*) to compare the sets of candidate loci obtained from different approaches: see BayeScan* (Foll & Gaggiotti, 2008) *and Bayenv* (Coop et al., 2010)*. A comparison of results allows for consolidation, as the accuracy of methods is known to differ (see, e.g.,* (Stucki et al., 2017)*). SamBada / R–SamBada* (Duruz et al., 2019) *gives reliable results when the population structure is weak, while LFMM2* (Caye et al., 2019) *is better suited to detect selection signatures in well-structured populations. Analyses of simulated data using, e.g., CDPOP* (Landguth et al., 2010) *is usually advised to demonstrate the effectiveness of the method before moving to the analysis of empirical data (see, e.g.,* (De Mita et al., 2013; Frichot et al., 2013; Stucki et al., 2017)*). GEONOMICS, a Python package, performs forward-time, individual-based, continuous-space population genomic simulations on complex landscapes* (Terasaki Hart, Bishop, & Wang, 2021)*. GEONOMICS includes several analytical steps using models of a landscape with one or more environmental layers (geotiff files as input), each of which can undergo environmental changes, as well as species having genomes with realistic architecture and associated phenotypes. Species undergo non-Wright–Fisher evolution in continuous space, with localized mating and mortality. The results produced are useful for a wide variety of theoretical and empirical purposes such as species conservation and management.*

4.5. Artificial Intelligence and Machine Learning Approaches

With advances in genomic technology and more sophisticated sensing systems, “big data” sets are being created and a large amount of data needs to be stored every day (Stephens et al., 2015). These data sets will potentially reveal changes in genomes that adapt animals to a wide range of conditions

and environments. However, the information is a mixture of homogeneous and heterogeneous data types where the relationships among parameters may be hidden or difficult to identify. Artificial Intelligence (AI) and Machine Learning (ML) methods are increasingly used to extract information from this type of data to overcome the limits of traditional linear models (250, 251) (see Box 6). ML and AI have not yet been fully applied to study adaptation to climate change in livestock; however, the role of big data and machine learning will become increasingly important for modern farming (Neethirajan, 2020).

ML methods have been used in the quest for regions associated with adaptation, in particularly to detect *de novo* mutations and selective sweeps for previously segregating variants in humans (Rees, Castellano, & Andrés, 2020). The S/HIC Deep Learning (DL) model has shown that most human mutations are neutral in populations, and that those conferring an adaptive advantage only rise in frequency when a change in the environment gives advantages to individuals carrying a particular mutation (D.R. Schrider & Kern, 2017). This approach has been used to identify genes associated with metabolism in a southern African ethnic groups using the SWIF(r) DL algorithm (Sugden et al., 2018). Variants of these genes arose thousands of years ago to store fat when food was scarce.

There are a few examples of the use of ML in livestock genetics and breeding (Nayeri, Sargolzaei, & Tulpan, 2019; Okser et al., 2014; Utsunomiya et al., 2020), and new DL genetic models are only just being tested (Arpanahi Abdollahi, Gianola, & Peñagaricano, 2020; Bellot, Campos, & Pérez-enciso, 2018; B. Li et al., 2018; Waldmann, 2018). The identification of SNPs directly associated with candidate genes affecting growth traits in Brahman cattle was more successful using ML Gradient Boosting Machine (GBM) than Random Forest statistical methods (B. Li et al., 2018). ML algorithms have been used together with RNA-Seq expression data to identify genes associated with feed efficiency in pigs, and to classify animals' phenotypic extreme for residual feed intake (Piles et al., 2019).

Box 6. Artificial Intelligence and Machine Learning.

Artificial Intelligence (AI) uses algorithms that automate the decision process (Kaluarachchi, Reis, & Nanayakkara, 2021)*, while Machine Learning (ML) uses AI to automatically learn complex relationships and patterns in data* (Helm et al., 2020; Jordan & Mitchell, 2015)*. ML algorithms may be unsupervised or supervised. The former explores the dataset structure without prior knowledge of data organization, while the latter uses prior knowledge to train the model and predict the outcome in a test dataset* (Daniel R. Schrider & Kern, 2018)*. ML algorithms are adapted to explore nonlinear relationships* (Harfouche et al., 2019)*. Deep learning (DL) creates multiple processing layers (neural networks), which mimic the structure of a human brain, to extract information and learn from the input data. DL is being used to discover intricate structures in large datasets* (Helm et al., 2020; LeCun, Bengio, & Hinton, 2015)*. However, the neural network models are a "black box" as they are hidden as they develop. Tools*

are being developed to dissect the layers of the models developed to understand the neural network process; one example are the saliency maps (Liu et al., 2019; Voosen, 2017).

ML methods mainly focus on prediction, while classical statistical methods rely on inference (T. Hu, Darabos, & Urbanowicz, 2020). *ML has been used to recognize the location of specific sequence elements (i.e., splice sites, promoters, etc.) and to combine genomic elements to identify and annotate genomic features, e.g., to identify UTR, introns, and exons, and to functionally annotate genes* (Rees et al., 2020). *For example, S/HIC (https://github.com/kern-lab/shIC) is an ML classifier developed to detect targets of adaptive natural selection from whole genome sequencing data.*

Efficient DL software tools such as Tensorflow and Keras Python libraries, and the availability of supercomputing using graphics processing unit technology (GPU), have opened the way to the integration of multi-omics big data with environmental variables.

Table 1. Software for genome-wide analyses.

Software	Method	Application	Ref.	Link
Arlequin	Tajima's D	Selection signatures	(Excoffier & Lischer, 2010)	http://cmpg.unibe.ch/software/arlequin35/
BayeScan	F_{ST}	Selection Signatures, Landscape genomics	(Foll & Gaggiotti, 2008)	http://cmpg.unibe.ch/software/BayeScan/
Bcftools	ROH	Selection signatures	(H. Li, 2011)	https://github.com/samtools/bcftools
DnaSP	Tajima's D and Fay and Wu's statistic	Selection signatures		http://www.ub.edu/dnasp/
Hapbin	EHH	Selection signatures	(Maclean, Chue Hong, & Prendergast, 2015)	https://github.com/evotools/hapbin
hapFLK	hapFLK	Selection signatures	(Fariello et al., 2013)	https://forge-dga.jouy.inra.fr/projects/hapflk
HierFstat (R package)	F_{ST}	Selection signatures	(Goudet, 2005)	https://cran.r-project.org/web/packages/hierfstat/index.html
KING	ROH	Selection signatures	(Manichaikul et al., 2010)	https://www.kingrelatedness.com/
PLINK	F_{ST}, ROH	GWAS, Selection Signatures	(C. C. Chang et al., 2015)	https://www.cog-genomics.org/plink/2.0/ https://www.cog-genomics.org/plink/
PopGenome	Tajima's D	Selection signatures	(Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014)	https://cran.r-project.org/web/packages/PopGenome/index.html
PoPoolation	Tajima's D	Selection signatures	(Kofler et al., 2011)	https://sourceforge.net/p/popoolation/wiki/Main/
rehh (R package)	EHH	Selection signatures	(Mathieu Gautier, Klassmann, & Vitalis, 2017)	https://cran.r-project.org/web/packages/rehh/index.html
Selscan	EHH	Selection signatures	(Szpiech & Hernandez, 2014)	https://github.com/szpiech/selscan

VariScan	Tajima's D	Selection signatures	(Vilella, Blanco-Garcia, Hutter, & Rozas, 2005)	http://www.ub.edu/softevol/variscan/
VCFtools	F_{ST}, Tajima's D	Selection signatures	(Danecek et al., 2011)	http://vcftools.sourceforge.net/
EMMAX	GWAS based on variance component model	GWAS	(H. M. Kang et al., 2010)	http://genetics.cs.ucla.edu/emmax
GCTA	GWAS based on genome-wide complex trait analysis	GWAS	(J. Yang, Lee, Goddard, & Visscher, 2011)	http://gump.qimr.edu.au/gcta
BayesR	Bayesian mixture model	GWAS	(Moser et al., 2015)	http://www.cnsgenomics.com/software/
MatSAM	Logistic regression	Landscape genomics	(Joost et al., 2008)	www.econogene.eu/software/sam/
Samβada, R.SamBada (R package)	GEA based on logistic regression/spatial autocorrelation	Landscape genomics	(Duruz et al., 2019; Stucki et al., 2017)	https://github.com/Sylvie/sambada/releases/tag/v0.8.3 https://cran.r-project.org/package=R.SamBada
BAYENV	GEA based on Bayesian regression	Landscape genomics	(Gunther & Coop, 2013)	https://gcbias.org/bayenv/
LFMM2 (R package)	GEA based on latent factor mixed models	Landscape genomics	(Caye et al., 2019; Frichot & Francois, 2015)	https://bcm-uga.github.io/lfmm/
SGLMM	GEA based on allele-environment association analysis	Landscape genomics	(Guillot & Vitalis, 2014)	-
BayPass	GEA corrected for the covariance structure among the population allele frequencies	Landscape genomics	(M. Gautier, 2015)	http://www1.montpellier.inra.fr/CBGP/software/baypass/
BAYESCENV	GEA based on F_{ST} genome-scan	Landscape genomics	(Villemereuil & Gaggiotti, 2015)	https://github.com/devillemereuil/bayescenv
LOSITAN	F_{ST}	Landscape genomics	(Antao et al., 2008)	https://mybiosoftware.com/lositan-1-0-0-selection-detection-workbench.html
PCAdmix	Supervised LAI	Local Ancestry Inference	(Brisbin et al., 2012)	https://sites.google.com/site/pcadmix/home
Tractor	LA-aware regression model	Local Ancestry Inference	(Atkinson et al., 2021)	https://github.com/eatkinson/Tractor
LAMP	LAI accounting for recombination	Local Ancestry Inference	(Pasaniuc et al., 2009)	http://lamp.icsi.berkeley.edu/lamp/
MOSAIC (R package)	Unsupervised LAI	Local Ancestry Inference	(Salter-Townshend & Myers, 2019)	https://maths.ucd.ie/~mst/MOSAIC/
RFMix	LAI based on conditional random field	Local Ancestry Inference	(Maples et al., 2013)	https://github.com/slowkoni/rfmix
Loter	LAI for species other than humans	Local Ancestry Inference	(Dias-Alves et al., n.d.)	https://github.com/bcm-uga/Loter
GHap (R package)	Unsupervised LAI	Local Ancestry Inference	(Utsunomiya et al., 2020)	https://cran.r-project.org/package=GHap
PSIKO2	Unsupervised LAI	Local Ancestry Inference	(Popescu & Huber, 2015)	https://www.uea.ac.uk/computing/psiko
SWIF(r)	Probabilistic method to detect selective sweeps	Deep Learning	(Sugden et al., 2018)	https://github.com/ramachandran-lab/SWIFr

5. Conclusions

To maintain animal welfare and as a consequence productivity and production efficiency, breeds have to be well adapted to the environmental conditions in which they are kept. Rapid climate change inevitably calls for the use of various countermeasures to manage animals appropriately. Temperature mitigation methods (shaded area, water wetting, ventilation, air conditioning) are possible solutions; however, these can only be used when animals are kept in shelters and are not applicable to range-type farming systems. Most structural solutions to control the environment of animals have a high cost, and many have energy requirements that further contribute to climate change. Therefore, addressing livestock adaptation by breeding animals that are intrinsically more tolerant to extreme conditions is a more sustainable solution. Decreasing stress and increasing animal welfare is important for farmers and the general public. Animals stressed by high temperatures may be less able to cope with other stressors such as pollutants, dust, restraint, social mixing, transport, etc., that further affect welfare and productivity. Innovation in sensors and linking these into the "internet of things" (IoT) to collect and exchange data is increasing our ability to record environmental variables and animal welfare status and provide input to systems dedicated to the control of environmental conditions and provision of early warning of discomfort in individual animals. In the longer term, collecting such data will contribute to understanding the genetics underpinning tolerance and adaptation to environmental and other stressors in order to select animals better suited to different conditions. The resulting increase in efficiency will have additional benefits in terms of reducing greenhouse gas emissions, particularly methane from ruminants, which currently make a significant contribution to climate change.

Breed substitution by introducing breeds known to have particular resilience, e.g., to drought, temperature extremes or disease, may be a solution. This approach would facilitate a rapid response to climate change, although it is not ideal as breeds more tolerant of hot climates generally have low productivity. Additionally, imported breeds may not adapt to local conditions such as available feed resources and disease challenge.

Crossbreeding between highly productive and heat tolerant breeds is an approach that is currently used in tropical areas including Australia, the southern USA and Brazil, where crossing productive taurine breeds with heat adapted indicine breeds facilitates improved production in extreme conditions. Selection of these cross-bred populations has produced stable breeds that show good productivity and adaptation, such as the Brangus from the USA (do Prado Paim et al., 2020) and the Australian Droughtmaster (Francis & Little, 1964). O'Neil et al (O'Neill, Swain, & Kadarmideen, 2010) have reviewed the use of crossbred lines in tropical high tick challenge areas of Australia. However,

crossbreeding programs should be properly planned, organised and monitored, as indiscriminate crosses may cause the genetic erosion of local breeds and the loss of their adaptation.

Accelerated selection for thermal tolerance and resilience to new endemic diseases is also a possible sustainable solution. In this case, genomics plays a key role together with phenotype recording and the collection of epidemiological and environmental data. Research is approaching the challenging task of identifying genes having adaptive value using a range of methods, including those described in this review. Specific variants of major genes exist in local genetic resources, as demonstrated by the SLICK mutation associated with heat tolerance. However, identifying causal genes and variants is difficult, requiring large data sets which are often not available or affordable for livestock, and a focused effort to refine and test candidate genes. Therefore, most studies have simply localized genetic effects to chromosomal regions or quantitative trait loci (QTL) in genome-wide association studies. Additionally, it is now clear that most adaptation traits have complex genetic control, making the genetic basis difficult to unravel. Nevertheless, markers having significant effects can be used in selection programmes using marker assisted selection or by weighting particular SNPs within QTL regions in genomic selection estimates. Although genomics is presently only scratching the surface of the control mechanism of these traits, comparison between methods, studies, breeds and even species is starting to reveal that morphology, energy and lipid metabolism, and the immune system are key factors in adaptation, with some genes being consistently identified as carrying variants modulating adaptation. The identification of these genes confirms the importance of the conservation of local genetic resources as reservoirs of useful alleles. The evaluation and improvement of these breeds or the transfer of adaptive variants into highly selected breeds are the next steps to better match livestock to harsh conditions while maintaining productivity. These steps may be accelerated by marker-assisted or genomic selection, and even more rapidly by novel tools such as gene editing where such approaches are socially accepted. Parallel breeding for adaptation to climate change and the mitigation of the impact of livestock on climate change is probably the hardest challenge that the livestock sector has ever faced, but it is now urgent. The challenge can only be won if research, industry, decision makers and funders join forces with the objective of satisfying the rights future generations to a healthy diet and a clean planet.

Chapter 5: Deciphering climate-mediated adaptation in European sheep

Barbato M, Vajana E, Selmoni O, Somenzi E, Eufemi E, Del Corvo M, Tixier-Boichard M, Joost S, Lazzari B, Colli L, Stella A, Ajmone Marsan P, and the IMAGE Consortium

Abstract

After livestock domestication and dispersal, environment-mediated selective pressure has shaped phenotypic variation and left specific signatures in the genome of locally adapted breeds. The identification of genes containing adaptive variation is of strategic importance for the livestock sector, especially to prioritise genomic resources for conservation and to guide selection in a period of rapid climate change. Among domesticated species, sheep (*Ovis aries*) have established a wide geographic range due to their rusticity, manageable size, adaptability to nutrient poor diets, and tolerance to highly diverse environments and climatic conditions. Here, we used high-density SNP genotyping data from 80 autochthonous sheep breeds, spanning from North Africa to Scandinavia to sample a wide range of climatic conditions. We investigated the structuring of genetic diversity, and combined gene-environment association and selection signature analyses to pinpoint genomic regions candidate for adaptation to different climates. Common signals between the two methods pointed at genes involved in adaptation to extreme environments (ARHGAP26, CMYA5) and energy management (STARD13, SOCS2). Further, we identified genes that play a role in modulating thermoregulation through fat deposition, skin thickness, and hair follicle differentiation as the homeobox genes HOXC11, HOXC12, and HOXC13. Finally, we argue that the expected increase in temperatures will represent the main adaptive challenge for the species in the next few decades, given its current genetic makeup. Our findings shed light on the genomic mechanisms of climate adaptation in sheep and provide molecular information useful to allow selection for improved resilience and welfare under climate change as well as to inform resource allocation for conservation.

Introduction

Climate change

The global request for livestock products is predicted to double by 2050 due to increased human population and higher standards of living, and animal production will need to increase accordingly to meet the demand (Bernabucci, 2019; Hoffmann, 2013). However, the effects of global climate change (GCC) are predicted to challenge livestock capacity to adapt to new conditions, and ultimately to threaten food security by reducing the performance of agricultural production systems (based on grazing, mixed farming systems, or industrialised systems), along with affecting ecosystem functioning

and services (Boettcher et al., 2015; Pauls, Nowak, Bálint, & Pfenninger, 2013; Thomas et al., 2004). Negative effects such as a change in the thermal comfort in tropical zones are becoming crucial also for livestock reared in temperate zones, where increasing temperatures are challenging animal welfare (Bernabucci, 2019). This is particularly true for small ruminants such as sheep and goat, where many physiological mechanisms are influenced by seasonality due to daylight length, air temperature, and precipitation (Tsartsianidou et al., 2021).

Adaptation

Genetic variation is fundamental for a species to adapt to climatic stress and thus maintain viability in the long term (Y.-H. Cao et al., 2020). Domestic species represent an ideal biological system to investigate the genetic bases of adaptation leveraging the availability of exhaustive genome sampling, well-annotated genomes, and increasingly precise environmental databases. Domestic species originated from circumscribed domestication centres, then colonised a wide range of new environments, and finally differentiated due to the combined action of natural and artificial selection, and genetic drift (Flori et al., 2019). The adaptation of livestock breeds to local climatic conditions plays a major role in contemporary agriculture as it reduces the environmental stress on animals, which in turn reflects on increased yields and reduced environmental impact of the production (Rojas-Downing, Nejadhashemi, Harrigan, & Woznicki, 2017).

The effect of increasing temperature and seasonal weather variability are known to vary by geographic region, animal species and production type (Tsartsianidou et al., 2021). The Mediterranean basin has been identified as one of the most vulnerable regions for climate change, with a predicted significant increase of average atmospheric temperature, thermal inter-annual variability, increased drought especially during summer, and land-use change (Sala et al., 2000; Segnalini, Bernabucci, Vitali, Nardone, & Lacetera, 2013; Tsartsianidou et al., 2021).

The presence of environmental heterogeneity promotes the genetic differentiation of local populations towards traits providing an advantage under the given local environmental conditions, regardless of the consequences of these traits for fitness in other habitats (Kawecki & Ebert, 2004). The acquisition of fitness-enhancing traits due to genotype *x* environment interaction is referred to as local adaptation (Barbato, Hailer, Orozco-terWengel, et al., 2017; Barbato et al., 2020; Cortellari, Barbato, Talenti, Bionda, Carta, et al., 2021; Kawecki & Ebert, 2004). Along with divergent natural selection, gene-flow and genetic drift can affect/modulate local adaptation, providing new variation (adaptive introgression) or removing putatively favourable variants, respectively (Barbato, Hailer, Orozco-terWengel, et al., 2017; Barbato et al., 2020; Hudson et al., 2020; Kawecki & Ebert, 2004; Stronen, Pertoldi, Iacolina, Kadarmideen, & Kristensen, 2019).

Local domestic populations are often managed through traditional husbandry and are generally better adapted to local environments than specialised and cosmopolitan breeds, where the high

management mitigates the selective pressure of the local environment. Among livestock, small ruminants such as sheep (*Ovis aries*) and goat (*Capra hircus*) are often reared within low-management systems (Alberto et al., 2018; Cortellari, Barbato, Talenti, Bionda, Carta, et al., 2021). Sheep domestication occurred in the Fertile Crescent between the Neolithic and the Upper Palaeolithic (Sanna et al., 2015; J.-D. Vigne, Carrère, Briois, & Guilaine, 2011; Zeder, 2008). After domestication sheep were spread across the world following human migration and trade and reared under different environmental, management, and selection conditions, leading to >1,400 breeds adapted to a wide range of climates, pathogen challenges, and production systems (Y.-H. Cao et al., 2020; Deniskova et al., 2018), hence representing an attractive biological model to study the genetic adaptation to the environment.

Currently, high-density SNP arrays are available for most domestic species (Bruford, Ginja, Hoffmann, Joost, Orozco-terWengel, et al., 2015). Combining genome-wide scans for regions under selection with the identification of the genes harboured within such regions proved effective in identifying both common patterns and specific strategies of genomic adaptation to environmental stressors in several livestock species (Joost et al., 2007; Passamonti et al., 2021; Vajana et al., 2018).

Herein, we investigated a novel, large georeferenced dataset of genome-wide genotypes of local sheep breeds spanning from Scandinavia to North Africa and Iran. We collected current climatic data and combined them with genotype information to detect the genomic regions responsible for climate adaptation using selection and landscape genomics analysis. Additionally, we evaluated the adaptive challenge that the European sheep breeds will face under the future climatic change.

Materials and Methods

Environmental characterization

Environmental clusters. To outline contrasting environmental regions and optimise the chances of including genetic variants underlying local adaptation in the working dataset, environmental information was derived at 1,616 sampling locations across Europe and North Africa (Figure 1) including nineteen bioclimatic descriptors (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) and altitude (Ryan et al., 2009) at 5 and 0.1 km resolution, respectively. Principal component analysis (PCA) was applied on the derived dataset to obtain a few synthetic variables accounting for the majority of the environmental variance in the European continent. Finally, hierarchical clustering on principal components (HCPC) was applied to individuate the main environmental clusters and assign each sampling site to a specific group. *Quantum GIS* (QGIS) v.2.18.13 was used to retrieve and manage environmental information(QGIS development team, 2009), and R v.3.5 was used to perform both PCA and HCPC (R Core Team, 2021).

Gene-environment association analysis. Values for a total of 56 environmental variables were retrieved at the sampling locations from the Worldclim v.2.1 dataset at 1 km resolution (Fick & Hijmans, 2017). The retrieved information included 19 bioclimatic variables, monthly precipitation, monthly minimum and maximum temperatures, as well as altitude. In order to quantify heat stress at sampling sites ((Bohmanova, Misztal, & Cole, 2007), monthly relative humidity, mean temperature and diurnal temperature range were derived from the Climate Research Unit database v.2.0 (spatial resolution: ~340 km^2; temporal resolution: 1961-1990; (New, Lister, Hulme, & Makin, 2002), and used to compute monthly temperature humidity index (THI) values following the Lallo et al. (Lallo et al., 2018) formula for ruminants. The average between the THI values of July and August (i.e., the months at highest heat stress in the European continent) were computed and used in the gene-environment association (GEA) analysis (see section 'Latent factor mixed models').

Climate change. To quantify aridity across the study area, the Thornthwaite aridity index was derived from ENVIREM database (Pascal O. Title & Bemmels, 2018) at ~5 km^2 resolution and for the years 1970-2000. Future climatic data were downloaded from Worldclim v.2.1 as based on the BCC-CSM2-MR global climate model and the shared socio-economic pathway (SSP) 5-8.5 (i.e., the worst-case global change scenario currently prospected) for the years 2081-2100. The Thornthwaite aridity index was projected through the 'envirem' R package following instructions available at http://envirem.github.io/ENVIREM_tutorial.html. The R package 'raster' v.3.3-13 (Hijmans, 2020) was used to manage geospatial information and to retrieve climatic features at sampling locations.

Genotyping and quality control

Samples were genotyped at ~600k single nucleotide polymorphisms (SNPs) with the Ovine Infinium HD SNP BeadChip (Illumina). Markers located on sex chromosomes or with unknown map positions, or with a call rate <95% and minor allele frequency (MAF) <1%, were removed. Pairwise genome-wide identity-by-descent (IBD) estimates were obtained through the method-of-moments and one individual per pair showing IBD >0.08 was removed. Linkage disequilibrium (LD) pruning was performed using the '--indep-pairwise' function in PLINK v.1.9 (C. C. Chang et al., 2015), where SNPs with r^2 >0.2 were removed from sliding windows of 2,000kb and a step size 200kb. Only breeds with more than five individuals were retained for subsequent analyses.

Latent factor mixed models

Gene-environment associations tests were devised based on the latent factor mixed model (LFMM) approach as implemented in the R package 'lfmm' v.2.0 (Caye et al., 2019). This package allows for a computationally-efficient implementation of GEA models which is based on two steps: first, a preliminary structure analysis is performed to determine the most likely number of ancestral

populations (*K*) in the genetic dataset to adequately account for the potentially confounding effect of demography in GEAs; then, latent (demographic) factors are estimated based on the genetic dataset through PCA, and GEAs are devised by correcting genotype-environment associations for *K* principal components within a multiple regression framework. Both environment and demography are treated as fixed effects, and effect sizes (i.e., regression coefficients) are estimated for each covariate. Under the null expectation of no spatial relationship among genotypes and environment, regression coefficients are assumed to be distributed following a Student distribution with degrees of freedom equal to $n-K-1$, n being the number of individuals used in the test. Here, we used the R package 'LEA' and its implementation of sNMF (sparse non-negative matrix factorization algorithms; (Frichot & Francois, 2015) to determine the most likely *K* as based on the cross-entropy (CE) criterion. This analysis was run with default parameterization from two to 50 ancestral populations assumed. Regularised least squares estimates of the effect sizes were obtained through the 'lfmm_ridge' function, where *K* demographic factors were used to correct GEA tests. *P-values* were obtained with the 'lfmm_test' function for all tests among loci and environmental covariates, and calibrated using the genomic control method. Finally, false discovery rate (FDR) control was applied separately for each environmental covariate, and *p*-values were translated to *q*-values through the R package 'qvalue' (Storey & Tibshirani, 2003). GEA tests were deemed to be statistically significant at a 5% nominal FDR cut-off.

hapFLK

As a complementary analysis to LFMM we performed a genome-wide scan of selective sweep between pools of breeds inhabiting contrasting eco-climatic regions for THI, temperature and precipitation. For each climatic variable, we selected those animals living in areas within the top and bottom 25% of the variable distribution, hence representing two pools of individuals experiencing extremely contrasting climatic conditions (high *vs* low THI, hot *vs* cold, and dry *vs* wet; Supplementary Fig S2).

To maximise the selection signals putatively related with environmental adaptation, while reducing the signals involved with species differentiation, we performed three analyses for each climatic variable using the individuals within the right tail of the variable distribution (TOP), the left tail (BOT), and pooling the two together (ALL).

For related breeds which developed in the same environment, the selection analysis will mostly pick up those signals differentiating the breeds, but should not as easily identify environment-driven selection as these breeds likely underwent parallel selection for that specific selective pressure. Hence, we interpreted those selection signals present in ALL but absent in either TOP or BOT as likely due to the contrasting environment, rather than breed differentiation. To identify selection sweeps we used hapFLK v1.4 (Fariello, Boitard, Naya, SanCristobal, & Servin, 2013). hapFLK detects selection based on haplotype frequency differences among populations and uses hierarchical structures of

sampled populations into consideration to discriminate selection from drift. The genome-wide hapFLK analysis was run on each chromosome separately using the clustering parameter −K 20 and the hapFLK scores fitted to a χ2 distribution to obtain *p*-values (Fariello et al., 2013) and translated to *q*-value using the R function 'qvalue'. As for the latent factor mixed models, loci having *q*-value <5% were considered significant.

Gene annotation

For each significant locus identified through LFMM or hapFLK, we selected all genes located inside or within a flanking region of significance (FRS). We defined a FRS as the genomic region flanking the focal locus both downstream and upstream. To determine each flanking region size, the halved distance between a SNP under selection and the closest SNP was used. The maximum flanking region size was set to 25 kbp (corresponding to the interlocus distance at which the LD halves) on both ends. The 'oaries_gene_ensembl' dataset from the Ensembl genome browser 105 database was inspected to search for annotated genes within the FRSs identified (Howe et al., 2021). Quests for gene annotation were automated through the 'getBM' function from the 'biomaRt' R package (Durinck, Spellman, Birney, & Huber, 2009) and results were further complemented with questing the Uniprot database (Bateman et al., 2021).

Climate change and adaptive challenge in sheep

We assumed the individuals sampled in extreme environments to carry genetic variants conferring adaptation to harsh conditions. Then, we defined two reference groups as composed by (i) sheep experiencing extreme temperatures in the warmest period of the year (i.e., BIO10>25°C), and (ii) sheep living in geographic areas at severe water deficiency (Thornthwaite aridity index >80; (Thornthwaite, 1948)). Animals exposed to climate change (hereafter, vulnerable population) were outlined by projecting temperature and aridity into the future and identifying the areas where climatic conditions will become as extreme as those currently experienced by the reference groups. Then, the subsets of SNPs found to be linked with thermal and drought stress were used to quantify genetic similarity between reference and vulnerable populations by computing pairwise identity-by-state (IBS) measures. Finally, a beta regression analysis was performed to test the potential association between IBS values and the predicted climatic stress. IBS and beta regression were computed using PLINK v.1.9 and the 'betamfx' function from the R package 'mfx' v.1.2.2, respectively (Cribari-Neto & Zeileis, 2010; Fernihough, 2014).

Results

Environmental characterization and quality control

Following HCPC analysis, eight environmental clusters were identified across Europe and North Africa (Figure 1). Pruning for MAF, genotype/individual call rates, IBD, LD, and removing individuals with missing environmental information left 167,834 SNPs and 1,112 individuals from 54 breeds.

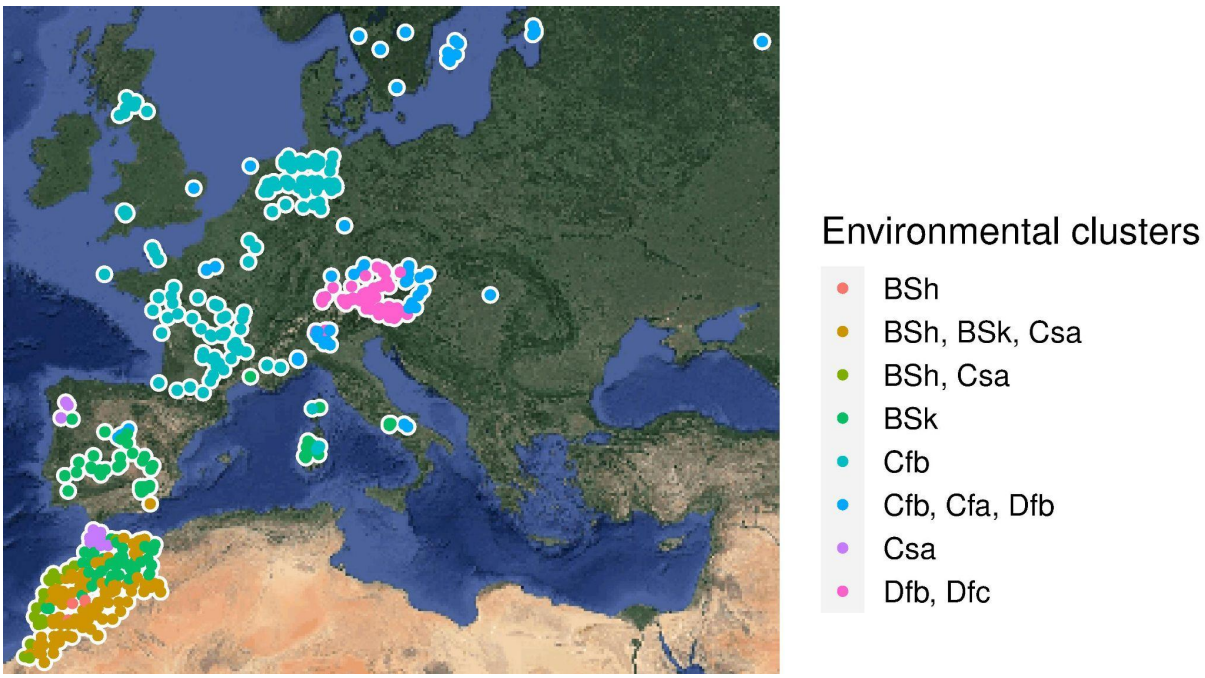


Figure 1. Sampling locations across Europe and North Africa and their environmental characterization. Sampling sites (i.e., individuals) belong to eight contrasting environmental clusters that correspond to specific climatic areas of the Köppen-Geiger classification (BSh: semi-desert; BSk: semi-arid; Csa: Mediterranean with dry summer; Cfa: humid subtropical; Cfb: oceanic; Dfb: humid continental; Dfc: subarctic).

Latent Factor Mixed Models

The cross-entropy criterion was unable to identify a univocal *K* to correct GEAs for neutral population structure (Supplementary figure 1). Hence, we chose to apply multiple *K*-based corrections in independent genome scans: *K*=19 (where the CE curve transitioned to plateau), *K*=28 (in proximity of the numerical minimum), and *K*=23 (intermediate value between the previous ones).

A total of 217, 175 and 157 SNPs were found to be involved into significant GEA models at the 5% FDR cut-off when correcting for nineteen, 23 and 28 latent factors, respectively (Table 1a). A core subset of 126 SNPs resulted to be shared across all analyses with 77.5% of the associations being recorded with precipitation-related variables, 16.3% with temperature-related environmental variables and 3.1% with heat stress and altitude, respectively.

In turn, functional annotation was found for 132, 100, and 93 genes after correcting for *K*=19, *K*=23 and *K*=28 latent factors, respectively (Table 1b). Among the annotated genes, 72 resulted to be shared by all analyses, with 80.6% of these genes showing an association with precipitation-related variables, 12.5% with temperature, 4.2% with heat stress, and 2.8% with both temperature and precipitation. No candidate gene was found to be related with adaptation to altitude.

Among the genes intercepted (Table 1c), two were previously associated with feed efficiency in ruminants: GFRAL and INSRR. GFRAL is a glial cell-derived neurotrophic factor (GDNF) receptor alpha-like. GFRAL is a brainstem-restricted receptor for food intake, energy expenditure and body weight in response to metabolic and toxin-induced stresses, with GFRAL knockout mice showing increased resistance to body weight loss (Hsu et al., 2017). Further, GFRAL has been found to be associated with carcass quality traits in cattle (Lee et al., 2012). INSRR also known as IRR (insulin receptor-related receptor), is a member of the insulin receptor family, which includes its homologs insulin-like growth factor receptor (IGF-IR) and has been associated with residual feed efficiency in cattle (Lam et al., 2021).

Genes previously associated with adaptation to climate were identified. CEP290 has been found associated with thermotolerance in cattle (Durbin, Lu, Yampara-Iquise, Miller, & Decker, 2020). Mutations in CEP290 distort photoreceptors, affecting an animal's ability to detect changes in seasons (Durbin et al., 2020). CEP290 was also found under selection in Alpine cattle breeds (Strillacci et al., 2020). GAS2L3 and LALBA were found associated with milk yield and quality in indigenous breeds living in harsh conditions (Abousoliman et al., 2021). Similarly, we found ATP5PO and ACBD6, which have been found to be selected in cattle inhabiting in tropical and cold climates, respectively (Kava et al., 2021).

Noticeably, eight out of the 72 genes identified have been described to play a role in fat deposition in sheep, namely: KCNS1, ZNF410, PCBP1, ALX4, PNPLA8, NDUFS4, NFIA, MKX, and LONRF1. KCNS1 and the Zinc finger protein 410 (ZNF410) have been previously associated with tail fat deposition in sheep (Salvatore Mastrangelo et al., 2019). Similarly, PCBP1 has been described for its association with tail type in sheep (Bakhtiarizadeh & Salami, 2019; Z. Yuan et al., 2017). ALX4 is known to bind HOX clusters proteins, is involved in the development of limbs and skeletal morphology (Fariello et al., 2013), and has also been reported in experiments differentiating thin *vs* fat tailed sheep (F. Zhao et al., 2020). A patatin-like phospholipase domain-containing gene (PNPLA8) was found associated with precipitation-related variables. PNPLA-like genes catalyse the hydrolysis of triglycerides, and function in triglyceride lipase metabolism in the body, especially in adipose tissue (Holmes & Rout, 2011). NDUFS4 has also been associated with fat formation and metabolism in sheep (Zhu, Cheng, Li, Liu, & Ma, 2021). NFIA, has been described as a positive transcriptional regulator of adipogenesis associated with brown adipose tissue (Hiraike et al., 2017). MKX and LONRF1 are known to play a role in cattle in the expression of adipocytes and intramuscular fat deposition, respectively (Hudson et al., 2020; Seong, Yoon, & Kong, 2016). The latter has also been described to be involved in seasonal fertility in sheep (Martinez-Royo et al., 2017). In fact, we identified several genes previously associated with fertility as GLIS3, GARNL3, NTRK1, and FIG4 (Mohammadi, Alijani, Rafat, & Abdollahi-Arpanahi, 2020;

K. Wang et al., 2021; Zehu Yuan et al., 2021, 2019) and reproductive seasonality (Lakhssassi et al., 2021).

Among the genes identified, four were previously associated with hair characteristics, as KRT80 and ZBTB8A, which have been associated with hair characteristics in alpaca, sheep, goat (Fernández Suárez, Gutiérrez Reynoso, & Ponce de León Bravo, 2019; H. Zhao et al., 2021), ITIH5 and RIPOR2 related to skin and follicle morphogenesis (S. Li et al., 2020).

We found five genes related with body structure and development, namely FOXP1 and SYNE1 and ITGA1 that play a role on body weight and muscle growth (Pasandideh, Rahimi-Mianji, & Gholizadeh, 2018) (Pasandideh et al., 2018) (Y. Fan et al., 2020) and CPEB4, the JNK1-associated membrane protein JKAMP, and the aforementioned ALX4 involved in bone, muscle, fat, bones and lung development in several livestock species (Fariello et al., 2013; Meng et al., 2017; Peng et al., 2021).

Two genes, EEF1A1 and RXFP2, well known for their association with horn presence in sheep (Y.-H. Cao et al., 2020; Wiedemar & Drögemüller, 2015), and NMI, that plays a role on gastrointestinal nematode resistance in sheep (Ibeagha-Awemu et al., 2016), were found. UEVLD, a UEV and lactate/malate dehydrogenase domain that regulates glucose concentration during lactation (Ha et al., 2015), and PAPPA2, which has been described to play a critical role during the formation of different patterns in sheep lambskin (T. Wu, Wang, Jin, Lv, & Sun, 2021) as well as to play a role in milk quality (Z.-H. Chen et al., 2021).

Table 1. Number of SNPs (a) and genes (b) identified by scanning the genome of 1,112 sheep from eight environmental clusters across Europe and North Africa through latent factor mixed models. Biological function is reported for the candidate genes highlighted together with the focal environmental variables associated (c).

(a) *Gene-environment association analysis (FDR=0.05)*

Variable type	Variable	*K*=19	*K*=23	*K*=28	Shared across *K*s	Shared by type
Heat stress	THI	15	14	4	4	
Temperature	Bioclim 1-11	9	8	11	8	0
	Monthly min. temperature	15	11	14	11	
	Monthly max. temperature	4	3	8	2	
Precipitation	Bioclim 12-19	149	115	102	86	50
	Monthly precipitation	111	91	77	64	
Topographic	Altitude	5	4	4	4	
Total		217	175	157	126	

(b) *Gene annotation*

Variable type	Variable	*K*=19	*K*=23	*K*=28	Shared across *K*s	Shared by type
Heat stress	THI	17	16	3	3	
Temperature	Bioclim 1-11	5	5	8	5	0
	Monthly min. temperature	13	5	9	5	
	Monthly max. temperature	3	2	3	1	
Precipitation	Bioclim 12-19	84	63	65	54	26
	Monthly precipitation	60	47	38	32	
Topographic	Altitude	0	0	0	0	
Total		132	100	93	72	

(c) *Genes IDs and biological functions*

Variable type	Variable	Gene ID	Protein description
Heat stress			
	THI	*GFRAL*	GDNF family receptor alpha like
		CRISP1	ShKT domain-containing protein
		ENSOARG00000026926	Uncharacterized protein
Temperature			
	Bioclim 1-11	*L3HYPDH*	Trans-3-hydroxy-L-proline dehydratase (EC 4.2.1.77) (Trans-L-3-hydroxyproline dehydratase)
		JKAMP	JNK1/MAPK8-associated membrane protein isoform X1
		PAPPA2	pappalysin-2
		CENPE	Uncharacterized protein
		ENSOARG00000023415	Uncharacterized protein
	Monthly min. temperature	*W5PI72*	ATP-dependent DNA helicase 2 subunit 1
		SPIRE2	KIND domain-containing protein
		LALBA	Alpha-lactalbumin (Lactose synthase B protein)
		RIPOR2	Rho family-interacting cell polarization regulator 2
		ENSOARG00000025636	Uncharacterized protein
	Monthly max. temperature	*UEVLD*	Ubiquitin-conjugating enzyme E2 variant 3 isoform X1
Precipitation			
	Bioclim 12-19	*ACBD6*	Acyl-CoA-binding domain-containing protein 6
		AK1	Adenylate kinase isoenzyme 1 (AK 1) (EC 2.7.4.3) (EC 2.7.4.6) (ATP-AMP transphosphorylase 1) (ATP:AMP phosphotransferase) (Adenylate monophosphate kinase) (Myokinase)
		CPEB4	Cytoplasmic polyadenylation element-binding protein 4 isoform X1
		ENG	Endoglin
		COL22A1	VWFA domain-containing protein
		CENPC	Centromere protein C
		FIG4	SAC domain-containing protein
		W5PLA8	60S ribosomal protein L17
		ARFGEF2	SEC7 domain-containing protein
		EXOC6B	Exocyst complex component
		RXFP2	G_PROTEIN_RECEP_F1_2 domain-containing protein
		ATP5PO	ATP synthase peripheral stalk subunit OSCP (ATP synthase subunit O, mitochondrial) (Oligomycin sensitivity conferral protein)
		MKX	Homeobox domain-containing protein
		GTSE1	GTSE1_N domain-containing protein
		TRMU	Mitochondrial tRNA-specific 2-thiouridylase 1 (EC 2.8.1.14)
		GAS2L3	GAS2-like protein 3
		ITIH5	Inter-alpha-trypsin inhibitor heavy chain H5
		EEF1A1	Elongation factor 1-alpha
		LRRIQ1	Leucine-rich repeat and IQ domain-containing protein 1 isoform X3
		MIA2	Melanoma inhibitory activity protein 2 isoform X6
		NFIA	Nuclear factor 1
		PDE5A	Phosphodiesterase (EC 3.1.4.-)
		SYNE1	Nesprin-1 isoform X4
		ZBTB8A	Zinc finger and BTB domain-containing protein 8A
		C6H4orf19	Uncharacterized protein
		PPFIA1	Uncharacterized protein
		ENSOARG00000018485	Uncharacterized protein
		ENSOARG00000026220	Uncharacterized protein
	Monthly precipitation	*ITGA1*	Integrin subunit alpha 1
		L3HYPDH	Trans-3-hydroxy-L-proline dehydratase (EC 4.2.1.77) (Trans-L-3-hydroxyproline dehydratase)
		FOXP1	forkhead box protein P1 isoform X2
		JKAMP	JNK1/MAPK8-associated membrane protein isoform X1
		LOC114115128	U6 spliceosomal RNA
		PLVAP	Plasmalemma vesicle-associated protein
	Monthly precipitation & Bioclim 12-19	*BCL11A*	B-cell lymphoma/leukemia 11A isoform X2
		ALX4	ALX homeobox 4
		KCNS1	Potassium voltage-gated channel modifier subfamily S member 1
		INSRR	Tyrosine-protein kinase receptor (EC 2.7.10.1)
		NTRK1	Tyrosine-protein kinase receptor (EC 2.7.10.1)
		LOC101115135	Phospholipid-transporting ATPase (EC 7.6.2.1)
		W5PZX3	PH domain-containing protein
		CEP290	CEP209_CC5 domain-containing protein
		KRT80	IF rod domain-containing protein
		GARNL3	GTPase-activating Rap/Ran-GAP domain-like protein 3 isoform X1
		GLIS3	Zinc finger protein GLIS3
		NDUFS4	NADH dehydrogenase [ubiquinone] iron-sulfur protein 4, mitochondrial (Complex I-18 kDa) (NADH-ubiquinone oxidoreductase 18 kDa subunit)
		NMI	N-myc-interactor isoform X2
		PCBP1	Poly(RC)-binding protein 1
		PNPLA8	Calcium-independent phospholipase A2-gamma isoform X1
		TPST2	Protein-tyrosine sulfotransferase (EC 2.8.2.20)
		WFDC5	WAP four-disulfide core domain protein 5
		ZNF410	Zinc finger protein 410 isoform X1
		W5NYY3	Uncharacterized protein
		ZBTB7C	Uncharacterized protein
		LONRF1	Uncharacterized protein
		W5PPS9	Uncharacterized protein
		W5PS47	Uncharacterized protein
		ENSOARG00000022426	Uncharacterized protein
		ENSOARG00000023636	Uncharacterized protein
		ENSOARG00000026409	Uncharacterized protein

hapFLK

We assessed selection signals due to environmental stressors in breeds inhabiting contrasting environments (hot *vs* cold, dry *vs* wet, high *vs* low THI). We recorded 38, 35 and 36 genes putatively under selection due to environmental stressors in the hot *vs* cold, dry *vs* wet, high *vs* low THI contrasts, respectively (Table 2; Figure 2).

RPLP0 (BTA17) a housekeeping gene that encodes a ribosomal protein, was found in each of the three contrasts. A genomic region in BTA2 spanning ~100kb identified in both the hot *vs* cold and dry *vs* wet contrasts harboured eight genes (FAM214B, STOML2, PIGO, FANCG, VCP, C2H9orf131, DNAJB5, PHF24). The heat shock protein family (Hsp40) member B5 (DNAJB5) gene encodes a member of the DNAJ heat shock protein 40 family of co-chaperone proteins (Lampis et al., 2018) and plays a vital role in the stress tolerance of immune cells, especially against heat stress (Vjestica, Zhang, Liu, & Oliferenko, 2013), with molecular chaperone and anti-apoptosis effects in the maintenance of immune cell survival and internal stability (Morimoto, 1993). Valosin-containing proteins (VCP) ensure the protection of naïve proteins during their transport within the cell and were found upregulated in heat-tolerant indicine cattle (Khan et al., 2021). The heat shock protein family H member 1 (HSPH1) has been previously associated with climate adaptation in sheep, playing a role in energy homeostasis, melanocyte stimulation and immune response (Howard et al., 2014; Xiangyang Miao, Luo, & Qin, 2015; Wollenberg Valero et al., 2014; Yurchenko et al., 2019). A selection block identified on chromosome 2 pointed at several members of the homeobox C gene family (HOXC4-6, HOXC8-13). Transcriptome analysis of HOXC genes in sheep found HOXC4, HOXC6, HOXC8 and HOXC9 primarily expressed in the visceral fat, whereas HOXC10-13 were most active in the tail fat (Abied, Bagadi, et al., 2020; D. Kang et al., 2017). ARHGAP26 affects growth and intramuscular fatty acid in pigs (Edea et al., 2017), but has been found associated with climatic variables in a sheep landscape genomic study (Ahbara et al., 2019). The cardiomyopathy associated 5 (CMYA5) gene was previously reported to be associated with the cardiovascular system, and found associated with high-altitude adaptation in goats (Song et al., 2016). The TSHR and GTF2A1 genes play a role in the photoperiod control of reproduction in chickens and in the seasonal reproduction of sheep and goats (F. Zhao et al., 2020). ADARB2, UBE2Q1, CHRNB2 and NR3C1 have been found associated with fat tail deposition in sheep (Salvatore Mastrangelo et al., 2019; Z. Yuan et al., 2017; F. Zhao et al., 2020). CHRNB2 was also previously associated with high-altitude adaptation in Tibetan sheep, along with STARD13, and SOCS2 (Wei et al., 2016).

Table 2. Genes identified by the selection signature contrast analysis. The Chromosome (Chr) physical position (bp), gene name are shown in the first three columns. The asterisks in the last three columns reflect the associated environmental contrast: temperature, precipitations, and THI (TEMP, PREC, THI, respectively.

Chr	bp	Gene	TEMP	PREC	THI
1	103.2-103.3	*IL6R, SHE*		*	
		UBE2Q1			*
		CHRNB2	*		
2	52.1-52.1	*GLIPR2*			*
2	53.1-53.1	*FAM214B, STOML2, PIGO, FANCG*		*	
2	53.1-53.2	*VCP, C2H9orf131, DNAJB5, PHF24*	*	*	
3	129.7-129.7	*SOCS2*		*	*
3	129.8-130	*CRADD*			*
3	132.2-132.2	*SMUG1*			*
		HOXC4, HOXC5, HOXC6	*	*	
3	132.4-132.4	*HOXC8, HOXC9, HOXC10, HOXC11, HOXC12, HOXC13*	*	*	
3	154.6-154.7	*WIF1*	*	*	
3	163-163.1	*RNF41, SMARCC2*			*
3	209.6-209.8	*FGF6, TIGAR, CCND2*			*
4	24.6-25	*CRPPA*	*		
5	51.2-51.9	*ARHGAP26, NR3C1*		*	
6	33.7-34.5	*CCSER1*			*
6	35.9-36.3	*FAM13A, PYURF/PIGY, HERC6*	*		*
7	89.3-89.5	*TSHR, GTF2A1*	*	*	
7	89.6-89.7	*STON2*		*	
10	25.4-25.5	*SPART*			*
10	28.2-28.4	*STARD13*			*
		KL		*	
10	28.8-28.9	*N4BP2L1, BRCA2, ZAR1L*	*		
10	30-30.1	*HSPH1*	*		*
10	29.3-29.5	*EEF1A1*			*
10	30-30.1	*HSPH1*	*		*
10	30.2-30.2	*TEX26*			*
10	30.3-30.3	*MEDAG*	*		*
10	30.4-30.4	*ALOX5AP*			*
10	36.1-36.4	*CRYL1, GJB6, GJB2, GJA3, ZMYM2*			*
11	27.8-27.8	*CCDC42, PIK3R6*		*	
11	41.7-41.8	*RAB5C, KCNH4*			*
12	37.2-37.3	*PRRC2C, MYOC*	*		
12	39.8-39.8	*MIIP*	*		
13	53-53.1	*PCMTD2, NPBWR2*		*	
13	62.6-62.9	*ZNF341, CHMP4B, RALY, EIF2S2*			*
14	14.1-14.1	*VPS9D1*			*
14	14.4-14.4	*SHCBP1*	*		
15	72.8-72.9	*CD82*		*	
17	61.5-61.7	*RPL6, HECTD4*	*		
17	62.1-62.2	*GCN1*		*	
17	62.2-62.2	*RPLP0*	*	*	*
17	62.2-62.2	*PXN*		*	
20	17.4-17.4	*VEGFA*	*		
20	50.4-50.4	*FOXC1*	*		
23	43.8-43.9	*RNMT*	*		

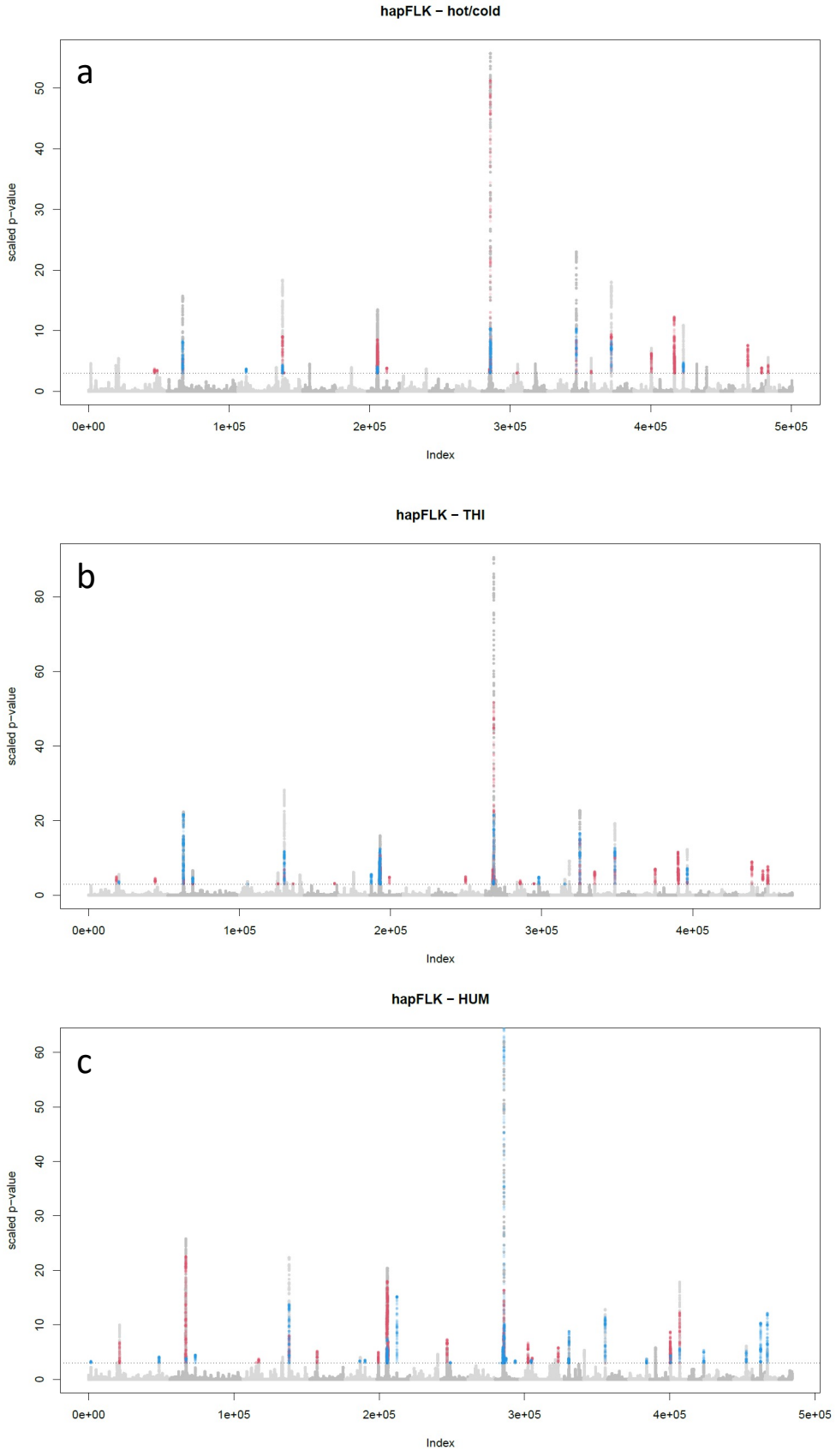


Figure 2. Manhattan plot of genome-wide hapFLK analysis performed on groups from contrasting eco-climatic regions for temperature (Fig.2a, hot vs cold), THI (Fig.2b, high vs low) and precipitation (Fig2c, dry vs wet).

Climate change and adaptive challenge

We tested the adaptive challenge faced by European sheep breeds in ~80 years, by evaluating the availability of the alleles we found associated with inhabiting harsh climates. Our results marked a lack of variability associated with adaptation to hot climate in most of the European breeds, as expressed by IBS values, with a maximum of ~36% IBS difference between the current genetic makeup and the one putatively necessary to face the predicted increase in temperature (Figure 3a-b). Furthermore, a negative association was found between climate change severity and IBS (inset of Figure 3b). Similarly, a systematic lack in variation linked with adaptation to drought was observed in the vulnerable populations with a maximum of ~34% of unshared alleles with the reference group. However, no significant change in IBS was observed among the vulnerable populations with increasing drought stress in this case (Figure 3c-d).

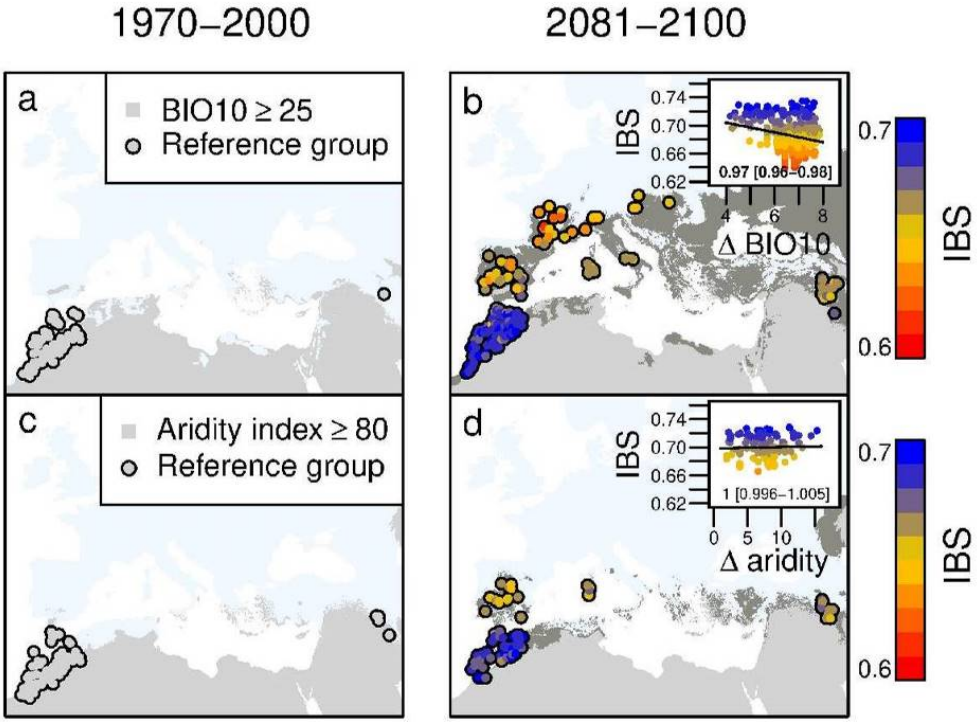


Figure 3. Predicted adaptive challenge for vulnerable sheep populations according to the BCC-CSM2-MR global climate model, SSP 5-8.5, and the candidate loci under divergent selection as highlighted by LFMM analysis. Reference populations have been defined as those currently experiencing extreme hot (a) and arid (c) conditions; vulnerable populations have been defined as those that will tackle the same extreme conditions as those currently experienced by the reference groups by 2100. Vulnerable populations are colored according to the proportion of adaptive alleles shared with the references (IBS). The higher the proportion of shared alleles the higher the IBS between reference and vulnerable populations. IBS was calculated based either on the loci associated with temperature-related variables (b) or on the loci associated with precipitation-related variables (d). A significant difference in IBS has been detected across the vulnerable populations as a function of the predicted change in temperature (b); conversely, no significant difference was found for the expected change in aridity (d). Odds ratios from beta regression analysis are reported along with confidence intervals and the regression lines.

Discussion

Climate change will have a dramatic impact on both extensive and intensive livestock systems (based on grazing, mixed farming systems, or industrialised systems). The productivity-driven selection of many commercial breeds contextually increased their sensitivity to fluctuations in the environment and feeding (reduced their plasticity). Local breeds typically lack the management infrastructure of commercial breeds but are instead adapted to the environment where they evolved. This is particularly true for sheep, which have been introduced to all inhabited continents following the post-domestication dispersal and often adapted to the most adverse environments (Alberto et al., 2018).

The effects of climate change have been described as a general increase in global temperature, combined with the intensification of extreme climatic events like extreme cooling, flooding and drought (Masson-Delmotte, 2017; Passamonti et al., 2021). Climatic modifications are altering pasture and forage crops seasonality and quality, alter the pregnancy rates of livestock (Amundson et al., 2006), and modifying the distribution of pathogens, vectors, and parasites (Bett et al., 2017). In the last decade genomic research focused on investigating the resilient and adapted local livestock and identifying the genomic make up responsible for ensuring long-term sustainability of livestock agri-food systems (Cortellari, Barbato, Talenti, Bionda, Carta, et al., 2021; Serranito et al., 2021; Tsartsianidou et al., 2021; Upadhyay et al., 2021). However, local livestock comprise several varieties, strains and ecotypes, and different databases enable different adaptive strategies to climate to be disentangled. Here, we sought to detect genomic regions putatively responsible for adaptation to climatic factors by means of GEA and selection signature analysis.

Our analysis highlighted several candidate genes already reported to be important in energy management, including feed efficiency and lipid metabolism and deposition. The ability to regulate energy homoeostasis (lipid and carbohydrate metabolism, appetite, thermogenesis) and the thermogenic activity associated with feeding are key adaptive features to survive in extreme climates. Feed efficiency reflects the ability of an animal to process food more efficiently and consequently thrive on low quality feeds (Bolormaa et al., 2013; Seif, Johnson, & Lippincott, 1979). Improved feed efficiency can help to optimise the energy expenditure towards growth and production, especially in harsh environmental conditions, which often present restricted food availability (Koluman Darcan & Silanikove, 2018). Adipose tissue is increasingly recognised to act as an endocrine gland, effectively emitting signals to regulate food intake and energy expenditure, ultimately orchestrating changes in energy balance and nutritional adaptation (Chilliard et al., 2000). Specifically, white adipose tissue can store energy in the form of lipids and serve as a long-term energy reserve to survive times of food scarcity, whereas brown adipose tissue contributes to both thermal homeostasis and energy balance by producing heat, and is widely recognised as a fundamental player in adaptation to cold (Bukowiecki, Collet, & Follea, 1982; Weldenegodguad et al., 2021).

Several of the genes related with fat deposition we detected have also been identified in research focused on sheep tail type differentiation. This is not surprising as fat-tailed sheep are mostly found in the Middle East, North Africa and Central and East Asia (S. Xu et al., 2017). The fat-tail phenotype emerged ~5,000 years ago as a human mediated adaptive response to harsh and challenging environmental conditions (Chilliard et al., 2000; Moradi, Nejati-Javaremi, Moradi-Shahrbabak, Dodds, & McEwan, 2012). The external localization of the fat allows better heat dissipation from the rest of the body, which becomes less insulated by the fat tissue. Additionally, the fat stored in the tail represents an energy store that can be mobilised in times of food scarcity (Veerasamy Sejian et al., 2017). Interestingly, one the adipose tissue-related genes we identified (NFIA) has been previously described as introgressed from argali (*Ovis ammon*) into Chinese domestic sheep breeds adapted to high altitude (Y.-H. Cao et al., 2020; Upadhyay et al., 2021). The substantial overlap between the genes we identified through GEA analysis and those previously reported in studies focused on sheep tail-type is striking and supports the role of adipose tissues in regulating energy homoeostasis and thermogenic activity as key adaptive strategies (Lv et al., 2014b; Weldenegodguad et al., 2021). Indeed, the tail type is a prominent morphological outcome of climate adaptation, easily identifiable by the first herders which selected larger, localised, fat reservoirs, whereas the adipose tissue in thin-tail phenotypes is more broadly distributed (Kalds et al., 2021).

We recorded several genes related to body shape and hair traits. Morphological adaptations as size and shape of the body, and skin and hair attributes can enhance the fitness of a population in a given climate (Abied, Bagadi, et al., 2020; Veerasamy Sejian et al., 2017). Sheep are small ruminants with a high surface area/body volume proportion; hence, more susceptible to extreme thermal stresses. Similarly, skin thickness and colour, and density of sweat glands, coupled with hair colour, length, density and diameter influence water loss due to evaporative cooling. Indeed, sheep breeds inhabiting arid and semiarid climates tend to have sparser and shorter fleeces, dense enough to screen solar radiation, whilst allowing effective cutaneous evaporative cooling (Veerasamy Sejian et al., 2017).

We investigated subsets of individuals inhabiting contrasting environments through selection signature analysis using hapFLK. Strikingly, our results highlighted several candidate genes involved in fat deposition. Further, among the genes identified in the selection test performed on the temperature contrasts we found genes of the heat shock family. Heat shock proteins (HSPs) act as chaperones and a basal expression is required for normal protein folding. Heat stress conditions increase the presence of misfolded proteins, which in turn increase the expression levels of HSPs (ref). HSPs would then bind to the hydrophobic side chains exposed by the unfolding thermally sensitive protein and prevent further denaturation. Then HSPs assists hte protein to refold into the correct

configuration, or guide the irreversibly damaged protein to enter the proteolytic cycle if folding is not possible (Veerasamy Sejian et al., 2017).

We detected no genes shared between the LFMM and hapFLK approaches (Table 1 and 2). This is in line with the usually low consensus observed between population differentiation-based approaches and GEA methods (Henriques et al., 2018; Postolache et al., 2021), possibly reflecting an antithetical vision of the role of population structure in shaping adaptive variation: a confounding factor to be removed to detect reliable signals of environmentally-driven spatially-divergent selection in the case of LFMM (Caye et al., 2019; Stucki et al., 2017) the eminent signal upon which to base our claims on natural selection across populations inhabiting contrasting habitats in the case of hapFLK (Luu, Bazin, & Blum, 2017) (Fariello et al., 2013). Therefore, LFMM and hapFLK could target genomic regions responding to different evolutionary processes and might be considered complementary rather than mutually validating approaches. Moreover, LFMM fits independent models for each locus, whereas hapFLK seeks for selection sweeps of entire haplotype blocks, ultimately smoothing out any single SNP signal which might instead be relevant in the LFMM modelling effort. Lastly, two different sample subsets were submitted to LFMM and hapFLK (all samples included in the case of LFMM, and a selection of samples living in contrasting environments in the case of hapFLK), which might lead LFMM to detect small-effect loci involved into polygenic adaptation and hapFLK to catch prominent signals of selection involving major genes.

Several European ovine populations will be affected by increased temperatures and drought in the next few decades. In particular, extreme habitat conditions like those currently experienced by sheep populations in Morocco are predicted to shift northwards and affect Iranian, Spanish, Italian, French and Balkan populations by the end of the century with poorly predictable effects for animal welfare and productivity (Figure 3b-d). Safeguarding locally adapted populations that hold unique adaptive alleles and introgressing adaptive variation to secure viability and production standards by ad-hoc breeding programs is key to mitigate the detrimental effects of climate change on livestock. Here, we identified the most vulnerable sheep populations among those analysed and estimated the amount of adaptive alleles shared with locally adapted Moroccan sheep. We detected a systematic lack of putatively adapted alleles in the vulnerable populations for the loci linked with both hot climates and aridity (Figure 3c-d). Furthermore, a significant decrease in genetic similarity was observed among locally adapted and vulnerable populations along a gradient of increasing thermal challenge (Figure 3b). This finding would point towards thermal stress as the main candidate for adaptive management through ad-hoc breeding programmes (Table 1c and 2). In particular, the *RIPOR2* and *PAPPA2* genes are both related to wintriness and involved with skin and follicle morphogenesis which might point towards their implication into physiological mechanisms of thermoregulation and homeostasis (S. Li

et al., 2020; T. Wu et al., 2021). The gene *UEVLD* was found in association with the maximal temperature in October (Figure 4), a period which usually corresponds to the females' oestrus and consequent lactation (Ha et al., 2015).

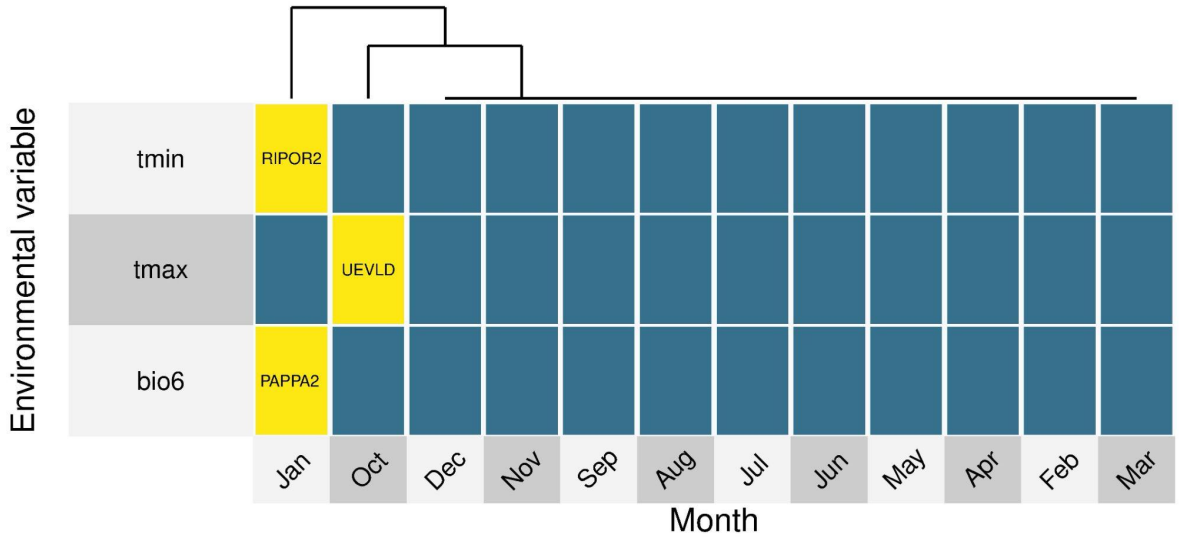


Figure 4. Three potential candidate genes for adaptive management against climate change (RIPOR2, UEVLD and PAPPA2). Associations are highlighted in yellow to locate the months (or seasons) of the year whose thermal conditions are possibly driving adaptive divergence across the studied populations.

Environmental characterization and sampling scheme

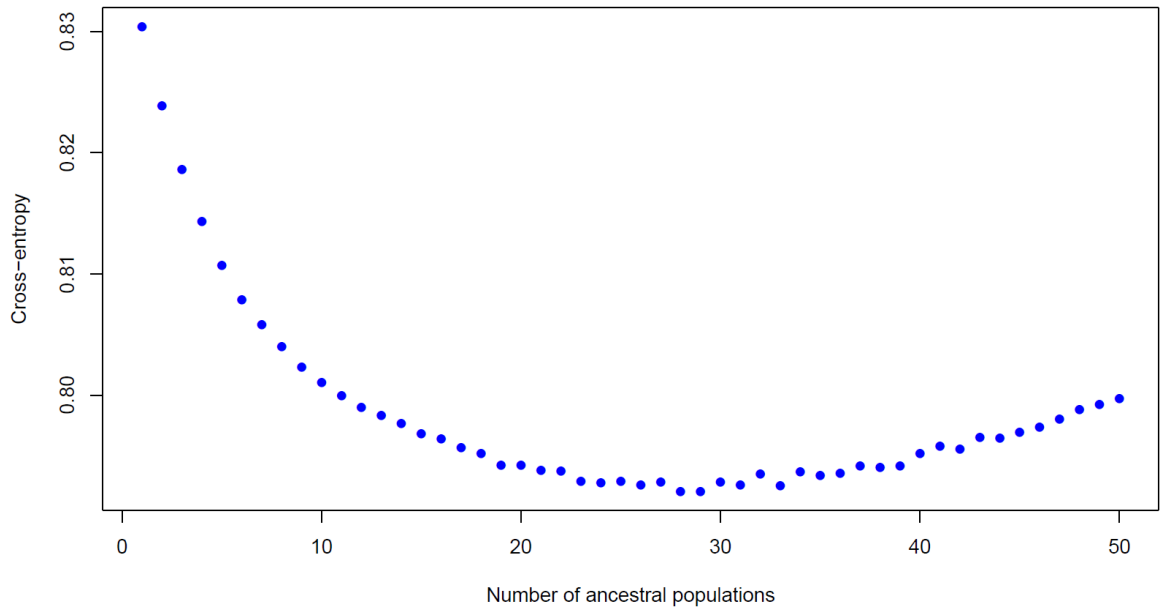
Genome representativity, spatial arrangement of sampling locations, and number of individuals sampled (sample size) are key parameters to discover reliable adaptive variants through GEA analyses. Here, we implemented an 'environmental representativeness-based' approach to characterise sampling sites, group them by habitat similarity, optimise the inclusion of environmental variability in the spatial arrangement of sampled individuals (Figure 1), and grant the associated genetic variability to be represented in the working dataset (Selmoni, Vajana, Guillaume, Rochat, & Joost, 2020). Furthermore, we coupled environmental representativeness with a comprehensive genome coverage (>160k SNPs) and large sample size (n>1,000) to gain high statistical power even in the presence of an underlying complex demography (expected true positive rate: ~100%; (Selmoni et al., 2020).

Conclusions and outlook

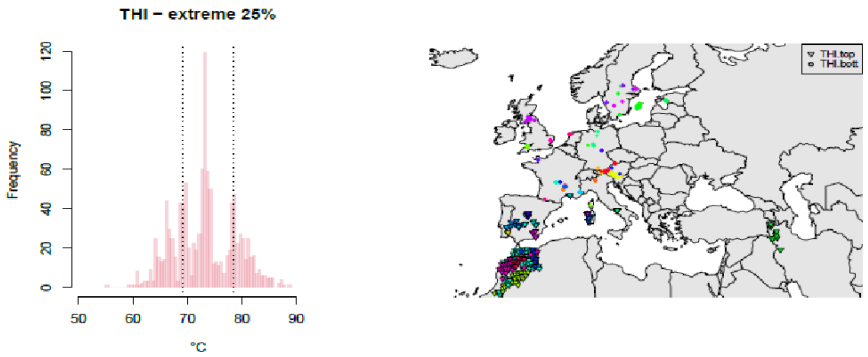
Climate change is predicted to have massive effects on sheep farming due to increased heat stress, food and water scarcity, and changing patterns in vector-borne and disease incidence. Here, we applied an integrated approach coupling gene-environment association and selection signature analyses to detect genomic signals associated with climate adaptation. We analysed several European local breeds inhabiting a wide range of environments. Local breeds are strongly intertwined and adapted with the environments where they developed. Our results showed a clear convergence on energy management as the primary driver for improved resilience against environmental stressors. In

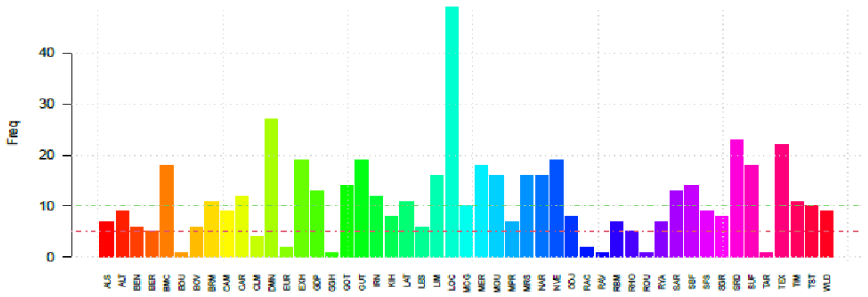
particular, we found adipose tissue metabolism to be extensively associated with the adaptation to climate, likely due to its central role in thermoregulation and energy storage. We highlighted genes which might become targets of dedicated physiological studies and targets for selection. The multifaceted adaptation strategy that emerged from our results includes low metabolic heat production through molecular optimisation and anatomic and morphologic structure as the key strategies for future adaptation to climate change.

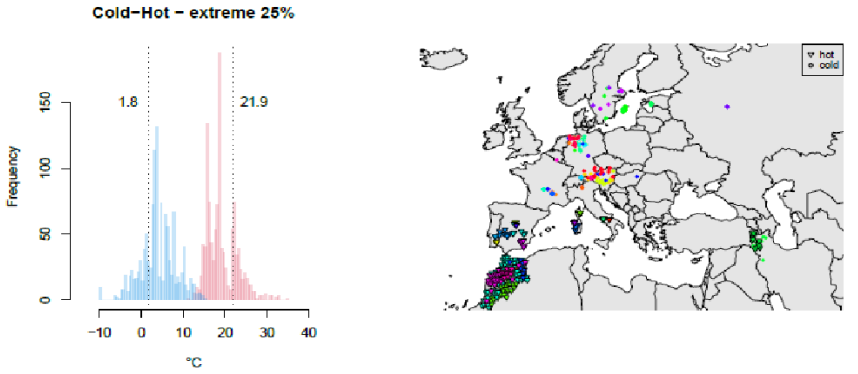
Supplementary Material

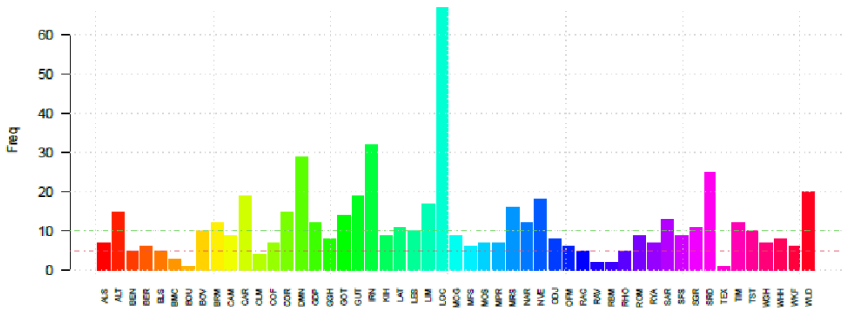


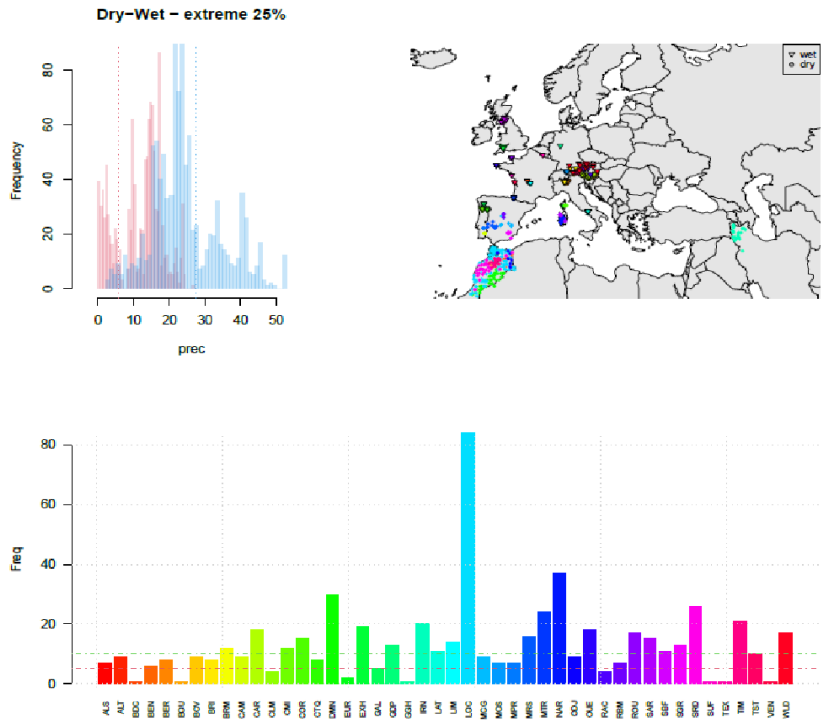
Supplementary figure 1. Cross-entropy (CE) decay observed in sNMF analysis as a function of increasing numbers of ancestral populations assumed (*K*). sNMF was run with default parametrization (one repetition per *K*, regularisation parameter equal to 10, tolerance error equal to $1 \cdot 10^{-5}$, 5% of masked genotypes when computing the cross-entropy criterion, 200 iterations) and highlighted a complex demographic scenario with a conspicuous number of ancestral populations associated with an unclear minum in the CE decay. To tackle such a demographic complexity, we decided to keep three different solutions for subsequent GEA analyses: a 'simplified' demographic scenario with 19 ancestral populations roughly corresponding to the knee in the CE curve; a 'complex' scenario with 28 ancestral populations corresponding to the CE numerical minimum; an 'intermediate' scenario with 23 ancestral populations.


THI – extreme 25%
Frequency
°C
THI.top
THI.bott


Freq


Cold−Hot – extreme 25%
1.8
21.9
Frequency
°C
hot
cold


Freq



Supplementary Figure S2. Sample selection for the selection signature contrasts.

Conclusions

Summary

In this thesis the following subjects have been addressed:

In Chapter 1 panels of ancestry informative markers (AIMs) were identified to assess admixture between feral and domestic sheep. Various sizes of AIM panels, accounting less than 100 SNPs, were able to identify the ancestry proportion with the same accuracy as a medium-density SNP array. These AIMs panels can be used to exclude hybrids from reintroduction in conservation programmes at a lower cost than the larger SNP panels. The two-step selection algorithm developed together with the species specific AIM panels could be used to detect hybrids in other wildlife and domestic species.

Chapter 2 investigated the evolutionary history and diversity of the Montecristo goat, an endangered feral population living on the homonymous island in the Tuscan Archipelago. Results highlighted that the *in situ* population shares its ancestry with breeds from the surrounding areas of Sardinia, Corsica, and Tuscany and carries signatures of an ancient bottleneck/founder effect but not of recent inbreeding. Conversely, the molecular diversity of the *ex situ* population seems to have been severely impacted by recent extensive inbreeding, probably the result of a sub-optimal conservation program design. These results can be considered as a starting point for the implementation of marker-assisted monitoring to preserve the genomic heritage of the feral goats of Montecristo.

In Chapter 3 individual ROH number and length in goat breeds from Northern and Central-southern Italy were assessed to reveal regions where loss of diversity indicate sub-optimal breding stratgies, and inbreeding. Additionally, the loss of heterozigosity allowed us to identify regions of the genome under selection. Results suggest that ROH assessment can be used to regularly check if selection in a population is leading to an increase in its average homozygosity and inbreeding, and thus indicate whether a fine-tuning of the breeding scheme is necessary.

Chapter 4 reviewed advances in the methodologies used for studying livestock genomes and the impact of environmental conditions on animal production. The review concluded that in the face of climate change, to ensure animal welfare and to maintain productivity, that livestock adaptation should be addressed by breeding animals that are intrinsically more tolerant to extreme conditions. This can be assisted by identifying genes controlling adaptation traits. Many of the favourable alleles for adaptation are found in local breeds confirming the importance of the conservation of genetic resources.

A practical example of an investigation seeking adaptation-related genes is given in chapter 5. SNP genotype data from 80 autochthonous sheep breeds, spanning North Africa to Scandinavia, allowed the identification of genotype-environment associations and selection signatures of regions with environmental adaptation gene. These findings can provide molecular information for selection and to guide resource allocation for conservation.

Biodiversity and adaptation of sheep and goats: limits and future directions

Preserving biodiversity in human-dominated environments, such as rural areas, not only involves the preservation of farm animal genetic resources, but also includes the responsibility of safeguarding wildlife genomic heritage. Hybridization between domestic animals and their wild counterpart is a multifaceted issue affecting several species (see Chapter 1 - Introduction). The approach proposed in Chapter 1 is to use genomic data to detect hybridization between wild and domestic species. Results highlighted that the program developed to simulate hybrid genomes generates simplified admixture patterns, with respect to those occurring in real populations. Thus, the software requires further refinement before being applied to guiding conservation programmes. The algorithm proposed in Chapter 1 to identify ancestry informative markers can be used to identify small marker sets and improve conservation strategies for biodiversity preservation. The Montecristo feral goat (Chapter 2) is a good example of the importance of using genomic information in the design and monitoring conservation plans. In this case was found that the *ex situ* conservation population did not adequately reflect the natural population. Sub-optimal management for several generations has led to extremely high levels of inbreeding and the consequent irreversible loss of the original genetic make-up. An ongoing monitoring of molecular diversity parameters would have helped to avoid the genetic erosion that occurred. The conservation value of the *ex situ nucleus* is therefore questionable. The assessment of inbreeding level in Italian goat breeds, as presented in Chapter 3, identified some breeds with high inbreeding levels and reduced population size. These results may help farmers and breeders' associations to adjust mating plans.

Livestock adaptation was investigated in this thesis from a theoretical point of view, through the Review in Chapter 4, and by assessing different approaches for analysing genomic data of goat (Chapter 3) and sheep (Chapter 5). These approaches led to the detection of diverse types of selective pressures shaping genomes. The analysis of genomic regions with loss of heterozygosity highlighted signatures of selection for productivity-related traits (such as fertility, growth and milk production) and environmental adaptation traits. Conversely, the landscape genomic approach used in Chapter 5, was specifically focused on the detection of adaptation related to environmental variables. Findings pinpointed that different approaches can give complementary information on adaptation. Taken together, outcomes from different investigations can give a more complete representation of the results of both human and environmental mediated pressures on small ruminant genomes. When the geographical coordinates of sampling are available, the detection of adaptation-related genes can be further improved, as seen in analysis of Italian goat breeds using a landscape genomic approach in Chapter 5. The further validation of these genes would help in elucidating genomic mechanisms of

climate adaptation and thus provide molecular information to guide the selection of animals for improved resilience under climate change.

General conclusions

The main objective of this thesis was to actively contribute to the characterization of sheep and goat genetic resources. This goal was achieved through the comprehensive molecular analyses of more than 80 different local sheep breeds, 30 local goats breeds and an endangered feral goat population. Taken together the findings presented in this thesis demonstrate how genomics provides a useful tool for livestock characterization and the evaluation of genetic resources. In addition, the genomic studies are a source of information on livestock biology and gene variants that may be useful to the improvement of highly selected breeds. Results underline the importance of the conservation of livestock genetic resources. To prevent genetic erosion, and the consequent irreversible loss of precious adaptive features, a detailed knowledge of the genetic make-up of breed is crucial. The ongoing monitoring of genomic diversity parameters across generations, and an understanding of the origin and history of endangered breeds is an essential starting point to correctly plan resettlement strategies. The information in this thesis contributes to the scientific community understanding of the diversity and uniqueness of small ruminants, and provides a guide to what is needed to ensure the preservation of their invaluable genomic heritage.

References

Abied, A., Bagadi, A., Bordbar, F., Pu, Y., Augustino, S. M. A., Xue, X., … Zhao, Q. (2020). Genomic Diversity, Population Structure, and Signature of Selection in Five Chinese Native Sheep Breeds Adapted to Extreme Environments. *Genes*, *11*(494), 1–16.

Abied, A., Xu, L., Sahlu, B. W., Xing, F., Ahbara, A., Pu, Y., … Ma, Y. (2020). Genome-wide analysis revealed homozygosity and demographic history of five chinese sheep breeds adapted to different environments. *Genes*, *11*(12), 1–16. doi:10.3390/genes11121480

Abousoliman, I., Reyer, H., Oster, M., Murani, E., Mohamed, I., & Wimmers, K. (2021). Genome-wide snp analysis for milk performance traits in indigenous sheep: A case study in the egyptian barki sheep. *Animals*, *11*(6), 1–9. doi:10.3390/ani11061671

Affymetrix Inc. (2020). Mitigating Sequencing Errors, Monomorphs, and Poor Performing Markers during de novo SNP Selection for Genotyping Applications. Retrieved January 17, 2022, from https://assets.thermofisher.com/TFS-Assets/LSG/brochures/mitigating_genotyping_appnote.pdf

Ahbara, A., Bahbahani, H., Almathen, F., Abri, M. Al, Agoub, M. O., Abeba, A., … Mwacharo, J. M. (2019). Genome-wide variation, candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep. *Frontiers in Genetics*, *10*(JAN), 1–21. doi:10.3389/fgene.2018.00699

Ajmone-Marsan, P. (2010). A global view of livestock biodiversity and conservation - GLOBALDIV. *Animal Genetics*, *41*(SUPPL. 1), 1–5. doi:10.1111/j.1365-2052.2010.02036.x

Al Kalaldeh, M., Gibson, J., Lee, S. H., Gondro, C., & Van Der Werf, J. H. J. (2019). Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. *Genetics Selection Evolution*, *51*(1), 1–18. doi:10.1186/s12711-019-0479-1

Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., … Pompanon, F. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, *9*(1), 813. doi:10.1038/s41467-018-03206-y

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. doi:10.1101/gr.094052.109

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology and Evolution*. doi:10.1016/S0169-5347(01)02290-X

Álvarez, I., Fernández, I., Traoré, A., Pérez-Pardal, L., Menéndez-Arias, N. A., & Goyache, F. (2020a). Ancient homozygosity segments in West African djallonké sheep inform on the genomic impact of livestock adaptation to the environment. *Animals*, *10*(7), 1–17. doi:10.3390/ani10071178

Álvarez, I., Fernández, I., Traoré, A., Pérez-Pardal, L., Menéndez-Arias, N. A., & Goyache, F. (2020b). Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. *Scientific Reports*, *10*(1), 1–13. doi:10.1038/s41598-020-59839-x

Ames, D. R., & Insley, L. W. (1975). Wind-Chill Effect for Cattle and Sheep. *Journal of Animal Science*, *40*(1), 161–165. doi:https://doi.org/10.2527/jas1975.401161x

Amundson, J. L., Mader, T. L., Rasby, R. J., & Hu, Q. S. (2006). Environmental effects on pregnancy rate in beef cattle. *Journal of Animal Science*, *84*(12), 3415–3420. doi:10.2527/jas.2005-611

An, B., Xia, J., Chang, T., Wang, X., Miao, J., Xu, L., … Gao, H. (2018). Genome-wide association study identifies loci and candidate genes for internal organ weights in simmental beef cattle. *Physiological Genomics*, *50*(7), 523–531. doi:10.1152/physiolgenomics.00022.2018

Angelici, F. M., Laurenti, A., & Nappi, A. (2009). A checklist of the mammals of small Italian islands. *Hystrix, Italian Journal of Mammalogy*, *20*(1), 3–27. doi:10.4404/hystrix-20.1-4429

Antao, T., Lopes, A., Lopes, R. J., Beja-pereira, A., & Luikart, G. (2008). LOSITAN : A workbench to detect molecular adaptation based on a F st -outlier method. *BMC Bioinformatics*, *5*, 1–5. doi:10.1186/1471-2105-9-323

Arpanahi Abdollahi, R., Gianola, D., & Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genetics Selection Evolution*, *52*, 1–15. doi:10.1186/s12711-020-00531-z

Atkinson, E. G., Maihofer, A. X., Kanai, M., Martin, A. R., Karczewski, K. J., Santoro, M. L., … Neale, B. M. (2021). Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nature Genetics*, *53*(2), 195–204. doi:10.1038/s41588-020-00766-y

Atlija, M., Arranz, J. J., Martinez-Valladares, M., & Gutiérrez-Gil, B. (2016). Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genetics Selection Evolution*, *48*(1), 1–16. doi:10.1186/s12711-016-0182-4

Baêta, F. C., Meador, N., Shanklin, & Johnson, H. D. (1987). Equivalent temperature index at temperatures above the thermoneu-tral for lactating dairy cows. In *Meeting of the American Society of Agricultural engineers*.

Bahbahani, H., Tijjani, A., Mukasa, C., & Wragg, D. (2017). Signatures of Selection for Environmental Adaptation and Zebu × Taurine Hybrid Fitness in East African Shorthorn Zebu, *8*(June). doi:10.3389/fgene.2017.00068

Bakhtiarizadeh, M. R., & Salami, S. A. (2019). Identification and expression analysis of long noncoding RNAs in fat-tail of sheep breeds. *G3: Genes, Genomes, Genetics*, *9*(4), 1263–1276. doi:10.1534/g3.118.201014

Bandelt, H. J., & Dress, A. W. M. (1992). A canonical decomposition theory for metrics on a finite set. *Advances in Mathematics*, *92*(1), 47–105. doi:10.1016/0001-8708(92)90061-O

Bansbach, C. E., Bétous, R., Lovejoy, C. A., Glick, G. G., & Cortez, D. (2009). The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes and Development*, *23*(20), 2405–2414. doi:10.1101/gad.1839909

Barbato, M., Hailer, F., Orozco-Terwengel, P., Kijas, J., Mereu, P., Cabras, P., … Bruford, M. W. (2017). Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Scientific Reports*, *7*(1), 1–13. doi:10.1038/s41598-017-07382-7

Barbato, M., Hailer, F., Upadhyay, M., Del Corvo, M., Colli, L., Negrini, R., … Ajmone-Marsan, P. (2020). Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. *Scientific Reports*, *10*(1), 1–11. doi:10.1038/s41598-020-57880-4

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015). SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, *6*(MAR), 1–6. doi:10.3389/fgene.2015.00109

Barker, J. S. F. (1990). A global protocol for determining genetic distances among domestic livestock breeds, *1980*, 501–508.

Barton, N. H. (2000). Genetic hitchhiking. *Trans. R. Soc. B*, *355*, 553–1562.

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., … Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. doi:10.1093/nar/gkaa1100

Baylis, M., & Githeko, A. K. (2006). *The effects of climate change on infectious diseases of animals. Report for the Foresight Project on Detection of Infectious Diseases.*

Baylis, M., & Risley, C. (2013). Infectious Diseases, Climate Change Effects on. In P. Kanki & D. Grimes (Eds.), *Infectious Diseases*. New York: Springer.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B.*, *263*, 1619–1623.

Belasco, E., Cheng, Y., & Schroeder, T. C. (2015). The Impact of Extreme Weather on Cattle Feeding Profits. *Journal of Agricultural and Resource Economics*, *40*, 285–305.

Bellot, P., Campos, G. D. L., & Pérez-enciso, M. (2018). Can Deep Learning Improve Genomic Prediction of complex human traits? *Genetics*, *210*(November), 809–819.

Bernabucci, U. (2019). Climate change: impact on livestock and how can we adapt. *Animal Frontiers*, *9*(1), 3–5. doi:10.1093/af/vfy039

Berry, D. P., & Kearney, J. F. (2011). Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. *Animal, The International Journal of Animal Biosciences*, *5*(8), 1162–1169. doi:10.1017/S1751731111000309

Bertolini, F, Galimberti, G., Schiavo, G., Mastrangelo, S., Gerlando, R. Di, Strillacci, M. G., … Fontanesi, L. (2018). Preselection statistics and Random Forest classification identify population informative single nucleotide polymorphisms in cosmopolitan and autochthonous cattle breeds, 12–19. doi:10.1017/S1751731117001355

Bertolini, Francesca, Cardoso, T. F., Marras, G., Nicolazzi, E. L., Rothschild, M. F., & Amills, M. (2018). Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution*, *50*(1), 1–12. doi:10.1186/s12711-018-0424-8

Bertolini, Francesca, Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., … Crepaldi, P. (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication 06 Biological Sciences 0604 Genetics. *Genetics Selection Evolution*, *50*(1), 1–24. doi:10.1186/s12711-018-0421-y

Bett, B., Kiunga, P., Gachohi, J., Sindato, C., Mbotha, D., Robinson, T., … Grace, D. (2017). Effects of climate change on the occurrence and distribution of livestock diseases. *Preventive Veterinary Medicine*, *137*(November 2015), 119–129. doi:10.1016/j.prevetmed.2016.11.019

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., … Smith, T. P. L. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, *49*(4), 643–650. doi:10.1038/ng.3802

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. Retrieved May 17, 2021, from https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references

Boettcher, P. J., Hoffmann, I., Baumung, R., Drucker, A. G., McManus, C., Berg, P., … Thompson, M. C. (2015). Genetic resources and genomics for adaptation of livestock to climate change. *Frontiers in Genetics*, *5*. doi:10.3389/fgene.2014.00461

Bohmanova, J., Misztal, I., & Cole, J. B. (2007). Temperature-Humidity Indices as Indicators of Milk Production Losses due to Heat Stress. *Journal of Dairy Science*, *90*(4), 1947–1956. doi:10.3168/jds.2006-513

Bolormaa, S., Pryce, J. E., Kemper, K., Savin, K., Hayes, B. J., Barendse, W., … Goddard, M. E. (2013). Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle. *Journal of Animal Science*, *91*(7), 3088–3104. doi:10.2527/jas.2012-5827

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., … Hayes, B. J. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics*, *50*, 362–367. doi:10.1038/s41588-018-0056-5

Braz, C. U., Rowan, T. N., Schnabel, R. D., & Decker, J. E. (2021). Genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. *Scientific Reports*, *11*(0123456789), 1–15. doi:10.1038/s41598-021-92455-x

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., … Bustamante, C. D. (2012). PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Human Biology*, *84*(4), 343–64. doi:10.3378/027.084.0401

Brondízio, E. S., Settele, J., Díaz, S., & Ngo, H. T. (eds). (2019). *IPBES (2019), Global assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. *Debating Nature's Value*. Retrieved from https://ipbes.net/global-assessment%0Ahttps://ipbes.net/global-assessment-report-biodiversity-ecosystem-services

Bruford, M. W., Ginja, C., Hoffmann, I., Joost, S., Wengel, P. O., Alberto, F. J., … Zhan, X. (2015). Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. *Frontiers in Genetics*, *6*(OCT), 1–11. doi:10.3389/fgene.2015.00314

Bruno, S., & Sauli, G. (1976). *Montecristo. Natura e Montagna* (Vol. n1).

Buggiotti, L., Yurchenko, A. A., Yudin, N. S., Vander Jagt, C. J., Vorobieva, N. V, Kusliy, M. A., … Larkin, D. M. (2021). Demographic History, Adaptation, and NRAP Convergent Evolution at Amino Acid Residue 100 in the World Northernmost Cattle from Siberia. *Molecular Biology and Evolution*, (May). doi:10.1093/molbev/msab078

Buitkamp, J., Filmether, P., Stear, M., & Epplen, J. (1996). Class I and class II major histocompatibility complex alleles are associated with faecal egg counts following natural, predominantly Ostertagia circumcincta infection. *Parasitology Research*, *82*, 693–696.

Bukowiecki, L., Collet, A. J., & Follea, N. (1982). Brown adipose tissue hyperplasia: A fundamental mechanism of adaptation to cold and hyperphagia. *American Journal of Physiology - Endocrinology and Metabolism*, *5*(6). doi:10.1152/ajpendo.1982.242.6.e353

Buontempo, C., Hutjes, R., Beavis, P., Berckmans, J., Cagnazzo, C., Vamborg, F., … Dee, D. (2020). Fostering the development of climate services through Copernicus Climate Change Service (C3S) for agriculture applications. *Weather and Climates Extremes*, *27*. doi:10.1016/j.wace.2019.100226

Burdick, N., Agado, B., White, J., Matheney, K., Neuendorff, D., Riley, D., & Al., E. (2011). Technical note: Evolution of exit velocity in suckling Brahman calves. *Journal of Animal Science*, *89*, 233–236.

Burren, A., Neuditschko, M., Signer-Hasler, H., Frischknecht, M., Reber, I., Menzi, F., … Flury, C. (2016). Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Animal Genetics*, *47*(6), 727–739. doi:10.1111/age.12476

Burrow, H. (2001). Variances and covariances between productive and adaptative traits and temperament in a composite breed of tropical beef cattle. *Livestock Production Science*, *70*, 213–233.

Café, L. M., Robinson, D. L., Ferguson, D. M., McIntyre, B. L., Geesink, G. H., & Greenwood, P. L. (2011). Cattle temperament: persistence of assessments and associations with productivity, efficiency, carcass and meat quality traits. *Journal of Animal Science*, *89*, 1452–1465.

Cai, Y., Fu, W., Cai, D., Heller, R., Zheng, Z., Wen, J., … Russo, C. (2020). Ancient Genomes Reveal the Evolutionary History and Origin of Cashmere-Producing Goats in China. *Molecular Biology and Evolution*, *37*(7), 2099–2109. doi:10.1093/molbev/msaa103

Calleja-Cabrera, J., Boter, M., Oñate-Sánchez, L., & Pernas, M. (2020). Root Growth Adaptation to Climate Change in Crops. *Frontiers in Plant Science*, *11*(May), 1–23. doi:10.3389/fpls.2020.00544

Cantor, R. M., Lange, K., & Sinsheimer, J. S. (2010). Prioritizing GWAS Results : A Review of Statistical Methods and Recommendations for Their Application. *The American Journal of Human Genetics*, *86*(1), 6–22. doi:10.1016/j.ajhg.2009.11.017

Cao, Y.-H., Xu, S.-S., Shen, M., Chen, Z.-H., Gao, L., Lv, F.-H., … Li, M.-H. (2020). Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Molecular Biology and Evolution*. doi:10.1093/molbev/msaa236

Cao, Y. H., Xu, S. S., Shen, M., Chen, Z. H., Gao, L., Lv, F. H., … Li, M. H. (2021). Historical Introgression from Wild Relatives Enhanced Climatic Adaptation and Resistance to Pneumonia in Sheep. *Molecular Biology and Evolution*, *38*(3), 838–855. doi:10.1093/molbev/msaa236

Cardoso, T. F., Amills, M., Bertolini, F., Rothschild, M., Marras, G., Boink, G., … Lenstra, J. A. (2018). Patterns of homozygosity in insular and continental goat breeds. *Genetics Selection Evolution*, *50*(1), 1–11. doi:10.1186/s12711-018-0425-7

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution*, *36*, 852–860.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics*, *19*(4), 220–234. doi:10.1038/nrg.2017.109

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*.

Chapman, S. C., Chakraborty, S., Dreccer, F., & Howden, M. (2012). Plant adaptation to climate change — opportunities and priorities in breeding. *Crop and Pasture Science*, *63*, 251–268.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*(4), 379–384. doi:10.1371/journal.pgen.0020064

Chen, Y., Yang, Y., Li, C., Li, R., Xiao, H., & Chen, S. (2020). Genetic diversity of TLR3 and TLR8 genes among five Chinese native cattle breeds from southwest China. *Livestock Science*, *232*(July 2019), 103895. doi:10.1016/j.livsci.2019.103895

Chen, Z.-H., Xu, Y.-X., Xie, X.-L., Wang, D.-F., Aguilar-Gómez, D., Liu, G.-J., … Li, M.-H. (2021). Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Communications Biology*, *4*(1), 1–15. doi:10.1038/s42003-021-02817-4

Cheng, J., Cao, X., Hanif, Q., Pi, L., Hu, L., Huang, Y., … Chen, H. (2020). Integrating Genome-Wide CNVs Into QTLs and High Confidence GWAScore Regions Identified Positional Candidates for Sheep Economic Traits. *Frontiers in Genetics*, *11*(June), 1–7. doi:10.3389/fgene.2020.00569

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., … Palmarini, M. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science*, *324*(5926), 532–536. doi:10.1126/science.1170587

Chilliard, Y., Ferlay, A., Faulconnier, Y., Bonnet, M., Rouel, J., & Bocquier, F. (2000). Adipose tissue metabolism and its role in adaptations to undernutrition in ruminants. *Proceedings of the Nutrition Society*, *59*(1), 127–134. doi:10.1017/S002966510000015X

Ciani, E., Crepaldi, P., Nicoloso, L., Lasagna, E., Sarti, F. M., Moioli, B., & Napolitano, F. (2013). Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds, 256–266. doi:10.1111/age.12106

Ciani, F., & Masseti, M. (1991). *Considerazioni sull'origine della popolazione ircina dell'isola di Montecristo, nel Mar Tirreno settentrionale. Atti Conv. Genet. Conserv. fauna Suppl. Ric. Biol. Selvaggina 18*.

Colli, L., Lancioni, H., Cardinali, I., Olivieri, A., Capodiferro, M. R., Pellecchia, M., … Achilli, A. (2015). Whole mitochondrial genomes unveil the impact of domestication on goat matrilineal variability. *BMC Genomics*, *16*(1), 1–12. doi:10.1186/s12864-015-2342-2

Colli, L., Milanesi, M., Talenti, A., Bertolini, F., Chen, M., Crisà, A., … Stella, A. (2018). Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genetics Selection Evolution*, *50*(1), 1–20. doi:10.1186/s12711-018-0422-x

Colli, L., Negrini, R., Nicoloso, L., & Crepaldi, P. (2014). Assessing The Spatial Dependence of Adaptive Loci in 43 European and Western Asian Goat Breeds Using AFLP Markers. *PLOS ONE*, *9*(1), 1–12. doi:10.1371/journal.pone.0086668

Collier, R. J., Baumgard, L. H., Zimbelman, R. B., & Xiao, Y. (2019). Heat stress : physiology of acclimation and adaptation. *Animal Frontiers*, *9*(1), 12–19. doi:10.1093/af/vfy031

Collin, F. D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., … Estoup, A. (2021). Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Molecular Ecology Resources*, *21*(8), 2598–2613. doi:10.1111/1755-0998.13413

Convention on Biological Diversity. (n.d.). Aichi Biodiversity Targets. Retrieved January 10, 2022, from https://www.cbd.int/sp/targets/#GoalB

Coop, G., Witonsky, D., Rienzo, A. Di, & Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, *185*(August), 1411–1423. doi:10.1534/genetics.110.114819

Copenhaver, G. P., Browne, W. E., & Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(1), 247–252. doi:10.1073/pnas.95.1.247

Cortellari, M., Barbato, M., Talenti, A., Bionda, A., Randi, E., Sarti, F. M., … Liotta, L. (2021). The climatic and genetic heritage of Italian goat breeds with genomic SNP data. *Scientific Reports*, *11*(1), 1–12. doi:10.1038/s41598-021-89900-2

Cortellari, M., Bionda, A., Negro, A., Frattini, S., Mastrangelo, S., Somenzi, E., … Crepaldi, P. (2021). Runs of homozygosity in the Italian goat breeds: impact of management practices in low-input systems. *Genetics Selection Evolution*, *53*(1). doi:10.1186/s12711-021-00685-4

Crepaldi, P., Corti, M., & Cicogna, M. (1999). Factors affecting milk production and prolificacy of Alpine goats in Lombardy (Italy). *Small Ruminant Research*, *32*(1), 83–88. doi:10.1016/S0921-4488(98)00156-4

Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, *34*(2), 1–24. doi:10.1201/9781315119403-7

Cui, D., Liang, S., Wang, D., & Liu, Z. (2021). Köppen-Geiger climate classification and bioclimatic variables. *Earth System Science Data*, (March), 1–34.

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*, *166*(1), 26–34. doi:10.1016/j.livsci.2014.05.034

da Silva, V. P. R., Araújo e Silva, R., Palmeira Cavalcanti, E., Campos Braga, C., Vieira de Azevedo, P., Singh, V. P., & Ricardo Rodrigues Pereira, E. (2010). Trends in solar radiation in NCEP/NCAR database and measurements in northeastern Brazil. *Solar Energy*, *84*, 1852–1862.

Daly, K. G., Delser, P. M., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., … Erek, C. M. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, *361*, 24–27.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi:10.1093/bioinformatics/btr330

Davenport, K. M., Bickhart, D. M., Worley, K., Murali, S. C., Cockett, N., Heaton, M. P., … Rosen, B. D. (2021). 42 An Improved, High-quality Ovine Reference Genome Assembly. *Journal of Animal Science*, *99*(Supplement_3), 23–24. doi:10.1093/jas/skab235.039

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge, United Kingdom: Cambridge University Press.

De Mita, S., Thuillet, A., Gay, L., Ahmadi, N., Manel, S., & Ronfort, J. (2013). Detecting selection along environmental

gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.*, *22*, 1383–1399.

Deniskova, T. E., Dotsev, A. V, Selionova, M. I., Kunz, E., Medugorac, I., Reyer, H., … Zinovieva, N. A. (2018). Population structure and genetic diversity of 25 Russian sheep breeds based on whole-genome genotyping. *Genetics Selection Evolution*, *50*(1), 29. doi:10.1186/s12711-018-0399-5

di Trana, A., Sepe, L., di Gregorio, P., Di Napoli, M. A., Giorgio, D., Caputo, A. R., & Claps, S. (2015). *The role of local sheep and goat breeds and their products as a tool for sustainability and safeguard of the mediterranean environment*. *The Sustainability of Agro-Food and Natural Resource Systems in the Mediterranean Basin*. doi:10.1007/978-3-319-16357-4_7

Dias-Alves, T., Mairal, J., & Blum, M. G. B. (n.d.). Loter: A Software Package to Infer Local Ancestry for a Wide Range of Species. doi:10.1093/molbev/msy126

Dikmen, S., Alava, E., Pontes, E., Fear, J. M., Dikmen, B. Y., Olson, T. A., & Hansen, P. J. (2008). Differences in Thermoregulatory Ability Between Slick-Haired and Wild-Type Lactating Holstein Cows in Response to Acute Heat Stress. *Journal of Dairy Science*, *91*(9), 3395–3402. doi:10.3168/jds.2008-1072

Dikmen, S., Khan, F. A., Huson, H. J., Sonstegard, T. S., Moss, J. I., Dahl, G. E., & Hansen, P. J. (2014). The SLICK hair locus derived from Senepol cattle confers thermotolerance to intensively managed lactating Holstein cows. *Journal of Dairy Science*, *97*(9), 5508–5520. doi:10.3168/jds.2014-8087

Dimauro, C., Cellesi, M., Steri, R., Gaspa, G., Sorbolini, S., Stella, A., & Macciotta, N. P. P. (2013). Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes, 377–382. doi:10.1111/age.12021

Dimauro, C., Nicoloso, L., Cellesi, M., Macciotta, N. P. P., Ciani, E., Moioli, B., … Crepaldi, P. (2015). Selection of discriminant SNP markers for breed and geographic assignment of Italian sheep. *Small Ruminant Research*, *128*, 27–33. doi:10.1016/j.smallrumres.2015.05.001

do Prado Paim, T., Hay, E. H. A., Wilson, C., Thomas, M. G., Kuehn, L. A., Paiva, S. R., … Blackburn, H. (2020). Genomic Breed Composition of Selection Signatures in Brangus Beef Cattle. *Frontiers in Genetics*, *11*(July), 1–10. doi:10.3389/fgene.2020.00710

Domestic Animal Diversity Information System (DAD-IS). (n.d.). Retrieved January 10, 2022, from https://www.fao.org/dad-is/en/

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., … Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nature Biotechnology*, *31*(2), 135–143. doi:10.1038/nbt.2478

Dopazo, J., Dress, A., & Von Haeseler, A. (1993). Split decomposition: A technique to analyze viral evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(21), 10320–10324. doi:10.1073/pnas.90.21.10320

Doro, M. G., Casu, G., Leoni, G. G., Naitana, S., Pirastu, M., Novelletto, A., & Fraticelli, F. (2016). The complete mitochondrial DNA sequence of the Montecristo goat. *Livestock Science*, *188*, 120–123. doi:10.1016/j.livsci.2016.04.011

Dowling, T. E., & Secor, C. L. (1997). The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*. doi:10.1146/annurev.ecolsys.28.1.593

Durbin, H. J., Lu, D., Yampara-Iquise, H., Miller, S. P., & Decker, J. E. (2020). Development of a genetic evaluation for hair shedding in American Angus cattle to improve thermotolerance. *Genetics Selection Evolution*, *52*(1), 1–14. doi:10.1186/s12711-020-00584-0

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–1191. doi:10.1038/nprot.2009.97

Duruz, S., Rochat, E., Selmoni, O., Vajana, E., Orozco, P., Joost, S., … Bruford, C. M. W. (2019). Rapid identification and interpretation of gene – environment associations using the new R . SamBada landscape genomics pipeline. *Mol. Ecol. Resour.*, *19*(March), 1355–1365. doi:10.1111/1755-0998.13044

Dutta, P., Talenti, A., Young, R., Jayaraman, S., Callaby, R., Jadhav, S. K., … Prendergast, J. G. D. (2020). Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication. *Nature Communications*, *11*(1). doi:10.1038/s41467-020-18550-1

Dýrmundsson, Ó. R., & Ninikowski, R. (2010). North European short-tailed breeds of sheep: A review. *Animal*, *4*(8), 1275–1282. doi:10.1017/S175173110999156X

E, G. X., Zhao, Y. J., & Huang, Y. F. (2019). Selection signatures of litter size in Dazu black goats based on a whole genome sequencing mixed pools strategy. *Molecular Biology Reports*, *46*(5), 5517–5523. doi:10.1007/s11033-019-04904-6

Edea, Z., Dadi, H., Dessie, T., & Kim, K. S. (2019). Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. *Genes and Genomics*, *41*, 973–981.

Edea, Z., Dadi, H., Dessie, T., Uzzaman, M. R., Rothschild, M. F., Kim, E. S., … Kim, K. S. (2018). Genome-wide scan reveals divergent selection among taurine and zebu cattle populations from different regions. *Animal Genetics*, *49*(6), 550–563. doi:10.1111/age.12724

Edea, Z., Hong, J. K., Jung, J. H., Kim, D. W., Kim, Y. M., Kim, E. S., … Kim, K. S. (2017). Detecting selection signatures between Duroc and Duroc synthetic pig populations using high-density SNP chip. *Animal Genetics*, *48*(4), 473–477. doi:10.1111/age.12559

Eigenberg, R. A., Brown-Brandl, T. M., Nienaber, J. A., & Hahn, G. L. (2005). Dynamic Response Indicators of Heat Stress in

Shaded and Non-shaded Feedlot Cattle, Part 2: Predictive Relationships. *Biosystems Engineering*, *91*, 111–118.
Escarcha, J. F., Lassa, J. A., & Zander, K. K. (2018). Livestock under climate change: A systematic review of impacts and adaptation. *Climate*, *6*(3), 1–17. doi:10.3390/cli6030054
Estrada-Reyes, Z. M., Tsukahara, Y., Amadeu, R. R., Goetsch, A. L., Gipson, T. A., Sahlu, T., … Mateescu, R. G. (2019). Signatures of selection for resistance to Haemonchus contortus in sheep and goats. *BMC Genomics*, *20*(1), 1–14. doi:10.1186/s12864-019-6150-y
European Committee for the Conservation of Nature and Natural Resources. (1986). Montecristo Island Nature Reserve - Application for the European Diploma. Retrieved June 11, 2020, from https://rm.coe.int/montecristo-island-nature-reserve-italy-application-for-the-european-d/16809a7ff3
Eusebi, P.G., Cortés, O., Carleos, C., Dunner, S., & J., C. (2018). Detection of selection signatures for agonistic behaviour in cattle. *J Anim Breed Genet.*, *135*, 170–177. doi:doi: 10.1111/jbg.12325
Eusebi, P.G., Sevane, N., Cortés, O., Contreras, E., Cañon, J., & Dunner, S. (2020). Aggressive behavior in cattle is associated with a polymorphism in the MAOA gene promoter. *Anim Genet*, *51*, 14–21.
Eusebi, Paulina G., Martinez, A., & Cortes, O. (2020). Genomic tools for effective conservation of livestock breed diversity. *Diversity*, *12*(1). doi:10.3390/d12010008
Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. doi:10.1111/j.1755-0998.2010.02847.x
Eydivandi, S., Roudbar, M. A., Ardestani, S. S., Momen, M., & Sahana, G. (2021). A selection signatures study among Middle Eastern and European sheep breeds. *Journal of Animal Breeding and Genetics*, (November 2020), 1–15. doi:10.1111/jbg.12536
F. Bertolini, G. Galimberti, D. G. C. (2015). Combined use of principal component analysis and random forests identify population-informative single nucleotide polymorphisms : application in cattle breeds. *Animal Breeding and Genetics*, *132*(ii), 346–356. doi:10.1111/jbg.12155
Fan, R., Gu, Z., Guang, X., Marín, J. C., Varas, V., González, B. A., … Dong, C. (2020). Genomic analysis of the domestication and post-Spanish conquest evolution of the llama and alpaca. *Genome Biology*, *21*(1), 1–26. doi:10.1186/s13059-020-02080-6
Fan, Y., Liang, Y., Deng, K., Zhang, Z., Zhang, G., Zhang, Y., & Wang, F. (2020). Analysis of DNA methylation profiles during sheep skeletal muscle development using whole-genome bisulfite sequencing. *BMC Genomics*, *21*(1), 1–15. doi:10.1186/s12864-020-6751-5
FAO. (2015). *the Second Report on the State of the World'S Animal Genetic Resources for Food and Agriculture. Edited by B.D. Scherf & D. Pilling. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome*. Retrieved from http://www.fao.org/3/a-i4787c/index.html
Faraut, T., Wu, C., Muzny, D. M., Li, Y., Zhang, W., Aken, B. L., … Worley, K. C. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, *344*(6188), 1168–1173.
Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, *193*(3), 929–941. doi:10.1534/genetics.112.147231
Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*(3), 1405–1413. doi:10.1093/genetics/155.3.1405
Felsenstein, J. (2004). PHYLIP (Phylogeny Inference Package) version 3.6.
Fernández Suárez, A. G., Gutiérrez Reynoso, G. A., & Ponce de León Bravo, F. A. (2019). Identificación bioinformática de Polimorfismos de Nucleótido Simple (PNSs) en genes candidatos para las características de la fibra en alpacas (Vicugna pacos). *Revista Peruana de Biología*, *26*(1), 087–094. doi:10.15381/rpb.v26i1.15911
Fernihough, A. (2014). mfx: Marginal Effects, Odds Ratios and Incidence Rate Ratios for GLMs.
Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *4*. doi:https://doi.org/10.1002/joc.5086
Fitak, R. R. (2021). OptM : estimating the optimal number of migration edges on population trees using Treemix . *Biology Methods and Protocols*, *6*(1), 1–6. doi:10.1093/biomethods/bpab017
Flori, L., Gonzatti, M. I., Thevenon, S., Chantal, I., Pinto, J., Berthier, D., … Gautier, M. (2012). A Quasi-Exclusive European Ancestry in the Senepol Tropical Cattle Breed Highlights the Importance of the slick Locus in Tropical Adaptation. *PLOS ONE*, *7*(5), 1–10. doi:10.1371/journal.pone.0036133
Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., … Gautier, M. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology*, *28*(5), 1009–1029. doi:10.1111/mec.15004
Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. doi:10.1534/genetics.108.092221
Fonseca, P. A. S., Suárez-Vega, A., Marras, G., & Cánovas, Á. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. *GigaScience*, *9*(12), 1–9. doi:10.1093/gigascience/giaa149
Fox, N. J., Glenn, M., Davidson, R. S., White, P. C. L., & Hutchings, M. (2015). Climate-driven tipping-points could lead to sudden, high-intensity parasite outbreaks. *R. Soc. Open Sci.*, *2*.
Francis, J., & Little, D. A. (1964). Resistance of Droughtmaster cattle to tick infestation and babesiosis. *Australian Veterinary Journal*, *40*, 247–253.

Freitas, P. H. F., Wang, Y., Yan, P., Oliveira, H. R., Schenkel, F. S., Zhang, Y., … Brito, L. F. (2021). Genetic Diversity and Signatures of Selection for Thermal Stress in Cattle and Other Two Bos Species Adapted to Divergent Climatic Conditions. *Frontiers in Genetics*, *12*(February).

Frichot, E., & Francois, O. (2015). LEA : An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*, 925–929. doi:10.1111/2041-210X.12382

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. doi:10.1093/molbev/mst063

Frkonja, A., Gredler, B., Schnyder, U., Curik, I., & So, J. (2012). Prediction of breed composition in an admixed cattle population, 696–703. doi:10.1111/j.1365-2052.2012.02345.x

Galal, S., Rasoul, F. A., Annous, M. R., & Shoat, I. (2005). *Small Ruminant Breeds of Egypt. Characterization of Small Ruminant Breeds in West Asia and North Africa Vol 2: North Africa.* Aleppo.

Gaston, K. J., & Spicer, J. I. (2013). *Biodiversity: an introduction*.

Gaughan, J. B., & Cawdell-Smith, A. J. (2017). Impact of climate change on livestock production and reproduction. In V. Sejian, J. Gaughan, L. Baumgard, & C. Prasad (Eds.), *Climate change impacts on livestock: adaptation and mitigation* (Springer, pp. 51–60). New Delhi.

Gaughan, J B, Mader, T. L., Holt, S. M., & Lisle, A. (2008). A New Heat Load Index for Feedlot Cattle. *Faculty Papers and Publications in Animal Science*, *613*, 226–233. doi:10.2527/jas.2007-0305

Gaughan, John B., Sejian, V., Mader, T. L., & Dunshea, F. R. (2019). Adaptation strategies: Ruminants. *Animal Frontiers*, *9*(1), 47–53. doi:10.1093/af/vfy029

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, *201*, 1555–1579.

Gautier, Mathieu, Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, *17*(1), 78–90. doi:10.1111/1755-0998.12634

Georges, M., Charlier, C., & Hayes, B. (2019). Harnessing genomic information for livestock improvement. *Nat Rev Genet*, *20*, 135–156.

Gibson, J., Morton, N. E., & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*, *15*(5), 789–795. doi:10.1093/hmg/ddi493

Gippoliti, S. (2016). The wild goat of Montecristo Island: Did it ever exist? *Mammalia*, *80*(2), 221–222. doi:10.1515/mammalia-2015-0078

Giuffra, E., Tuggle, C. K., & FAANG Consortium. (2018). Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annual Review of Animal Biosciences*, *7*, 65–88.

Gobikrushanth, M., Purfield, D. C., Canadas, E. R., Herlihy, M. M., Kenneally, J., Murray, M., … Butler, S. T. (2019). Anti-Müllerian hormone in grazing dairy cows: Identification of factors affecting plasma concentration, relationship with phenotypic fertility, and genome-wide associations. *Journal of Dairy Science*, *102*(12), 11622–11635. doi:10.3168/jds.2019-16979

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews*, *10*(juNe), 381–391. doi:10.1038/nrg2575

Goodchild, M. F. (1992). Geographical information science. *International Journal of Geographical Information Systems*, *6*(1), 31–45. doi:10.1080/02693799208901893

Gorbach, D. M., Makgahlela, M. L., Reecy, J. M., Kemp, S. J., Baltenweck, I., Ouma, R., & Mwai, O. (2010). Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya, *127*, 348–351. doi:10.1111/j.1439-0388.2010.00864.x

Gorssen, W., Meyermans, R., Janssens, S., & Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genetics Selection Evolution*, *53*(1), 1–10. doi:10.1186/s12711-020-00599-7

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. doi:10.1111/j.1471-8286.2004.00828.x

Groenen, M. A. M., Archibald, A. ., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., … Kim, H. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, *491*, 393–398. doi:10.1038/nature11622

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., … Weigend, S. (2010). Genetic diversity in farm animals - A review. *Animal Genetics*, *41*(SUPPL. 1), 6–31. doi:10.1111/j.1365-2052.2010.02038.x

Guan, Y. (2014). Detecting structure of haplotypes and local ancestry. *Genetics*, *196*(3), 625–642. doi:10.1534/genetics.113.160697

Guillot, G., & Vitalis, R. (2014). Detecting correlation between allele frequencies and environmental variables as a signature of selection . A fast computational approach for genome-wide studies. *Spatial Statistics*, *8*, 145–155. doi:10.1016/j.spasta.2013.08.001

Gunther, T., & Coop, G. (2013). Robust Identification of Local Adaptation from allele frequencies. *Genetics*, *195*(September), 205–220. doi:10.1534/genetics.113.152462

Guo, J., Tao, H., Li, P., Li, L., Zhong, T., Wang, L., … Zhang, H. (2018). Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Scientific Reports*, *8*(1), 1–11. doi:10.1038/s41598-018-28719-w

Ha, N. T., Gross, J. J., Van Dorland, A., Tetens, J., Thaller, G., Schlather, M., … Simianer, H. (2015). Gene-based mapping and

pathway analysis of metabolic traits in dairy cows. *PLoS ONE*, *10*(3), 5–9. doi:10.1371/journal.pone.0122325
Hahn, G. (1996). Global warming and potential impacts on cattle and swine in tropical and temperate areas. In *Proceedings of the Brazilian Congress of Biometeorology*.
Hahn, G., Mader, T., Spiers, D., Gaughan, J., Nienaber, J., Eigenberg, R., … Hugenford, L. (2001). Heat wave impacts on feedlot cattle: considerations for improved environmental management. In *Livestock Environment International Symposium Proceedings*.
Halder, I., Shriver, Ã. M., Thomas, M., Fernandez, J. R., & Frudakis, T. (2008). A Panel of Ancestry Informative Markers for Estimating Individual Biogeographical Ancestry and Admixture From Four Continents : Utility and Applications, *29*(February), 648–658. doi:10.1002/humu.20695
Hamlyn-Hill, F. (2012). Improving temperament: effects on productivity and meat quality. Beef CRC.
Hannemann, F., Bichet, A., Ewen, K. M., & Bernhardt, R. (2007). Cytochrome P450 systems—biological variations of electron transport chains,. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1770*, 330–344.
Hansen, P. . (2004). Physiological and cellular adaptations of zebu cattle to thermal stress. *Animal Reproduction Science*, *82–83*, 349–360. doi:10.1016/J.ANIREPROSCI.2004.04.011
Harfouche, A. L., Jacobson, D. A., Kainer, D., Romero, J. C., Harfouche, A. H., Scarascia Mugnozza, G., … Altman, A. (2019). Accelerating Climate Resilient Plant Breeding by Applying Next-Generation Artificial Intelligence. *Trends in Biotechnology*, *37*, 1217–1235.
Hatzikotoulas, K., Gilly, A., & Zeggini, E. (2014). Using population isolates in genetic association studies. *Brief. Funct. Genomics*, *13*, 371–377.
Hayes, B. (2013). Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). *Methods Mol. Biol.*, *1019*, 149–169.
Hayes, B. J., Lewin, H. A., & Goddard, M. E. (2013). The future of livestock breeding : genomic selection for efficiency , reduced emissions intensity , and adaptation. *Trends in Genetics*, *29*(4), 206–214. doi:10.1016/j.tig.2012.11.009
He, S., Di, J., Han, B., Chen, L., Liu, M., & Li, W. (2020). Genome-wide scan for runs of homozygosity identifies candidate genes related to economically important traits in Chinese merino. *Animals*, *10*(3), 1–13. doi:10.3390/ani10030524
Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618. doi:10.1111/mec.12415
Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., … Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Curr Rev Musculoskelet Med*, *13*, 69–76.
Henriques, D., Wallberg, A., Chávez-Galarza, J., Johnston, J. S., Webster, M. T., & Pinto, M. A. (2018). Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula. *Scientific Reports*, *8*(1), 1–14. doi:10.1038/s41598-018-29469-5
Herbut, P., Angrecka, S., & Walczak, J. (2018). Environmental parameters to assessing of heat stress in dairy cattle — a review. *International Journal of Biometeorology*, *62*, 2089–2097.
Herd, R. M., & Bishop, S. C. (2000). Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. *Livestock Production Science*, *63*, 111–119.
Hidosa, D., & Guyo, M. (2017). Journal of Fisheries & Climate Change Effects on Livestock Feed Resources : A Review. *Journal of Fisheries and Livestock Production*, *5*(4), 5–8. doi:10.4172/2332-2608.1000259
Hiendleder, S., Kaupe, B., Wassmuth, R., & Janke, A. (2002). Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1494), 893–904. doi:10.1098/rspb.2002.1975
Hijmans, R. J. (2020). raster: Geographic Data Analysis and Modeling. R package version 3.3-13.
Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. doi:10.1002/joc.1276
Hill, D., & Wall, E. (2015). Dairy cattle in a temperate climate: The effects of weather on milk yield and composition depend on management. *Animal*, *9*, 138–149.
Hiraike, Y., Waki, H., Yu, J., Nakamura, M., Miyake, K., Nagano, G., … Kadowaki, T. (2017). NFIA co-localizes with PPARγ and transcriptionally controls the brown fat gene program. *Nature Cell Biology*, *19*(9), 1081–1092. doi:10.1038/ncb3590
Hoffmann, I. (2013). Adaptation to climate change--exploring the potential of locally adapted breeds. *Animal : An International Journal of Animal Bioscience*, *7 Suppl 2*, 346–362. doi:10.1017/S1751731113000815
Holmes, R. S., & Rout, U. K. (2011). Comparative studies of vertebrate beta integrin genes and proteins: Ancient genes in vertebrate evolution. *Biomolecules*, *1*(1), 3–31. doi:10.3390/biom1010003
Hoppe, S., Brandt, H. R., Nig, S. K., Erhardt, G., & Gauly, M. (2010). Temperament traits of beef calves measured under field condi-tions and their relationships to performance. *Journal of Animal Science*.
Horwitz, L. K., & Bar-Gal, G. K. (2006). The Origin and Genetic Status of Insular Caprines in the Eastern Mediterranean: A Case Study of Free-Ranging Goats (Capra aegagrus cretica) on Crete. *Human Evolution*, *21*(2), 123–138. doi:10.1007/s11598-006-9015-8
Howard, J. T., Kachman, S. D., Snelling, W. M., Pollak, E. J., Ciobanu, D. C., Kuehn, L. A., & Spangler, M. L. (2014). Beef cattle body temperature during climatic stress: A genome-wide association study. *International Journal of Biometeorology*, *58*(7), 1665–1672. doi:10.1007/s00484-013-0773-5
Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., … Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891. doi:10.1093/nar/gkaa942

Hsu, J. Y., Crawley, S., Chen, M., Ayupova, D. A., Lindhout, D. A., Higbee, J., … Allan, B. B. (2017). Non-homeostatic body weight regulation through a brainstem-restricted receptor for GDF15. *Nature*, *550*(7675), 255–259. doi:10.1038/nature24042

Hu, T., Darabos, C., & Urbanowicz, R. (2020). Editorial: Machine Learning in Genome-Wide Association Studies. *Frontiers in Genetics*, *11*. doi:https://doi.org/10.3389/fgene.2020.593958

Hu, X.-J., Yang, J., Xie, X.-L., Lv, F.-H., Cao, Y.-H., Li, W.-R., … Li, M.-H. (2019). The Genome Landscape of Tibetan Sheep Reveals Adaptive Introgression from Argali and the History of Early Human Settlements on the Qinghai–Tibetan Plateau. *Molecular Biology and Evolution*, *36*(2), 283–303. doi:10.1093/molbev/msy208

Hu, X. J., Yang, J., Xie, X. L., Lv, F. H., Cao, Y. H., Li, W. R., … Li, M. H. (2019). The genome landscape of tibetan sheep reveals adaptive introgression from Argali and the history of early human settlements on the Qinghai-Tibetan Plateau. *Molecular Biology and Evolution*, *36*(2), 283–303. doi:10.1093/molbev/msy208

Hudson, N. J., Reverter, A., Griffiths, W. J., Yutuc, E., Wang, Y., Jeanes, A., … Greenwood, P. L. (2020). Gene expression identifies metabolic and functional differences between intramuscular and subcutaneous adipocytes in cattle. *BMC Genomics*, *21*(1), 1–23. doi:10.1186/s12864-020-6505-4

Hulsegge, B., Calus, M. P. L., Windig, J. J., Hoving-Bolink, A. H., Maurice-van Eijndhoven, M. H. T., & Hiemstra, S. J. (2013). Selection of SNP from 50K and 777K arrays to predict breed of origin in cattle. *Journal of Animal Science*, *91*(11), 5128–5134. doi:10.2527/jas.2013-6678

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. doi:10.1093/molbev/msj030

Iacolina, L. (2018). Hybridisation in European ungulates : an overview of the current status , causes , and consequences. doi:10.1111/mam.12140

Igoshin, A., Yudin, N., Aitnazarov, R., Yurchenko, A. A., & Larkin, D. M. (2021). Whole-Genome Resequencing Points to Candidate DNA Loci Affecting Body Temperature under Cold Stress in Siberian Cattle Populations.

Igoshin, A. V, Yurchenko, A. A., Belonogova, N. M., Petrovsky, D. V, Aitnazarov, R. B., Soloshenko, V. A., … Larkin, D. M. (2019). Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations, *20*(Suppl 1).

Illumina Inc. (2016). Infinium® iSelect® Custom Genotyping Assays. Guidelines for Using the DesignStudio™ Microarray Assay Designer Software to Create and Order Custom Arrays.

International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*, 695–777. doi:10.1038/nature03154

International Human Genome Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(February), 860–921.

Islam, R., Liu, X., Gebreselassie, G., Abied, A., Ma, Q., & Ma, Y. (2020). Genome-wide association analysis reveals the genetic locus for high reproduction trait in Chinese Arbas Cashmere goat. *Genes and Genomics*, *42*(8), 893–899. doi:10.1007/s13258-020-00937-5

ISTAT. (2010). ISTAT. 6° Censimento agricoltura.

Jaton, C., Schenkel, F. S., Sargolzaei, M., Cánova, A., Malchiodi, F., Price, C. A., … Miglior, F. (2018). Genome-wide association study and in silico functional analysis of the number of embryos produced by Holstein donors. *Journal of Dairy Science*, *101*(8), 7248–7257. doi:10.3168/jds.2017-13848

Jelinski, D. E. (1997). On genes and geography: A landscape perspective on genetic variation in natural plant populations. *Landscape and Urban Planning*, *39*(1), 11–23. doi:10.1016/S0169-2046(97)00071-6

Jin, L., Mao, K., Li, J., Huang, W., Che, T., Fu, Y., … Li, M. (2018). Genome-wide profiling of gene expression and DNA methylation provides insight into low-altitude acclimation in Tibetan pigs. *Gene*, *5*, 522–532.

Joost, S., Bonin, A., Bruford, M., Despres, L., Concord, C., Erhardt, G., & Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, *16*.

Joost, S., Kalbermatten, M., & Bonin, A. (2008). Spatial analysis method (sam): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources*, *8*, 957–960.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*, 255–260.

Kaback, D. (1996). Chromosome-size dependent control of meiotic recombination in humans. *Nature Genetics*, *13*, 20–21.

Kaback, D. B., Barber, D., Mahon, J., Lamb, J., & You, J. (1999). Chromosome size-dependent control of meiotic reciprocal recombination in Saccharomyces cerevisiae: The role of crossover interference. *Genetics*, *152*(4), 1475–1486. doi:10.1093/genetics/152.4.1475

Kalds, P., Luo, Q., Sun, K., Zhou, S., Chen, Y., & Wang, X. (2021). Trends towards revealing the genetic architecture of sheep tail patterning: Promising genes and investigatory pathways. *Animal Genetics*, *52*(6), 799–812. doi:10.1111/age.13133

Kaluarachchi, T., Reis, A., & Nanayakkara, S. (2021). A Review of Recent Deep Learning Approaches in Human-Centered Machine Learning. *Sensors*, *21*, 2514.

Kang, D., Zhou, G., Zhou, S., Zeng, J., Wang, X., Jiang, Y., … Chen, Y. (2017). Comparative transcriptome analysis reveals potentially novel roles of Homeobox genes in adipose deposition in fat-tailed sheep. *Scientific Reports*, *7*(1), 1–17. doi:10.1038/s41598-017-14967-9

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., … Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, *42*, 348–354.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-auza, R. W., … Kessler, M. (2017). Data Descriptor :

Climatologies at high resolution for the earth ' s land surface areas. *Nature Publishing Group*, *4*, 1–20. doi:10.1038/sdata.2017.122
Kava, R., Peripolli, E., Berton, M. P., Lemos, M., Lobo, R. B., Stafuzza, N. B., … Baldi, F. (2021). Genome-wide structural variations in Brazilian Senepol cattle, a tropically adapted taurine breed. *Livestock Science*, *253*(September). doi:10.1016/j.livsci.2021.104708
Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. doi:10.1111/j.1461-0248.2004.00684.x
Keane, O. M., Zadissa, A., Wilson, T., Hyndman, D. L., Greer, G. J., Baird, D. B., … McEwan, J. C. (2006). Gene expression profiling of Naïve sheep genetically resistant and susceptible to gastrointestinal nematodes. *BMC Genomics*, *7*, 1–12. doi:10.1186/1471-2164-7-42
Kenyon, F., Sargison, N. D., Skuce, P. J., & Jackson, F. (2009). Sheep helminth parasitic disease in south eastern Scotland arising as a possible consequence of climate change. *Veterinary Parasitology*, *163*, 293–297.
Kern, C., Wang, Y., & Chitwood, J. (2018). Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. *BMC Genomics*, *19*(684).
Kersbergen, P., van Duijn, K., Kloosterman, A. D., den Dunnen, J. T., Kayser, M., & De Knijff, P. (2009). Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genetics*, *10*, 69. doi:10.1186/1471-2156-10-69
Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Neto, L. R., Cristobal, M. S., … Dalrymple, B. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, *10*(2). doi:10.1371/journal.pbio.1001258
Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., Wilson, P., … Raadsma, H. (2009). A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLOS ONE*, *4*(3), 1–13. doi:10.1371/journal.pone.0004668
Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., & Rothschild, M. F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, *116*(3), 255–264. doi:10.1038/hdy.2015.94
Kim, K., Kwon, T., Dessie, T., Yoo, D. A., Mwai, O. A., Jang, J., … Kim, H. (2020). The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nature Genetics*, *52*(October). doi:10.1038/s41588-020-0694-2
Kim, Y., & Neilsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics*, *167*(3), 1513–1524. doi:10.1534/genetics.103.025387
Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., Mckeigue, P. M., & Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE*, *5*(11), 1–7. doi:10.1371/journal.pone.0013996
Koch, R. M., Swiger, L., Chambers, D., & Gregory, K. (1963). Efficiency of feed use in Beef Cattle. *Journal of Animal Science*, *22*, 486–494.
Kofler, R., Orozco-terWengel, P., de Maio, N., Pandey, R. V., Nolte, V., Futschik, A., … Schlötterer, C. (2011). Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE*, *6*(1). doi:10.1371/journal.pone.0015925
Koluman Darcan, N., & Silanikove, N. (2018). The advantages of goats for future adaptation to Climate Change: A conceptual overview. *Small Ruminant Research*, *163*(February 2017), 34–38. doi:10.1016/j.smallrumres.2017.04.013
Kominakis, A., Hager-Theodorides, A. L., Zoidis, E., Saridaki, A., Antonakos, G., & Tsiamis, G. (2017). Combined GWAS and 'guilt by association'-based prioritization analysis identifies functional candidate genes for body size in sheep. *Genetics Selection Evolution*, *49*(1), 1–16. doi:10.1186/s12711-017-0316-3
Koren, S., Rhie, A., Walenz, B. P., Dilthey, A. T., Bickhart, D. M., Kingan, S. B., … Adam, M. (2018). Complete assembly of parental haplotypes with trio binning. *BioRxiv*, 1–26.
Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., … Ã, M. F. S. (2008). Ancestry Informative Marker Sets for Determining Continental Origin and Admixture Proportions in Common Populations in America Human Mutation. doi:10.1002/humu.20822
Kristensen, T. N., Hoffmann, A. A., Pertoldi, C., & Stronen, A. V. (2015). What can livestock breeders learn from conservation genetics and vice versa? *Frontiers in Genetics*, *5*(FEB), 1–12. doi:10.3389/fgene.2015.00038
Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., & Scott, J. K. (2012). CliMond : global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, *3*, 53–64. doi:10.1111/j.2041-210X.2011.00134.x
Kumar, H., Panigrahi, M., Chhotaray, S., Pal, D., Bhanuprakash, V., Shandilya, R., … Bhushan, B. (2019). Identification of breed-specific SNP panel in nine different cattle genomes ., *30*(1), 78–81.
Lakhssassi, K., Lahoz, B., Sarto, P., Iguácel, L. P., Folch, J., Alabart, J. L., … Calvo, J. H. (2021). Genome-wide association study demonstrates the role played by the cd226 gene in rasa aragonesa sheep reproductive seasonality. *Animals*, *11*(4). doi:10.3390/ani11041171
Lallo, C. H. O., Cohen, J., Rankine, D., Taylor, M., Cambell, J., & Stephenson, T. (2018). Characterizing heat stress on livestock using the temperature humidity index (THI)—prospects for a warmer Caribbean. *Regional Environmental Change*, *18*(8), 2329–2340. doi:10.1007/s10113-018-1359-x
Lam, S., Miglior, F., Fonseca, P. A. S., Gómez-Redondo, I., Zeidan, J., Suárez-Vega, A., … Cánovas, A. (2021). Identification of functional candidate variants and genes for feed efficiency in Holstein and Jersey cattle breeds using RNA-

sequencing. *Journal of Dairy Science*, *104*(2), 1928–1950. doi:10.3168/jds.2020-18241
Lampis, A., Carotenuto, P., Vlachogiannis, G., Cascione, L., Hedayat, S., Burke, R., … Braconi, C. (2018). MIR21 Drives Resistance to Heat Shock Protein 90 Inhibition in Cholangiocarcinoma. *Gastroenterology*, *154*(4), 1066-1079.e5. doi:10.1053/j.gastro.2017.10.043
Landguth, E., Cushman, S., Schwartz, M., McKelvey, K., Murphy, M., & G, L. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Mol Ecol*, *19*, 4179–4191.
Lau, C. L., Smythe, L. D., Craig, S. B., & Weinstein, P. (2010). Climate change, flooding, urbanisation and leptospirosis: fuelling the fire? *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *104*, 631–638.
Lean, J. L., & Rind, D. H. (2009). How will Earth ' s surface temperature change in future decades ? *Geophysical Research Letters*, *36*(June 2009), 1–5. doi:10.1029/2009GL038932
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.
Lee, S. H., Van Der Werf, J., Lim, D. J., Park, E. W., Gondro, C., Yoon, D., … Thompson, J. (2012). Genome wide QTL mapping to identify candidate genes for carcass traits in Hanwoo (Korean Cattle). *Genes and Genomics*, *34*(1), 43–49. doi:10.1007/s13258-011-0081-6
Lees, A. M., Sejian, V., Wallage, A. L., Steel, C. C., Mader, T. L., Lees, J. C., & Gaughan, J. B. (2019). The Impact of Heat Load on Cattle. *Animals*, *9*, 1–20.
Legge, T. (1996). The beginnings of caprine domestication. In D. R. Harris (Ed.), *The Origins and Spread of Agriculture and Pastoralism in Eurasia* (pp. 238–262). New York: Smithsonian Institution Press.
Lenstra, J. a. (2005). Evolutionary and demographic history of sheep and goats suggested by nuclear, mtDNA and Y chromosome markers. *The Role of Biotechnology for the Characterization and Conservation of Crop, Forestry, Animal and Fishery Genetic Resources. International Workshop, Turin, Italy, 5-7 March 2005*, 1–4.
Leroy, G. (2014). Inbreeding depression in livestock species: Review and meta-analysis. *Animal Genetics*, *45*(5), 618–628. doi:10.1111/age.12178
Leroy, Gregoire, Baumung, R., Boettcher, P., Besbes, B., From, T., & Hoffmann, I. (2018). Animal genetic resources diversity and ecosystem services. *Global Food Security*, *17*(August 2017), 84–91. doi:10.1016/j.gfs.2018.04.003
Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., & Paschou, P. (2011). Tracing Cattle Breeds with Principal Components Analysis Ancestry Informative SNPs, *6*(4). doi:10.1371/journal.pone.0018007
Li, B., Zhang, N., Wang, Y.-G., George, A. W., Reverter, A., & Li, Y. (2018). Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Frontiers in Genetics*, *9*(237). doi:https://doi.org/10.3389/fgene.2018.00237
Li, H. (2011). A statistical framework for SNP calling, mutation discovery, as- sociation mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*, 2987–2993.
Li, R., Li, C., Chen, H., Li, R., Chong, Q., Xiao, H., & Chen, S. (2020). Genome-wide scan of selection signatures in Dehong humped cattle for heat tolerance and disease resistance. *Animal Genetics*, *51*(2), 292–299. doi:10.1111/age.12896
Li, S., Chen, W., Zheng, X., Liu, Z., Yang, G., Hu, X., & Mou, C. (2020). Comparative investigation of coarse and fine wool sheep skin indicates the early regulators for skin and wool diversity. *Gene*, *758*(March), 144968. doi:10.1016/j.gene.2020.144968
Liang, Z., Bu, L., Qin, Y., Peng, Y., Yang, R., & Zhao, Y. (2019). Selection of Optimal Ancestry Informative Markers for Classification and Ancestry Proportion Estimation in Pigs. *Frontiers in Genetics*, *10*, 183. doi:10.3389/fgene.2019.00183
Librado, P., Gamba, C., Gaunitz, C., Der Sarkissian, C., Pruvost, M., Albrechtsen, A., … Al., E. (2017). Ancient genomic changes associated with domestication of the horse. *Science*, *356*, 442–445.
Lima-ribeiro, M. S., Varela, S., & Oliveira, G. De. (2015). EcoClimate : a database of climate data from multiple models for past , present , and future for macroecologists and biogeographers ECOCLIMATE : A DATABASE OF CLIMATE DATA FROM MULTIPLE MODELS FOR PAST , PRESENT , AND FUTURE FOR MACROECOLOGISTS AND BIOGEO. *Biodiversity Informatics*, *10*(November), 0–21. doi:10.17161/bi.v10i0.4955
Littlejohn, M. D., Henty, K. M., Tiplady, K., Johnson, T., Harland, C., Lopdell, T., … Davis, S. R. (2014). Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nature Communications*, 1–8. doi:10.1038/ncomms6861
Liu, Y., Wang, D., He, F., Wang, J., Joshi, T., & Xu, D. (2019). Phenotype Prediction and Genome-Wide Association Study Using Deep Convolutional Neural Network of Soybean. *Frontiers in Genetics*, *10*.
Liu, Y., Weyrich, L., & Llamas, B. (2020). More Arrows in the Ancient DNA Quiver: Use of Paleoepigenomes and Paleomicrobiomes to Investigate Animal Adaptation to Environment. *Mol. Biol. Evol.*, *37*, 307–319.
Lorenzini, R., Cabras, P., Fanelli, R., & Carboni, G. L. (2011). Wildlife molecular forensics: Identification of the Sardinian mouflon using STR profiling and the Bayesian assignment test. *Forensic Science International: Genetics*, *5*(4), 345–349. doi:10.1016/j.fsigen.2011.01.012
Low, W. Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D. M., … Williams, J. L. (2020). Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nature Communications*, *11*, 1–14. doi:10.1038/s41467-020-15848-y
Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. doi:10.1111/1755-0998.12592
Lv, F. H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., … Marsan, P. A. (2014a). Adaptations to climate-mediated selective pressures in sheep. *Molecular Biology and Evolution*, *31*(12), 3324–3343. doi:10.1093/molbev/msu264

Lv, F. H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., … Marsan, P. A. (2014b). Adaptations to climate-mediated selective pressures in sheep. *Molecular Biology and Evolution*, *31*(12), 3324–3343. doi:10.1093/molbev/msu264

Macciotta, N. P. P., Colli, L., Cesarani, A., Ajmone-Marsan, P., Low, W. Y., Tearle, R., & Williams, J. L. (2021). The distribution of runs of homozygosity in the genome of river and swamp buffaloes reveals a history of adaptation, migration and crossbred events. *Genetics Selection Evolution*, *53*(1), 1–21. doi:10.1186/s12711-021-00616-3

Maclean, C. A., Chue Hong, N. P., & Prendergast, J. G. D. (2015). Hapbin: An efficient program for performing haplotype-based scans for positive selection in large genomic datasets. *Molecular Biology and Evolution*, *32*(11), 3027–3029. doi:10.1093/molbev/msv172

Madalena, F. E. (2002). DAIRY ANIMALS | Bos indicus Breeds and Bos indicus × Bos taurus Crosses. In *Encyclopedia of Dairy Sciences* (pp. 576–585).

Mader, T., Davis, S., Gaughan, J., & Brown-brandl, T. (2003). Wind speed and solar radiation adjustments for the temperature-humidity index.

Mader, T. L., Davis, M. S., & Brown-Brandl, T. (2006). Environmental factors influencing heat stress in feedlot cattle. *Journal of Animal Science*, *84*.

Mader, T. L., Johnson, L. J., Gaughan, J. B., Mader, T. L., Johnson, L. J., & Gaughan, J. B. (2010). A comprehensive index for assessing environmental stress in animals. *Journal of Animal Science*, *88*, 2153–2165. doi:10.2527/jas.2009-2586

Mallet, J. (2005). Hybridization as an invasion of the genome, *20*(5). doi:10.1016/j.tree.2005.02.010

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, *18*(4), 189–197. doi:10.1016/S0169-5347(03)00008-9

Manfredi, M. T., Di Cerbo, A. R., Zanzani, S., & Stradiotto, K. (2010). Breeding management in goat farms of Lombardy, northern Italy: Risk factors connected to gastrointestinal parasites. *Small Ruminant Research*, *88*(2–3), 113–118. doi:10.1016/j.smallrumres.2009.12.018

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. doi:10.1093/bioinformatics/btq559

Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics*, *93*(2), 278–288. doi:10.1016/j.ajhg.2013.06.020

Mariasegaram, M., Chase, C. C. J., Chaparro, J. X., Olson, T. A., Brenneman, R. A., & Niedz, R. P. (2007). The slick hair coat locus maps to chromosome 20 in Senepol-derived cattle. *Animal Genetics*, *38*, 54–59.

Marino, R., Capoferri, R., Panelli, S., Minozzi, G., Strozzi, F., Trevisi, E., … Williams, J. L. (2016). Johne's disease in cattle: an in vitro model to study early response to infection of Mycobacterium avium subsp paratuberculosis using RNA-seq. *Molecular Immunology*, *91*, 259–271.

Martinez-Royo, A., Alabart, J. L., Sarto, P., Serrano, M., Lahoz, B., Folch, J., & Calvo, J. H. (2017). Genome-wide association studies for reproductive seasonality traits in Rasa Aragonesa sheep breed. *Theriogenology*, *99*, 21–29. doi:10.1016/j.theriogenology.2017.05.011

Martinez, A. M., Gama, L. T., Canon, J., Ginja, C., Delgado, J. V., Dunner, S., & Al, E. (2012). Genetic footprints of Iberian cattle in America 500 years after the arrival of Columbus. *PLOS ONE*, *7*.

Masseti, M. (2009). The wild goats Capra aegagrus Erxleben, 1777 of the Mediterranean Sea and the Eastern Atlantic Ocean islands. *Mammal Review*, *39*(2), 141–157. doi:10.1111/j.1365-2907.2009.00141.x

Masseti, M. (2016). The wild goat, Capra aegagrus Erxleben, 1777, of the island of Montecristo (Northern Tyrrhenian Sea, Italy): Does it still exist? *Mammalia*, *80*(2), 125–141. doi:10.1515/mammalia-2014-0168

Masson-Delmotte, V. (2017). Towards the IPCC Special Report on Global Warming of 1.5◦C. *Geophysical Research Abstracts*. Retrieved from https://www.ipcc.ch/site/assets/uploads/sites/2/2019/05/SR15_SPM_version_report_LR.pdf

Masson-Delmotte, V., Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., Shukla, P. R., … Waterfield, T. (2018). *IPCC, 2018: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of cli*.

Mastrangelo, S., Moioli, B., Ahbara, A., Latairish, S., Portolano, B., Pilla, F., & Ciani, E. (2019). Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Animal Production Science*, *59*(5), 835–848. doi:10.1071/AN17753

Mastrangelo, Salvatore, Di Gerlando, R., Sardina, M. T., Sutera, A. M., Moscarelli, A., Tolone, M., … Portolano, B. (2021). Genome-wide patterns of homozygosity reveal the conservation status in five italian goat populations. *Animals*, *11*(1510), 1–11. doi:10.3390/ani11061510

Mastrangelo, Salvatore, Moioli, B., Ahbara, A., Latairish, S., Portolano, B., Pilla, F., & Ciani, E. (2019). Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Animal Production Science*, *59*(5), 835. doi:10.1071/AN17753

Mastrangelo, Salvatore, Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., & Portolano, B. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution*, *49*(1), 1–10. doi:10.1186/s12711-017-0360-z

Mazzullo, G., Rifici, C., Cammarata, F., Caccamo, G., Rizzo, M., Piccione, G., … Practicioner, V. (2014). Effect of different environmental conditions on some haematological parameters in cow. *Ann. Anim. Sci.*, *14*(4), 947–954. doi:10.2478/aoas-2014-0049

McHugo, G. P., Browett, S., Randhawa, I. A. S., Howard, D. J., Mullen, M. P., Richardson, I. W., … MacHugh, D. E. (2019). A Population Genomics Analysis of the Native Irish Galway Sheep Breed. *Frontiers in Genetics*, *10*(October), 1–13. doi:10.3389/fgene.2019.00927

Mchugo, G. P., Dover, M. J., & Machugh, D. E. (2019). Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. *BMC Biology*, 1–20.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., … Wilson, J. F. (2008). Runs of Homozygosity in European Populations. *American Journal of Human Genetics*, *83*(3), 359–372. doi:10.1016/j.ajhg.2008.08.007

Mdladla, K., & Dzomba, E. F. (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

Meng, Q., Wang, K., Liu, X., Zhou, H., Xu, L., Wang, Z., & Fang, M. (2017). Identification of growth trait related genes in a Yorkshire purebred pig population by genome-wide association studies. *Asian-Australasian Journal of Animal Sciences*, *30*(4), 462–469. doi:10.5713/ajas.16.0548

Mereu, P., Pirastru, M., Barbato, M., Satta, V., Hadjisterkotis, E., Manca, L., … Leoni, G. G. (2019). Identification of an ancestral haplotype in the mitochondrial phylogeny of the ovine haplogroup B. *PeerJ*, *2019*(10), 1–18. doi:10.7717/peerj.7895

Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, *11*(jANuARy), 31–46. doi:10.1038/nrg2626

Meuwissen, T. H. E., Sonesson, A. K., Gebregiwergis, G., & Woolliams, J. A. (2020). Management of Genetic Diversity in the Era of Genomics. *Frontiers in Genetics*, *11*(August), 1–16. doi:10.3389/fgene.2020.00880

Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*, 1819–1829.

Meyermans, R., Gorssen, W., Buys, N., & 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), 1–14. doi:10.1186/s12864-020-6463-x

Miao, X., Luo, Q., Zhao, H., & Al., E. (2016). Co-expression analysis and identification of fecundity-related long non-coding RNAs in sheep ovaries. *Sci. Rep.*, *6*.

Miao, Xiangyang, Luo, Q., & Qin, X. (2015). Genome-wide analysis reveals the differential regulations of mRNAs and miRNAs in dorset and small tail han sheep muscles. *Gene*, *562*(2), 188–196. doi:10.1016/j.gene.2015.02.070

Michailidou, S., Tsangaris, G. T., Tzora, A., Skoufos, I., Banos, G., Argiriou, A., & Arsenos, G. (2019). Analysis of genome-wide DNA arrays reveals the genomic population structure and diversity in autochthonous Greek goat breeds. *PLOS ONE*, *14*, 1–28.

Mignon-grasteau, S., Boissy, A., Bouix, J., Faure, J., Fisher, A. D., Hinch, G. N., … Beaumont, C. (2005). Genetics of adaptation and domestication in livestock B. *Livestock Production Science*, *93*, 3–14. doi:10.1016/j.livprodsci.2004.11.001

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

Miller, J. M. (2012). Genomic consequences of genetic rescue in an insular population of bighorn sheep (Ovis canadensis), 1583–1596. doi:10.1111/j.1365-294X.2011.05427.x

Miron B. Kursa, W. R. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistics*, *36*(11), 1–13.

Miyagawa, Y., Nishimura, H., Tsujimura, A., Matsuoka, Y., Matsumiya, K., Okuyama, A., … Tanaka, H. (2005). Single-nucleotide polymorphisms and mutation analyses of the TNP1 and TNP2 genes of fertile and infertile human male populations. *Journal of Andrology*, *26*(6), 779–786. doi:10.2164/jandrol.05069

Mohammadi, A., Alijani, S., Rafat, S. A., & Abdollahi-Arpanahi, R. (2020). Genome-Wide Association Study and Pathway Analysis for Female Fertility Traits in Iranian Holstein Cattle. *Annals of Animal Science*, *20*(3), 825–851. doi:10.2478/aoas-2020-0031

Molotsi, A., Cloete, S., Taylor, J. F., & Whitacre, L. (2018). Identification of selection signatures in South African sheep popula-tions using HAPFLK and Bayesian Fst approaches Identification of selection signatures in South African sheep populations using HAPFLK and Bayesian Fst approaches. ,. In *Proceedings of the World Congress on Genetics Applied to Livestock* (pp. 1–5.).

Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., & McEwan, J. C. (2012a). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genetics*, *13*(1), 10. doi:10.1186/1471-2156-13-10

Morimoto, R. I. (1993). Cells in Stress: Transcriptional Activation of Heat Shock. *Science*, *259*(5100), 1409–1410. Retrieved from http://www.jstor.org/stable/2880783

Morton, J. F. (2007). The impact of climate change on smallholder and subsistence agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19680–19685. doi:10.1073/pnas.0701855104

Moser, G., Lee, S. H., Hayes, B. J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2015). Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLOS ONE*, *11*.

Mundel, T., Bunn, S. J., Hooper, P. L., & Jones, D. A. (2007). The effects of face cooling during hyperthermic exercise in man : evidence for an integrated thermal , neuroendocrine and behavioural response. *Exp. Physiol*, *92*, 187–195. doi:10.1113/expphysiol.2006.034934

Mwacharo, J. M., Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B. A., & Rothschild, M. F. (2017). Genomic

footprints of dryland stress adaptation in Egyptian fat-Tail sheep and their divergence from East African and western Asia cohorts. *Scientific Reports*, *7*(1), 1–10. doi:10.1038/s41598-017-17775-3

Naderi, S., Rezaei, H. R., Taberlet, P., Zundel, S., Rafat, S., Naghash, H., … Pompanon, F. (2007). Large-Scale Mitochondrial DNA Analysis of the Domestic Goat Reveals Six Haplogroups with High Diversity. *PLOS ONE*, (10), 1–12. doi:10.1371/journal.pone.0001012

Nandolo, W., Mészáros, G., Banda, L. J., Gondwe, T. N., Lamuno, D., Mulindwa, H. A., … Sölkner, J. (2019). Timing and extent of inbreeding in African goats. *Frontiers in Genetics*, *10*(JUN), 1–11. doi:10.3389/fgene.2019.00537

Nannini, D., Calistri, P., Giovannini, A., Di Ventura, M., Cafiero, M. A., Ferrari, G., … Caporale, V. (2004). Health management of large transhumant animal populations and risk of bluetongue spread to disease-free areas. *Veterinaria Italiana*, *40*(4), 707–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20422619

National Research Council. (1981). Effect of enviroment on utrient requirement of domestic animals. In National Research Council (Ed.), *Subcommittee on Environmental Stress* (National A). Washington D. C.

Nayeri, S., Sargolzaei, M., & Tulpan, D. (2019). A review of traditional and machine learning methods applied to animal breeding. *Anim. Health. Res. Rev.*, *20*, 31–46.

Neethirajan, S. (2020). Sensing and Bio-Sensing Research The role of sensors , big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, *29*(June), 1–8. doi:10.1016/j.sbsr.2020.100367

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12 (I)), 3321–3323. doi:10.1073/pnas.70.12.3321

Nei, Masatoshi. (1978). Estimation of Average Heterozygosity and genetic distance from a small number of individuals. *Genetics*, 583–590.

New, M., Lister, D., Hulme, M., & Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate Research*, *21*(1), 1–25. doi:10.3354/cr021001

Nicolazzi, E. L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., … Stella, A. (2015). SNPchiMp v.3: Integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics*, *16*(1), 1–6. doi:10.1186/s12864-015-1497-1

Nicoloso, L., Bomba, L., Colli, L., Negrini, R., Milanesi, M., Mazza, R., … Crepaldi, P. (2015). Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genetics Selection Evolution*, *47*(1), 1–10. doi:10.1186/s12711-015-0140-6

Noce, S., Caporaso, L., & Santini, M. (2020). A new global dataset of bioclimatic indicators. *Scientific Data*, *7*(Cmcc), 1–12. doi:10.1038/s41597-020-00726-5

Nomura, K., Ishii, K., Dadi, H., Takahashi, Y., Minezawa, M., Cho, C. Y., … Amano, T. (2012). Microsatellite DNA markers indicate three genetic lineages in East Asian indigenous goat populations. *Animal Genetics*, *43*(6), 760–767. doi:10.1111/j.1365-2052.2012.02334.x

Nothnagel, M., Lu, T. T., Kayser, M., & Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human Molecular Genetics*, *19*(15), 2927–2935. doi:10.1093/hmg/ddq198

O'Mara, F. P. (2012). The role of grasslands in food security and climate change. *Annals of Botany*, *110*, 1263–1270. doi:10.1093/aob/mcs209

O'Neill, C., Swain, D., & Kadarmideen, H. (2010). Evolutionary process of Bos taurus cattle in favourable versus unfavourable environments and its implications for genetic selection. *Evol Appl*, *3*, 422–433.

Okser, S., Pahikkala, T., Airola, A., Salakoski, T., Ripatti, S., & Aittokallio, T. (2014). Regularized Machine Learning in the Genetic Prediction of Complex Traits. *PLoS Genetics*, *10*(11). doi:10.1371/journal.pgen.1004754

Olden, J. D., Poff, N. L., Douglas, M. R., Douglas, M. E., & Fausch, K. D. (2004). Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution*, *19*(1), 18–24. doi:10.1016/j.tree.2003.09.010

Oliveira, P. S. N. De, Cesar, A. S. M., Nascimento, M. L., Chaves, A. S., Tizioto, P. C., Tullio, R. R., … Ontology, G. (2014). Identification of genomic regions associated with feed efficiency in Nelore cattle, 1–10.

Oliveira, R., Randi, E., Mattucci, F., Kurushima, J. D., Lyons, L. A., & Alves, P. C. (2015). Toward a genome-wide approach for detecting hybrids: Informative SNPs to detect introgression between domestic cats and European wildcats (Felis silvestris). *Heredity*, *115*(3), 195–205. doi:10.1038/hdy.2015.25

Olson, T. A., Lucena, C., Chase, C. C. J., & Hammond, A. C. (2003). Evidence of a major gene influencing hair length and heat tolerance in Bos taurus cattle. *Journal of Animal Science*, *81*, 80–90.

Olwoch, J. M., Reyers, B., Engelbrecht, F. A., & Erasmus, B. F. N. (2008). Climate change and the tick-borne disease, Theileriosis (East Coast fever) in sub-Saharan Africa. *Journal of Arid Environments*, *72*, 108–120.

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

Orrù, L., Catillo, G., Napolitano, F., De Matteis, G., Scatà, M. C., Signorelli, F., & Moioli, B. (2009). Characterization of a SNPs panel for meat traceability in six cattle breeds. *Food Control*, *20*(9), 856–860. doi:10.1016/j.foodcont.2008.10.015

Ortiz-Bobea, A., Knippenberg, E., & Chambers, R. G. (2018). Growing climatic sensitivity of U.S. Agriculture linked to technological change and regional specialization. *Science Advances*, *4*(12), 1–9. doi:10.1126/sciadv.aat4343

Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., … Richards, M. B. (2012). Mitochondrial DNA signals of late glacial recolonization of europe from near eastern refugia. *American Journal of Human Genetics*, *90*(5), 915–924. doi:10.1016/j.ajhg.2012.04.003

Paredes-Sanchez, F. A., Sifuentes-Rincon, A. M., Casas, E., Arellano-vera, W., Parra-bracamonte, G. M., Riley, D. G., …

Randel, R. D. (2020). Novel genes involved in the genetic architecture of temperament in Brahman cattle. *PLOS ONE*, *15*, 1–13. doi:10.1371/journal.pone.0237825

Pariset, L., Joost, S., Ajmone Marsan, P., Valentini, A., & Ec, E. C. (2009). Landscape genomics and biased F ST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics*, *8*, 1–8. doi:10.1186/1471-2156-10-7

Pariset, L., Joost, S., Gargani, M., & Valentini, A. (2012). Landscape Genomics in Livestock. *Analysis of Genetic Variation in Animals*, (Box 7). doi:10.5772/32927

Park, S. D. E., Magee, D. A., McGettigan, P. A., Teasdale, M. D., Edwards, C. J., Lohan, A. J., … MacHugh, D. E. (2015). Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biology*, *16*(1), 1–15. doi:10.1186/s13059-015-0790-2

Pasandideh, M., Rahimi-Mianji, G., & Gholizadeh, M. (2018). A genome scan for quantitative trait loci affecting average daily gain and Kleiber ratio in Baluchi Sheep. *Journal of Genetics*, *97*(2), 493–503. doi:10.1007/s12041-018-0941-9

Pasaniuc, B., Sankararaman, S., Kimmel, G., & Halperin, E. (2009). Inference of locus-specific ancestry in closely related populations. *Bioinformatics (Oxford, England)*, *25*(12), i213-21. doi:10.1093/bioinformatics/btp197

Paschino, P., Stocco, G., Dettori, M. L., Pazzola, M., Marongiu, M. L., Pilo, C. E., … Vacca, G. M. (2020). Characterization of milk composition, coagulation properties, and cheese-making ability of goats reared in extensive farms. *Journal of Dairy Science*, *103*(7), 5830–5843. doi:10.3168/jds.2019-17805

Paschou, P., Drineas, P., Lewis, J., Nievergelt, C. M., Nickerson, D. A., Smith, J. D., … Ziv, E. (2008). Tracing Sub-Structure in the European American Population with PCA-Informative Markers, *4*(7). doi:10.1371/journal.pgen.1000114

Paschou, P., Ziv, E., Burchard, E. G., Choudhry, S., Rodriguez-cintron, W., Mahoney, M. W., & Drineas, P. (2007). PCA-Correlated SNPs for Structure Identification in Worldwide Human Populations, *3*(9). doi:10.1371/journal.pgen.0030160

Pasquariello, R., Fernandez-Fuertes, B., Strozzi, F., Mazza, R., Lonergan, P., Gandolfi, F., & Williams, J. L. (2017). Profiling bovine blastocyst microRNAs using deep sequencing. *Journal of Reproduction, Fertility and Development*, *29*, 1545–1555.

Passamonti, M. M., Somenzi, E., Barbato, M., Chillemi, G., Colli, L., Joost, S., … Ajmone-Marsan, P. (2021). The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock. *Animals*, *11*(10), 2833. doi:10.3390/ani11102833

Pauls, S. U., Nowak, C., Bálint, M., & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, *22*(4), 925–946. doi:10.1111/mec.12152

Pavan, M. (1971). *Montecristo riserva naturale*. *Istituto Entomologia agraria di Pavia/Tipografia Meroni, Albese (Como).*

Pelliccioni Raganella, E., Lazzaro, L., Gotti, C., & Baccetti, N. (2015). *PIANO DI GESTIONE E CONSERVAZIONE DELLA CAPRA DI MONTECRISTO : SINTESI DEL CONTESTO E AZIONI*. Retrieved from https://www.restoconlife.eu/wordpress/wp-content/uploads/2015/06/Piano-di-gestione-della-capra-di-Montecristo.pdf

Pendleton, A., Shen, F., Taravella, A., Emery, S., Veeramah, K., Boyko, A., & JM., K. (2018). Comparison of village dog and wolf ge-nomes highlights the role of the neural crest in dog domestication. *BMC Biology*, *16*.

Peng, S., Shi, S., Tao, G., Li, Y., Xiao, D., Wang, L., … Xiao, J. (2021). JKAMP inhibits the osteogenic capacity of adipose-derived stem cells in diabetic osteoporosis by modulating the Wnt signaling pathway through intragenic DNA methylation. *Stem Cell Research and Therapy*, *12*(1), 1–15. doi:10.1186/s13287-021-02163-6

Pereira, F., Queirós, S., Gusmão, L., Nijman, I. J., Cuppen, E., Lenstra, J. A., … Amorim, A. (2009). Tracing the history of goat pastoralism: New clues from mitochondrial and y chromosome DNA in North Africa. *Molecular Biology and Evolution*, *26*(12), 2765–2773. doi:10.1093/molbev/msp200

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., & Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics*, *48*(3), 255–271. doi:10.1111/age.12526

Pezzopane, R. M., Carlos, A., Bernardi, C., Azenha, M. V., Anch, P. P., Bosi, C., … Esteves, N. (2020). Production and nutritive value of pastures in integrated livestock production systems: shading and management effects. *Scientia Agricola*, *77*.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*(7), 1929–1936. doi:10.1093/molbev/msu136

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11), 17. doi:10.1371/journal.pgen.1002967

Piles, M., Lozano, C. F., Galilea, M. V., Rodríguez, O. G., Sánchez, J. P., Torrallardona, D., … Quintanilla, R. (2019). Machine learning applied to transcriptomic data to identify genes associated with feed efficiency in pigs. *Genetics Selection Evolution*, *51*, 1–15. doi:10.1186/s12711-019-0453-y

Pitt, D., Bruford, M. W., Barbato, M., Orozco-terWengel, P., Martínez, R., & Sevane, N. (2019). Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evolutionary Applications*, *12*(1), 105–122. doi:10.1111/eva.12641

Pitt, D., Sevane, N., Nicolazzi, E. L., MacHugh, D. E., Park, S. D. E., Colli, L., … Orozco-terWengel, P. (2019). Domestication of cattle: Two or three events? *Evolutionary Applications*, *12*(1), 123–136. doi:10.1111/eva.12674

Popescu, A.-A., & Huber, K. T. (2015). PSIKO2: a fast and versatile tool to infer population stratification on various levels in GWAS. *Bioinformatics (Oxford, England)*, *31*(21), 3552–4. doi:10.1093/bioinformatics/btv396

Porto-neto, L. R., Bickhart, D. M., Landaeta-hernandez, A. J., Utsunomiya, Y. T., Pagan, M., Jimenez, E., … Stella, A. (2018). Convergent Evolution of Slick Coat in Cattle through Truncation Mutations in the Prolactin Receptor. *Frontiers in Genetics*, *9*(February), 1–8. doi:10.3389/fgene.2018.00057

Porto-neto, L. R., Reverter, A., & Prayaga, K. C. (2014). The Genetic Architecture of Climatic Adaptation of Tropical Cattle. *PLOS ONE*, *9*, 1–22. doi:10.1371/journal.pone.0113284

Postolache, D., Oddou-Muratorio, S., Vajana, E., Bagnoli, F., Guichoux, E., Hampe, A., … Vendramin, G. G. (2021). Genetic signatures of divergent selection in European beech (Fagus sylvatica L.) are associated with the variation in temperature and precipitation across its distribution range. *Molecular Ecology*, *30*(20), 5029–5047. doi:10.1111/mec.16115

Preserving the Biodiversity of the "Il Felcetone" Farm in Tuscany. (n.d.). Retrieved May 10, 2021, from https://www.nandoandelsaperettifoundation.org/en/page.php/page.php?project=512

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. doi:10.1093/genetics/155.2.945

QGIS development team. (2009). QGIS geographic information system. Open source geospatial foundation project. Retrieved January 29, 2022, from http://www.qgis.org/

R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. doi:10.1108/eb003648

Randi, E, Tosi, G., Toso, S., Lorenzini, R., & Fusco, G. (1990). Genetic variability and conservation problems in Alpine ibex, domestic and feral goat populations (genus Capra). *Zeitschrift Für Säugetierkunde*, *55*(6), 413–420.

Randi, Ettore, Hulva, P., Fabbri, E., Galaverni, M., Galov, A., Kusak, J., … Caniglia, R. (2014). Multilocus detection of wolf x dog hybridization in Italy, and guidelines for marker selection. *PLoS ONE*, *9*(1). doi:10.1371/journal.pone.0086409

Rashamol, V. P., Sejian, V., Pragna, P., Lees, A. M., Bagath, M., Krishnan, G., & Gaughan, J. B. (2019). Prediction models , assessment methodologies and biotechnological tools to quantify heat stress response in ruminant livestock. *International Journal of Biometeorology*. doi:https://doi.org/10.1007/s00484-019-01735-9 REVIEW PAPER Prediction

Raven, L.-A., Cocks, B. G., & Hayes, B. J. (2014). Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics*, *15*.

Rees, J. S., Castellano, S., & Andrés, A. M. (2020). The Genomics of Human Local Adaptation. *Trends in Genetics*, *36*, 415–428.

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance, 767–779.

Rezaei, H. R., Naderi, S., Chintauan-Marquier, I. C., Taberlet, P., Virk, A. T., Naghash, H. R., … Pompanon, F. (2010). Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution*, *54*(2), 315–326. doi:10.1016/j.ympev.2009.10.037

Rhymer, J. M., & Simberloff, D. (1996). Extinction By Hybridization and Introgression. *Annual Review of Ecology and Systematics*, *27*(1), 83–109. doi:10.1146/annurev.ecolsys.27.1.83

Riley, D. G., C.A., G. ., Boldt, C. R., Funkhouser, R. R., Herring, A. D., Riggs, P. K., … Sanders, J. O. (2016). Crossbred steer temperament as yearlings and whole genome association of steer temperament as yearlings and calf temperament post-weaning. *J Anim Sci.*, *94*, 1408–1414. doi:doi: 10.2527/jas.2015-0041.

Rodríguez-Ramilo, S. T., Elsen, J. M., & Legarra, A. (2019). Inbreeding and effective population size in French dairy sheep: Comparison between genomic and pedigree estimates. *Journal of Dairy Science*, *102*(5), 4227–4237. doi:10.3168/jds.2018-15405

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

Ropiquet, A., & Hassanin, A. (2006). Hybrid origin of the Pliocene ancestor of wild goats. *Molecular Phylogenetics and Evolution*, *41*(2), 395–404. doi:10.1016/j.ympev.2006.05.033

Rotter, R., & Van de Geijn, S. C. (1999). Climate Change Effects On Plant Growth, crop yield and livestock. *Climatic Change*, *43*, 651–681.

Ruane, J., & Sonnino, A. (2006). Background document to the e-mail conference on the role of biotechnology for the characterization and conservation of crop, forest, animal and fishery genetic resources in developing countries. In The role of biotechnology in exploring and protecting agr. *FAO Rome*, 151–172.

Russell, T., Cullingham, C., Kommadath, A., Stothard, P., Herbst, A., & Coltman, D. (2019). Development of a Novel Mule Deer Genomic Assembly and Species-Diagnostic SNP Panel for Assessing Introgression in Mule Deer, White-Tailed Deer, and Their Interspecific Hybrids. *G3: Genes|Genomes|Genetics*, *9*(3), g3.200838.2018. doi:10.1534/g3.118.200838

Rust, W., Holman, I., Bloomfield, J., Cuthbert, M., & Corstanje, R. (2019). Understanding the potential of climate teleconnections to project future groundwater drought. *Hydrology and Earth System Sciences*, *23*, 3233–3245.

Ryan, W. B. F., Carbotte, S. M., Coplan, J. O., O'Hara, S., Melkonian, A., Arko, R., … Zemsky, R. (2009). Global multi-resolution topography synthesis. *Geochemistry, Geophysics, Geosystems*, *10*(3). doi:10.1029/2008GC002332

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., … Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*(6909), 832–837. doi:10.1038/nature01140

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913–918. doi:10.1038/nature06250

Saif, R., Henkel, J., Jagannathan, V., Drögemüller, C., Flury, C., & Leeb, T. (2020). The LCORL Locus Is under Selection in Large-Sized Pakistani Goat Breeds. *Genes*, *11*(168).

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., … Wall, D. H. (2000). Global biodiversity

scenarios for the year 2100. *Science*, *287*(5459), 1770–1774. doi:10.1126/science.287.5459.1770
Salter-Townshend, M., & Myers, S. (2019). Fine-scale inference of ancestry segments without prior knowledge of admixing groups. *Genetics*, *212*(3), 869–889. doi:10.1534/genetics.119.302139
Sandrucci, A., Bava, L., Tamburini, A., Gislon, G., & Zucali, M. (2019). Management practices and milk quality in dairy goat farms in Northern Italy. *Italian Journal of Animal Science*, *18*(1), 1–12. doi:10.1080/1828051X.2018.1466664
Sanna, D., Barbato, M., Hadjisterkotis, E., Cossu, P., Decandia, L., Trova, S., … Mereu, P. (2015). The First Mitogenome of the Cyprus Mouflon (Ovis gmelini ophion): New Insights into the Phylogeny of the Genus Ovis. *PloS One*, *10*(12), e0144257. doi:10.1371/journal.pone.0144257
Santana, M. H. A., Utsunomiya, Y. T., Neves, H. H. R., Gomes, R. C., Garcia, J. F., Fukumasu, H., … Ferraz, J. B. S. (2014). Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. *BMC Genetics*, *15*, 1–8. doi:10.1186/1471-2156-15-21
Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., & Mishra, B. P. (2020). Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livestock Science*, *241*(July), 104257. doi:10.1016/j.livsci.2020.104257
Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., & Mishra, B. P. (2021). Genome-wide analysis of genetic diversity and selection signatures in three Indian sheep breeds. *Livestock Science*, *243*(November 2020), 104367. doi:10.1016/j.livsci.2020.104367
SAS Institute Inc. 1. (n.d.). JMP®, Version 16. Retrieved October 8, 2021, from https://www.jmp.com/support/notes/35/282.html
Schrider, D.R., & Kern, A. D. (2017). Soft Sweeps Are the Dominant Mode of Adaptation in the Human Genome. *Molecular Biology and Evolution*, *34*, 1863–1877. doi:https://doi.org/10.1093/molbev/msx154
Schrider, Daniel R., & Kern, A. D. (2018). Supervised Machine Learning for Population Genetics: A New Paradigm. *Trends in Genetics*, *34*(4), 301–312. doi:10.1016/j.tig.2017.12.005
Schubert, R., Andaleon, A., & Wheeler, H. E. (2020). Comparing local ancestry inference models in populations of two- And three-way admixture. *PeerJ*, *8*, 1–19. doi:10.7717/peerj.10090
Segnalini, M., Bernabucci, U., Vitali, A., Nardone, A., & Lacetera, N. (2013). Temperature humidity index scenarios in the Mediterranean basin. *International Journal of Biometeorology*, *57*(3), 451–458. doi:10.1007/s00484-012-0571-5
Seif, S. M., Johnson, H. D., & Lippincott, A. C. (1979). The effects of heat exposure (31°C) on Zebu and Scottish Highland cattle. *International Journal of Biometeorology*, *23*(1), 9–14. doi:10.1007/BF01553372
Sejian, Veerasamy, Bhatta, R., Gaughan, J., Malik, P. K., Naqvi, S. M. K., & Lal, R. (2017). *Sheep production adapting to climate change*. *Sheep Production Adapting to Climate Change*. doi:10.1007/978-981-10-4714-5
Selli, A., Ventura, R. V, Fonseca, P. A. S., Buzanskas, M. E., Andrietta, L. T., Balieiro, J. C. C., & Brito, L. F. (2021). Detection and Visualization of Heterozygosity-Rich Regions and Runs of Homozygosity in Worldwide Sheep Populations. *Animals*, *11*(2696), 1–24.
Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., & Joost, S. (2020). Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. *Molecular Ecology Resources*, *20*(1), 154–169. doi:10.1111/1755-0998.13095
Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., … Cassandro, M. (2020). The genetic heritage of Alpine local cattle breeds using genomic SNP data. *Genetics Selection Evolution*, *52*(1), 1–12. doi:10.1186/s12711-020-00559-1
Seong, J., Yoon, H., & Kong, H. S. (2016). Identification of microRNA and target gene associated with marbling score in Korean cattle (Hanwoo). *Genes and Genomics*, *38*(6), 529–538. doi:10.1007/s13258-016-0401-y
Serranito, B., Cavalazzi, M., Vidal, P., Taurisson-Mouret, D., Ciani, E., Bal, M., … Da Silva, A. (2021). Local adaptations of Mediterranean sheep and goats through an integrative approach. *Scientific Reports*, *11*(1), 1–17. doi:10.1038/s41598-021-00682-z
Sharma, A., Seop, J., Dang, C. G., Sudrajad, P., & Kim, H. C. (2015). Stories and Challenges of Genome Wide Association Studies in Livestock — A Review, *28*(10), 1371–1379.
Shukla, P. R., Skea, J., Buendia, E. C., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D. C., … Malley, J. (2019). *IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*.
Sillmann, J., Kharin, V. V, Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble : Part 1 . Model evaluation in the present climate. *JOURNAL OF GEOPHYSICAL RESEARCH*, *118*(July 2012), 1716–1733. doi:10.1002/jgrd.50203
Sillmann, J., Kharin, V. V, Zwiers, F. W., Zhang, X., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble : Part 2 . Future climate projections. *JOURNAL OF GEOPHYSICAL RESEARCH*, *118*(July 2012), 2473–2493. doi:10.1002/jgrd.50188
Silva, D. B. S., Fonseca, L. F. S., Pinheiro, D. G., Magalhães, A. F. B., Muniz, M. M. M., Ferro, J. A., … Albuquerque, L. G. (2020). Spliced genes in muscle from Nelore Cattle and their association with carcass and meat quality. *Scientific Reports*, *10*, 1–13. doi:10.1038/s41598-020-71783-4
Simonsen, K. L., Churchill, G. A., & Aquadro, C. F. (1995). Properties of Statistical Tests of Neutrality for DNA Polymorphism Data. *Genetics*, *141*, 413–429.
Singh, A., Mehrotra, A., Gondro, C., Romero, A. R. da S., Pandey, A. K., Karthikeyan, A., … Kumar, A. (2020). Signatures of Selection in Composite Vrindavani Cattle of India. *Frontiers in Genetics*, *11*(December), 1–7.

doi:10.3389/fgene.2020.589496
Somenzi, E., Ajmone-Marsan, P., & Barbato, M. (2020). Identification of ancestry informative marker (AIM) panels to assess hybridisation between feral and domestic sheep. *Animals*, *10*(4). doi:10.3390/ani10040582
Song, S., Yao, N., Yang, M., Liu, X., Dong, K., Zhao, Q., … Jiang, L. (2016). Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus). *BMC Genomics*, *17*(1), 1–12. doi:10.1186/s12864-016-2449-0
Spagnesi, M., Toso, S., & Masseti, M. (2003). *Dati paleontologici e archeozoologici*. *Fauna d'Italia. Mammalia III. Carnivora – Artiodactyla. Edizioni Calderini del Il Sole 24 ORE Edagricole, Bologna.*
St-Pierre, N. R., Cobanov, B., & Schnitkey, G. (2003). Economic Losses from Heat Stress by US Livestock Industries. *Journal of Dairy Science*, *86*(31), E52–E77. doi:10.3168/jds.S0022-0302(03)74040-5
Stella, A., Nicolazzi, E. L., Van Tassell, C. P., Rothschild, M. F., Colli, L., Rosen, B. D., … Zhang, W. (2018). AdaptMap: Exploring goat diversity and adaptation. *Genetics Selection Evolution*, *50*(1), 1–7. doi:10.1186/s12711-018-0427-5
Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., … Robinson, G. E. (2015). Big Data: Astronomical or Genomical? *PLoS Biology*, *13*.
Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9440–9445. doi:10.1073/pnas.1530509100
Strillacci, M. G., Vevey, M., Blanchet, V., Mantovani, R., Sartori, C., & Bagnato, A. (2020). The genomic variation in the aosta cattle breeds raised in an extensive alpine farming system. *Animals*, *10*(12), 1–18. doi:10.3390/ani10122385
Stronen, A. V., Pertoldi, C., Iacolina, L., Kadarmideen, H. N., & Kristensen, T. N. (2019). Genomic analyses suggest adaptive differentiation of northern European native cattle breeds. *Evolutionary Applications*, *12*(6), 1096–1113. doi:10.1111/eva.12783
Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., Colli, L., Masembe, C., … Joost, S. (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*, *17*(5), 1072–1089. doi:10.1111/1755-0998.12629
Sugden, L. A., Atkinson, E. G., Fischer, A. P., Rong, S., Henn, B. M., & Ramachandran, S. (2018). Localization of adaptive variants in human genomes using averaged one-dependence estimation. *Nat Commun*, *9*, 1–14. doi:https://doi.org/10.1038/s41467-018-03100-7
Szpiech, Z. A., & Hernandez, R. D. (2014). Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827. doi:10.1093/molbev/msu211
Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., & Ajmone-Marsan, P. (2008). Are cattle, sheep, and goats endangered species? *Molecular Ecology*, *17*(1), 275–284. doi:10.1111/j.1365-294X.2007.03475.x
Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics Society of America*.
Talenti, A., Dreger, D. L., Frattini, S., Polli, M., Marelli, S., Harris, A. C., … Crepaldi, P. (2018). Studies of modern Italian dog populations reveal multiple patterns for domestic breed evolution. *Ecology and Evolution*, *8*(5), 2911–2925. doi:10.1002/ece3.3842
Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467–484. doi:10.1038/s41576-019-0127-1
Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E. G., & Risch, N. J. (2007). Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *The American Journal of Human Genetics*, *81*(3), 626–633. doi:10.1086/520769
Tang, H., Coram, M., Wang, P., Zhu, X., & Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. *American Journal of Human Genetics*, *79*(1), 1–12. doi:10.1086/504302
Terasaki Hart, D. E., Bishop, A. P., & Wang, I. J. (2021). Geonomics : Forward-Time , Spatially Explicit , and Arbitrarily Complex Landscape Genomic Simulations. *Mol. Biol. Evol.* doi:10.1093/molbev/msab175
The Bovine genome Sequencing and Analysis Consortium. (2009). The Genome Sequence of Taurine cattle: a window to ruminant biology and evolution. *Science*, *324*, 522–529.
Thépaut, J., Dee, D., Engelen, R., & Pinty, B. (2018). The Copernicus Programme and its Climate Change Service. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1591–1593).
Thom, E. C. (1959). The discomfort index. *Weatherwise*, *12*.
Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., … Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*(6970), 145–148. doi:10.1038/nature02121
Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical Review*, *38*(1), 55. doi:10.2307/210739
Thornton, P. K., van de Steeg, J., Notenbaert, A., & Herrero, M. (2009). The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. *Agricultural Systems*, *101*(3), 113–127. doi:10.1016/j.agsy.2009.05.002
Title, P.O., & Bemmels, B. (2017). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*. doi:https://doi.org/10.1111/ecog.02880
Title, Pascal O., & Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, *41*(2), 291–307. doi:10.1111/ecog.02880
Toghiani, S., Hay, E. H., Roberts, A., & Rekaya, R. (2020). Impact of cold stress on birth and weaning weight in a composite

beef cattle breed. *Livestock Science*, *236*.
Tonteri, A., Vasemagi, A., Lumme, J., & Primmer, C. R. (2010). Beyond MHC: signals of elevated selection pressure on Atlantic salmon (Salmo salar) immune-relevant loci. *Molecular Ecology*, *19*.
Tortereau, F., Weisbecker, J., Marcon, D., Bouvier, F., & François, D. (2020). Genetic parameters for feed efficiency in Romane rams and responses to single-generation selection. *Animal, The International Journal of Animal Biosciences*, *14*(4), 681–687. doi:10.1017/S1751731119002544
Toschi, A. (1953). *Note sui vertebrati dell'Isola di Montecristo. Zoologia Applicata alla Caccia*. Bologna: Tip. Compositori.
Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., … Zhao, S. (2014). Design and characterization of a 52K SNP chip for goats. *PLoS ONE*, *9*(1), 8. doi:10.1371/journal.pone.0086227
Trail, J. C. M., Gregory, K. E., Marples, H. J. S., & Kakonge, J. (1985). Comparison of Bos Taurus-Bos Indicus Breed Crosses with Straightbred Bos Indicus Breeds of Cattle for Maternal and Individual Traits. *Journal of Animal Science*, *60*, 1181–1187.
Tsartsianidou, V., Kapsona, V. V., Sánchez-Molano, E., Basdagianni, Z., Carabaño, M. J., Chatziplis, D., … Banos, G. (2021). Understanding the seasonality of performance resilience to climate volatility in Mediterranean dairy sheep. *Scientific Reports*, *11*(1), 1889. doi:10.1038/s41598-021-81461-8
Tubiello, F. N., Soussana, J. F., & Howden, S. M. (2007). Crop and pasture response to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19686–19690. doi:10.1073/pnas.0701728104
United Nations. (1992). Convention on biological diversity. *Diversity*, 30. doi:10.1146/annurev.ento.48.091801.112645
Upadhyay, M., Kunz, E., Sandoval-Castellanos, E., Hauser, A., Krebs, S., Graf, A., … Medugorac, I. (2021). Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Molecular Ecology*, *30*(24), 6701–6717. doi:10.1111/mec.16184
Utsunomiya, Y. T., Milanesi, M., Barbato, M., Utsunomiya, A. T. H., Sölkner, J., Ajmone-Marsan, P., & Garcia, J. F. (2020). Unsupervised detection of ancestry tracks with the GHap R package. *Methods Ecol Evol*, *2041–210X*. doi:https://doi.org/10.1111/2041-210X.13467
Utsunomiya, Y. T., Milanesi, M., & Fortes, M. R. S. (2019). Genomic clues of the evolutionary history of Bos indicus cattle. *Animal Genetics*, *50*, 557–568. doi:10.1111/age.12836
Vajana, E., Barbato, M., Colli, L., Milanesi, M., Rochat, E., Fabrizi, E., … Ajmone-Marsan, P. (2018). Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Frontiers in Genetics*, *9*, 13. doi:10.3389/fgene.2018.00385
Vallée, A., Daures, J., Arendonk, J. A. M. van, & Bovenhuis, H. (2016). Genome-wide association study for behavior, type traits, and muscular development in Charolais beef cattle. *Journal of Animal Science*, *94*, 2307–2316.
Vanvanhossou, S. F. U., Scheper, C., Dossa, L. H., Yin, T., Brügemann, K., & König, S. (2020). A multi-breed GWAS for morphometric traits in four Beninese indigenous cattle breeds reveals loci associated with conformation, carcass and adaptive traits. *BMC Genomics*, *21*(1), 1–16. doi:10.1186/s12864-020-07170-0
Vega, G. C., Pertierra, L. R., & Olalla-tárraga, M. Á. (2017). Data Descriptor : MERRAclim , a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Scientific Data*, *4*, 1–12.
Verdugo, M. P., Mullin, V. E., Scheu, A., Mattiangeli, V., Daly, K. G., Delser, P. M., … Fulton, D. (2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science*, *176*(July), 173–176.
Vigne, J.-D., Carrère, I., Briois, F., & Guilaine, J. (2011). The Early Process of Mammal Domestication in the Near East. *Current Anthropology*, *52*(S4), S255–S271. doi:10.1086/659306
Vigne, J. (2011). The origins of animal domestication and husbandry : A major change in the history of humanity and the biosphere. *Comptes Rendus - Biologies*, *334*(3), 171–181. doi:10.1016/j.crvi.2010.12.009
Vilella, A. J., Blanco-Garcia, A., Hutter, S., & Rozas, J. (2005). VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics*, *21*(11), 2791–2793. doi:10.1093/bioinformatics/bti403
Villemereuil, P. De, & Gaggiotti, O. E. (2015). A new F ST -based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*, 1248–1258. doi:10.1111/2041-210X.12418
Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, *47*, 97–120. doi:10.1146/annurev-genet-111212-133526
Vjestica, A., Zhang, D., Liu, J., & Oliferenko, S. (2013). Hsp70-Hsp40 Chaperone Complex Functions in Controlling Polarized Growth by Repressing Hsf1-Driven Heat Stress-Associated Transcription. *PLoS Genetics*, *9*(10). doi:10.1371/journal.pgen.1003886
Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*(3), 0446–0458. doi:10.1371/journal.pbio.0040072
VonHoldt, B. M., P., J. P., Earl, D. A., Parker, H. G., Ostrander, E. A., & Wayne, R. K. (2013). Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping, 80–88. doi:10.1007/s00335-012-9432-0
Voosen, P. (2017). How AI detectives are cracking open the black box of deep learning. *Science*. doi:https://doi.org/10.1126/science.aan7059
Waldmann, P. (2018). Approximate Bayesian neural networks in genomic prediction. *Genet. Sel. Evol.*, *50*.
Wang, K., Liu, X., Qi, T., Hui, Y., Yan, H., Qu, L., … Pan, C. (2021). Whole-genome sequencing to identify candidate genes for litter size and to uncover the variant function in goats (Capra hircus). *Genomics*, *113*(1), 142–150. doi:10.1016/j.ygeno.2020.11.024

Wang, L., Cai, B., Zhou, S., Zhu, H., Qu, L., Wang, X., & Al, E. (2017). RNA-seq reveals transcriptome changes in goats following myostatin gene knockout. *PLOS ONE*, *12*.

Wang, X., Ju, Z., Jiang, Q., Zhong, J., Liu, C., Wang, J., … Huang, J. (2021). Introgression, admixture, and selection facilitate genetic adaptation to high-altitude environments in cattle. *Genomics*, *113*(3), 1491–1503. doi:10.1016/j.ygeno.2021.03.023

Wang, Z., Zhang, H., Yang, H., Wang, S., Rong, E., Pei, W., … Wang, N. (2014). Genome-wide association study for wool production traits in a Chinese merino sheep population. *PLoS ONE*, *9*(9), 3–10. doi:10.1371/journal.pone.0107101

Wei, C., Wang, H., Liu, G., Zhao, F., Kijas, J. W., Ma, Y., … Du, L. (2016). Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. *Scientific Reports*, *6*(June 2015), 1–11. doi:10.1038/srep26770

Weinberg, P. J. (2002). Capra cylindricornis. *Mammalian Species*, *695*, 1–9.

Weldenegodguad, M., Pokharel, K., Niiranen, L., Soppela, P., Ammosov, I., Honkatukia, M., … Kantanen, J. (2021). Adipose gene expression profiles reveal insights into the adaptation of northern Eurasian semi-domestic reindeer (Rangifer tarandus). *Communications Biology*, *4*(1). doi:10.1038/s42003-021-02703-z

Wheeler, T., & Reynolds, C. (2013). Predicting the risks from climate change to forage and crop production for animal feed. *Animal Frontiers*, *3*(1), 36–41. doi:10.2527/af.2013-0006

Wiedemar, N., & Drögemüller, C. (2015). A 1.8-kb insertion in the 3′-UTR of RXFP2 is associated with polledness in sheep. *Animal Genetics*, *46*(4), 457–461. doi:10.1111/age.12309

Wilkinson, S., Wiener, P., Archibald, A. L., Law, A., Schnabel, R. D., Mckay, S. D., … Ogden, R. (2011). Evaluation of approaches for identifying population informative markers from high density SNP Chips.

Wilson, A. J., & Mellor, P. S. (2009). Bluetongue in Europe: past, present and future. *Philosophical Transactions of the Royal Society of London. B*, *364*.

Wollenberg Valero, K. C., Pathak, R., Prajapati, I., Bankston, S., Thompson, A., Usher, J., & Isokpehi, R. D. (2014). A candidatemultimodal functional genetic network for thermal adaptation. *PeerJ*, *2014*(1). doi:10.7717/peerj.578

Worku, A., & Tadesse, Y. (2017). A Review on Population Structure, Genetic Diversity Analysis, Genetic Distance between Population and Genetic Singularity in Livestock. *Advances in Life Science and Technology*, *54*(April), 1–6. Retrieved from www.iiste.org

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, (15), 323–354.

Wu, D.-D., Ding, X.-D., Wang, S., Wójcik, J. M., Zhang, Y., Tokarska, M., … Zhang, Y.-P. (2018). Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nature Ecology & Evolution*, *2*(7), 1139–1145. doi:10.1038/s41559-018-0562-y

Wu, J., Liu, Y., & Zhao, Y. (2021). Systematic Review on Local Ancestor Inference From a Mathematical and Algorithmic Perspective. *Frontiers in Genetics*, *12*, 639877. doi:10.3389/fgene.2021.639877

Wu, T., Wang, S., Jin, Q., Lv, X., & Sun, W. (2021). PAPPA2 promote the proliferation of dermal papilla cells in Hu Sheep (Ovis aries) by regulating IGFBP5. *Genes*, *12*(10). doi:10.3390/genes12101490

Wykes, D. L. (2004). Robert Bakewell (1725-1795) of Dishley: Farmer and livestock improver. *Agricultural History Review*, *52*(1), 38–55.

Xu, L., Bickhart, D. M., Cole, J. B., Schroeder, S. G., Song, J., Van Tassell, C. P., … Liu, G. E. (2015). Genomic Signatures Reveal New Evidences for Selection of Important Traits in Domestic Cattle. *Molecular Biology and Evolution*, *32*, 711–725.

Xu, S., Ren, X., Yang, G., Xie, X., Zhao, Y., Zhang, M., … Li, M. (2017). Genome-wide association analysis identifies the genetic basis of fat deposition in the tails of sheep (Ovis aries). *Animal Genetics*, *48*(5), 560–569. doi:10.1111/age.12572

Yang, B., Cui, L., Enciso, M. P., Traspov, A., Crooijmans, R. P. M. A., Zinovieva, N., … Knorr, C. (2017). Genome - wide SNP data unveils the globalization of domesticated pigs. *Genetics Selection Evolution*, *49*, 1–15. doi:10.1186/s12711-017-0345-y

Yang, I., & Kim, S. (2015). Analysis of Whole Transcriptome Sequencing Data: Workflow and Software. *Genomics Inform.*, *13*, 119–125.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*, 76–82. doi:https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, Ji, Li, W. R., Lv, F. H., He, S. G., Tian, S. L., Peng, W. F., … Liu, M. J. (2016). Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Molecular Biology and Evolution*, *33*(10), 2576–2592. doi:10.1093/molbev/msw129

Yuan, Z., Liu, E., Liu, Z., Kijas, J. W., Zhu, C., Hu, S., … Wei, C. (2017). Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Animal Genetics*, *48*(1), 55–66. doi:10.1111/age.12477

Yuan, Zehu, Ge, L., Sun, J., Zhang, W., Wang, S., Cao, X., & Sun, W. (2021). Integrative analysis of Iso-Seq and RNA-seq data reveals transcriptome complexity and differentially expressed transcripts in sheep tail fat. *PeerJ*, *9*, 1–22. doi:10.7717/peerj.12454

Yuan, Zehu, Zhang, J., Li, W., Wang, W., Li, F., & Yue, X. (2019). Association of polymorphisms in candidate genes with the litter size in two sheep breeds. *Animals*, *9*(11). doi:10.3390/ani9110958

Yurchenko, A. A., Daetwyler, H. D., Yudin, N., Schnabel, R. D., Vander Jagt, C. J., Soloshenko, V., … Larkin, D. M. (2018). Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Scientific Reports*, *8*(1), 1–16. doi:10.1038/s41598-018-31304-w

Yurchenko, A. A., Deniskova, T. E., Yudin, N. S., Dotsev, A. V., Khamiruev, T. N., Selionova, M. I., … Larkin, D. M. (2019). High-density genotyping reveals signatures of selection related to acclimation and economically important traits in

15 local sheep breeds from Russia. *BMC Genomics*, *20*(Suppl 3), 1–19. doi:10.1186/s12864-019-5537-0

Zanichelli, F., Giannini, F., De Pietro, F., & Puppo, F. (2014). *I Quaderni del Parco, documenti tecnici volume 2 PROGETTO LIFE+ MONTECRISTO 2010, Eradicazione di componenti florofaunistiche aliene invasive e tutela di specie e habitat nell'Arcipelago Toscano*. Retrieved from https://www.mite.gov.it/sites/default/files/archivio/allegati/life/life_pubblicazione_montecristo2010.pdf

Zappaterra, M., Gioiosa, S., Chillemi, G., Zambonelli, P., & Davoli, R. (2020). Muscle transcriptome analysis identifies genes involved in ciliogenesis and the molecular cascade associated with intramuscular fat content in Large White heavy pigs. *PLOS ONE*, *15*(5), 1–29. doi:10.1371/journal.pone.0233372

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 11597–11604. doi:10.1073/pnas.0801317105

Zeder, M. A. (2015). Core questions in domestication research. *Proceedings of the National Academy of Sciences*, *112*(11), 3191–3198. doi:10.1073/pnas.1501711112

Zeder, M. A., & Hesse, B. (2000). The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science*, *287*(5461), 2254–2257. doi:10.1126/science.287.5461.2254

Zeng, J., De Vlaming, R., Wu, Y., Robinson, M. R., Lloyd-Jones, L. R., Yengo, L., … Yang, J. (2018). Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics*, *50*(5), 746–753. doi:10.1038/s41588-018-0101-4

Zhang, Y., Xue, X., Liu, Y., Abied, A., Ding, Y., Zhao, S., … Zhao, Q. (2021). Genome-wide comparative analyses reveal selection signatures underlying adaptation and production in Tibetan and Poll Dorset sheep. *Scientific Reports*, *11*(1), 1–12. doi:10.1038/s41598-021-81932-y

Zhao, F., Deng, T., Shi, L., Wang, W., Zhang, Q., Du, L., & Wang, L. (2020). Genomic Scan for Selection Signature Reveals Fat Deposition in Chinese Indigenous Sheep with Extreme Tail Types. *Animals*, *10*(5), 773. doi:10.3390/ani10050773

Zhao, H., Zhu, S., Guo, T., Han, M., Chen, B., Qiao, G., … Yang, B. (2021). Whole-genome re-sequencing association study on yearling wool traits in Chinese fine-wool sheep. *Journal of Animal Science*, *99*(9), 1–12. doi:10.1093/jas/skab210

Zhao, W., Ma, Y., Chen, A., Fu, L., & Zhang, Y. (2019). Robust sliding mode control for Markovian jump singular systems with randomly changing structure. *Applied Mathematics and Computation*, *349*, 81.96.

Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., … Xu, Y. (2020). The origin of domestication genes in goats. *Science Advances*, *6*, 1–13.

Zhu, C., Cheng, H., Li, N., Liu, T., & Ma, Y. (2021). Isobaric Tags for Relative and Absolute Quantification-Based Proteomics Reveals Candidate Proteins of Fat Deposition in Chinese Indigenous Sheep With Morphologically Different Tails. *Frontiers in Genetics*, *12*(November), 1–10. doi:10.3389/fgene.2021.710449