UNIVERSITÀ DEGLI STUDI DEL MOLISE

Department of Agricultural, Environmental and Food Sciences

PhD Course in:

AGRICULTURE TECHNOLOGY AND BIOTECHNOLOGY

(CURRICULUM: ANIMAL WELFARE, BIOTECHNOLOGY AND QUALITY OF ANIMAL PRODUCTION)

(CYCLE XXXV)

Related disciplinary scientific section: AGR/17 (Zootecnia Generale e Miglioramento Genetico)

PhD thesis:

Genetic insights for sustainable livestock farming through bioinformatic analysis of SNP data

Coordinator of the PhD Course: Prof. Giuseppe Maiorano
Supervisor: Prof. Fabio Pilla
Co-Supervisor: Dr. Gabriele Senczuk

PhD Student: Christian Persichilli
167266

ACADEMIC YEAR 2021/2022

Index:

1. Introduction
2. What is Sustainability?
    2.1. Sustainability in Agriculture
    2.2. Sustainability in Animal Farming
    2.3. Role of Animal Breeding and Genetics in the Sustainability of Livestock Farming
3. From Genetics to Genomics in Animal Breeding
    3.1. Molecular Markers
        3.1.1. Markers Based on Sequencing
        3.1.2. SNPs
    3.2. Genomics
4. Bioinformatics and its Applications in the Study of Animal Breeding and Genetics
    4.1. The Study of Genetic Diversity
        4.1.1. Indexes of Genetic Diversity
        4.1.2. Structure of Populations
        4.1.3. Events Influencing Genetic Diversity
    4.2. GWAS and Signatures of Selection
        4.2.1. GWAS
        4.2.2. Signatures of Selection
    4.3. Advances in Selective Breeding
5. Aim of the Thesis
6. Articles
    6.1. Articles Included in the Thesis
    6.2. Additional Research
7. Discussion & Conclusion

1. Introduction

Millions of people around the world depend on animal farming for the production of milk, meat, fiber, and working power. Because of the quick rise in the worldwide demand for animal products, the sustainability of livestock farming is daily questioned by public opinion. Because of this, pressure is mounting on the livestock industry to increase productivity while reducing its environmental impact. To achieve this, it is necessary to develop sustainable livestock production techniques that can raise animal productivity while reducing the environmental impact of animal production.

Animal breeding and genetics are essential disciplines for improving the efficiency, sustainability, and welfare of animal production systems. Our ability to comprehend the genetic basis of animal traits and to hasten genetic improvement has been revolutionized by advancements in genomics, such as high-throughput genotyping technologies and bioinformatics techniques. Bioinformatics is the application of computational techniques to analyze biological data, such as DNA sequences, gene expression profiles, phenotypic records, and genomic annotations. In particular, bioinformatics has developed into a powerful tool for analyzing and interpreting genomic data, making it simpler to identify genes linked to important traits and enabling the creation of more effective and sustainable breeding plans. The use of bioinformatic tools for animal breeding and genetics aims to address some of the current challenges and opportunities in livestock genetics research, such as identifying genes and variants associated with economically important traits, estimating genetic parameters and breeding values using genomic information, developing genomic selection strategies for complex traits, and exploring the genetic diversity and structure of livestock populations. The thesis will use various bioinformatics approaches, such as genome-wide association studies (GWAS), signatures of selection, genetic prediction models, and population genomics methods, to analyze data from different livestock species (cattle and sheep) and production systems.

The main goal of this work is to investigate the genetic aspects of livestock production sustainability using bioinformatics tools and methods. The findings of this study will advance knowledge of the genetic basis of significant livestock traits and shed light on how genomics and bioinformatics may be used to improve the sustainability of animal production. The thesis will contribute to the advancement of knowledge and innovation in animal breeding and genetics for sustainable livestock farming through the use of bioinformatics.

2. What is Sustainability?

Throughout history, there have been numerous discussions about the meaning and definitions of the word "sustainability." But in the 20th century, worries about the environment and society led to the development of the contemporary understanding of sustainability. The concept gained widespread attention after the publication of the report "Our Common Future", also known as “The Brundtland Report”, at the United Nations World Commission on Environment and Development in 1987 (Brutland, 1987). In her report, Gro Harlem Brutland, a Norwegian physician and politician, described for the first time the concept of “sustainable development” as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (Brundtland, 1987). The idea of sustainable development is opposed to the policies of economic growth that went on from the end of WWII to that moment (Du Pisani, 2006). The dream of economic growth led to an extractive attitude among the main economic powers towards the primary resources in order to improve the welfare of the present generations. In fact, economic growth theories assume that unlimited growth and consumption are possible. This unlimited growth consequently influences the growth of the needs of the population, and this would make it possible to create, as a consequence, more jobs and welfare. This theory has been proven to be far from the truth and detrimental to both the environment and the economies of the less developed countries (Du Pisani, 2006; Kidd, 1992; Rodrik, 2014).

Since the publication of the Brutland Report, sustainability and sustainable development have become the goal and the guiding principle for numerous individuals, organizations, and public institutions all over the globe (Sneddon et al., 2006). The report's definition of sustainable development and emphasis on striking a balance between economic, social, and environmental factors had an international impact on the creation of many policies and programs. The four pillars of sustainable development, as described in the Brutland Report, are:

1. Environmental sustainability, is intended as the set of actions that lead to the conservation and protection of the natural environment and its resources, including air, water, and land. It involves minimizing pollution, reducing greenhouse gas emissions, conserving natural resources, and protecting ecosystems. Environmental sustainability entails conserving biodiversity, using natural resources wisely, cutting down on waste production, and supporting sustainable development.
2. Economic sustainability, this pillar involves creating economic systems that are resilient, efficient, and equitable. This includes promoting economic growth, creating jobs, and reducing poverty. It also involves reducing economic inequality and

ensuring that economic activities do not harm the environment or compromise social sustainability.

3. Social Sustainability, refers to the well-being of people and communities, including their health, safety, and quality of life. It involves promoting social justice, equality, diversity, and inclusion, while also fostering strong social connections, cultural diversity, and civic engagement. Social sustainability entails building just and equitable societies that provide for everyone's basic requirements, such as access to food, shelter, healthcare, education, and work. Additionally, it entails advocating for human rights, cultural diversity, and societal cohesion.
4. Cultural sustainability means preserving and promoting cultural diversity, heritage, and traditions. It includes protecting the rights of indigenous peoples, promoting cultural exchange and dialogue, and supporting cultural activities that enhance social cohesion and well-being.

Each one of these aspects of sustainability has a key role in the perspective of "sustainable development" (Brundtland, 1987).

The Brundtland Report had a significant influence on political choices made at the international level, including those pertaining to agriculture. One of the most significant impacts of the Brundtland Report was the 1992 United Nations Conference on Environment and Development (UNCED), also known as the "Earth Summit," held in Rio de Janeiro, Brazil. The summit brought together world leaders, environmentalists, and other stakeholders to discuss urgent environmental issues and work toward a more sustainable future. The meeting featured representatives from 172 countries and resulted in the adoption of Agenda 21, a comprehensive action plan to support sustainable development. The Rio Conference resulted in the production of three documents and two legally binding agreements (Quarrie, 1992). The three documents that resulted from the Earth Summit are:

1. Agenda 21, includes specific recommendations related to agriculture, forestry, biodiversity conservation, climate change, and other environmental issues. The agenda included suggestions to reach the objective of sustainable agriculture, including promoting sustainable farming methods, safeguarding soil and water resources, and preserving genetic diversity in crops and livestock (*Agenda 21 | Department of Economic and Social Affairs*, s.d.).
2. The Rio Declaration on Environment and Development is a non-binding declaration that outlines the principles of sustainable development and the rights and responsibilities of nations to protect the environment (UNCED, 1992a).

3. The Forest Principles, which outline international standards for the management, conservation, and sustainable development of forests worldwide. The key principles of this document include recognizing the intrinsic value of forests, promoting sustainable forest management, respecting the rights of indigenous peoples and local communities, promoting international cooperation and coordination, and committing to reducing deforestation and forest degradation. (UNCED, 1992c)

The two legally binding agreements are:

1. The Convention on Biological Diversity (CBD) is an international agreement that seeks to ensure the equitable and fair distribution of the advantages associated with the use of genetic resources as well as to advance the conservation and sustainable use of biodiversity. The CBD recognizes the importance of biological diversity for the well-being of present and future generations and for the functioning of ecosystems. The treaty defines biological diversity 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, between species and of ecosystems." (United Nations, 1992). Additionally, the CBD promotes the responsible use of biological resources, including traditional knowledge, for societal and economic advancement. Developing national strategies and action plans to advance the sustainable use of biological diversity, including through the advancement of biotechnology and the use of genetic resources, is a requirement of the CBD-subscribing entities. With the purpose of promoting food security, eradicating poverty, and fostering sustainable development, the CBD places a strong emphasis on the conservation and ethical use of domestic animal genetic resources through more sustainable practices in breeding plans aimed at preserving. their genetic diversity. The CBD promotes local people's participation in decision-making processes and acknowledges the critical role that indigenous and local communities play in the preservation and sustainable use of domestic animal genetic resources. In order to assist in the preservation and sustainable use of animal genetic resources, the CBD also promotes research and development in the fields of animal breeding and genetics. In fact, the CBD stipulates that the benefits accruing from the use of these resources must be distributed fairly and equally while acknowledging the sovereignty of states over their own genetic resources. CBD has been ratified by 196 countries and the European Union, making it one of the most widely ratified treaties in history (United Nations, 1992).

2. The United Nations Framework Convention on Climate Change (UNFCCC), an international treaty that was signed by 154 countries. The UNFCCC aims to address the problem of global climate change by reducing greenhouse gas emissions and promoting adaptation to the impacts of climate change. (UNCED, 1992b)

2.1. Sustainability in Agriculture

Sustainability in agriculture involves balancing the production of food, fiber, and other agricultural products with the conservation of natural resources, the protection of biodiversity, and the reduction of environmental impacts. Agricultural practices can have a significant negative impact on the environment, including soil erosion, water pollution, and greenhouse gas emissions. Therefore, achieving environmental sustainability in agriculture is essential to ensuring the long-term viability of our food systems and the health of the planet. All four pillars of sustainability, including the environmental, economic, social, and cultural aspects of sustainability, should be taken into consideration in order to have a more sustainable agriculture and achieve a sustainable development of the agricultural sector (Velten et al., 2015).

To reach the objective of a more environmentally sustainable agriculture, many actions are necessary. As an example, to prevent their depletion and long-term harm, natural resources must be used responsibly. To accomplish this, it is necessary to integrate waste reduction, recycling, and resource reuse strategies into the agricultural system, along with the transition to the use of renewable energy sources like solar, wind, and hydroelectric power, in spite of the fossil fuels still used today. The steps necessary to address pollution and waste include lowering greenhouse gas emissions, improving air and water quality, reducing the use of pesticides, and producing less waste, to name just a few (Abraham & Pingali, 2020).

By the way, while working on environmental sustainability, economic sustainability should be kept in mind. Economic sustainability in agriculture is the capacity of agricultural systems to maintain economic viability over the long term while ensuring that social and environmental goals are met. Agriculture can achieve economic sustainability by implementing techniques and plans that increase output and profit while minimizing negative effects on the environment and social justice (Brodt et al., 2011). The diversification of farms, the effective use of resources, the production of value-added goods rather than just primary goods, and the increased access of farmers to markets are all strategies for achieving economic sustainability in agriculture. By diversifying their livestock and crops, farmers can increase their income and support the sustainability of their business while reducing the risks of monoculture and gaining access to a wider range of markets. Meanwhile, minimizing waste and optimizing the use of primary

resources such as water, fertilizers and pesticides can reduce costs and increase the profitability of the farm while also reducing negative environmental impacts. Furthermore, by adding value to their products—for example, by producing their own cheese instead of selling their milk to third parties—farmers can increase their income and help build more resilient local economies. This, combined with promoting the access of the farmers to the markets by forming cooperatives or other types of producer groups, may help them negotiate better prices and increase their bargaining power (Arru et al., 2021; Brodt et al., 2011; Peng et al., 2022; Valliant et al., 2017).

Social sustainability in agriculture refers to the ability of agricultural systems to maintain and enhance social well-being and equity, both within and beyond agricultural communities. This includes ensuring fair treatment for farmers and farm workers, the involvement of local communities in decision-making processes, and that the redistribution of the agricultural system benefits society. Fair labor practices, community involvement, supplying small farmers with equal access to resources, ensuring food security and nutrition, and rural development are some methods for achieving social sustainability in agriculture (Martínez-Castillo, 2016; McGreevy et al., 2022; Nguyen, s.d.; Nicolosi et al., 2021). This demands fair treatment of agricultural workers, safe working conditions, competitive pay, and opportunities for professional advancement. Agricultural planning should respect and take into account local communities' traditional knowledge and cultural practices (Gordillo & Jerónimo, 2013). A priority for agricultural systems should be the production of affordable, wholesome food for all members of society, especially those who are food insecure. A larger rural development strategy should include social well-being and fighting poverty in rural areas as well as the cultural aspect of the agricultural system and the protection of different traditional practices (HLPE, 2020).

Cultural sustainability in agriculture refers to the ability of agricultural systems to maintain and enhance the cultural values, traditions, and knowledge that are associated with farming and agricultural communities. This entails appreciating and respecting the cultural diversity of farming communities, fostering the transmission of traditional knowledge from one generation to the following, and promoting traditional knowledge and practices. Indigenous and conventional farming methods, which have frequently developed over many years and have shown to be highly effective at adapting to local environmental conditions, should be supported by agricultural systems. Local groups should be involved in decision-making procedures pertaining to agricultural development, and agricultural planning should respect and incorporate local groups' cultural norms and practices (Jamal & Dredge, 2014; Neef & Neubert, 2011; Selman, 2004; Thrupp, 1989). Adopting all the strategies for environmental, economic, social, and cultural sustainability, the objectives of sustainable development may be reached, meeting the

needs of the present without compromising the ability of future generations to meet their own needs (Annan-Diab & Molinari, 2017).

2.2. Sustainability in Livestock Farming

Within agriculture, animal agriculture is a significant contributor to environmental impacts. Animal products like meat, milk, and eggs are produced extensively, which may have a negative effect on the environment by causing greenhouse gas emissions, water pollution, and habitat destruction. With emissions estimated at 7.1 gigatonnes CO2-eq annually, or 14.5% of human-induced greenhouse gas emissions, the livestock sector has a sizable impact on climate change. Emissions from the production of beef and milk from cattle make up the majority of the sector's emissions, at 41 and 20 percent, respectively. Pig and poultry meat and eggs account for 9 and 8% of the sector's emissions, respectively. As a result of this production's strong projected growth, higher emission shares and volumes are anticipated over time. The production and processing of feed, which accounts for 39% of sector emissions, and enteric fermentation from ruminants, which accounts for 45% of emissions, are the two main sources of emissions. Manure storage and processing account for 10% of the total (FAO, 2013).

The present level of animal production is the result of the evolution of human diets, which sees consumers consuming more animal products as their income rises. Around 330 million cattle and buffaloes are raised for meat production, along with 65 billion chickens, 1.5 billion pigs, 1 billion goats, and lambs. (FAOSTAT, 2016). The same study notes that there are almost 234 million cows used for milk production and 7.6 billion laying hens used in the egg industry. According to Herrero et al. (2016), the livestock industry serves the livelihoods and food security of nearly 1.3 billion people in developing nations, accounting for up to 50% of the global agricultural GSP. Along with the increase in human population, livestock farming has become significantly more intensive due to the use of concentrated feed, pharmaceuticals, and vaccinations, as well as improvements in infrastructure and feed efficiency (Ramankutty et al., 2018). Industrial-scale farming, which takes place in facilities known as concentrated animal feeding operations (CAFOs), is the most well-known and frequently used method of raising animals. But according to Tullo et al. (2019), this system may cause the contamination of water and soil with hazardous substances found in the manure, such as excretory nitrogen, phosphorus, organic matter, and fecal microbes. On the other hand, despite the fact that the term is being used to describe many different aspects of system change, intensification (production per unit area) is frequently linked to increased farm yields (Dumont et al., 2018). Along with its environmental costs, intensive livestock farming's high efficiency helps reduce hunger and malnutrition for

billions of people (Dumont et al., 2018; Foley et al., 2005; Godfray et al., 2010). In a de-intensified farming system, more land needs to be farmed in order to maintain a certain level of production because a drop in production "intensity" is likely to result in a drop in yield (Borlaug, 2007). It becomes necessary in this context to renew intensification in a more sustainable manner. Ecological intensification and sustainable intensification both play a growing role in the management and development of agriculture and forests in this environment by reducing the negative effects of agricultural intensification, increasing livestock and crop diversification, and intensifying ecosystem services (Campbell et al., 2014; Raj et al., 2021). All of this promotes ecological sustainability by increasing soil fertility, agricultural yield, and food security. To make it possible, it is necessary to encourage the use of improved ecological intensification methodologies, such as intensifying the agroecosystem and using different farming systems (e.g., agriculture, agroforestry, and other farming systems) (Raj et al., 2021).

In summary, a variety of steps can be taken to achieve a sustainable intensification of the livestock industry. A more sustainable form of animal farming passes by an overall sustainable management of the farm and the whole sector. A higher efficiency in feed production, the promotion of animal welfare, with the consequent reduction in the use of antibiotics and anthelmintic drugs, proper management of animal waste, and breeding programs deed to preserve genetic diversity while enhancing the production efficiency are all actions that can together lead to the sustainable development of the livestock farming sector (Broom, 2018; Hocquette et al., 2014; Lange, 2022; Stern et al., 2005; Tälle et al., 2019).

2.3. Role of Animal Breeding and Genetics in the Sustainability of Livestock Farming

In this context, animal breeding and genetics play a crucial role in sustainable development. In fact, through proper breeding plans, it is possible to improve animal productivity and efficiency, obtaining, as an example, higher growth rates, better feed conversion, and improved disease resistance (Martyniuk, 2021). More resource-efficient animals result in a reduction in the use of resources such as land and feed and lower the environmental impact per unit of animal product and, as a consequence, of animal agriculture in general. Moreover, an increase in animal efficiency for production has an effect on the reduction of greenhouse emissions in the livestock sector. The reduction in the amount of feed needed to produce a given amount of animal protein translates into lower environmental costs in terms of greenhouse gas emissions for forage cultivation and also lower emissions due to enteric fermentations per unit of final product (MacLeod et al., 2019). Furthermore, breeding plans are becoming more sensitive to animal welfare. By selecting animals that are healthier, more robust, and more resilient to disease,

animal breeders can improve animal welfare and reduce the need for antibiotics and other medications in animal production systems. This aspect is fundamental to reaching the goal of sustainability in animal production. In fact, the use of drugs and medication in farmed animals results in pollution of water and soil through their dejections or adjunctive costs for their treatment, which costs would inevitably fall on farmers and society, resulting in both cases in a lack of sustainability both from an environmental and economic point of view. In addition, animal genetics is useful to protect genetic diversity (Colditz et al., 2016; Farstad, 2018; Flint & Woolliams, 2007; Hovi et al., 2003). The study of population genetics, helps to preserve biological diversity at different levels by identifying local populations, identifying the molecular differences that make each population different from others and more adapted to certain environments and conditions, and planning breeding schemes aimed at preserving these differences (Biscarini et al., 2015; Hoffmann, 2010). This information may give indications to breeders to preserve local populations, integrating more classical field approaches like the collection of pedigrees, the phenotyping of the animals in the breeds, and organizing the breeds' farmers for a proper and standardized data collection (Windig & Engelsma, 2010; Herrero-Medrano et al., 2015; van Marle-Köster et al., 2021).

3. From Genetics to Genomics in Animal Breeding

Animal breeding started during the Neolithic era, when humans caught the first animals and started to control their reproduction. At the beginning, the process was totally based on phenotypes: the animals that fit the needs of the farmers were chosen as reproductors. Over time, methods built on the scientific method have been used in conjunction with the empirical approach to enhance it. In recent years, we have transitioned from so-called mass and morphological selection, which was achieved by selecting the animals with the best appearances as reproducers, to a quicker genetic selection, which was accomplished by estimating the animals' reproductive value (also known as Estimated Breeding Values) for certain zootechnical phenotypes of interest (Gianola & Rosa, 2015). The study of genetics, and consequently genetics applied to animal science, has undergone a revolution as a result of the discovery of the DNA molecule in 1953 (Watson & Crick, 1953), and the subsequent developments in molecular biology have led from genetics to the era of genomics through important steps as outlined below.

3.1. Molecular Markers

A molecular marker can be defined as a genomic locus detectable with probes or primers that, by virtue of this genomic locus, distinguishes in a characteristic and unequivocal way the stretch of

the genome and the regions closely associated with the 5' and 3' ends (Mondini & Pagnotta, 2015). The use of molecular markers offers several advantages over other methods of analyzing genetic variation:

- Molecular markers are not influenced by environmental factors, meaning that they can be used to identify genetic factors that contribute to phenotypic differences without the need to control for environmental effects. However, some markers can detect variations due to environmental factors such as methylation.
- Molecular markers can cover any part of the genome, including transcribed and non-transcribed regions, introns, and regulatory regions. They can be also used to analyze extranuclear DNA such as mitochondrial DNA (mtDNA), which is inherited maternally and is useful to reconstruct an animal's phylogeny.
- Molecular markers can be used to identify genetic differences between individuals that are genetically similar but phenotypically indistinguishable
- Analyzing molecular markers requires small amounts of tissue and can be performed very early (prenatal analysis).
- Molecular markers are traceable in any tissue/organ (with an exception for methylation markers, because methylation is different in the various tissues).

The characteristics desired in a molecular marker may also vary based on the objectives of the experiment. For instance, molecular markers are advantageous for characterizing a genomic locus because markers are inherited in a Mendelian fashion (Marwal & Gaur, 2020). Controlling the inheritance of a marker means controlling at the same time the inheritance of both the regions that contain it and those closely associated with it (genes of interest because they are responsible for diseases, production-related traits, etc.). For this use, the molecular marker must be as close as possible to the gene or to the studied genomic region because the closer it is, the more chances it has of being inherited with it thanks to the linkage phenomenon. The farther it is from the gene, the higher the chance of crossing over. If we use molecular markers for other purposes, other characteristics may be interesting (Benavides & Guénet, 2012; Lowe & Bruce, 2019; Salgotra & Stewart. 2020). An example may be the bacterial 16S ribosomal RNA (rRNA) gene, a popular molecular marker for identifying bacterial species. It has developed into a helpful tool for bacterial identification due to the growing sequence information databases. Between the 16S and 23S rRNA genes, the internal transcribed spacer (ITS) region offers more clarity for species-level identification. The 16S rRNA gene, however, has drawbacks, including the fact that it exists in numerous copies within a bacterial genome and does not offer enough resolution for species-level identification (Hassler et al., 2022).

The way through molecular markers has been opened by allozymes. The use of allozymes as genetic markers applied to population genetics was first described in two papers published in the late 1960s (Harris, 1966; Lewontin & Hubby, 1966). Electrophoresis was used to separate proteins extracted from different individuals of the same species, finding that some of the proteins had different forms (or "alleles") that were encoded by different versions of the same gene. These protein variants, or allozymes, have since been used as genetic markers to study patterns of genetic variation within and between populations. One of the advantages of allozyme analysis is that it requires no a priori information about a species. Since all such tissues include enzymes, the same procedures can be applied without modification to any fresh tissue sample from an animal, plant, or microbial culture. Allozyme analysis allowed researchers to estimate heterozygosity at multiple loci, which provided a more comprehensive picture of genetic diversity than single-locus measures. In addition to heterozygosity, other genetic diversity indices such as allelic richness, F_{ST}, and gene diversity were also estimated using allozyme data. These indices provided valuable information about the distribution and abundance of genetic variation within and between populations. Until that moment, these measures required a pedigree for calculation; otherwise, they were estimated based on phenotypic variation, which fails to capture genetic differences among phenotypically similar individuals (Allendorf, 2017). This biochemical marker gained significance as it enabled the measurement of genetic diversity in populations lacking pedigrees, such as local livestock breeds and wild populations.

In 1977, Fredrick Sanger introduced the first DNA sequencing technique (Sanger et al., 1977). The Sanger method achieved immense success due to its simplicity, accuracy, and widespread accessibility. Consequently, it served as the primary DNA sequencing technique for decades, providing the initial opportunity to amass a substantial volume of genetic data. . In less than thirty years, this technology gave the ability to reach many goals in molecular sciences, making a strong contribution to the complete sequencing of the human genome by the Human Genome Project in 2003 (Collins et al., 2003a). Its development and implementation with the Polymerase Chain reaction (PCR) made it possible to discover most of the molecular markers employed since then. The creation of PCR marked a development in the creation of molecular markers. Primers are brief synthetic oligonucleotides that are specifically created to anneal to unique regions of the DNA template and enable its amplification by PCR. PCR-based techniques take advantage of this ability. One benefit of the PCR-based markers is that, unlike Southern blotting, they can be quickly and inexpensively visualized using agarose or acrylamide gel electrophoresis or by capillary sequencer matrices.

Among the first molecular markers used to study genetic diversity and population structure were mitochondrial DNA (mtDNA), which consists of genetic material found in the mitochondria of cells and inherited maternally (M. M. K. Nass & Nass, 1963; S. Nass & Nass, 1963). The first use of mtDNA as a genetic marker in population genetics was reported in 1979 (Avise, Giblin-Davidson, et al., 1979; Avise, Lansman, et al., 1979; Brown & Wright, 1979). Its analysis involves sequencing a portion of the mtDNA genome, which can be used to identify differences in the DNA sequence among individuals or populations. These differences, or mutations, can be used to reconstruct the evolutionary history of populations and species and to infer patterns of migration and gene flow over time. One advantage of using mtDNA as a genetic marker is that it provides a direct record of maternal lineages over time, which can be useful for studying the evolutionary history of populations and species. Furthermore, it is haploid and is not subject to recombination like nuclear DNA. For these reasons, it has the ability to provide a more straightforward picture of population structure and gene flow than other markers. mtDNA has been used for various purposes in conservation genetics, including identifying evolutionarily significant units (ESUs) and management units (MUs), reconstructing phylogenies and phylogeographic patterns, estimating population sizes and demographic histories, and detecting hybridization events between species. However, it is important to note that mtDNA represents only a small portion of an organism's genome, and for this reason, it is poorly useful to study production traits of livestock species.

During the past decades, a vast number of different methodologies relying on PCR-based markers have been developed (RAPD, AFLP, et cetera), and their use in animal breeding has been largely employed (Camargo, 2018). Primarily, the genetic markers based on PCR were based on arbitrary primers, which were designed to bind to arbitrary regions of the template DNA and lead to the amplification of various DNA regions in various samples. An example of molecular markers based on PCR are the Random Amplified Polymorphic DNA (RAPD) markers. These kinds of markers gave for the first time the opportunity to perform fingerprinting analyses, allowing to assess the similarities between and within animals of the same population and to perform the traceability of animal products. This technique had the advantage of inspecting multiple loci at a glance, but the experiments were scarcely reproducible. On the other hand, markers based on semi-arbitrary primers (primers are designed to bind to a particular region of the template DNA), allow for more specificity and reproducibility in the amplification of target regions. The first example of a semi-arbitrary primer was the Amplified Fragment Length Polymorphism (AFLP) (Vos et al., 1995). These markers have the advantage of being able to detect co-dominant polymorphisms, combining the benefits of both RFLPs and RAPD.

Furthermore, they were highly polymorphic and reproducible, but since the results were highly influenced by the operator, they were scarcely replicable, putting a limit on their use in population genetics because of the scarce shareability of the generated data.

Nowadays, among the most notable and widely used markers are SSRs and SNPs. Simple Sequence Repeats (SSRs) or microsatellites are short tandem repeats (STRs) of 2–6 nucleotides that vary in number among individuals. SSRs are highly polymorphic, reproducible, and informative, but they require prior knowledge of the sequence and location of the repeat motifs. SSR markers belong to a class of markers that identify polymorphisms at the level of repeated DNA sequences. SSR polymorphism is of the codominant single-locus type and is generated by variation in the number of repeats in the microsatellite region. On the other hand, the flanking regions that are commonly conserved are used to design oligos for amplifying only that locus, resulting in a specific PCR product. One advantage of using microsatellites as genetic markers is that they are highly variable and can provide a high level of resolution for studying population structure and gene flow. Additionally, because microsatellites are found throughout the genome, they can provide a more comprehensive picture of genetic variation than other types of markers. SSRs have many applications, including genetic mapping, parentage analysis, and population genetics. However, it is important to note that microsatellite analysis can be time-consuming and expensive compared to other types of genetic analysis. Additionally, because microsatellites are highly variable, they can be prone to genotyping errors or mutations that can complicate data interpretation. For this reason, in order to use microsatellites for population genetics, it is necessary to include in the study as many loci as possible (Feldman et al., 1997; Goldstein & Pollock, 1997). This obstacle was circumvented by the development and application of Multiplex PCR to SSR amplification. This technique made it possible to analyze multiple loci together, making possible the use of SSR data as multilocus markers (Hayden et al., 2008). Thanks to their extremely high levels of variability, SSRs are processed quickly and have the capacity to isolate a large number of loci, which makes them an effective marker for identifying variations between populations that are closely related (Bruzzone et al., 2003; Iamartino et al., 2005).

3.1.1. Markers Based on Sequencing

Next-generation sequencing (NGS) markers have revolutionized the field of animal breeding and genetics, providing powerful tools for studying genetic variation, identifying polymorphisms, and conducting population-level analyses (Danvey et al., 2011; Kulski, 2016).

NGS markers are a class of molecular markers based on sequencing. Their bases are founded on an implementation of the Sanger method called "Shotgun sequencing", which consists of the random fragmentation of the DNA into smaller segments that are individually sequenced and merged together into reads. With the aid of software, the reads are then put together in order to reconstruct the whole genome of the studied organism (Danvey et al., 2011; Kulski, 2016). In fact, the Human Genome Project sequenced the first fully human genome in 2003 using this technique (Barranco, 2021). Next-generation sequencing (NGS) markers have transformed the field of animal breeding and genetics, providing powerful tools for studying genetic variation because of their ability to identify multiple polymorphisms of many types in one analysis. High-throughput DNA sequencing is made possible by NGS, creating vast volumes of sequence data that may be utilized to locate and describe SNPs. The SNP markers found through NGS data processing are then used to design SNP beadchips, which are subsequently manufactured and placed on the surface of microarray chips (Goodwin et al., 2016; Kulski, 2016).

Furthermore, with the use of NGS, it is now possible to sequence whole genomes at a fraction of the cost and time required by earlier sequencing techniques. Improving the possibility of sequencing the genomes of a wide range of animal species, revealing important details about genetic diversity and evolution (Goodwin et al., 2016; Goldfeder et al., 2017).

3.1.2. SNPs

SNP (Single Nucleotide Polymorphism) was the molecular marker employed in the studies reported in this thesis. Currently, SNP (Single Nucleotide Polymorphism) markers are the most commonly used molecular markers in animal breeding and genetics. These markers are widely distributed throughout the genome; for instance, the SNP density in humans is 1 every 100–300 base pairs. Markers that are so close to one another have high probabilities to be in linkage among them and to nearby genes and other genomic regions (Ke et al., 2008). For this reason, SNPs represent ideal markers for population genetics analyses, phylogenies, signatures of selection, and genome-wide association studies (GWAS), allowing the identification of SNPs associated with specific traits or diseases (Nicolazzi et al., 2015).

SNPs are typically biallelic, meaning that there are usually only two possible alleles at each locus (for example, A/G or G/T) (Brookes, 1999). This makes genotyping relatively easy and cost-effective, as only two probes are needed to detect each SNP. SNPs are also typically codominant, meaning that both alleles are expressed equally in heterozygotes. This makes it easy to distinguish between homozygous and heterozygous genotypes (Vignal et al., 2002).

Based on their location in the genome, SNPs can be classified as regulatory SNPs and coding SNPs, but SNPs can also be found in non-coding regions like introns and intergenic regions. Regulatory SNPs (rSNPs) are located in regulatory regions of the genome and can affect gene expression. Coding SNPs (cSNPs) are instead located in exons, the coding regions of genes, and can affect the amino acid sequence of the encoded protein (Khlestkina & Salina, 2006). cSNPs can be further subdivided into synonymous SNPs, which do not change the amino acid sequence, and non-synonymous SNPs, which change the amino acid sequence, eventually causing functional consequences, such as creating a premature stop codon or altering the structure of the protein; an example could be the case of the SNPs found in casein genes in cattle (Meier et al., 2019). The majority of SNPs in the human genome are found in non-coding areas such as introns or intergenic regions (Sud et al., 2017). These SNPs can change epigenetic modifications or the structure of the chromatin to affect the expression of target genes. As a result, non-coding SNPs have a role in cell migration, invasion, apoptosis, and proliferation.

SNP markers can be identified using a variety of methods, including de novo sequencing of the genome or transcriptome or using existing sequence data from databases. Once identified, SNPs must be validated to ensure that they are true polymorphisms and not sequencing errors. The validation of SNPs is usually performed by confronting genotypes revealed by different methods or by sequencing the region containing the SNP with a reliable technique, such as the aforementioned Sanger method (Van Tassell et al., 2008). Validated SNPs can then be genotyped in populations of interest using a variety of methods, including PCR-based assays and microarray chips.

Microarray genotyping is the most commonly used technology for SNP genotyping. Its success is due to the ability to genotype an ever-increasing number of samples with an ever-increasing number of markers at an ever-lower price. A microarray for SNP genotyping works by using the principle of hybridization between complementary DNA strands.

A microarray chip for SNP genotyping typically consists of a glass slide or a silicon wafer with thousands or millions of microscopic wells or beads, each of which contains many copies of an oligonucleotide probe that is complementary to a particular SNP allele. The SNP-specific probes are synthesized on the surface of the chip using a process called photolithography.

The DNA samples to be genotyped are first tested through spectrophotometry or fluorometry to assess their quality and quantity, and then, in order to generate multiple copies of the target DNA region containing the SNP of interest, they are amplified with a PCR-free method to avoid (G+C)-bias in the sequence (Kozarewa et al., 2009), and then fluorescently labeled using dyes that are specific to each allele. The labeled samples are then hybridized to the microarray chip,

which allows the SNP-specific probes to bind to their complementary SNP alleles. The chip is then washed to remove any unbound DNA, and the fluorescence intensities of the bound, labeled DNA are measured using a scanner.

The genotypes of the samples are then determined based on the relative fluorescence intensities of the labeled alleles. Typically, each SNP allele is associated with a different dye, which allows the relative intensities of the two dyes to be used to infer the genotype of the sample at that SNP locus. The raw intensity data is then processed using software that uses statistical algorithms to determine the genotypes of each sample at each SNP locus.

Overall, the process of SNP genotyping using microarray chips is a highly sensitive, accurate, and high-throughput method for genotyping large numbers of SNPs in many samples simultaneously. It has revolutionized the field of genetics and enabled researchers to study the genetic basis of complex traits and diseases in a way that was previously not possible.

SNP markers have many applications in animal breeding and genetics. They can be used to identify genes and genetic variants associated with specific traits, such as milk production or disease resistance. They can also be used to estimate genetic parameters, such as heritability and genetic correlations, which are important for breeding value estimation and selection. In addition, SNP markers can be used for parentage testing and pedigree reconstruction, which are important for maintaining genetic diversity and preventing inbreeding.

3.2. Genomics

Genes are the hereditary units that contain the instructions for constructing proteins, which control the behavior of cells and the physiology of an organism. The study of genes, genetic variation, and heredity in living beings is the subject of the branch of biology known as genetics (Noble & Noble, 2023). Genomics, instead, examines an organism's entire genome as well as how its genes interact with one another and with the environment in which it lives. The genome is an organism's complete set of genetic information (Youssef et al., 2019). This information in animals is stored in the DNA molecule. It provides all of the information the organism requires to function, including all of its genes and non-coding regions (Youssef et al., 2019). Genomics entails the examination of this DNA to find genes, genetic variants, and other elements that influence an organism's traits. Genomics is a relatively new scientific discipline that has revolutionized the fields of animal breeding and genetics. With the help of genomics, we are now able to pinpoint particular genes and genetic polymorphisms that are linked to desirable traits, including production traits, health traits, and the ability to adapt to one or many environments (Collins et al., 2003b; Medema et al., 2021).

The introduction of high-throughput sequencing technologies, such as SNP array technology, made the application of genomics in animal breeding and genetics simpler, giving the opportunity to obtain a vast amount of genomic data at an affordable cost (Singh & Singh, 2015; Zhao et al., 2019; Johnsson, 2023). This data may later be used for many applications, including the association of certain markers with Quantitative Trait Loci (QTL) or genes involved in the resistance to pathogens or other environmental conditions (Poczai et al., 2013; Hasan et al., 2021). Another application of genomics to quantitative genetics is to use its potential to estimate breeding values and have more effective breeding plans (Singh et al., 2018; Johnsson, 2023). Definitely, the comparison of a large amount of genomic data between different breeds may aid in the genetic characterization and identification of local breeds and help plan their conservation. All of these activities are fundamental for the sustainable development of animal agriculture; breeding more efficient, healthy, and well-adapted animals to the farming environments helps to reduce emissions and the price per unit of the final product while responding to the increasing demand for animal products (Rothschild & Plastow, 2014). Furthermore, the conservation of local populations gives the opportunity to preserve rare genetic variants that can help to deal with the ever-changing climate and other environmental conditions. In addition, the conservation of this genetic variability may help us keep breeding for selective purposes in the future (Segelbacher et al., 2022). Genomics, however, faces challenges in its application: Genes alone do not provide a complete picture. Phenotypes, the observable traits and characteristics of an organism, are crucial for a comprehensive understanding (Goddard & Hayes, 2009). Local breeds often struggle to establish effective breeding programs due to a lack of data infrastructure. Without organized data collection and coordination among farmers, the applications of genomics remain insufficient for the genetic management of livestock populations (Ngene et al., 2017). Anyway, when other kinds of data have not been collected, the genomic analysis of local populations gives the opportunity to have indications about the status of the genetic diversity in a population and to program future breeding strategies (Eusebi et al., 2022; Tolone et al., 2022). With these assumptions, it is easy to assert that genomics can help address food insecurity in the developing world by identifying signatures of selection associated with adaptations to environmental variables and reducing susceptibility to major disease agents (Biscarini et al., 2015). This can lead to more efficient livestock production, better and more authentic food products, and increased animal health and welfare (Tait-Burkard et al., 2018).

4. Bioinformatics and its Applications in the Study of Animal Breeding and Genetics

Bioinformatics is the application of computer technology to the management of biological data. Decoding the complete DNA and RNA sequences of numerous organisms was the initial focus of bioinformatics. Large databases of biological information from microorganisms, plants, animals, and humans, including nucleotide and amino acid sequences, protein domains, and protein structures, are currently created, maintained, and managed by bioinformatics. Researchers from all over the world have access to this enormous amount of data for their work and can submit new, revised data every day. Currently, the field of bioinformatics involves the management, analysis, and interpretation of numerous types of data, such as nucleotide and amino acid sequences, protein domains, and protein structures (Singh et al., 2018).

The study of bioinformatics entails the analysis and interpretation of biological data using computational tools and techniques. In the field of livestock genetics, bioinformatics is used to analyze and decipher genomic data in the form of genotypes or whole genome sequences (Bayat, 2002).

Among the main applications of bioinformatics to genomic data are genome annotation and comparative genomics studies. Furthermore, bioinformatics has turned into a crucial tool for genomic selection and breeding (Bayat, 2002; Chen & Coppola, 2018).

Bioinformatics can play a significant role in animal breeding and genetics, particularly in the analysis of genomic data. With advances in genomics technologies, it is now possible to generate vast amounts of genetic data from animals, including information on their DNA sequence, gene expression, and genetic variation (McGuire et al., 2020; Chakraborty et al., 2022). Furthermore, bioinformatic tools and methods can be used to analyze and interpret these data, providing valuable insights into the genetic basis of traits of interest and informing breeding decisions. The results of these studies contribute to enhancing the sustainability of the livestock farming sector (Gutiérrez-Reinoso et al., 2023; Zhao et al., 2023). Through its applications in animal breeding, bioinformatics has a role in enhancing the efficiency of production, improving animal welfare, reducing the use of antibiotics and anthelmintic drugs, and reducing the impact of production in terms of pollutants, greenhouse gas emissions, and the use of primary resources (Singh et al., 2018). The applications of bioinformatics in the study of genetic diversity contribute to the sustainability of the animal farming sector by helping to design breeding strategies helpful to the conservation of local breeds that may carry rare genetic variants involved in the adaptation to certain environmental variables and detect them through the scan for signatures of selection (Biscarini et al., 2015; Osei-Amponsah et a., 2019; Ben-Jemaa et al., 2021; Demir et al., 2022). Finally, through Genome Wide Association Studies, bioinformatics is involved in the

sustainability of the livestock sector, helping to associate specific genomic regions with traits of interest like health and welfare-related traits and production traits (Brandenburg et al., 2022.; John et al., 2022: Shao et al., 2022).

4.1. The Study of Genetic Diversity

Thanks to bioinformatics, the genetic diversity of livestock populations can be analyzed both within and between them using bioinformatics tools. Breeders can reduce the risk of inbreeding and preserve genetic diversity by having a better understanding of genetic diversity, increasing the long-term sustainability of animal production. Since the 1960s, the race for genetic improvement and the industrialization of animal production have led to the loss of genetic diversity, both in terms of the number of breeds and in terms of genetic variability within breeds. As a consequence, 20% of livestock in local populations are at risk of extinction all over the world. In order to create effective management and conservation strategies for farm animal genetic resources, genomic tools have been used to evaluate genetic diversity, population structure, and relationships among breeds (Rischkowsky & Pilling, 2007). Studying and understanding the residual genetic diversity of local populations is fundamental to planning breeding strategies designed for the conservation and management of animal genetic resources.

In the area of genetics and its applications to animal breeding, we can distinguish between two schools. The first, which we can refer to as the "molecular tradition," dates back to Mendel and aims to identify and characterize from a biochemical perspective those elements that make up the genome in the hope of one day manipulating it for our advantage. The second, which can be credited to Galton and which we can refer to as statistical tradition, looks at how the genomic information manifests itself in quantitative traits through the correlations between relatives with the aim of causing a genetic-economic change in productive traits. These two visions of animal breeding are not kept separate, and they may blend or divide based on the many goals and achievements (Blasco & Toro, 2014).

4.1.1 Indexes of Genetic Diversity

To measure out genetic diversity within a population, we can use bioinformatic tools to calculate indices of genetic diversity. In this section, I will describe the most commonly used of these measures.

- Effective population size (Ne) is the size of a hypothetical population that experiences the same level of random genetic drift as the actual population and whose inbreeding rate is the same as that of the actual population but without migration or selection. Ne serves as

a gauge for a population's genetic diversity, adaptability, and general health. By measuring the degree of linkage disequilibrium or computing different statistics like the genetic variance, the inbreeding coefficient, or the magnitude of drift, one can estimate the variability of the population's genes and determine the effective population size using genetic markers. Among the most commonly used bioinformatic tools used to calculate Ne is SNeP (Scenario Ne Estimation Program), a coalescent-based software for estimating effective population size (Ne) from genetic data. In order to calculate the effective population size at various points in the past, the SNeP simulates the coalescent process, which traces the sampled individuals' ancestry backward in time. Several parameters, including the mutation rate, population size, migration rates, and demographic scenarios, are input into the software by the user to model the coalescent process. Another popular software for this purpose is GoNe. GONe is a user-friendly Windows program for calculating effective size (Ne) in populations with overlapping generations. To take age structure in populations into account, it modifies the temporal method according to Jorde-Ryman. For this method to work, estimates of the birth and survival rates for each age group as well as measurements of the allele frequencies across two or more successive cohorts are needed. Ne is estimated at every locus and across all loci for each interval between succeeding cohorts. Additionally, three different genetic drift estimators produce Ne estimates (Fs, Fc, and Fk).

- Heterozygosity is an indicator of genetic diversity within a population. It shows the percentage of individuals in a population that have different alleles of a particular gene. Expected heterozygosity (He) and observed heterozygosity (Ho) are the two main measures of heterozygosity. Based on the genotyping data, the following formulas can be used to estimate He and Ho:
    - He = [n/(n-1)] [1 - Σ(pi^2)] where n is the number of individuals genotyped and pi is the frequency of each allele.
    - Ho = number of heterozygous individuals / total number of individuals genotyped.
- The index that measures the similarity of one or multiple alleles across the genome is called homozygosity. Homozygosity can be described as the opposite of heterozygosity. If the two alleles at a certain locus are identical, they are said to be homozygous. The proportion of homozygous individuals at a given locus or across the genome can be calculated using the following formula:

Homozygosity = Number of homozygous individuals / Total number of individuals genotyped.

If the homozygous genotype is denoted by AA, the heterozygous genotype is Aa, and the homozygous recessive genotype is aa, then the proportion of homozygotes for a given locus can be estimated from the frequency of each genotype using the following formula:

H = [nAA + 0.25*nAa - 0.5*n(aa)] / n

where n is the total number of individuals genotyped, and n(AA), n(Aa), and n(aa) are the counts of each genotype.

- The inbreeding coefficient calculates the likelihood that two alleles at a specific locus in an individual are identical by descent, meaning that they were passed down from a common ancestor. The coefficient of inbreeding can be calculated using different metrics, such as Fis, Fit, F_{ST}, and Froh.
    - Fis: The inbreeding coefficient within a subpopulation, calculated as (He-Ho)/He. Fis ranges from -1 to 1, with positive values indicating excess homozygosity and negative values indicating excess heterozygosity.
    - Fit: The inbreeding coefficient within an individual, calculated as 1- (Ho/He). Fit ranges from 0 to 1, with higher values indicating higher levels of inbreeding.
    - Froh: The inbreeding coefficient based on runs of homozygosity (ROH) in the genome of an individual, calculated as the proportion of the genome covered by ROH. Runs of homozygosity (ROH) are a measure of the length and frequency of stretches of homozygous DNA in an individual or population, which can provide insights into recent inbreeding events. Froh ranges from 0 to 1, with higher values indicating higher levels of inbreeding.
    - F_{ST}: is the inbreeding coefficient among subpopulations, or fixation index. It is calculated as (Ht-Hs)/Ht, where Ht is the total heterozygosity and Hs is the average heterozygosity within subpopulations. Fst ranges from 0 to 1, with higher values indicating more genetic differentiation among subpopulations. This measure is useful to detect Signatures of Selection and is the base for Genome Wide Association Studies.

4.1.2. Structure of Populations

Inferring population structure in livestock species allows us to appreciate the diversity within and between the studied populations, inspect the distribution of genetic variation within and between populations, and provide insights into the history, demography, and adaptation of livestock

species. Thanks to bioinformatic tools, we are able to use SNP data to assess the genetic diversity, relatedness, and differentiation between populations. For this purpose, various statistical methods are available to analyze the genotype data and infer the structure of livestock populations.

- *Methods based on multivariate analysis:*

  One of the most commonly used statistical techniques used to infer the structure of livestock populations is Principal Component Analysis (PCA). This multivariate approach allows one to visualize the overall variance in decreasing dimensions in a multi-dimensional space. PCA can show the main genetic variation sources as well as the level of population differentiation or admixture. Unfortunately, PCA results can be subject to bias and are easily manipulated (Elhaik, 2022). An established bias to which this technique is subject is due to non-random missing data caused by the variability in the quality of the samples (Yi & Latch, 2022).

  A method similar to PCA is Multidimensional Scaling (MDS). This method differs from PCA because it can incorporate genetic distance matrices and produce more accurate representations of the relationships between populations. MDS can also reveal patterns of isolation-by-distance and gene flow between populations.

- *Distance based methods:*

  One relevant approach used to reconstruct the genetic relationship among individuals is the use of distance-based methods. Distance matrices can then be used to build phylogenetic trees or networks using different algorithms, such as neighbor-joining and neighbor-net. These methods can be used with a variety of datasets because, being distance-based, they do not rely on particular models of evolution.

  The distance matrices can be calculated with different methods based on the desired event we want to observe. The two most commonly used methods to calculate a distance matrix are the Reinholds' method, which only takes into account the genetic drift that separates the pairs of populations from their common ancestor, and the Nei's method, which adds to the formula the effect of mutations.

  One of the most popular methods based on distance matrices is Neighbor-joining, a bottom-up algorithm that iteratively joins the closest pairs of populations into a tree, starting with a distance matrix between all pairs of individuals or populations. The process is repeated until all pairs have been merged into a tree, with each step updating the distance between the merged pair to reflect the average distance to the other pairs.

The resulting tree can be shown as a dendrogram, with the labels representing the individuals or populations and the branches lengths depicting the genetic distances.

Neighbor-net, a more adaptable technique, builds networks rather than trees and can represent reticulation or hybridization events that a tree does not adequately catch. Neighbor-net begins with a distance matrix as well and incrementally adds splits or edges to reduce any conflicts or inconsistencies in the distances. The resulting network can be represented graphically as a circle or rectangle, with the nodes standing in for individuals or populations and the edges for genetic distances or splits. The distance matrices used for this method are basically the same as those used for Neighbor-join.

- *Methods based on clustering algorithms:*

  Often, the methods based on distance matrices are not sufficient to take into account all the variables involved in the differentiation and phylogeny within and between populations. There are many different clustering algorithms (e.g., STRUCTURE, STRUCTURAMA, DAPC, and other methods based on both linear and Bayesian models); in this section, a focus will be put on ADMIXTURE and TREEMIX, the methods mostly used for the research described in this thesis. Admixture analysis is a technique that determines the amounts of admixture or ancestry among populations and can spot introgressed or hybrid individuals. The popular admixture analysis program ADMIXTURE employs the maximum likelihood method to calculate the ideal number of ancestral populations and the admixture coefficients. In order to infer the population structure and admixture proportions using SNP data, the software tool ADMIXTURE is often used (Alexander & Lange, 2011). ADMIXTURE's algorithm is based on maximum likelihood estimation and optimizes the likelihood function using a block relaxation strategy. The ADMIXTURE algorithm begins with a preliminary estimate (k) of the mixing ratios and allele frequencies for each population or ancestral group. To increase the likelihood of the observed genotype data, it then iteratively updates the admixture proportions and allele frequencies while limiting their sums to one for the admixture proportions and the number of individuals for the allele frequencies. Block by block, a group of SNPs that are in linkage equilibrium or are weakly associated are updated to reflect the current admixture proportions. The expectation-maximization (EM) technique is used to update the admixture proportions within each block. This approach iteratively calculates the predicted genotype probabilities given the current parameter estimates (k) and then maximizes the likelihood by updating the admixture proportions. Similar updates are made to the allele frequencies, but this time, each block corresponds to a

single SNP. The allele frequencies are updated within each block using a closed-form solution that is dependent on the current admixture ratios. Once convergence (after a maximum number of iterations) is reached, the ADMIXTURE algorithm iterates between the admixture proportions and allele frequencies. Based on the variation in the parameter estimations, or the likelihood function between iterations, the convergence is evaluated. The predicted allele frequencies for each SNP, the estimated admixture proportions for each person or group, and the probability of the model are all included in the output of the ADMIXTURE algorithm. The cross-validation error, which assesses the predictive efficacy of the model on a subset of the data not used for training, may also be included in the output.

TREEMIX uses a tree-based method to perform admixture analysis. Its purpose is to reconstruct the evolutionary history and migration patterns between populations and identify population splits and admixture events. TREEMIX uses a maximum likelihood approach, similar to the algorithm of ADMIXTURE, to fit a tree model to the given genotype data and can incorporate population covariates and migration edges.

4.1.3. Events Influencing Genetic Diversity

Many events can influence the genetic diversity within and between livestock populations, either reducing or increasing the genetic variability of the studied populations.

- *Genetic drift:*

  One of the mechanisms of evolution that can alter the genetic make-up of a population is known as "genetic drift." It is definable as the random variation in the frequency of genetic variants or alleles in a population through time, caused by chance events rather than by natural selection. Due to the higher influence of chance events on the frequency of genetic variations in small populations, genetic drift can be especially significant in these populations. There are many models that can describe different events causing genetic drift. The two main models of events causing genetic drift are:

  - The Bottleneck effect: It occurs when a population experiences a sharp decline in size, whether as a result of a natural disaster, a disease outbreak, or human intervention, the remaining individuals may have different genetic frequencies than the initial population. In the event of a bottleneck, the population's genetic diversity may decline and specific genetic variants may become more or less prevalent as a result of random events. An example of the bottleneck effect is what we can observe in improved breeds of small sizes, where the systematic

selection of productive traits led to the loss of genetic variability in the loci involved in the expression of selected phenotypes.

- The Founder effect: Describes the genetic drift that happens when a small number of members of a larger population move to a new area and start a new population. The genetic diversity of the original population may not be fully represented in the new population. Because of chance events that occurred during the formation of the new population, it is possible that it has less genetic diversity and different allele frequencies than the original population.

Often in livestock species, these two effects may be observed happening together. An example may be the case of small populations kept for conservation purposes, which are often isolated from external populations; in this case, we can observe a founder effect, but it can be added to bottleneck events that could happen when the operators try to plan the breeding for some morphological trait selected in the reproductor, as an example of morphological descriptors of the breed. Genetic drift, among other evolutionary processes like mutation, migration, and artificial and natural selection, can collectively reshape populations' genetic make-up and have an effect on their genetic diversity, structure, and capacity for adaptation over time.

- *Mutation:*

  Mutation is a spontaneous change in the DNA sequence of an organism's genome. It can introduce new genetic variants or alleles into a population, which can contribute to genetic diversity. Mutation rates vary among species and can be influenced by various factors such as environmental conditions, genetic background, and exposure to mutagens.

- *Gene flow:*

  The movement of individuals or genetic material (such as gametes) between populations is referred to as "gene flow." It can happen as a result of population movement or dispersal and result in the exchange of genetic variations between populations. Since it introduces new genetic variations or changes the frequency of existing variants, gene flow can promote genetic diversity within populations. One typical mechanism of gene flow in livestock farming is artificial reproduction, which is often used to cross individuals from different breeds and areas; a typical example may be the artificial breeding of swine and dairy cattle.

- *Selection:*

  Genetic diversity in livestock species can be impacted by natural selection, artificial selection, and other types of selection. Changes in allele frequencies result from natural

selection acting on the genetic variants already present in a population and favoring those that offer an advantage for adaptation to a particular environment. Changes in allele frequencies over time can also be caused by artificial selection, which is guided by human selection criteria and in which particular traits are preferentially chosen in spite of others.

- *Introgression:*

  The term "genetic introgression" describes how interbreeding can transfer genetic material from one group or species to another. This can happen spontaneously or as a result of artificial crossbreeding or hybridization operations. Populations' genetic diversity may be impacted by genetic introgression, which can introduce new genetic variations or change the frequency of already-present variants. Sometimes interbreeding between different but morphologically similar populations is practiced in order to improve the genetic variability in populations that suffer from strong genetic drift. After the crossbreeding, the crossbred population is bred again with purebred males in order to recover the original morphology of the population, but with the introgression of some of the genetic material derived from the crossbreed.

4.2. GWAS and Signatures of Selection

One of the utilities of genotypes obtained through SNP beachips, is to study the consanguinity of the animals, the structure of the populations, or to calculate the frequency of the genotypes (Fst). Through this last piece of information, it is possible to identify those markers whose frequency varies significantly with the variation of the phenotypic values (selection signatures) or to carry out genome-wide association studies (GWAS) that allow us to associate the variation in the Fst value of a given SNP and its significance with a specific trait. The comparison between the results of these analyses and the reference genome from which the genotype was obtained permits us to identify the genes involved in the expression of complex traits. This information can be used to implement population selection with the search and diagnosis of specific genotypes, which implement the information provided by EBVs (Gouveia et al., 2014).

4.2.1. GWAS

Genome-wide association studies (GWASs) could be used as a result of the development of high-density SNP arrays for the majority of livestock species. Testing the SNPs for relationships with traits is the fundamental idea of GWAS. The explanation for statistically significant SNP trait connections is that the SNP is in linkage disequilibrium (LD) with the gene that causes the

trait and that the gene and the SNP are closely related (Schmid & Bennewitz, 2017). This is possible because the level of LD depends on how far apart two chromosomal locations are from one another. Although genomic selection has been successful, finding the genes for complex traits is still a major problem in animal genetics. Goddard et al. (2016) provided useful guidelines for the use of GWAS to study complex traits: Improving genomic selection is the first step. The results of the GWAS can also advance our understanding of how traits are expressed biologically. Genes discovered by GWAS can be used to generate and test hypotheses about the synthesis of traits. This is particularly relevant for novel features like feather picking in laying hens, greenhouse gas emissions in ruminants, or nutrient efficiency that may someday be included in the selection goal or that may be managed by specialized medications or feeding regimens. Lastly, the findings of GWAS can shed light on the genetic makeup of the quantitative trait; for example, we may be concerned about the number of genes that regulate the genetic variance, the magnitude of their effects, the significance of nonadditive genetic effects, and other factors.

A measurement of the non-random relationship between alleles at several loci is called linkage disequilibrium (LD). Genome-wide association studies (GWAS) frequently employ the r^2 measure of LD to find genetic variations linked to complex characteristics. Given that N_e is the effective population size and c is the rate of recombination between the two loci, the expectation of r2 may be written as E(r2) = 1/(1+4Ne*c). The likelihood that a crossover event will take place between two loci during meiosis is determined by the recombination rate between them; for this reason, the closer two loci are, the lower the probability that the linkage is "broken."

Single-marker models

In genome-wide association studies (GWAS), the single-marker model is a statistical model that is employed to examine the relationship between a single genetic marker and a quantitative characteristic. One SNP at a time is fitted in this model, typically using a mixed linear model. The gene substitution effect is represented by the SNP effect, and an estimated SNP regression coefficient is used. This parametrization usually uses the equation shown below:

$y_i = \mu + b_j x_{ij} + u_i + e_i$.

In this case, y_i is the trait record for individual “i” stands for the fixed mean (assuming there are no other fixed effects), b_j is the regression coefficient for SNP j to be estimated, x_{ij} is the number of copies of the allele with the minor frequency at the SNP for each individual i (coded as 0, 1, or 2), u_i is random polygenic effects, and e i is random residual effects. There are some drawbacks to single-marker GWASs, including insufficient power to detect small-effect mutations and difficulties separating real relationships from false positives brought on by

multiple testing. Yet they are still an effective tool for locating potential genes and regions linked to complex traits in animal breeding.

Bayesian multi-marker models

Bayesian multi-marker models are statistical frameworks that are applied in genome-wide association studies (GWAS) to examine the relationship between a number of genetic markers and a quantitative characteristic. The impacts of many variations can be simultaneously captured by these models, which also take into consideration the correlation structure among markers. Stringent criteria are required with dense SNP data to avoid an inflation of type-I error due to the high amount of multiple testing. Moreover, due to incomplete linkage disequilibrium (LD), it is likely that the influence of a gene is only partially captured by a single marker and may instead be better explained collectively by the SNPs that surround the gene. When modeling complicated genetic architectures, Bayesian approaches offer a flexible framework that can take into account previously collected data on genetic effects, such as functional annotations or pathway details. These models can also accommodate imputed genotypes and missing data. The Bayesian LASSO (Least Absolute Shrinkage and Selection Operator) model, which employs a prior distribution that promotes sparsity in the regression coefficients, is an illustration of a Bayesian multi-marker model. As a result, the model will only include a portion of the markers with significant impacts, while the remaining markers will have coefficients that are very near zero. Another example is the Bayesian variable selection model, which chooses variables based on their posterior probabilities by sampling from the posterior distribution of regression coefficients using Markov Chain Monte Carlo (MCMC) algorithms. In contrast to single-marker models, Bayesian multi-marker models have been demonstrated to increase power and accuracy in finding associations, particularly for characteristics with low heritability or complex genetic architectures. However, they demand additional processing power and Bayesian statistical knowledge in order to take the prior assumptions into careful consideration. Furthermore, the results obtained through this models should be validated using independent datasets

4.2.2. Signatures of Selection

Selection signatures are genomic regions that show evidence of being under positive or negative selection in a population. These regions can be identified by analyzing patterns of genetic variation across the genome and comparing them to neutral expectations. Positive selection signatures indicate that certain alleles or haplotypes have increased in frequency due to their beneficial effects on fitness or production traits, while negative selection signatures suggest that certain variants have been removed from the population due to their deleterious effects. The

detection of selection signatures can provide insights into the evolutionary history, adaptation, and breeding history of livestock populations. There are several methods for the detection of selection signatures in livestock genomes.

Single site method

The method of single-site population differentiation is frequently employed to identify selection fingerprints in the genomes of livestock. In order to determine which regions are being selected, this strategy examines the frequency of alleles or SNPs between populations. A typical statistic to assess the level of population divergence at a single locus is the fixation index (FST). Higher values of FST denote stronger differentiation and range from 0 (no differentiation) to 1 (full differentiation). Positive selection is indicated by highly variable allele frequencies between populations at any particular locus (higher FST values), whereas negative selection is suggested by low FST values. Wright's FST, on the other hand, may result in an overestimation of FST for small sample sizes since it assumes infinite population sizes. In livestock genomics research, single-site population differentiation techniques are frequently employed to pinpoint genomic areas that are being selected for attributes like disease resistance, meat quality, and milk output. This approach has significant drawbacks, such as its vulnerability to demographic phenomena like founder effects and bottlenecks, which can alter allele or haplotype frequencies in a manner akin to selection signatures. Therefore, it's crucial to employ a variety of techniques and strategies when looking for selection signatures in genomic data. A bayesian model designed for the detection of signatures of selection on the base of a single marker FST is the one implemented in the widely used software Bayescan (Foll & Gaggiotti, 2008).

Haplotype based Methods

Haplotype-based differentiation techniques are a type of selection signature analysis that employs patterns of linkage disequilibrium (LD) between several SNPs within a haplotype block to pinpoint regions under selection. These techniques are predicated on the idea that selective sweeps produce lengthy haplotypes with high frequencies and reduced LD with nearby loci.

HapFLK, OmegaPlus, and SweepFinder2 are a few examples of approaches for haplotype-based distinction. HapFLK is a technique that estimates haplotype frequencies and identifies genomic areas under selection using a hierarchical Bayesian model. A composite likelihood ratio test is used by OmegaPlus, another technique, to identify selective sweeps based on haplotype frequencies. Using a hidden Markov model, the SweepFinder2 approach calculates the likelihood of selective sweeps based on haplotype homozygosity. Among the most commonly used haplotype-based tests are the methods based on integrated haplotype homozygosity (iHH). iHH is determined by integrating the extended haplotype homozygosity (EHH) of a chosen allele over

a predetermined distance on both sides of the core SNP. The SNP with the greatest EHH value within a particular region is typically considered to be the core SNP. Each population's iHH values are calculated individually, and variations between populations' iHH values can be used to identify genomic regions with various patterns of haplotype diversity, which may be a sign of selection. As indicated by positive values of iHH, the selected allele has a high frequency and long-range LD, which are traits of selective sweeps, whereas negative values show a low frequency and short-range LD, which are traits of balancing selection. Among the most popular tests derived from iHH is cross-population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007). This approach uses data from several populations' haplotypes and is less prone to SNP ascertainment bias. By comparing the extended haplotype homozygosity (EHH) of a chosen allele between two populations, the XP-EHH method can identify genomic areas that are under selection. To calculate XP-EHH between two populations (e.g. A and B), first, the iHH values for each population are calculated independently by integrating the EHH of the entire sample in the population. The XP-EHH score is directional, and positive and negative scores suggest that selection occurred in populations A and B, respectively. The ability of XP-EHH to identify selection signatures linked to recent positive selection or local adaptation has been demonstrated (Saravanan et al., 2020). Demographic factors, like migration and population structure, may also have an impact on the method's outcomes and result in false-positive findings. Similar to XP-EHH, RSB compares the distributions of EHHs between pairs of populations at various loci, as opposed to XP-EHH, which focuses on comparing haplotype homozygosity between two populations at particular loci. The iHS (integrated haplotype score) method, which is based on EHH, aids in determining in which population we see the selective sweep. The iHS statistic is calculated at each SNP in a population by comparing the EHH of the chosen allele to that of the alternate allele. The EHH for the chosen allele indicates whether recent positive selection has occurred on the chosen allele if it is higher than the EHH for the alternative allele (Sabeti et al., 2005). To standardize iHS values across SNPs with different allele frequencies, developed a procedure that uses average and standard deviation values over all SNPs with similar frequencies. However, this approach has low power when the frequency of the selected allele is high. To address this issue, Tang et al. (2007) developed a similar procedure called EHHS (site-specific EHH), which compares EHHS profiles between populations to detect signals of positive selection in genomic regions with unusually high Rsb values. EHHS calculates a weighted average of EHH at both alleles for each SNP in each population and then compares these profiles between populations to identify regions with high levels of haplotype homozygosity.

Haplotype-based differentiation techniques have been proven to be more effective at detecting selection signals than single-locus testing, in particular for soft or incomplete selective sweeps. However, demographic factors like population structure and migration may also have an impact on these methodologies, leading to false-positive results. Subsequently, it's crucial to carefully select the best technique for a specific study and take into account a variety of supporting data while interpreting the outcomes.

4.3. Advances in Animal Breeding

Genetic improvement in livestock has contributed to obtaining, from generation to generation, increasingly specialized and efficient animals. In fact, the improved breeds make it possible to optimize the use of raw materials necessary to obtain a certain quantity of product. This improvement is extremely evident in dairy breeds (Brito et al., 2021), but can also be observed in beef cattle breeds (ANABIC, 2022; Canestrari, 2005).

In a few decades, we have passed from the so-called mass and morphological selection, obtained by choosing the “best-looking” animals as reproducers, to a more rapid genetic selection obtained through the estimation of the reproductive value of the animals (also called Estimated Breeding Values) for certain phenotypes of productive interest (Gianola & Rosa, 2015). Among the statistical techniques that have revolutionized the genetic improvement of livestock are the BLUP indices. They allow for "unbiased" estimates of the reproductive value of the animals using the values of the phenotypes, collected by breed controllers and experts, to estimate what is the contribution that the genetics of the animals makes to the phenotypic variance within their population. This is thanks to the information about the relatives of the animals obtained from the pedigree, from which it is then possible to estimate the values concerning the heritability of the characters and their frequency in the population. In recent years, the accuracy of these estimates has improved significantly thanks to the advent of genomics (van der Werf, 2013).

With the advent of genomics, large-scale DNA information is available to predict performance potential. With the advent of this technology, the estimate of the EBVs of animals has become increasingly accurate, as the use of genomic data allows for more precise estimates of some parameters compared to the sole use of the pedigree (relationships, heritability, genotypic frequencies). In recent years, the use of genomics has become increasingly widespread, thanks to the reduction in the costs of the analyses necessary to obtain the genotypes of animals. This is thanks to new technologies such as SNP chips that allow for rapid analysis of multiple samples (up to more than 1,000 subjects at a time) at different densities (8-18k low density, ~50k medium density, >70k high density). The costs of these analyses vary according to the density of the chip,

i.e., according to how many SNP markers will compose the genotypes obtained from them. The chips of different densities, however, contain a certain quantity of markers in common, which allows us, thanks to statistics and bioinformatics techniques, to impute the genotypes of the animals. That is, to estimate the presence or absence of a SNP marker on a specific locus, one can exploit the genotypic information of its relatives (both direct and indirect) and pedigree; one can obtain high-density genotypes even for those animals genotyped with cheaper chips. This allows associations to rationalize the costs of genomics by genotyping only a reference population with the most expensive chips (Nicolazzi et al., 2013). BLUP assessments using genomic data can be of two types: GBLUP (multiple-step) or ssGBLUP (single-step). They are called GBLUP or multi-step when the relationships between the genotyped animals are calculated through the genomic data obtained with the SNPchips, for which a second analysis (post-processing) will be necessary to also use the information of the non-genotyped animals through the calculation of the classic parentage based on pedigree values. The main advantage of this approach is that the cost is greatly reduced as only breeding candidates and animals of high genetic value, such as approved sires, are genotyped. The traditional BLUP assessment is kept unchanged, and genomic selection can be done using additional analyses. However, these methods are not without disadvantages, as genomic values are generated only for simple models (for example, single-trait, non-maternal models), and this does not meet the needs of field reality. Furthermore, only genotyped animals are included in the model. The lack of possibility of making a multiple-trait evaluation requires the construction of pseudo-phenotypes, which are expensive to calculate and whose accuracy is linked to approximate algorithms. Another genomic assessment methodology is the single-step Genomic BLUP (ssGBLUP). Starting with the notion that relationships derived from pedigrees can be regarded expected while those derived from genotypes can be considered observed, non-genotyped animals' relationships will be supplemented by the genotypes of their genotyped relatives (Legarra et al., 2009; Lourenco et al., 2020).

5. Aim of the Thesis

As reported in the introduction, the sustainability of livestock production is one of the major challenges that animal production will face in the near future. Among the different approaches that can be used to face this challenge, the genetic one is certainly relevant. Within this, two strategies are possible, one based on the use of local breeds adapted to specific environments and the other on selection for greater sustainability of cosmopolitan breeds. In recent years, to implement these strategies, powerful tools have become available that concern, on the one hand,

the possibility of easily obtaining a large amount of genetic data and, on the other hand, the possibility of analyzing it with appropriate bioinformatic tools.

This thesis aims to describe the applications of bioinformatic tools to analyze the SNP data in order to enhance the sustainability of livestock farming. The work focuses on two aspects of animal breeding and genetics: the study of biodiversity, intended as genetic diversity within and between populations, and studies about the detection of loci under selection in livestock through signatures of selection. The studied populations are cattle and sheep breeds farmed in Italy, including both local and cosmopolitan breeds.

The first part will focus on biodiversity by examining, using bioinformatic tools used to manage SNP data, the genetic diversity of a local endangered sheep breed, Pagliarola, as well as the transboundary cattle breed, Holstein-Friesian.

The Pagliarola breed will be investigated to ascertain its genetic variability, structure, relationship, and introgression with other breeds. Knowledge about the genome of this local endangered breed is of fundamental importance for its conservation and exploitation. This matter is very important from the perspective of a sustainable development of livestock farming because the conservation of local breeds may help to preserve rare genetic variants that are useful to deal with climate change and other environmental alterations. It is also known that local breeds are often linked to local farming and traditional products; for this reason, their conservation is important not only from the perspective of ecological and economic sustainability but also plays a role in social and cultural sustainability.

Based on imputed SNP data, the genetic diversity of the cosmopolitan dairy cattle breed Holstein-Friesian will be inspected within the Italian population selected for different breeding goals and compared with the North American population. A study of the signatures of selection among the different Holstein populations has been performed. The study of the biodiversity of a cosmopolitan, highly selected breed in order to detect the existence of loci that can determine a better adaptation to different environments. With this information, the selection can be tailored according to the environment.

The second part will focus on the identification of loci in a selected dairy sheep breed that can determine resistance to the parasites. In this section of the thesis, the study of Fecal Egg Count (FEC) phenotypes in Comisana dairy sheep (local to Sicily) that led to the development of a model aimed at the calculation of breeding values for the resistance to gastrointestinal nematodes (GIN) in sheep will be described. The genomic regions involved in the resistance mechanism and their role in the animals' physiology have been identified by comparing the most and least

resistant reproductors to GIN using SNP data. This has been done by identifying the genes involved in metabolic pathways that were differentially expressed in the two compared groups.

6. Articles

6.1. Articles Included in the Thesis

The articles presented in this section are an integral part of my Ph.D. thesis project. The first article in which I depicted, together with my co-authors, the characterization of the Pagliarola local population has been published in the international journal “Italian Journal of Animal Science” (Persichilli et al., 2021). The second article, describing a comparison within the breed of the Italian Holstein based on different breeding goals and the parallel confrontation with samples from the North-American stock, is in press in the “Journal of Dairy Science" (Persichilli et al., 2023a). The last article in which Estimated Breeding Values for Fecal Egg Count have been used as a metric for genetic parasite resistance in order to catch genome-wide associations to parasite resistance is still in submission to an international journal (Persichilli et al., 2023b).

Italian Journal of Animal Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjas20

Genome-wide diversity of Pagliarola sheep residual population and its conservation implication

Christian Persichilli, Gabriele Senczuk, Loi Pasqualino, Debora Agata Anzalone, Riccardo Negrini, Paolo Ajmone Marsan, Licia Colli, Fabio Pilla & Salvatore Mastrangelo

To cite this article: Christian Persichilli, Gabriele Senczuk, Loi Pasqualino, Debora Agata Anzalone, Riccardo Negrini, Paolo Ajmone Marsan, Licia Colli, Fabio Pilla & Salvatore Mastrangelo (2021) Genome-wide diversity of Pagliarola sheep residual population and its conservation implication, Italian Journal of Animal Science, 20:1, 1695-1705, DOI: 10.1080/1828051X.2021.1970033

To link to this article: https://doi.org/10.1080/1828051X.2021.1970033

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 26 Oct 2021.

Submit your article to this journal

Article views: 944

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjas20

ITALIAN JOURNAL OF ANIMAL SCIENCE
2021, VOL. 20, NO. 1, 1695–1705
https://doi.org/10.1080/1828051X.2021.1970033

Taylor & Francis
Taylor & Francis Group

PAPER

OPEN ACCESS Check for updates

Genome-wide diversity of Pagliarola sheep residual population and its conservation implication

Christian Persichilli[a], Gabriele Senczuk[a], Loi Pasqualino[b], Debora Agata Anzalone[b], Riccardo Negrini[c], Paolo Ajmone Marsan[c], Licia Colli[c], Fabio Pilla[a] and Salvatore Mastrangelo[d]

[a]Dipartimento di Agraria, Ambientale e Scienze dell'alimentazione, University of Molise, Campobasso, Italy; [b]Facoltà di Veterinaria, University of Teramo, Campobasso, Italy; [c]Department of Animal Science Food and Nutrition, Nutrigenomics and Proteomics Research Centre – PRONUTRIGEN, and Biodiversity and Ancient DNA Research Centre, Università Cattolica del Sacro Cuore, Piacenza, Italy; [d]Dipartimento di Scienze Agrarie, Alimentari e Forestali, University of Palermo, Palermo, Italy

ABSTRACT

Local breeds represent an underestimated resource in terms not only of their important cultural and economical role in marginal areas, but also because they often own a potential genetic pool well adapted to extreme conditions. This fact is of increasing interest, especially when considering climate global challenges where peculiar and uncommon traits could be advantageous. In this study, we genotyped 24 individuals belonging to the small residual Pagliarola sheep population using the OvineSNP50K array, in order to compare its genomic architecture with other 21 Italian local breeds. Moreover, we performed the fixation index (F_{ST}) outlier analysis to identify genes most differentiated between Pagliarola and Merino-derived Italian breeds. All population genetic analyses highlighted that Pagliarola breed represents a distinct genetic unit showing a genetic relationship with Merino-derived breeds underlying the close cultural connection between these breeds related to transhumance. However, Pagliarola breed resulted to be divided into two different sub-populations, named Pagliarola 1 (PAG-1) and Pagliarola 2 (PAG-2). Genetic diversity indices and inbreeding estimated from runs of homozygosity (F_{ROH}) indicated that while PAG-1 showed among the highest values, PAG-2 turned out strongly inbred, probably as a consequence of a founder effect. The F_{ST} outlier analysis identified the 5 most differentiated single nucleotide polymorphisms, two of which mapping within known genes (*MACF1* and *GLIS1*) reported to be linked to feed efficiency and local adaption. All these information strongly suggest that proper conservation measures should be implemented in order to recover the Pagliarola population and its fundamental local products.

HIGHLIGHTS

- All population genetic analyses highlighted that Pagliarola breed represent a distinct genetic unit, and clusters with the Merino-derived Italian breeds.
- The breed resulted to be divided into two different sub-populations.
- For the two Pagliarola sub-populations, the genetic diversity indices suggest a very different genetic makeup.
- Results from the Bayesian approach identified two know genes (*MACF1*and *GLIS1*) related to feed efficiency in livestock species and local adaptation.
- The information generated in this study strongly suggests that proper conservation measures should be implemented in order to recover the Pagliarola population.

ARTICLE HISTORY
Received 13 April 2021
Revised 12 August 2021
Accepted 16 August 2021

KEYWORDS
Local sheep; population structure; genetic diversity; SNP; candidate genes

Introduction

During the last century, erosion of livestock genetic resources was observed as the result of a massive replacement of low-productivity local breeds with highly productive ones. These local breeds are often reduced to populations having a small effective size, implying difficulties to manage inbreeding and within-breed genetic diversity. In recent years, an increasing interest in recovering and preserving local breeds has taken place (e.g. Ciani et al. 2014; Mastrangelo et al. 2018). Local breeds are the result of a combination of natural and artificial selection carried on over centuries

CONTACT Dr Gabriele Senczuk g.senczuk@unimol.it Dipartimento di Agraria, Ambientale e Scienze dell'alimentazione, University of Molise, Campobasso, Italy

Supplemental data for this article can be accessed here.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

which shaped animals to adapt and produce in a variety of different environments (Ajmone-Marsan 2010; Ajmone-Marsan et al. 2014). In fact, local populations often possess valuable traits such as disease resistance, high fertility, good maternal qualities, longevity and adaptation to harsh conditions and poor-quality feed, all desirable qualities for low-input, sustainable agriculture. In addition, local breeds are a reservoir of diversity that may turn out useful to face the upcoming climate change and disease outbreaks and contribute to conserve local culture and traditional products (Ciani et al. 2013).

With a total of 263, Italy is one of the European countries with the highest number of local livestock breeds. Among these 84 are sheep breeds, 17 of which are already extinct (Bittante 2011). Italy has a long history of sheep breeding and, despite a dramatic contraction in numbers, still raise several local sheep breeds that may represent a unique source of genetic diversity (Ciani et al. 2013, 2014; Mastrangelo et al. 2017).

The Pagliarola is an ancient sheep population farmed in central Italy, mainly in the Abruzzo region (Marchi and Mascheroni 1925; Fabrizio 1928; Bonadonna 1976), together with other small local populations such as Bariscianese, Pomarancina, Perugina, Pecora dell'Amiata, Todina and Vissana. This sheep finds its origins in the Apennines mountain range, and is identified as the ancestral Appenninica sheep population. Even if these small local populations were never selected to improve productive traits such as meat (Marchi and Mascheroni 1925; Bonadonna 1976; Sarti et al. 2002), they were raised for their adaptation to the Apennines environment and for their low nutritional requirements. In particular, Pagliarola was resident in the mountain villages all year round (Renieri et al. 1984). During the severe winter season, the breed was usually supplemented with straw (''paglia'') and this derived their actual name. Pagliarola is a dual-purpose breed; it is reared both for its milk and the resultant local dairy products ('Pecorino di Farindola', 'Pecorino del Matese', and 'Formaggio di Pietracatella') and for its meat ('Agnello del Centro Italia IGP', made from the meat of the lambs of the local sheep breeds of Apennines, including Pagliarola). Ceccobelli et al. (2016) reported that only a few small Pagliarola flocks were still surviving in very hard environmental conditions, thus, an appropriate breeding program in order to recover Pagliarola population would be desirable. More recently, this population has been counted for a total of 25 individuals (21 ewes and 4 rams) in the territory of Barisciano (AQ) and samples of semen from two Pagliarola rams were collected and preserved in order to establish a genetic bank of the population (Anzalone et al. 2018). This is the last estimation of the population consistency.

The availability of sheep genome-wide SNP panels allows providing background information concerning genome structure in local and cosmopolite breeds (e.g. Kijas et al. 2012; Ciani et al. 2014). Genetic diversity is a key measure for the monitoring of genetic parameters that are important for the prevention of genetic erosion, inbreeding and other deleterious processes that may lead to population extinction. In addition, an investigation of genomic variation is an important prerequisite to maintain its integrity and to ensure appropriate conservation strategies in the presence of genetic structure that is not easily identifiable with morphological data. Moreover, considering the challenging of the climate changes, the identification of genetic stock which might harbour potential genes well adapted to extreme environment, should be a topic of increasing importance (Hoffmann 2010; Biscarini et al. 2015).

The aim of this work was to study the genetic diversity and population structure of Pagliarola sheep in the Italian context using data generated from Illumina OvineSNP50K array. Moreover, a genome-scan analytical approach was performed to identify genomic regions harbouring genes conceivably associated to the breed aptitude. This paper highlights how genomic survey is a valuable tool in order to advise possible reintroduction and conservation plans. Indeed, contrariwise to other local sheep for which conservation programs, supported by several authorities (PSRN, ASSONAPA), have a longstanding strategy to manage zootechnical biodiversity (genomic characterisation, semen collection and reconstitution), fewer efforts have been made on small residual local breeds that instead represent an important reservoir for genes adapted to specific environmental conditions.

Materials and methods

DNA sampling, genotyping and quality control

Blood samples were collected from 24 individuals of Pagliarola sheep. Genomic DNA was extracted using the Genelute Mammalian Genomic DNA Kit (Sigma). The concentration of extracted DNA was assessed with NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). All animals were genotyped for 54,241 single nucleotide polymorphisms (SNPs) using the Illumina OvineSNP50K array. The raw data have been merged using the software PLINK

(Purcell et al. 2007) with the genotypic data of 21 Italian sheep breeds assembled by the Italian Sheep Consortium (BIOVITA) (Ciani et al. 2014). The genotypes of Merinizzata Italiana breed have been retrieved from the dataset used in Ciani et al. (2015). In addition to these Italian breeds, the Ile de France sheep (IDF) was considered due to its likely cross-breeding with the Pagliarola breed. Genotypic data of IDF breed were available from Rochus et al. (2018). After filtering for minor allele frequency (–maf 0.05), missing genotype call rate (–geno 0.01), linkage disequilibrium (r^2 0.02) and missngness per individual (–mind 0.1), a final dataset consisting of 29,970 SNPs was obtained and one Pagliarola individual has been removed.

Analysis of genetic relationships and population structure among breeds

Genetic structure of breeds has been inspected by using the maximum likelihood clustering approach as implemented in the software ADMIXTURE (Alexander and Lange 2011) with K values ranging from 2 to 24. The results have been plotted using the *membercoef.circos* function in the R package BITE (Milanesi et al. 2017).

To explore the genetic differentiation among breeds, we performed a multidimensional scaling plot (MDS) based on pairwise identical-by-state (IBS) distances using the software PLINK (–cluster, mds-plot 3) and visualised using the R package *ggplot2* (Wickham 2011).

In order to reconstruct genomic relationships among all populations and to investigate the presence of patterns of population splitting and migration events, a maximum likelihood tree has been generated using the software TREEMIX (Pickrell and Pritchard 2012). For this analysis we also used the genotypic data of Soay sheep (Kijas et al. 2012) as an outgroup, and allowing migration edges from 1 to 10. The best predictor for the number of migration edges was selected using the *optM* function in the R package OptM (Fitak 2018).

Genetic diversity indices

Genetic diversity indices such as observed (Ho) and expected (He) heterozygosity and minor allele frequency (MAF) have been estimated using PLINK, while trends in historical effective population size (Ne) for the 13th generation (presumed to be contemporary Ne), based on linkage disequilibrium (LD), have been estimated using the software SNeP v1.1. (Barbato et al. 2015).

In addition, to have an estimate of the inbreeding within and among breeds, we performed runs of homozygosity (ROH) using PLINK and the following parameters: a sliding window of 50 SNPs across the genome; the minimum number of consecutive SNPs included in an ROH was 50; the minimum length of an ROH was set to 4000 kb; the maximum gap between consecutive homozygous SNPs was 1000 kb; a density of one SNP per 100 kb; and a maximum of one SNP with missing genotype and up to one heterozygous genotype were allowed in an ROH. The inbreeding coefficient (F) on the basis of ROH (F_{ROH}) for each animal was calculated as follows:

$$F_{ROH} = L_{ROH}/L_{aut}$$

where L_{ROH} is the total length of all ROH in the genome of an individual and Laut is the specified length of the autosomal genome covered by SNPs on the chip (2452.06 Mb).

F_{ST}-outlier analysis

Based on the genetic structure results, the F_{ST}-outlier approach as implemented in the BAYESCAN software (Foll and Gaggiotti 2008) was adopted to identify loci involved in the differentiation between the Pagliarola 1 (see Results and discussion section) and the other Merino-derived Italian breeds (GDP, SPV and MER), using the default parameters. To control the number of false positives, significant markers were defined by applying a q-value threshold of 0.05 and using the 0.9999 SNPs of F_{ST} percentile distribution (Moscarelli et al. 2021). Chromosomal coordinates for each SNP were obtained using the sheep reference genome Oar 3.1.

Results and discussion

High-throughput genotyping arrays have greatly facilitated the study of genetic structure in livestock species, but whereas much effort has been devoted to study population structure within dominant commercial breeds, local breeds, which are less widely used, are generally understudied (Beynon et al. 2015). This study provides the first overview of population structure estimates of the Pagliarola sheep from a genome-wide perspective.

Analysis of genetic relationships and population structure among breeds

The ADMIXTURE analysis, MDS and TREEMIX were used to visualise and explore the genetic relationships between Pagliarola and the other sheep breeds.

Population structure inferred using the software ADMIXTURE, with *K* ranging from 2 to 24, showed the lowest CV error at $K = 18$ (Supplementary Materials, Figure S1). At this *K* value, the ADMIXTURE components highlighted that all breeds have their own genetic identity, with the exception of Biellese and Bergamasca breeds which appear to be indistinguishable and Comisana and Pinzirita which share a large proportion of genetic components (Mastrangelo et al. 2014) (Figure 1). At $K = 2$, Ile de France breed first differentiated while Merinizzata showed high level of admixture when compared with all the other Italian native breeds. At $K = 3$, Alpagota, Biellese and Bergamasca separated, while the Pagliarola breed resulted to be divided into two different sub-populations which are clearly distinguishable at $K = 4$, and hereinafter named Pagliarola 1 (PAG-1, $n = 14$) and Pagliarola 2 (PAG-2, $n = 9$). It should be noted that despite the number of individuals belonging to the two Pagliarola subgroups was quite small, it has been empirically demonstrated that for population structure analyses, the patterns observed using six randomly extracted animals per breed closely mirrors those inferred from 20 to 24 animals per breed (Gaouar et al. 2017).

These two genetic clusters (PAG-1 and PAG-2) match with the two different farms where the samples were taken. Therefore, the genetic structure detected for Pagliarola sheep could be due to the introgression of genes from other breeds (PAG-1) and/or to geographical isolation for a long time with genetic drift

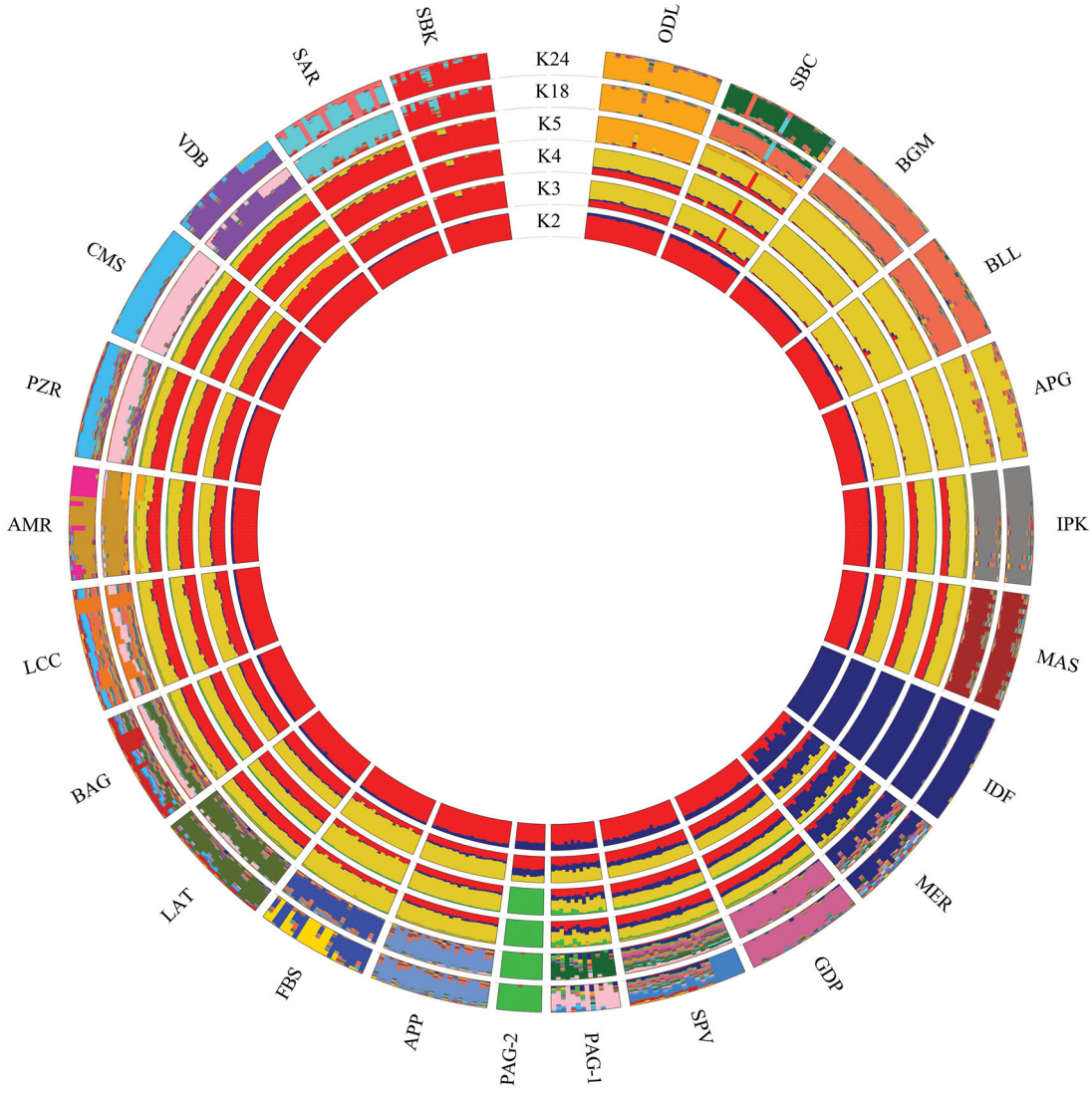


Figure 1. ADMIXTURE analysis showing the genetic structure of the analysed breeds. For full definition of breeds see Table 1.

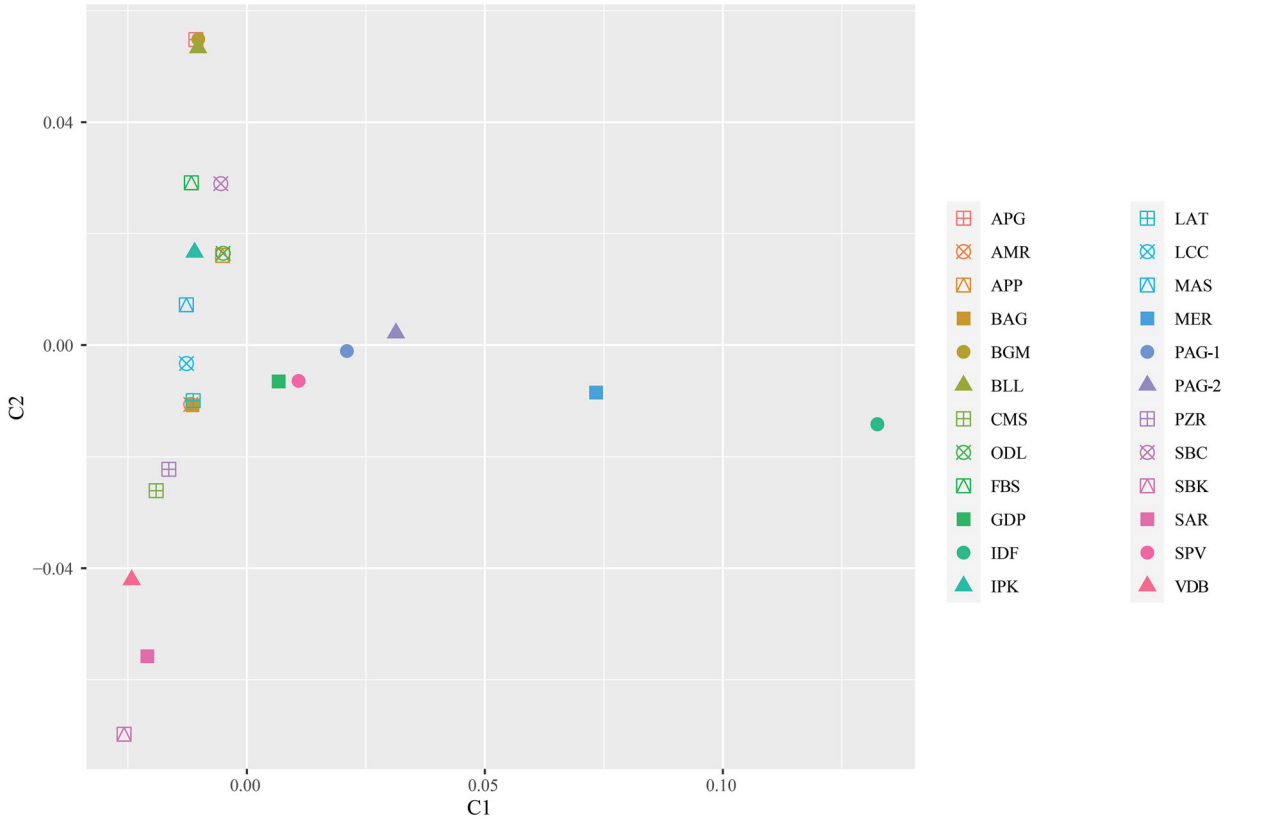


Figure 2. Multidimensional Scaling Plot (MDS) based on identity-by-state distance matrix indicating breed differentiation. Each dot corresponds to a different individual. For full definition of breeds see Table 1.

(PAG-2). These results were also in agreement with a previous study conducted on the genetic structure and relationship using microsatellite markers (Ceccobelli et al. 2016).

Component 1 of the MDS plot reveals a clear differentiation of breeds from central Italy (MER, PAG-1, PAG-2, SPV and GDP) including the Ile de France breed which as expected is shown as the most differentiated. On the other hand, component 2 reflects a latitudinal genetic cline with breeds from northern Italy located mostly in the upper portion of the plot, while breeds from southern Italy and Sardinia cluster in the lower quadrant of the plot (Figure 2). This strong phylogeographic pattern of sheep breeds has been highlighted in previous studies at both worldwide (Kijas et al. 2009, 2012) and regional level (Ciani et al. 2014; Mastrangelo et al. 2017). Such a geographic structuring has been suggested to be the outcome of a combination of adaptive processes driven by natural selection and historical admixture prompted by human selection. Moreover, the observed phylogeographic pattern mirror the distribution of human genetic diversity (Brisighelli et al. 2012; Anagnostou et al. 2017) underling the close connection between livestock diversity and historical human migrations/barriers as highlighted in other livestock species (Senczuk et al. 2021). For the Pagliarola sheep, these results confirmed the findings based on the ADMIXTURE, where, the population formed two separated and differentiated clusters.

The TREEMIX phylogram shows a clear distribution of clusters according to the geographic origin of breed (Figure 3). Indeed, we observe a group including all breeds from Central-Eastern Alps and Central Apennines (IPK, APG, BGM, BLL, APP, FBS), a second group including Western Alps (ODL and SBC), a group including Central Apennines breeds (GDP, SPV, MER, PAG-1, PAG-2) and another group including breeds from Central and Southern Italy (LAT, BAG, AMR), Sicily and Apulia (LCC, CMS, PZR, VDB) and Sardinia (SAR, SBK). In this context, Pagliarola clusters with the Merino-derived Italian breeds (Figure 3). This result is not surprising considering the close connection between Pagliarola and some Merino-derived Italian breeds. In fact, up to the 1950s of the last century, the Pagliarola was traditionally reared in the same flocks with Gentile di Puglia, though the management conditions of the two breeds were different (Renieri et al. 1984). This topology remained reasonably constant through iterations. The linear method implemented in the OptM (R Package OptM) function indicated a major changing point in the log likelihood at four

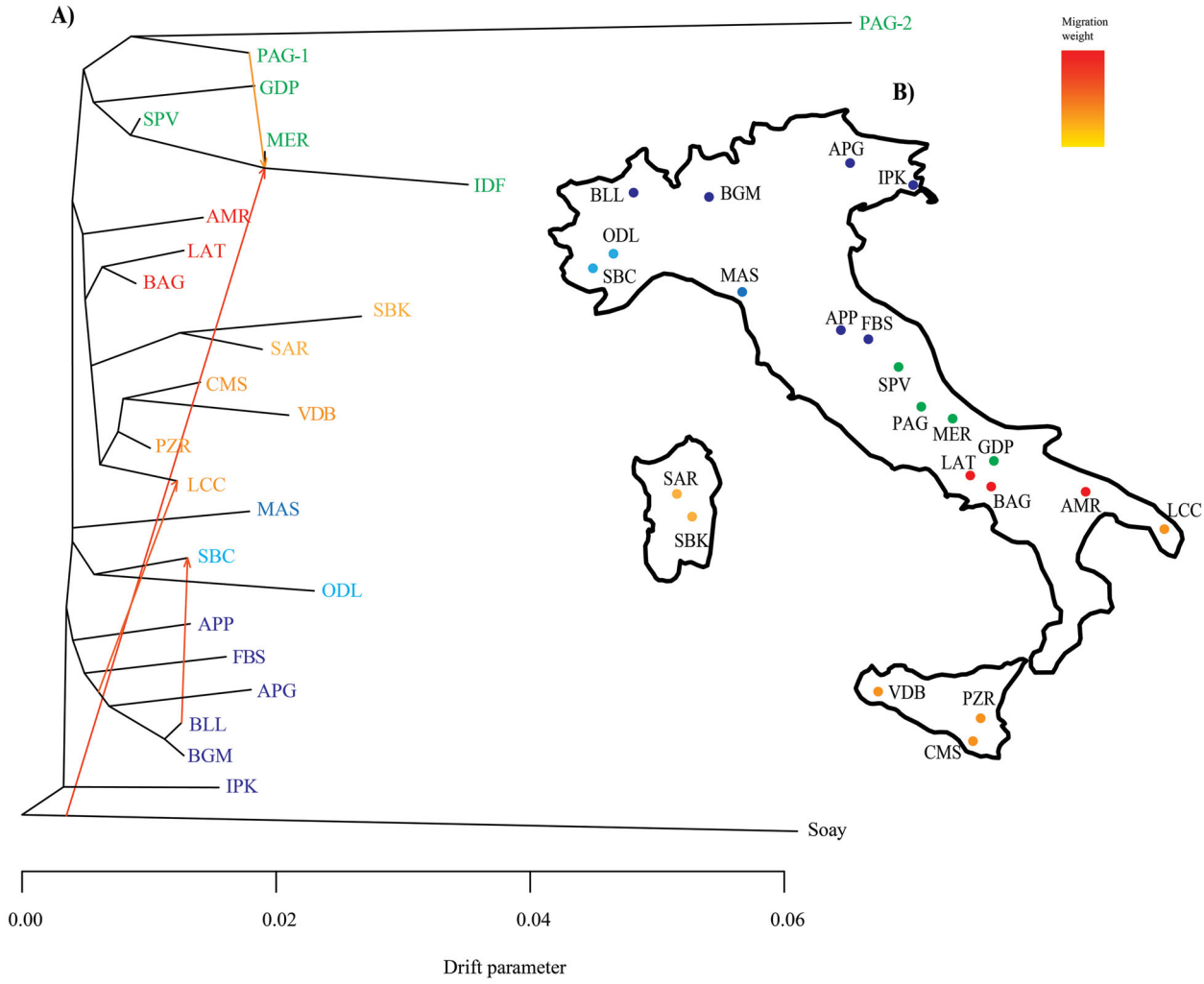


Figure 3. TREEMIX analysis with the most supported number of migration events with their relative weights (A). Geographic location of the native Italian breeds used in this study. Colours depict the main geographic cluster as highlighted in the TREEMIX analysis. For full definition of breeds see Table 1.

migration events (Supplementary Materials, Figure S2). A strong migration event is shown between Soay and the node separating Merinizzata and Ile de France breeds. This migration event is probably linked to the Northern European origins of both Ile de France and Soay breeds. On the other hand, the close relationship of Merinizzata with these Northern European breeds is due to its known admixed origin between local Merino-derived breeds and exotic Merino meat breeds including Ile de France (Morbidini et al. 1995; Sarti and Panella 1999). Another strong migration edge is shown between the base of the crown including Alpagota, Biellese and Bergamasca and the Leccese breeds. This result is in accordance with documented livestock exchange between Southern and Northern regions in order to improve production (Ciani et al. 2014). An additional migration event is highlighted between Biellese and Sambucana breeds. In this context, the use of Biellese to improve meat yield in other autochthonous breeds including Sambucana is well known (Bigi and Zanon 2008). Finally, another migration event is shown between Pagliarola (PAG-1) and the node including Merinizzata and Ile de France breeds indicating some gene exchange between these breeds. Even if Pagliarola has been considered as one of the local populations of the Apennines (Marchi and Mascheroni 1925; Bonadonna 1976) from which Appenninica breed has been selected after crosses with meat selected breeds (Sarti et al. 2002), neither migration events or drift parameters indicated a genetic relationship between these two breeds.

Genetic diversity indices

Genetic diversity indices, which are key parameters in the genetic management of populations, are reported in Table 1. The results revealed that Pagliarola 1 and Pagliarola 2 tended to display the highest and the

Table 1. Genetic diversity indices for all the analysed breeds.

Breed	N	Code	Ho ± SD	He ± SD	MAF ± SD	Ne	F_{ROH}
Alpagota	24	APG	0.350 ± 0.172	0.351 ± 0.148	0.268 ± 0.145	93	0.031
Altamurana	23	AMR	0.360 ± 0.166	0.358 ± 0.144	0.275 ± 0.143	82	0.045
Appenninica	24	APP	0.362 ± 0.173	0.359 ± 0.142	0.275 ± 0.141	90	0.029
Bagnolese	23	BAG	0.366 ± 0.155	0.372 ± 0.132	0.287 ± 0.137	116	0.036
Bergamasca	24	BGM	0.362 ± 0.163	0.362 ± 0.139	0.278 ± 0.14	130	0.014
Biellese	22	BLL	0.355 ± 0.168	0.362 ± 0.140	0.278 ± 0.141	123	0.022
Comisana	24	CMS	0.374 ± 0.164	0.363 ± 0.139	0.279 ± 0.14	118	0.008
Fabrianese	23	FBS	0.367 ± 0.174	0.352 ± 0.147	0.269 ± 0.144	75	0.031
Gentile di Puglia	24	GDP	0.346 ± 0.172	0.351 ± 0.148	0.268 ± 0.145	73	0.045
Istrian Pramenka	24	IPK	0.366 ± 0.171	0.356 ± 0.145	0.273 ± 0.143	91	0.019
Laticauda	24	LAT	0.364 ± 0.160	0.365 ± 0.137	0.281 ± 0.139	110	0.033
Leccese	25	LCC	0.345 ± 0.154	0.366 ± 0.137	0.281 ± 0.139	102	0.068
Massese	24	MAS	0.355 ± 0.170	0.350 ± 0.149	0.268 ± 0.145	88	0.032
Merinizzata	24	MER	0.371 ± 0.165	0.366 ± 0.137	0.282 ± 0.140	95	0.027
Ovino delle Langhe	24	ODL	0.342 ± 0.174	0.342 ± 0.153	0.261 ± 0.148	80	0.052
Pagliarola 1	14	PAG-1	0.391 ± 0.182	0.361 ± 0.137	0.276 ± 0.139	48	0.024
Pagliarola 2	9	PAG-2	0.274 ± 0.233	0.254 ± 0.194	0.192 ± 0.167	20	0.133
Pinzirita	24	PZR	0.361 ± 0.156	0.369 ± 0.134	0.284 ± 0.138	116	0.044
Sambucana	24	SBC	0.361 ± 0.164	0.362 ± 0.140	0.278 ± 0.14	104	0.024
Sardinian Ancestral Black	20	SBK	0.356 ± 0.188	0.332 ± 0.156	0.252 ± 0.149	65	0.046
Sardinian White	24	SAR	0.352 ± 0.175	0.350 ± 0.148	0.268 ± 0.144	98	0.025
Sopravissana	24	SPV	0.360 ± 0.162	0.370 ± 0.133	0.286 ± 0.137	111	0.052
Valle del Belice	24	VDB	0.339 ± 0.166	0.346 ± 0.150	0.264 ± 0.146	79	0.067
Ile de France	23	IDF	0.343 ± 0.179	0.336 ± 0.158	0.256 ± 0.150	101	0.031

Observed (Ho) and expected (He) heterozygosity, average minor allele frequency (MAF), effective population size (Ne), standard deviation (SD) and mean inbreeding coefficient based on runs of homozygosity (F_{ROH}).

lowest variability, respectively, than other Italian breeds. However, we cannot rule out the presence of an ascertainment bias effect due to the use of small samples size in the sub-populations (Bedhiaf-Romdhani et al. 2020). Pagliarola 1 showed the highest H_o (0.391 ± 0.182), whereas Pagliarola 2 showed the lowest one (0.274 ± 0.233). Similar outcomes are observed for expected heterozygosity (H_e = 0.361 ± 0.137; H_e = 0.254 ± 0.194 for Pagliarola 1 and 2, respectively). The lowest mean values of MAF are recorded for Pagliarola 2 (0.192 ± 0.167) and the highest for Bagnolese (0.287 ± 0.137), while Pagliarola 1 showed an intermediate value (0.276 ± 0.139).

The effective population size (N_e) estimates for each breed are also reported in Table 1. The minimum N_e value was found for Pagliarola 2 (N_e = 20), the maximum value was found for Bergamasca (N_e = 130), while Pagliarola 1 showed a small N_e of 48. For the two sub-populations of Pagliarola breed, N_e was smaller (average 34) than for all other breeds (average 94.5). Consistent with our results, Mastrangelo et al. (2017), reported a contemporary effective population size of 25 in the Barbaresca sheep, which, like Pagliarola, is an endangered population that has experienced a dramatic census contraction.

Individual genomic inbreeding have been evaluated using ROH analysis as reported in Figure 4. The highest mean value of F_{ROH} is observed in Pagliarola 2 (0.133), followed by Leccese (0.068) and Valle del Belice (0.067), while the lowest values are observed in Comisana, Bergamasca and Pagliarola 1 (0.008, 0.014 and 0.024, respectively) (Table 1). For the Italian sheep breeds, similar low values were previously reported (Mastrangelo et al. 2018), in particular for small and endangered breeds, as Altamurana and Barbaresca. The total number of ROH and the total genomic length for each animal of the two Pagliarola sub-populations are shown in Figure 5. For Pagliarola 1, the majority of the animals clustered near to the origin of coordinates with a total length <100 Mb. For the Pagliarola 2, there were individuals with a larger number of ROH (from 20 to 70) and with the total length of genome in ROH more than 500 Mb. Therefore, Pagliarola 2 showed ROH patterns typically produced by inbreeding or strong population bottlenecks. Generally, in the farming system of local sheep breeds, the farmers use rams born in the flock and the exchange of animals is quite unusual. This leads to an increase of inbreeding within the population due to the uncontrolled mating between genetically related individuals (Mastrangelo et al. 2012). Therefore, in these small populations, such as Pagliarola breed, directional mating and the use of selected rams can be considered as encouraging breeding practices to prevent higher increases of inbreeding in order to preserve genetic diversity.

For the two Pagliarola sub-populations, the genetic diversity indices suggest a very different genetic makeup. Indeed, all genetic diversity indices indicate that Pagliarola 2 sub-population could be the result of

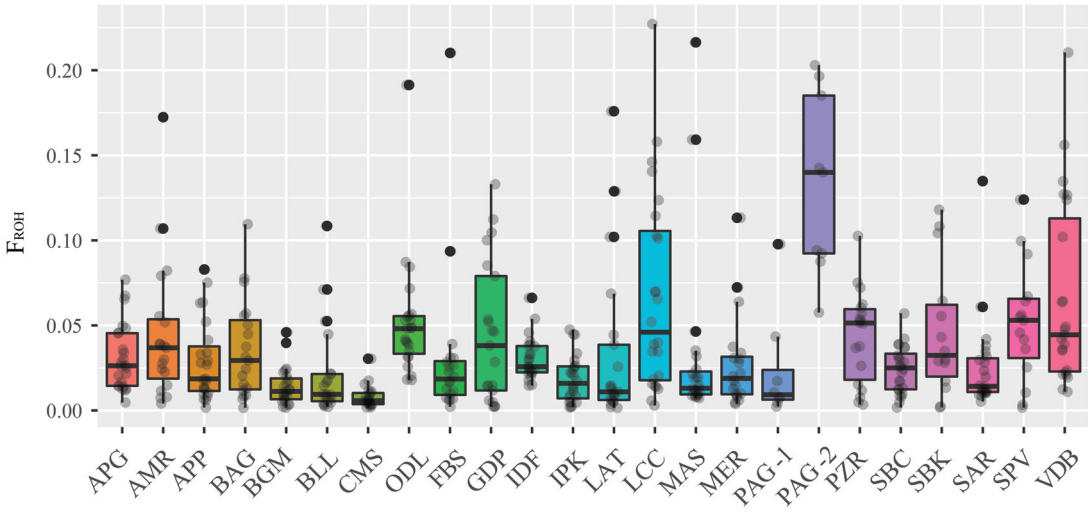


Figure 4. Genomic inbreeding values estimated from runs of homozygosity (F_{ROH}) of each breed. For full definition of breeds see Table 1.

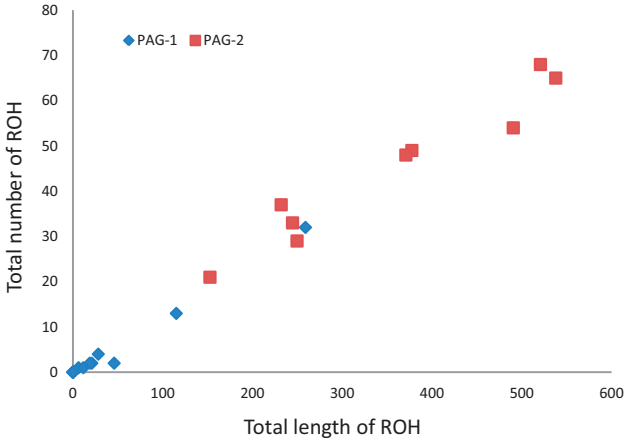


Figure 5. Relationship between the total number of runs of homozygosity (ROH) segments (y-axis) and the total length (Mb) of genome in such ROH (x-axis) for each individual of the two Pagliarola sub-populations: Pagliarola 1 (PAG-1) and Pagliarola 2 (PAG-2).

a strong genetic drift, as underlined by high level of homozygosity (F_{ROH}) and a low value of effective population size (N_e) and heterozygosity (Table 1 and Figure 4), confirming the results on the genetic structure above reported. One potential explanation for such a trend is a combination of a founder effect facilitated by a small genetic pool from which the Pagliarola 2 sub-population may have originated. Strong differences in terms of genetic drift are not an exception in livestock and have been repeatedly reported also in other species (Lawson Handley et al. 2007; Ciani et al. 2015; Nicoloso et al. 2015; Lenstra et al. 2017; Senczuk et al. 2020; Moscarelli et al. 2021). Moreover, this small genetic background could have been even amplified by possible attempts to keep the subpopulation as 'pure' as possible. Unlike Pagliarola 2, Pagliarola 1 showed the highest genetic diversity values when compared with all the other native Italian breeds, likely because of its admixed origin. Such a result is very important to underline that an urgent recovery of this population is needed in order to not lose a marked genetic potential.

Table 2. Results of the BayeScan analysis showing outlier single nucleotide polymorphisms (SNPs) identified between Pagliarola 1 and Merino-derived Italian breeds (Gentile di Puglia, Sopravissana and Merinizzata).

OAR	SNP	Position (bp)	q-value	F_{ST}	Gene
1	s58309.1	13889004	0.024	0.168	*MACF1*
1	s57436.1	28009525	0.004	0.193	*GLIS1*
1	DU342117_350.1	296260626	0.003	0.168	–
10	OAR10_61194426.1	61194426	0.005	0.181	–
15	OAR15_3688056.1	3688056	0.002	0.168	–

When markers were located within genes, these are indicated. OAR: *Ovis aries* chromosome number; SNP: single nucleotide polymorphisms.

F_{ST}-outlier analysis

The genome-wide distribution of pair-wise F_{ST} values was used to identify highly differentiated SNPs between Pagliarola 1 (considered to be the most likely original genetic stock) and the other three Merino-derived Italian breeds (GDP, SPV and MER). Results from the Bayesian approach identified a total of 5 outlier SNPs (Table 2) located on OAR01, OAR10 and OAR15. Among the significant markers, two SNPs on OAR01 mapped within know genes (*MACF1*and *GLIS1*). The *MACF1* gene plays crucial role in biological processes related to feed efficiency in livestock species (Reyer et al. 2017; Lam et al. 2018). The *GLIS1* is reported as a candidate gene for feed efficiency (Cockrum et al. 2012) and local adaption (Sweet-Jones

et al. 2021) in sheep. These results are consistent with the phenotypic characteristics of Pagliarola sheep that have excellent abilities to adapt to harsh environments and with the ability to exploit poor-quality feed compared to the other Merino-derived Italian breeds, for which instead the transhumance was carried out.

Conclusion

This study has reported, for the first time, genetic diversity and population structure estimates from a genome-wide perspective in the Pagliarola sheep. The genetic structure analyses show that Pagliarola breed own a distinct and unique genetic background composed of two different genetic groups with a common origin but characterised by a different genomic make up. Indeed, Pagliarola 2 showed a marked genomic inbreeding, low values of both heterozygosity and effective population size, suggesting that a strong genetic drift occurred in this population. On the contrary, Pagliarola 1 showed low genomic inbreeding values and high values of heterozygosity.

These results stress the importance of genetic investigations before specific conservation strategies, including a possible reintroduction of the original genetic pool, can be implemented. Indeed, basing on our outcomes we strongly encourage a reintroduction from the original genetic stock of Pagliarola 1 population. Moreover, concerning the genetic relationships with other local breed from the Apennines, although Pagliarola has been considered as one of the ancestral populations from which Appenninica breed derived, we found a closer relationship with Merino-derived breeds. Such a genetic relatedness might be the outcome of the close cultural connection related to transhumance. Indeed, despite Pagliarola breeds was not an active transhumant breed was part of the same herds during the non-transhumant periods. Finally, the BayeScan results demonstrate the importance to preserve the Pagliarola genome as genetic reservoir of alleles particularly adapted to low-input farming system. The information generated in this study strongly suggests that proper conservation measures should be implemented in order to recover the Pagliarola population and its fundamental local products.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical approval

All procedures involving animal sample collection followed the recommendation of directive 2010/63/EU. Sampling was carried out by trained veterinarians within the frame of vaccination campaigns, hence no permission from the animal research ethics committee was necessary. Veterinarians adhered to standard procedures and relevant national guidelines to ensure appropriate animal care.

Funding

PL laboratory is partially funded by the Project 'Demetra' (Dipartimenti di Eccellenza 2018–2022, CUP_C46C1800 0530001, MIUR).

ORCID

Paolo Ajmone Marsan http://orcid.org/0000-0003-3165-4579
Fabio Pilla http://orcid.org/0000-0002-1781-994X
Salvatore Mastrangelo http://orcid.org/0000-0001-6511-1981

Data availability statement

The data that support the findings of this study are available upon request from the corresponding author.

References

Ajmone-Marsan P. 2010. A global view of livestock biodiversity and conservation-GLOBALDIV. Anim Genet. 41(Suppl 1):1–5.

Ajmone-Marsan P, Colli L, Han JL, Achilli A, Lancioni H, Joost S, Crepaldi P, Pilla F, Stella A, Taberlet P, et al. 2014. The characterization of goat genetic diversity: towards a genomic approach. Small Rumin Res. 121(1):58–72.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinf. 12(1):1–6.

Anagnostou P, Capocasa M, Dominici V, Montinaro F, Coia V, Destro-Bisol G. 2017. Evaluating mtDNA patterns of genetic isolation using a re-sampling procedure: a case study on Italian populations. Ann Hum Biol. 44(2):140–148.

Anzalone DA, Palazzese L, Iuso D, Martino G, Loi P. 2018. Freeze-dried spermatozoa: an alternative biobanking option for endangered species. Anim Reprod Sci. 190: 85–93.

Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet. 6:109.

Bedhiaf-Romdhani S, Baazaoui I, Ciani E, Mastrangelo S, Sassi MB. 2020. Genetic structure of Tunisian sheep breeds as inferred from genome-wide SNP markers. Small Rumin Res. 191:106192.

Beynon SE, Slavov GT, Farré M, Sunduimijid B, Waddams K, Davies B, Haresign W, Kijas J, MacLeod IM, Newbold CJ,

et al. 2015. Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. BMC Genet. 16(1):65.

Bigi D, Zanon A. 2008. Atlante delle razze autoctone. Bov Equini Ovicaprini Suini Allev Ital Edagricole.

Biscarini F, Nicolazzi EL, Stella A, Boettcher PJ, Gandini G. 2015. Challenges and opportunities in genetic improvement of local livestock breeds. Front Genet. 6:33.

Bittante G. 2011. Italian animal genetic resources in the Domestic Animal Diversity Information System of FAO. Ital J Anim Sci. 10(2):e29.

Bonadonna T. 1976. Etnologia zootecnica. Torino: UTET.

Brisighelli F, Blanco-Verea A, Boschi I, Garagnani P, Pascali VL, Carracedo A, Capelli C, Salas A. 2012. Patterns of Y-STR variation in Italy. Forensic Sci Int Genet. 6(6):834–839.

Ceccobelli S, Lorenzo PD, Panella F, Lasagna E, Sarti FM. 2016. Morphological and genetic characterisation of Pagliarola breed and its genetic relationships with other three indigenous Italian sheep breeds. Ital J Anim Sci. 15(1):47–54.

Ciani E, Ciampolini R, D'Andrea M, Castellana E, Cecchi F, Incoronato C, d'Angelo F, Albenzio M, Pilla F, Matassino D, et al. 2013. Analysis of genetic variability within and among Italian sheep breeds reveals population stratification and suggests the presence of a phylogeographic gradient. Small Rumin Res. 112(1-3):21–27.

Ciani E, Crepaldi P, Nicoloso L, Lasagna E, Sarti FM, Moioli B, Napolitano F, Carta A, Usai G, D'Andrea M, et al. 2014. Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. Anim Genet. 45(2):256–266.

Ciani E, Lasagna E, D'Andrea M, Alloggio I, Marroni F, Ceccobelli S, Delgado Bermejo JV, Sarti FM, Kijas J, Lenstra JA, et al. 2015. Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. Genet Sel Evol. 47(1):64.

Cockrum RR, Pickering NK, Anderson RM, Hyndman DL, Bixley MJ, Dodds KG, Stobart RH, McEwan JC, Cammack KM. 2012. Identification of single nucleotide polymorphisms associated with feed efficiency in rams. In: Conference: Western section American society of Animal Science [place unknown]; p. 79.

Fabrizio P. 1928. La pecora del Molise [Sheep of Molise]. Italia Agricola. 1965(6):283–295.

Fitak RR. 2021. optM: an R package to optimize the number of migration edges using threshold models. (in preparation)

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.

Gaouar SB, Lafri M, Djaout A, El-Bouyahiaoui R, Bouri A, Bouchatal A, Maftah A, Ciani E, Da Silva AB. 2017. Genome-wide analysis highlights genetic dilution in Algerian sheep. Heredity. 118 (3):293–301.

Hoffmann I. 2010. Climate change and the characterization, breeding and conservation of animal genetic resources. Anim Genet. 41(s1):32–46.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LRP, Cristobal MS, Servin B, McCulloch R, Whan V, Gietzen K, et al. 2012. Genome-Wide Analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10(2):e1001258.

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, et al. 2009. A Genome Wide Survey of SNP variation reveals the genetic structure of sheep breeds. PLOS One. 4(3):e4668.

Lam S, Miglior F, Fonseca P, Seymour D, Asselstine V, Brito L, Schenkel F, Cánovas A. 2018. Identification of variants associated with divergent feed efficiency groups using multiple RNA-sequencing datasets from dairy and beef cattle.In: Proc World Congr Genet Appl to Livest Prod. Vol. Electronic. Auckland, New Zealand; p. 773.

Lawson Handley L-J, Byrne K, Santucci F, Townsend S, Taylor M, Bruford MW, Hewitt GM. 2007. Genetic structure of European sheep breeds. Heredity. 99(6):620–631.

Lenstra JA, Tigchelaar J, Biebach I, Hallsson JH, Kantanen J, Nielsen VH, Pompanon F, Naderi S, Rezaei H-R, Saether N, et al. 2017. Microsatellite diversity of the Nordic type of goats in relation to breed conservation: how relevant is pure ancestry? J Anim Breed Genet. 134(1):78–84.

Marchi E, Mascheroni E. 1925. Zootecnia speciale. Torino: UTET.

Mastrangelo S, Ciani E, Sardina MT, Sottile G, Pilla F, Portolano B. 2018. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Anim Genet. 49(1):71–81.

Mastrangelo S, Di Gerlando R, Tolone M, Tortorici L, Sardina MT, Portolano B. 2014. Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. BMC Genet. 15(1):108.

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

Mastrangelo S, Sardina MT, Riggio V, Portolano B. 2012. Study of polymorphisms in the promoter region of ovine β-lactoglobulin gene and phylogenetic analysis among the Valle del Belice breed and other sheep breeds considered as ancestors. Mol Biol Rep. 39(1):745–751.

Milanesi M, Capomaccio S, Vajana E, Bomba L, Garcia JF, Ajmone-Marsan P, Colli L. 2017. BITE: an R package for biodiversity analyses. bioRxiv.:181610.

Morbidini L, Panella F, Pollidori P, Sarti DM, Sarti FM. 1995. Verifica delle caratteristiche riportate nello standard della razza Merinizzata Italiana da carne, [A check of traits carried in the standard of Italian Merinos breed [sheep]]. Allev Ovini E Caprini Italy.

Moscarelli A, Sardina MT, Cassandro M, Ciani E, Pilla F, Senczuk G, Portolano B, Mastrangelo S. 2021. Genome-wide assessment of diversity and differentiation between original and modern Brown cattle populations. Anim Genet. 52(1):21–31.

Nicoloso L, Bomba L, Colli L, Negrini R, Milanesi M, Mazza R, Sechi T, Frattini S, Talenti A, Coizet B, et al. 2015. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. Genet Sel Evol. 47:62.

Pickrell J, Pritchard J. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8(11):1002967.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al.

2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559–575.

Renieri C, Silvestrelli M, Casciotti D, Ronchi B. (Perugia U (Italy) I di PA). 1984. La popolazione ovina Pagliarola: studio demografico, [The 'Pagliarola' sheep population. Demographic study]. Vergaro Italy. 9.

Reyer H, Shirali M, Ponsuksili S, Murani E, Varley PF, Jensen J, Wimmers K. 2017. Exploring the genetics of feed efficiency and feeding behaviour traits in a pig line highly selected for performance characteristics. Mol Genet Genomics. 292(5):1001–1011.

Rochus CM, Tortereau F, Plisson-Petit F, Restoux G, Moreno-Romieux C, Tosser-Klopp G, Servin B. 2018. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. BMC Genomics. 19(1):71.

Sarti DM. 2002. L'Appenninica. Passato – presente – futuro [ovini], [The Appenninica. Past – present – future [sheep]]. Allev Ovini E Caprini Italy. 19(3):4.

Sarti DM, Panella F, Pauselli M, Sarti FM, Lasagna E. 2002. L'allevamento degli ovini, [Sheep farming]. Il Divulg Agenzia Territ Sostenibilita Aliment Agro-Ambient Ed Energ Ed Bologna Italy.

Sarti FM, Panella F. 1999. Study of the most suitable traits to use in linear morphological evaluation of Merinizzata Italiana sheep. Proceeding 13th National Congress ASPA Piacenza Italy. p. 206–208.

Senczuk G, Mastrangelo S, Ajmone-Marsan P, Becskei Z, Colangelo P, Colli Ferretti L, Karsli T, Lancioni H, Lasagna E, Marletta D, et al. 2021. On the origin and diversification of Podolian cattle breeds: testing scenarios of European colonization using genome-wide SNP data . Genet Sel Evol. 53(1):48.

Senczuk G, Mastrangelo S, Ciani E, Battaglini L, Cendron F, Ciampolini R, Crepaldi P, Mantovani R, Bongioni G, Pagnacco G, et al. 2020. The genetic heritage of Alpine local cattle breeds using genomic SNP data. Genet Sel Evol. 52(1):40.

Sweet-Jones J, Yurchenko AA, Igoshin AV, Yudin NS, Swain MT, Larkin DM. 2021. Resequencing and signatures of selection scan in two Siberian native sheep breeds point to candidate genetic variants for adaptation and economically important traits. Anim Genet. 52(1):126–131.

Wickham H. 2011. ggplot2. WIREs Comp Stat. 3(2):180–185.

J. Dairy Sci. 106:5537–5553
https://doi.org/10.3168/jds.2022-22159
© 2023, The Authors. Published by Elsevier Inc. and Fass Inc. on behalf of the American Dairy Science Association®.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Exploring genome-wide differentiation and signatures of selection in Italian and North American Holstein populations

Christian Persichilli,[1] Gabriele Senczuk,[1]* Salvatore Mastrangelo,[2] Maurizio Marusi,[3] Jan-Thijs van Kaam,[3] Raffaella Finocchiaro,[3] Marika Di Civita,[1] Martino Cassandro,[3,4] and Fabio Pilla[1]
[1]Department of Agricultural, Environmental and Food Sciences, University of Molise, Via A. De sanctis, 86100 Campobasso (CB), Italy
[2]Department of Agricultural, Food and Forest Sciences, University of Palermo, Viale delle Scienze, 90128 Palermo (PA), Italy
[3]National Association of Italian Holstein, Brown and Jersey Breeders, Via Bergamo, 292, 26100 Cremona (CR), Italy
[4]Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy

ABSTRACT

Among Italian dairy cattle, the Holstein is the most reared breed for the production of Parmigiano Reggiano protected designation of origin cheese, which represents one of the most renowned products in the entire Italian dairy industry. In this work, we used a medium-density genome-wide data set consisting of 79,464 imputed SNPs to study the genetic structure of Italian Holstein breed, including the population reared in the area of Parmigiano Reggiano cheese production, and assessing its distinctiveness from the North American population. Multidimensional scaling and ADMIXTURE approaches were used to explore the genetic structure among populations. We also investigated putative genomic regions under selection among these 3 populations by combining 4 different statistical methods based either on allele frequencies (single marker and window-based) or extended haplotype homozygosity (EHH; standardized log-ratio of integrated EHH and cross-population EHH). The genetic structure results allowed us to clearly distinguish the 3 Holstein populations; however, the most remarkable difference was observed between Italian and North American stock. Selection signature analyses identified several significant SNPs falling within or closer to genes with known roles in several traits such as milk quality, resistance to disease, and fertility. In particular, a total of 22 genes related to milk production have been identified using the 2 allele frequency approaches. Among these, a convergent signal has been found in the *VPS8* gene which resulted to be involved in milk traits, whereas other genes (*CYP7B1*, *KSR2*, *C4A*, *LIPE*, *DCDC1*, *GPR20*, and *ST3GAL1*) resulted to be associated with quantitative trait loci related to milk yield and composition in terms of fat and protein percentage. In contrast, a total of 7 genomic regions were identified combining the results of standardized log-ratio of integrated EHH and cross-population EHH. In these regions candidate genes for milk traits were also identified. Moreover, this was also confirmed by the enrichment analyses in which we found that the majority of the significantly enriched quantitative trait loci were linked to milk traits, whereas the gene ontology and pathway enrichment analysis pointed to molecular functions and biological processes involved in AA transmembrane transport and methane metabolism pathway. This study provides information on the genetic structure of the examined populations, showing that they are distinguishable from each other. Furthermore, the selection signature analyses can be considered as a starting point for future studies in the identification of causal mutations and consequent implementation of more practical application.
Key words: cattle, Holstein, genetic structure, selection signature, candidate genes

INTRODUCTION

With a total population size exceeding 65 million, Holstein cattle are the most widespread dairy breed in the world (FAO, 2018). Their origin can be traced back to the Dutch provinces of North Holland and Friesland, and Schleswig-Holstein in Northern Germany where they have had a long history of dual-purpose breeding. During the 1870s and 1880s it was exported to the United States where it was strongly selected to increase milk yields (Theunissen, 2012). Along with the growing milk demand occurred during the twentieth century, the North American stock was reused to improve milk production in many European countries (Cassandro, 2014; Hulsegge et al., 2022). Although on the one hand this has led to a significant global increase in milk productivity, on the other, recent genetic studies have highlighted a worrying reduction in the

Received April 5, 2022.
Accepted February 7, 2023.
*Corresponding author: g.senczuk@unimol.it

genetic variability of many traditional Holstein populations because of intense directional selection (Doekes et al., 2018; Makanjuola et al., 2020). Contrariwise, the maintenance of a good level of genetic diversity represents an important challenge also in highly productive systems especially when considering the ever-changing market requirements.

The Italian Friesian population was historically documented since around 1870 (Balasini, 2002), however, during the first half of the 20th century, the North American stock was gradually introduced especially in northern regions to improve the dairy industry (Porter et al., 2016). The Italian Holstein herd-book was established in 1956 by the National Association of Italian Friesian Cattle Breeders. The National Association of Italian Friesian Cattle Breeders, today the Italian Holstein, Brown and Jersey National Breeders Association, still holds the herd-book and establishes the selection objectives (ANAFIBJ, 2021a). Currently, among Italian dairy cattle, the Italian Holstein is the most reared breed for the production of Parmigiano Reggiano protected designation of origin cheese. Consequently, in recent times, the use of specific breeding values such as the "Production, Functionality and Type index," the "Economic - Functional Index," and the "Cheesemaking and Sustainability Index – Parmigiano Reggiano" (ANAFIBJ, 2021b), together with specific genotypes such as the homozygosity of the Bκ-CN allele (Chessa et al., 2020), has become an essential tool for breeders. Although local farmers have been selecting for casein and protein percentage for at least the last 2 decades, only since 2016 has the National Association defined the cheesemaking and sustainability index—Parmigiano Reggiano, which is specifically thought for the herds that produce milk that will be sold to Parmigiano Reggiano producers (Visentin et al., 2018; Marusi et al., 2019). As far as we are aware, this may have an impact on the genetic diversity of the Holstein breed because distinct selection criteria used within the same breed may affect its population structure (Andersson and Georges, 2004; Ablondi et al., 2021). A recent high-resolution representation of the genome-wide diversity and population structure of Italian local cattle breeds, including Holstein Friesian was provided, using a medium-density SNP array by Mastrangelo et al. (2018). Moreover, despite the intense genetic exchange of genetic material between Italy and the rest of the world (notably the United States), the specific environmental conditions and breeding system could have resulted in a genetic differentiation between the Italian and North American populations.

Aims of this study are to explore the genetic structure of the Italian Holstein assessing its distinctiveness from the North American population, and to identify whether the Italian Holstein population used to produce Parmigiano Reggiano is also distinguishable from other Italian Holstein not used for such a production. In doing so, we also performed a selection signature analysis by combining different statistical methods based either on allele frequencies ($\mathbf{F_{ST}}$) or extended haplotype homozygosity [**EHH**; standardized log-ratio of integrated EHH (**Rsb**) and cross-population EHH (**XP-EHH**)]. The incorporation of different methods is crucial to investigate selection signatures at different time resolutions. Indeed, even though F_{ST} methods are best suited to detect more ancient evidence of selection, the EHH-based methods are more suited to detect very recent signals of selection. Definitely, here we show that the use of multiple approaches is an essential element for the identification of selection signatures involved in phenotypic variation important to better understand the evolutionary processes that underlie traits that have been exposed to both natural and artificial selection (Saravanan et al., 2020).

MATERIALS AND METHODS

No animals were used in this study, and ethical approval for the use of animals was thus deemed unnecessary.

Source of the Genotypes and Quality Control

Based on pedigree and herd-book information, 280 samples of Holstein cattle breed were selected, taking into account their kinship to have pure representatives of each population and avoiding close kinship among samples. According to this criterion 3 groups have been generated: one including only individuals from the United States and Canada (**acHOL**), a second including Italian Holstein samples not used to produce Parmigiano Reggiano (**itHOL**), and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano (**prHOL**). Genotype data come from the genomic database of the Italian Holstein, Brown and Jersey National Breeders Association, including Italian genotyped animals, genotypes from the Intercontinental consortium (the United States, Canada, Italy, the United Kingdom, Switzerland), and genotypes provided by other parties.

To harmonize the data set, as they come from different SNPs panels, we performed an imputation approach using the software Pedimpute (Nicolazzi et al., 2013). We obtained 86,841 imputed SNPs that were then mapped against the ARS-UCD1.2 reference genome (Rosen et al., 2020). The software PLINK (Purcell et al., 2007;

Chang et al., 2015) was used to filter for minor allele frequency (maf 0.05) and missing genotype call rate (geno 0.01) obtaining a final data set of 79,464 SNPs.

Analysis of Genetic Diversity and Population Structure

For each population, basic genetic diversity indices including average of the observed and expected heterozygosity and minor allele frequency were calculated with PLINK (Purcell et al., 2007; Chang et al., 2015). The inbreeding coefficient based on runs of homozygosity was estimated using the software PLINK with parameters described in Moscarelli et al. (2021), whereas trends in effective population size were estimated using the software GONe (Santiago et al., 2020). This approach allows inferring a reliable demographic history of contemporary population within the past 100 generations, given the extent of the observed linkage disequilibrium (**LD**) between SNPs. Due to the likely presence of rather recent migrants among populations, the recombination rate was set to 0.01 while other parameters have been left as default.

To explore the genetic variation within and among groups, after calculating pairwise identical-by-state distances using the software PLINK (–cluster mds-plot 3), a multidimensional scaling plot was generated using the R package ggplot2 (Wickham, 2011). In addition, a maximum-likelihood clustering approach, as implemented in the software ADMIXTURE (Alexander and Lange, 2011) with K values (given ancestral populations) ranging from 2 to 5, has been used to assess population stratification. The results have been plotted and ordered following the sample grouping using the *membercoef.plot* function in the R package BITE (Milanesi et al., 2017). Using the membercoef.cv function from the same package, the best values of K have been also assessed.

Selection Signature Through Genome-Wide Allele Frequency Approaches

To identify possible genomic regions under selection among the 3 Holstein groups, 2 different F_{ST} approaches were used: (i) a genome-wide F_{ST} outlier method using the software BayeScan (Foll and Gaggiotti, 2008) and (ii) a genome window-based analysis using weighted F_{ST} average according to Weir and Cockerham (1984) method.

For both methods, the choice to select a reasonable region also putatively under selection was based by screening the extent of LD among markers by performing a LD decay plot and assessing when the r^2 reaches a stationary (around 120 kb, Supplemental Figure S1; https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023).

For the single locus pairwise F_{ST} outlier analysis we retained SNPs based on the top 0.05% F_{ST} value (99.95th percentile). We then considered a region of ± 60 kb around the significant SNPs to retain genes and explore their function.

For the F_{ST} window-based approach, after retaining windows with at least 4 SNPs, the top 0.05% F_{ST} (99.95th percentile) windows were considered as significant. According to the LD decay we screened across windows of 120 kb with a step of 60 kb using the software VCFtools (Danecek et al., 2011). For all comparisons, Manhattan plots have been generated using the *manhattan* function of the R package qqplot (Turner, 2018).

Information on the annotated genes within windows and SNPs were obtained from the Genome Data Viewer using the ARS-UCD1.2 as reference genome.

Selection Signature Through Extended Haplotype Homozygosity

Two additional methods based on XP-EHH (Sabeti et al., 2002) and Rsb (Tang et al., 2007) were also used. Both approaches are based on (site-specific) EHH and relative integrals (iES and inES) between populations that in turn provide reliable pictures of the ongoing selection acting on chromosomes even in the absence of a polarization between ancestral and derived alleles (Klassmann and Gautier, 2022).

To conduct these analyses the phased haplotypes obtained using the software Pedimpute (Nicolazzi et al., 2013) were imported in the *rehh* package (Gautier and Vitalis, 2012) to calculate the haplotype related statistics.

For both Rsb and XP-EHH approaches, candidate regions were selected as those containing at least 4 significant SNPs within 1-Mb sliding windows, overlapping 10 kb, and retaining all SNPs exceeding the threshold of $-\log10$ (P-value) = 2.5.

Finally, all annotated genes within a region were retrieved from the Genome Data Viewer using the ARS-UCD1.2 as reference genome.

Quantitative Trait Loci and Functional Enrichment Analysis

For both F_{ST} and EHH analyses, we used the R package *Genomic Annotation in Livestock for positional candidate loci* (*GALLO*; Fonseca et al., 2020), to perform an enrichment analysis. The QTL annotation file (*.gff*) was imported from Animal QTL Database (Cattle QTLdb, 2021; Hu et al., 2013, 2022), using

Table 1. Genetic diversity indices estimated for the 3 Holstein cattle populations: one including only individuals from the United States and Canada (acHOL), a second including Italian Holstein samples not used to produce Parmigiano Reggiano (itHOL), and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano (prHOL)

Population	Code	No.	Ho ± SD	He ± SD	F_{ROH}[1]	Ne[2]
Italian Holstein	itHOL	74	0.373 ± 0.160	0.354 ± 0.144	0.102	43.39
Parmigiano Reggiano Holstein	prHOL	105	0.363 ± 0.138	0.368 ± 0.135	0.125	89.05
North American Holstein	acHOL	101	0.342 ± 0.171	0.332 ± 0.160	0.168	39.97

[1]F_{ROH} = genomic inbreeding coefficient (F) derived from runs of homozygosity (ROH).
[2]Ne = effective population size.

ARS_UCD1.2 as reference genome. Concerning EHH regions, all the subsequent analyses were conducted only on the shared haplotypes detected by the 2 approaches (XP-EHH and Rsb). The proportion of QTL categories was calculated and plotted in pie charts using the function *plot_qtl_info*. Subsequently, enrichment analysis was conducted using the chromosome-based method on both the significant F_{ST} markers (taking 60 kb windows upstream and downstream) and within the selected windows-based and EHH regions. To compare the observed and the expected number of QTL for each trait, a bootstrap approach through 1,000 iterations of random sampling was used. Adjusted P-values (***Padj*-value**) were calculated based on the Bonferroni correction. Only chromosome-enriched traits with a *Padj*-value below the significance threshold of 0.05 were plotted.

Finally, with the aim of intercepting interactions between genes, the software ToppGene (Chen et al., 2009) was used to explore potential biological functions and pathways. As a gene list we included all genes falling within significantly enriched QTL as revealed by GALLO, either considering the F_{ST}-based or EHH-based approaches. The table containing the training set used for the gene prioritization analysis is available in Supplemental Table S1 (https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023). For the Gene Ontology (**GO**) terms and pathways enrichment analysis, we used a random sampling size of 5,000 and a minimum feature count of 2 as test parameters. The analysis was performed for each comparison separately, selecting GO terms and Kyoto Encyclopedia of Genes and Genomes pathways based on P-values adjusted for Bonferroni (*Padj*-value < 0.05).

RESULTS

Analysis of Genetic Relationships and Population Structure

Genetic diversity indices for the 3 Holstein populations are reported in Table 1. The itHOL showed the highest level of observed heterozygosity and the lowest level of inbreeding, the acHOL showed an opposite pattern, whereas the prHOL showed intermediate values. Concerning contemporary effective population size (***Ne***), the prHOL population showed the highest value whereas the acHOL showed the lowest (Table 1). Trends in population size as shown by exploring *Ne* over the last 100 generations indicated a similar pattern with a general decrease of all populations starting between 25 and 20 generations ago. However, the Italian Holstein (itHOL and prHOL) showed higher values of *Ne* compared with the North American Holstein (acHOL; Supplemental Figure S2; https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023).

The multidimensional scaling plot is reported in Figure 1. The first 2 dimensions cumulatively account for 14.18% of the total variance. The results show separation among the 3 Holstein populations; however, a slight overlap can be also observed among them.

Analogous results are given by the ADMIXTURE analysis which showed the most reliable repartition at value of K = 3 (Figure 2). Indeed, at this genetic structuring the acHOL group is clearly distinguishable from all the Italian Holsteins showing a predominance of the red component. However, 2 main groups corresponding to the itHOL and prHOL populations also showed a marked genetic distinctiveness. In particular, the itHOL showed a dominance of the blue component whereas in the prHOL group we observed a higher diversity in terms of genetic components, although the yellow one results the more frequent.

Selection Signature Through Genome-Wide Allele Frequency Approaches

Among the 3 sets of comparisons, a total of 120 outlier SNPs were detected using the single marker Bayesian approach whereas 140 SNPs were identified using the F_{ST} window-based approach (Figure 3). The list of all significant SNPs is reported in Supplemental Table S2 (https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023). Among these, we found 15 SNPs encompassing 6 chromosomes (BTA1, BTA3, BTA14, BTA16, BTA18, and BTA20) that were coincident

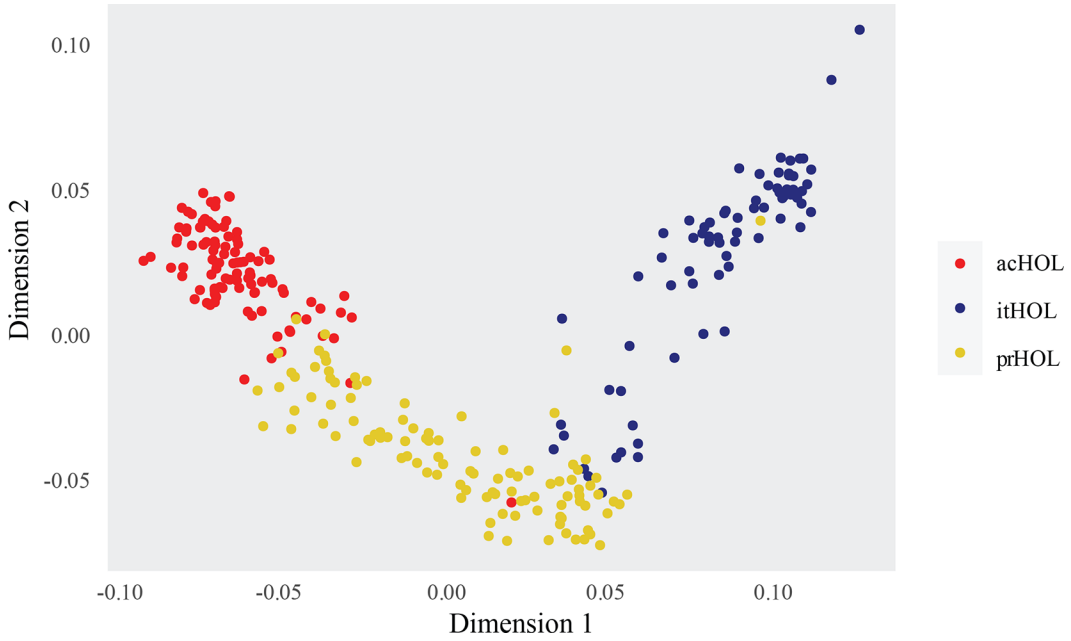


Figure 1. Multidimensional scaling plot showing the genetic relationships of the analyzed Holstein populations as reported in Table 1. acHOL = only individuals from the United States and Canada; itHOL = Italian Holstein samples not used to produce Parmigiano Reggiano; prHOL = only Italian Holstein cattle used to produce Parmigiano Reggiano..

between the 2 approaches. Within these candidate regions, 11 annotated genes were located (*VPS8*, *P2RY1*, *BTG2*, *CHIT1*, *MYBPH*, *CHI3L1*, *PPP1R15B*, *PIK-3C2B*, *ZNF404*, *ZNF45*, and *ZNF226*).

In the BayeScan analysis the F_{ST} values for the top 0.05% SNP ranged from 0.260 to 0.335, from 0.125 to 0.183, and from 0.110 to 0.132 in the itHOL versus acHOL, prHOL versus acHOL, and prHOL versus itHOL, respectively. For each comparison we found 40 top markers (0.05% F_{ST}) encompassing 25, 27, and 43 annotated genes, located in 12, 9, and 8 autosomes, in the comparisons itHOL versus acHOL, prHOL versus acHOL, and prHOL versus itHOL, respectively (Table 2). Interestingly, 5 genes (*UNC13C*, *MAPK6*, *LEO1*, *ZFPM1*, *KCNMA1*) were found close to significant markers in 2 comparisons (itHOL vs. acHOL and prHOL vs. acHOL), whereas a single gene (*C12H13orf42*) was found close to a significant SNP in the comparison between prHOL and acHOL.

Considering the F_{ST} window-based approach, we found 12 significant windows for each comparison (Supplemental Table S2). However, several windows were partially overlapped and were therefore merged (Table 3). In addition, only 7, 6, and 5 windows contained annotated genes in the itHOL versus acHOL, prHOL versus acHOL, and prHOL versus itHOL comparisons, respectively (Table 3). In the comparison between the itHOL and acHOL, prHOL and acHOL, and prHOL and itHOL, 9, 8, and 15 annotated genes located in 7, 5, and 4 autosomes were found. Also, in this case, overlapping significant SNPs in the comparisons itHOL versus acHOL and prHOL versus acHOL mapped within known genes (*VPS8* and *FER*; Table 3). In addition, within these 2 comparisons, the genes *VPS8*, *FER*, and *GRIK3* were detected in more than one significant window.

Selection Signature Through Extended Haplotype Homozygosity

The 2 EHH-based analyses between populations revealed a total of 29 candidate regions putatively under selection (Figure 4). Among those, 11 were obtained using the XP-EHH whereas 18 were obtained using the Rsb approach, respectively (Supplemental Table S2). We found a total of 7 overlapping genomic regions between the 2 EHH-based analyses, specifically 1 in the itHOL versus acHOL comparison, 3 in the prHOL versus acHOL comparison, and 3 in the prHOL versus itHOL comparison (Table 3).

Within the overlapping candidate regions identified by the 2 EHH-based analyses, a total of 266 annotated genes were found (Table 3). Among those, 52, 47, and 187 annotated genes were found within the itHOL versus acHOL, the itHOL versus prHOL, and the prHOL versus acHOL comparisons, respectively. In contrast, just a single genomic region located on the BTA10 was found in both the itHOL versus acHOL and the itHOL

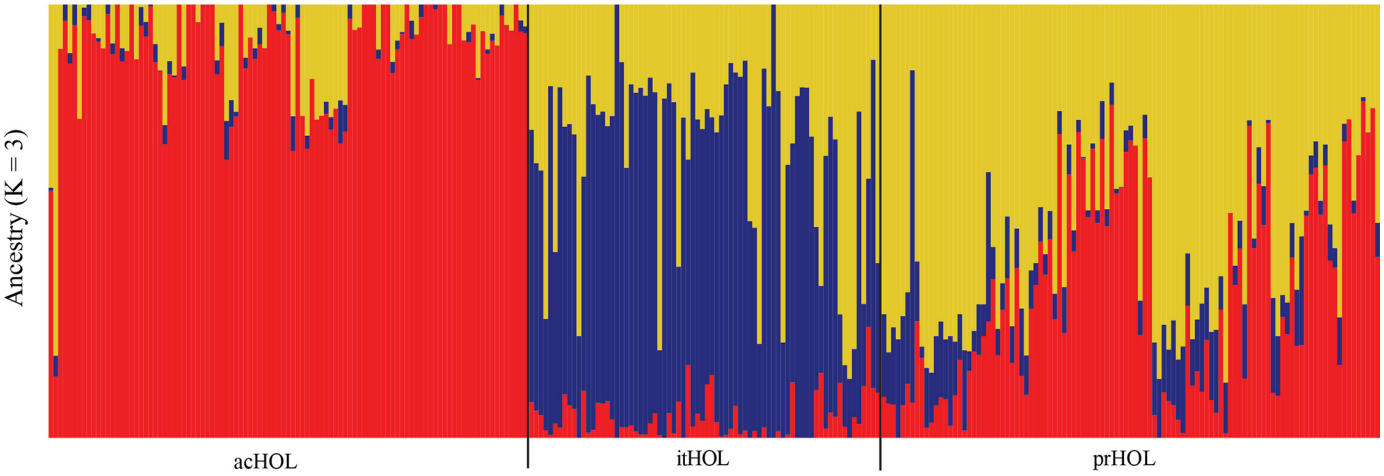


Figure 2. Bar plot of the ADMIXTURE analysis considering the most reliable number of given ancestral populations (K values; K = 3). acHOL = only individuals from the United States and Canada; itHOL = Italian Holstein samples not used to produce Parmigiano Reggiano; prHOL = only Italian Holstein cattle used to produce Parmigiano Reggiano.

versus prHOL comparisons. In this region, we found 20 overlapping annotated genes (*MGAT2*, *DNAAF2*, *MIR6517*, *POLE2*, *KLHDC1*, *KLHDC2*, *NEMF*, *ARF6*, *VCPKMT*, *SOS2*, *L2HGDH*, *ATP5S*, *CDKL1*, *MAP4K5*, *ATL1*, *SAV1*, *NIN*, *ABHD12B*, *PYGL*, and *TRIM9*).

Quantitative Trait Loci and Functional Enrichment Analysis

The proportion of the QTL categories found within the candidate genomic regions identified with BayeScan, F_{ST} window-based, and EHH-based approaches are reported in Supplemental Figures S3, S4, and S5, respectively (https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023). Both F_{ST} window-based and EHH-based approaches showed on average the higher proportion of QTL related to milk, whereas the Bayesian method the lowest. These results are also in line with the QTL enrichment analysis which showed the majority of significant milk-related QTL in both the F_{ST} windows-based and the EHH-based approaches. Indeed, for these 2 methods, most of the significantly enriched QTL (*Padj*-value < 0.05) were related to milk type.

In the F_{ST} windows-based approach, among the significantly enriched milk type QTL (all detected with *Padj*-value < 0.05) we found the glycosylated κ-CN percentage within the selection signal located on BTA7 detected in the itHOL versus acHOL comparison, the κ-CN percentage and milk fat percentage within the selection signals located on BTA3 and BTA14 in the prHOL versus acHOL comparison, and the somatic cell score and 305-d milk yield within the selection signal on BTA18 in the itHOL versus prHOL comparison (Supplemental Figure S4).

Concerning the EHH-based approach, among the significantly enriched milk type QTL (all detected with *Padj*-value < 0.05) we found milk protein percentage, milk fat content, and milk caproic acid content on BTA10 in the itHOL versus acHOL comparison, milk linoleic acid content and milk casein percentage within the selection signals located on BTA19 and BTA18 in the prHOL versus acHOL comparison, and milk phosphorus content and milk α-S1-CN percentage within the selection signal on BTA7 in the itHOL versus prHOL comparison (Supplemental Figure S5).

Finally, in the Bayesian method, the significantly enriched QTL were related to health, reproduction, and meat and carcass while no significantly enriched milk type QTL were found (Supplemental Figure S3).

Most of the significantly enriched GO terms were found in the comparison itHOL versus prHOL and included 10 GO molecular functions and 5 GO biological processes, all linked to AA transmembrane transporter activities (Supplemental Table S3; https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023). However, just 2 GO terms (GO:0008157 and GO:0005601) were found significantly enriched in the itHOL versus acHOL and prHOL versus acHOL, respectively.

Considering pathways, we only found one significantly enriched Kyoto Encyclopedia of Genes and Genomes term (MAP00680) involved in *methane metabolism* in the comparison between itHOL and prHOL.

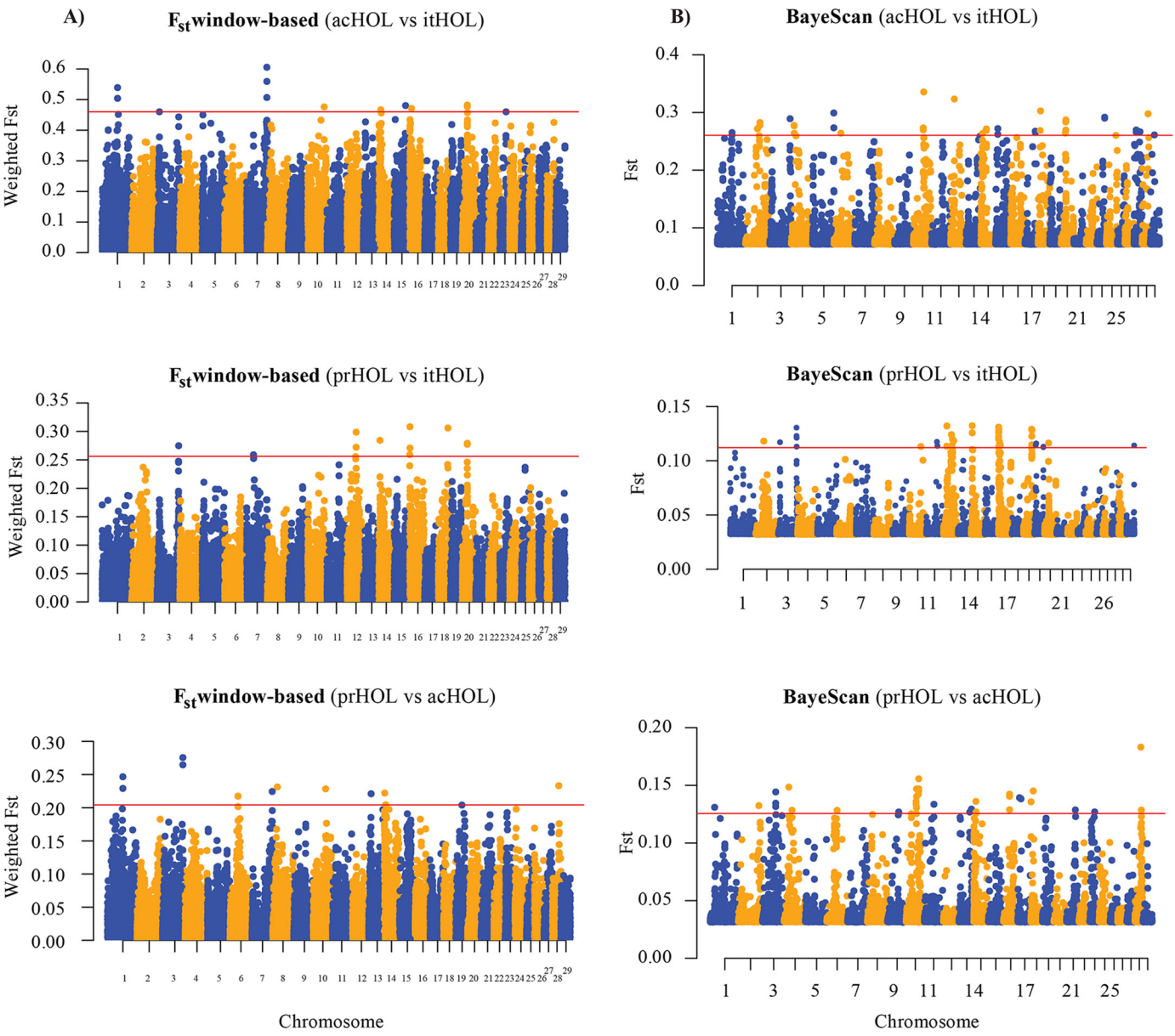


Figure 3. Manhattan plot of statistical method based on allele frequencies (F_{ST}) calculated with an F_{ST} window-based approach (A) between acHOL and itHOL, prHOL and itHOL, and prHOL and acHOL, or the Bayesian outlier approach (B) between acHOL and itHOL, prHOL and itHOL, and prHOL and acHOL. acHOL = only individuals from the United States and Canada; itHOL = Italian Holstein samples not used to produce Parmigiano Reggiano; prHOL = only Italian Holstein cattle used to produce Parmigiano Reggiano. The horizontal red line indicates the threshold based on the 0.05% F_{ST} (99.95th percentile).

DISCUSSION

Maintaining a good level of genetic variability represents an important challenge in highly productive systems where artificial selection for particular traits have markedly affected the genomic variability of many cosmopolitan breeds (Ma et al., 2019; Ablondi et al., 2022). The availability of SNP arrays has greatly improved the power of genome-wide studies for a deep investigation of genetic diversity, allowing the identification of genomic regions under selection in cattle (e.g., Qanbari et al., 2010; Zhao et al., 2015; Ben-Jemaa et al., 2020; Mastrangelo et al., 2020; Abbasi-Moshaii et al., 2023).

Our study shows that both the regular Italian Holstein and Italian Holstein used to produce Parmigiano Reggiano are clearly differentiated from the North American Holstein. In particular, we found that Italian Holstein used for the production of Parmigiano Reggiano seems to retain higher genetic diversity in terms

Table 2. List of significant SNPs detected using the BayeScan analysis for the 3 Holstein cattle populations: one including only individuals from the United States and Canada (acHOL), a second including Italian Holstein samples not used to produce Parmigiano Reggiano (itHOL), and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano (prHOL)[1]

itHOL vs. acHOL			prHOL vs. acHOL			prHOL vs. itHOL		
BTA	SNP	Gene	BTA	SNP	Gene	BTA	SNP	Gene
1	BTA-38502-NO-RS	***VPS8****	2	BOVINEHD0200033369	***IRS1***, *RHBDD1*	3	BOVINEHD0300004169	*FCRL4**
1	ARS-BFGL-NGS-14913	***VPS8****	3	BTB-01071906	***NEGR1****	3	ARS-BFGL-NGS-113768	***P2RY1***
2	ARS-BFGL-NGS-94696	***TMEFF2***	3	BTA-107777-NO-RS	***NEGR1****	10	BOVINEHD1000020414	*DACT1*
2	ARS-BFGL-NGS-102243	***TMEFF2****	3	ARS-BFGL-NGS-119369	***NEGR1****	11	ARS-BFGL-NGS-77753	***CAPN13***
2	BOVINEHD0200023298	***TMEFF2****	4	BTB-00171304	***HDAC9***	12	BOVINEHD1200006184	*C12H13orf42**
4	BTA-85377-NO-RS	*UMAD1**, ***RF00001***	4	BTA-70407-NO-RS	***RELN****	16	UA-IFASA-9087	***CHI3L1****, ***MYBPH***, ***ADORA1***
4	BTA-85374-NO-RS	*UMAD1*, ***RF00001***	6	HAPMAP54653-RS29025767	***KLHL5****	16	HAPMAP54719-RS29023014	*CHIT1*, *CHI3L1*
4	HAPMAP50489-BTA-85396	***NXPH1****, *TRNAC-GCA*	10	BTB-00430000	***UNC13C****	16	ARS-BFGL-NGS-49806	*CHIT1**
10	BTB-00430000	***UNC13C****	10	BOVINEHD1000016688	***UNC13C****	16	BOVINEHD1600000206	*BTG2*
10	BOVINEHD1000016688	***UNC13C****	10	ARS-BFGL-NGS-57640	***MYO5C****	16	BOVINEHD1600000210	*BTG2*
10	ARS-BFGL-NGS-55539	***MAPK6***, *LEO1*	10	ARS-BFGL-NGS-55539	***MAPK6***, *LEO1*	16	BOVINEHD1600000513	***PPP1R15B***
12	BOVINEHD1200006184	*C12H13orf42**	10	ARS-BFGL-NGS-54405	***TMOD3****	16	BOVINEHD1600000531	***PPP1R15B***, *PIK3C2B**
14	BOVINEHD1400008930	***CYP7B1***	10	HAPMAP54378-RS29019795	*GLDN*	16	BOVINEHD1600000544	*PIK3C2B**
14	BOVINEHD1400009020	*RF00100*	10	BTB-00434073	***DACT1***, ***KIAA0586***	16	BTA-39751-NO-RS	***GLRX2***, ***TROVE2****, *UCHL5**
15	HAPMAP33827-BES10_CONTIG357_1268	***KSR2****	10	BOVINEHD1300017663	*REM1*, ***HM13***	18	BTA-43728-NO-RS	*ITPKC**, *COQ8B*, *SNRPA*, *C18H19orf54*
17	BOVINEHD1700017220	*FBXW8**	13	HAPMAP47901-BTA-26507	***PTPRT****	18	BTA-23545-NO-RS	*CYP2S1*, ***AXL****, *HNRNPUL1*
17	ARS-BFGL-NGS-101538	***GSE1****	13	ARS-BFGL-NGS-32582	***PTPRT****	18	ARS-BFGL-NGS-2444	***LIPE***, *CNFN*, ***MEGF8***
18	BOVINEHD1800004542	***ZFPM1****	13	BOVINEHD1800000314	***VAC14****	18	ARS-BFGL-NGS-7137	***GSK3A****, ***DEDD2***, ***ZNF526***, *ERF*
18	BTA-96575-NO-RS	***gzmA****, ***CDC20B***	18	BOVINEHD1800004542	***ZFPM1****	18	ARS-BFGL-NGS-99872	***ZNF45****, ***ZNF404***, ***ZNF226***
20	BTA-56906-NO-RS	***GFOD1***, ***TBC1D7****	18	BOVINEHD2100014321	***FBXO33****	19	ARS-BFGL-NGS-77925	*TSPOAP1*, *MIR142*, ***SUPT4H1***, ***RNF43***
23	HAPMAP40134-BTA-56904	***TBC1D7***, ***PHACTR1***	23	HAPMAP47328-BTA-56087	***C4A***, *C2*, *CFB*, ***NELFE***, *DXO*, *STK19*	19	HAPMAP43381-BTA-86488	***BCAS3****
27	BOVINEHD2700009358	*NSD3**	28	BOVINEHD2800008965	***KCNMA1****	19	BOVINEHD1900015990	***SAP30BP***, ***RECQL5****, ***SMIM5***, *SMIM6*
28	BOVINEHD2800008965	***KCNMA1****				19	BOVINEHD1900016000	***SAP30BP***, ***RECQL5****, ***SMIM5***, *SMIM6*, ***MYO15B***
						20	BOVINEHD2000007525	***ARL15****

[1]Only markers falling within the gene (*) or including genes in a ±60-kb region are shown. In bold are indicated the genes with potential role in productivity traits in cattle. Genes overlapping between BayeScan and the statistical method based on allele frequencies (F_{ST}) window-based approaches are underlined.

Table 3. List of significant windows detected using the F_{ST} window-based approach and their relative genes for the 3 Holstein cattle populations: one including only individuals from the United States and Canada (acHOL), a second including Italian Holstein samples not used to produce Parmigiano Reggiano (itHOL), and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano (prHOL)[1]

itHOL vs. acHOL					prHOL vs. acHOL				prHOL vs. itHOL				
BTA	Start	End	no. Gene		BTA	Start	End	no. Gene		BTA	Start	End	no. Gene	
1	82260001	82440000	5	***VPS8****| 1	82260001	82440000	5	***VPS8****	3	114900001	115020000	5	***P2RY1***
3	12240001	12360000	5	***KIRREL1****, *TRNAG-CCC*	3	108840001	109020000	4	***GRIK3****	12	48420001	48540000	5	*KLF12**
7	107940001	108300000	4	***FER****	7	107940001	108060000	4	***FER****	16	960001	1080000	6	***CHI3L1****, ***MYBPH***, *CHIT1**, *BTG2*
10	87780001	87900000	4	*ESRRB**	8	24540001	24660000	4	***SLC24A2****	16	2100001	2280000	5	***PPP1R15B****, *PIK3C2B**, *PLEKHA6*, *MDM4*, *TRNAK-UUU*
14	10560001	10680000	6	***ASAP1****, *FAM49B**	14	2520001	2640000	4	*MROH5*, ***PTP4A3***, ***GPR20****	18	52020001	52140000	4	***ZNF404****, *ZNF983**, *ZNF45**, ***ZNF226****
15	61620001	61740000	4	***DCDC1****	14	7860001	7980000	6	***ST3GAL1****					
16	10440001	10560000	4	*HMGB1P1*										

[1]Only windows containing annotated genes are shown. When the marker falls within the gene a * is annotated. In bold are indicated genes with potential role in productivity traits in cattle. Genes overlapping between BayeScan and the statistical method based on allele frequencies (F_{ST}) window-based approaches are underlined.

of genomic admixture whereas both the North American and Italian groups seem to be characterized by a more homogeneous pattern. Such a pattern is probably due to stronger genetic drift caused by the early implementation of genomic selection in acHOL, and later in itHOL, as also detected in previous works (Makanjuola et al., 2020; Hulsegge et al., 2022). In contrast, the higher genomic admixture highlighted in the Holstein population used for the production of Parmigiano Reggiano might reflect the past use of the American stock to implement the breeding selection strategy linked to dairy productive traits. In addition, our results are also in line with other studies which found a certain distinctiveness of some Italian Holstein cows reared for other cheese production chains (Ablondi et al., 2021). These outcomes are also confirmed by all genetic diversity indices which indicates higher inbreeding and lower observed heterozygosity in the North American population which is well known to have been exposed to intense genomic selection. To strengthen these considerations are the effective population size estimates which underline a high contemporary value for the prHOL and lower values for both itHOL and acHOL. In general, the average inferred contemporary *Ne* for the 3 populations resulted comparable with those observed in others worldwide cosmopolitan breeds (Senczuk et al., 2020b). Finally, trends in *Ne* show considerable drops from 25 generations ago in all the analyzed populations but in general, the Italian Holstein populations (itHOL and prHOL) show higher values. Even thoough the inferred sudden *Ne* decline may reflect a recent increase in the genetic drift due to the strong selection, these results should be taken with a certain caution as very recent drastic drops might be related to bias estimates either in structured populations characterized by low migration rate or in population with recent migrants (Santiago et al., 2020).

Selection Signature Through Genome-Wide Allele Frequency Approaches

The analytical approach using F_{ST} has been recently applied in cattle for the identification of genomic regions involved in phenotypic differences (Mastrangelo et al., 2019a,b; Senczuk et al., 2020a; Moscarelli et al., 2021; Bertolini et al., 2021). The genome-wide information in comparative F_{ST} was used to detect signatures of selection that diverge among the 3 groups of Holstein cattle considered in this study. Several genes with known roles in traits related to milk quality, resistance to disease and immune response, and fertility have been identified (Table 2). The fact that we do not find a great overlap in significant SNPs between the 2 approaches in almost all comparisons is not surprising considering the 2 dif-

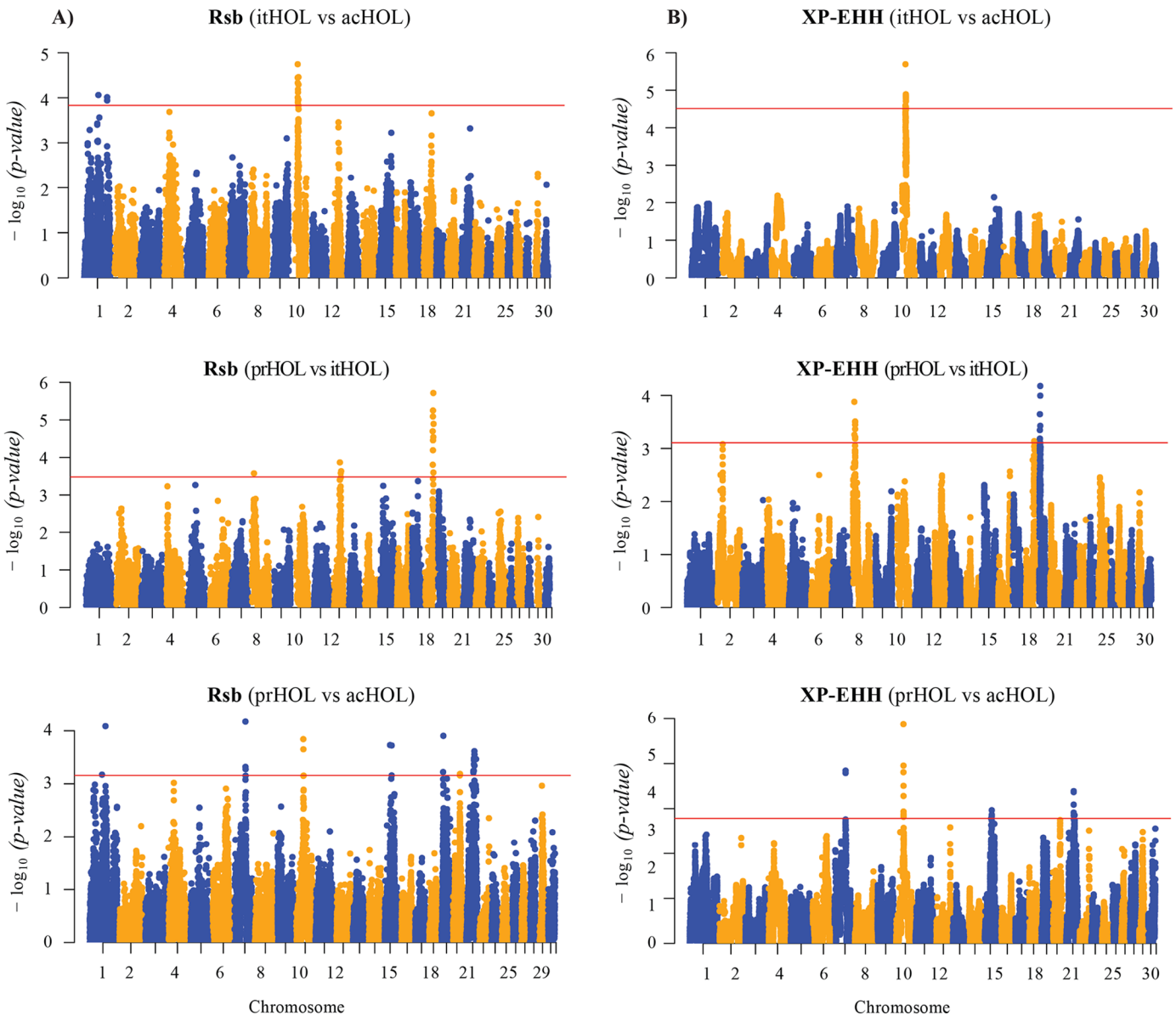


Figure 4. Manhattan plot of the $-\log_{10}$ (*P*-value) calculated between pairs of the 3 Holstein populations using the statistical method based on extended haplotype homozygosity [ratio of extended haplotype homozygosity segments between populations (Rsb; A) and the cross-population extended haplotype homozygosity (XP-EHH; B)] approaches. acHOL = only individuals from the United States and Canada; itHOL = Italian Holstein samples not used to produce Parmigiano Reggiano; prHOL = only Italian Holstein cattle used to produce Parmigiano Reggiano. The horizontal red line indicates the threshold based on the 0.05% $\log_{10}$ (*P*-value) (99.95th percentile).

ferent schemes in selecting SNPs we used. Indeed, for the F_{ST} windows-based approach we used a weighted F_{ST} average, retaining only windows with at least 4 SNPs (Weir and Cockerham, 1984). However, a total of 11 genes (*VPS8*, *P2RY1*, *BTG2*, *CHIT1*, *MYBPH*, *CHI3L1*, *PPP1R15B*, *PIK3C2B*, *ZNF404*, *ZNF45*, and *ZNF226*) still showed concordant results between the 2 approaches. Moreover, most of these genes have been shown to have a role in different productive traits in cattle (Pimentel et al., 2011; Du et al., 2019; Mohammadi et al., 2020; Illa et al., 2021; Soares et al., 2021; Yang et al., 2021). For example, the *VPS8* gene overlapped with the bovine QTL region of milk traits (Taye et al., 2017), the gene *ZNF226* is associated with growth and carcass quality QTL (MacHugh et al., 1997), and *CHI3L1* is related to the innate immune response during bacterial inflammation of the mammary gland (Breyne et al., 2018).

The BayeScan analysis detected 86 unique annotated genes falling within the upstream and downstream 60

kb from the significant SNPs. Several of these genes were detected in more than one comparison (*UNC13C*, *MAPK*, *LEO1*, *ZFPM1*, *KCNMA1*, and *C12H13orf42*). Among the 86 annotated genes, 55 have been shown to play a role in productive traits (in bold Tables 2 and 3) including genes related to milk production (*GSK3A*, *SUPT4H1*, *RNF43*, *VPS8*, *KSR2*, *ZFPM1*, *KCNMA1*, *TMEFF2*, *CYP7B1*, *RELN*, *ZFPM1*, *HDAC9*, *LIPE*, *FBXO33*, *C4A*, *IRS1*, and *KIAA0586*). Notably, among the 18 milk production associate genes, 16 have been found in the comparisons including the acHOL. For example, the *PRL/miR-183/IRS1* pathway has been shown to be involved in the regulation of milk fat metabolism (Jiao et al., 2020), whereas the gene *KCNMA1* has been predicted to have a functional interaction with the κ-CN gene (*CSN3*), thus possibly affecting coagulation and properties of milk, especially curd and cheese production (Patel and Chauhan, 2018). Interestingly, this latter gene has been found to be significant in both comparisons including the acHOL. In addition, several of these genes (*ZFPM1*, KIAA0586, *VPS8*, *C4A*, *CYP7B1*, and *KSR2*) mapped within genomic regions with QTL associated with milk production and yield (Pimentel et al., 2011; Raschia et al., 2020; Illa et al., 2021). Finally, the genes *SUPT4H1* and *RNF43* have been found within flanking regions of associated SNPs for lactation persistency (Do et al., 2017).

Considering the F_{ST} windows-based approach, we found 29 annotated genes with 2 genes shared in the 2 comparisons including the acHOL (*VPS8* and *FER*). Whereas the *VPS8* gene is involved in milk traits (Taye et al., 2017), the *FER* gene encodes a tyrosine kinase with a putative role in the regulation of innate immune response (McCafferty et al., 2002). Among the 29 genes, 15 have been previously detected to be associated with productive traits including milk yield and quality (*DCDC1*, *VPS8*, *PTP4A3*, *ST3GAL1*, *GPR20*, and *SLC24A2*). Among them, the gene *GPR20* has been found involved in the production of milk medium-chain fatty acids in a GWS framework (Freitas et al., 2020). Furthermore, the *SLC24A2* and *DCDC1* genes have been argued to be associated with milk production (da Cruz et al., 2021; Rajawat et al., 2022). Finally, transcripts of the gene *PTP4A3* have been detected to be downregulated in the mammary glands of dairy cows during early lactation (Lin et al., 2019).

Selection Signature Through Extended Haplotype Homozygosity

The genome-wide scan addressed to find putative genomic regions under selection using EHH approaches (XP-EHH and Rsb), identified a total of 29 genomic regions putatively under selection (Supplemental Table S2). However, in the haplotypic approach we did not find any common signal with the 2 F_{ST}-based approaches. This is a common observation in genome scans suggesting that detection of selection depends on the method used. In fact, F_{ST} and EHH-based approaches should be considered as complementary because of the different time scale resolution in which these methods are able to detect evidence of selection (Maiorano et al., 2018). In general, F_{ST} approaches are more appropriate to investigate evidence of selection in the distant past as being related to intermediate window size and therefore more sensitive to historical processes (Sabeti et al., 2006). In contrast, the EHH-based methods are best suited to detect more recent positive selection as they need a relatively large window size to distinguish allele fixation from noise (Mizuno et al., 2010). Therefore, combining complementary approaches is a growing promising strategy to identify a wider spectrum of possible selection signatures (Liu et al., 2013; Ben-Jemaa et al., 2020). A total of 7 genomic regions containing 266 annotated genes were identified combining the results of Rsb and XP-EHH, providing good evidence that these signals are not artifacts but really genomic regions under selection. Among the 52 annotated genes found within the itHOL versus acHOL comparison, most were related to reproductive and dairy traits. In particular, 15 genes have been previously shown to be associated with milk production (*ATL1*, *SAV1*, *FRMD6*, *GNG2*, *RTRAF*, *PCLAF*, *CSNK1G1*, *DAPK2*, *HERC1*, *FBXL22*, *USP3*, *APH1B*, *LACTB*, *C2CD4B*, and *VPS13C*; Marete et al., 2018; Deng et al., 2019; Klein et al., 2021; Korkuć et al., 2021; Vohra et al., 2021; Atashi et al., 2022; Nizamuddin et al., 2022). Noteworthy, this genomic region on BTA10 has been also found significant in a recent genome-wide association study for milk production traits, which reported the association of 6 genes (*FRMD6*, *GNG2*, *RTRAF*, *PCLAF*, *CSNK1G1*, and *DAPK2*) with SCC, milk protein yield, and daily milk yield (Atashi et al., 2022). In addition, the *DAPK2* gene has been also argued to be related to milk fat percentage in another recent genome-wide association study (Vohra et al., 2021).

Within the itHOL versus prHOL comparison, 5 genes located on BTA7, BTA10, and BTA20, have been shown to be associated with milk traits (*SAV1*, *ATL1*, *ANXA6*, *SLC36A2*, and *SLC36A1*) in previous studies (Bionaz and Loor, 2011; Kim et al., 2013; Zhou et al., 2019; Klein et al., 2021). In this comparison, 20 genes on BTA10 were also detected as significant in the itHOL versus acHOL comparison (see underlined genes in the Table 4) where 2 of these (*SAV1* and *ATL1*), are found to be milk-related (Klein et al., 2021). Furthermore, *SLC36A2* appeared to be associated with 6 milk protein traits (αs1-CN, αs2-CN, β-CN, κ-CN, protein

Table 4. List of significant genomic regions detected by both XP-EHH and Rsb approaches for the 3 Holstein cattle populations: one including only individuals from the United States and Canada (acHOL), a second including Italian Holstein samples not used to produce Parmigiano Reggiano (itHOL), and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano (prHOL)[1]

BTA	Start (Mb)	End (Mb)	Gene (no.)	Length (Mb)	Gene
itHOL vs. acHOL					
10	40.96	48.45	52	7.49	***MGAT2***, ***DNAAF2***, ***MIR6517***, ***POLE2***, ***KLHDC1***, ***KLHDC2***, *NEMF*, ***ARF6***, ***VCPKMT***, ***SOS2***, ***L2HGDH***, ***ATP5S***, ***CDKL1***, ***MAP4K5***, ***ATL1***, ***SAV1***, *NIN*, ***ABHD12B***, ***PYGL***, ***TRIM9***, *TMX1*, ***FRMD6***, ***GNG2***, ***RTRAF***, ***NID2***, ***PTGDR***, ***PLEKHO2***, ***PIF1***, ***RBPMS2***, ***OAZ2***, ***ZNF609***, ***TRIP4***, ***PCLAF***, ***CSNK1G1***, ***PPIB***, *SNX22*, ***SNX1***, *FAM96A*, ***DAPK2***, ***HERC1***, ***FBXL22***, ***USP3***, ***CA12***, ***APH1B***, ***RAB8B***, ***RPS27L***, ***LACTB***, ***TPM1***, ***TLN2***, *MIR190A*, ***C2CD4B***, ***VPS13C***
itHOL vs. prHOL					
7	61.890	63.88	22	1.99	***NDST1***, ***SYNPO***, ***MYOZ3***, ***RBM22***, ***DCTN4***, ***SMIM3***, *ZNF300*, ***GPX3***, ***TNIP1***, ***ANXA6***, ***CCDC69***, ***GM2A***, ***SLC36A3***, ***SLC36A2***, ***SLC36A1***, ***FAT2***, ***BERV-K1***, ***SPARC***, ***ATOX1***, ***G3BP1***, ***GLRA1***, ***NMUR2***
10	41.69	43.94	24	2.25	***RPS29***, ***RN7SL1***, ***LRR1***, ***RPL36AL***, ***MGAT2***, ***DNAAF2***, ***MIR6517***, ***POLE2***, ***KLHDC1***, ***KLHDC2***, *NEMF*, ***ARF6***, ***VCPKMT***, ***SOS2***, ***L2HGDH***, ***ATP5S***, ***CDKL1***, ***MAP4K5***, ***ATL1***, ***SAV1***, *NIN*, ***ABHD12B***, ***PYGL***, ***TRIM9***
20	42.28	44.48	1	2.2	***CDH6***
acHOL vs. prHOL					
8	16.82	19.8	1	2.98	***TUSC1***
18	49.18	53.56	153	4.38	***DLL3***, ***SELENOV***, *EID2*, ***DYRK1B***, *FBL*, ***FCGBP***, ***PSMC4***, ***MAP3K10***, ***TTC9B***, ***CNTD2***, ***AKT2***, *C18H19orf47*, ***PLD3***, ***HIPK4***, *PRX*, ***SERTAD1***, ***SERTAD3***, ***BLVRB***, ***SPTBN4***, ***SHKBP1***, ***LTBP4***, ***NUMBL***, ***COQ8B***, ***ITPKC***, ***C18H19orf54***, ***SNRPA***, ***MIA***, ***RAB4B***, ***EGLN2***, ***CYP2B6***, ***CYP2S1***, ***AXL***, ***HNRNPUL1***, ***CCDC97***, ***TGFB1***, ***B9D2***, ***TMEM91***, ***EXOSC5***, ***BCKDHA***, ***B3GNT8***, *DMAC2*, ***ERICH4***, ***CEACAM1***, ***LIPE***, ***CNFN***, ***MEGF8***, ***TMEM145***, ***PRR19***, ***PAFAH1B3***, *CIC*, *ERF*, ***GSK3A***, ***ZNF526***, ***DEDD2***, ***POU2F2***, ***ZNF574***, ***GRIK5***, ***ATP1A3***, ***RABAC1***, ***ARHGEF1***, ***CD79A***, ***RPS19***, ***DMRTC2***, ***LYPD4***, ***CXCL17***, ***CD177***, ***TEX101***, ***BSP3***, ***BSP5***, ***BSP1***, ***LYPD3***, ***PHLDB3***, ***ETHE1***, ***ZNF575***, ***XRCC1***, ***IRGQ***, ***ZNF576***, *ZNF428*, *SRRM5*, ***CADM4***, ***PLAUR***, *IRGC*, ***SMG9***, ***KCNN4***, ***LYPD5***, ***ZNF283***, ***ZNF404***, ***ZNF45***, ***ZNF226***, ***ZNF227***, *ZNF233*, *ZNF235*, ***ZNF112***, ***ZNF285***, *ZNF180*, ***CEACAM20***, ***IGSF23***, *PVR*, ***CEACAM19***, ***CEACAM16***, ***BCL3***, ***CBLC***, ***BCAM***, ***NECTIN2***, ***TOMM40***, ***APOE***, ***APOC4***, ***APOC2***, ***CLPTM1***, ***RELB***, ***CLASRP***, ***ZNF296***, ***GEMIN7***, ***PPP1R37***, ***NKPD1***, ***TRAPPC6A***, ***BLOC1S3***, ***EXOC3L2***, ***MARK4***, ***CKM***, ***KLC3***, ***ERCC2***, ***PPP1R13L***, ***CD3EAP***, ***ERCC1***, ***FOSB***, ***RTN2***, ***PPM1N***, ***VASP***, ***OPA3***, *GPR4*, ***EML2***, *MIR330*, ***GIPR***, ***SNRPD2***, ***QPCTL***, ***FBXO46***, ***BHMG1***, ***SIX5***, ***DMPK***, ***DMWD***, ***RSPH6A***, ***SYMPK***, ***FOXA3***, ***IRF2BP1***, ***MYPOP***, ***NANOS2***, ***NOVA2***, ***CCDC61***, ***MIR769***, ***PGLYRP1***, ***IGFL1***, ***HIF3A***
19	15.45	19.07	33	3.62	*TMEM132E*, ***CCL1***, ***CCL8***, ***CCL11***, ***CCL2***, ***ASIC2***, ***SPACA3***, ***TMEM98***, ***MYO1D***, ***CDK5R1***, ***PSMD11***, ***ZNF207***, *C19H17orf75*, *RHBDL3*, ***RHOT1***, ***RNF135***, ***ADAP2***, ***TEFM***, ***ATAD5***, ***CRLF3***, *SUZ12*, ***UTP6***, ***MIR2332***, *COPRS*, ***MIR365–2***, *MIR2333*, ***MIR193A***, ***RAB11FIP4***, ***EVI2A***, ***EVI2B***, ***OMG***, ***NF1***, ***WSB1***

[1]In bold are indicated genes with potential role in productivity traits in cattle whereas genes detected in more than one comparison are underlined.

yield, and protein percentage) suggesting a potential pleiotropic effect for protein composition traits and coagulation (Zhou et al., 2019).

Finally, in the comparison prHOL versus acHOL, we found 14 genes associated with milk production traits (*ZNF112*, *CBLC*, *SHKBP1*, *PGLYRP1*, *RAB4B*, *TMEM145*, *FOXA3*, *APOC2*, *BLVRB*, *TMEM98*, *MYO1D*, *RHOT1*, *ADAP2*, and *UTP6*) which mapped on BTA8, BTA18, and BTA19. Among these, Sablik et al. (2020) demonstrated how a mutation in *PGLYRP1* increased milk and protein production during first lactation in Holstein Friesian cattle, whereas *FOXA3* was observed to be associated with milk and protein yield in several studies (Kolbehdari et al., 2009; Valdez-Torres et al., 2020).

Quantitative Trait Loci and Functional Enrichment Analysis

Among comparisons and methods used, most of the proportion of QTL annotation was related to the class milk traits. In fact, the F_{ST} windows-based and EHH-based approaches showed the highest fraction of QTL related to milk traits in all comparisons. For example, in the F_{ST} windows-based approach, the enrichment analysis highlights significantly enriched QTL related to milk glycosylated κ-CN percentage on BTA7 (itHOL vs. acHOL); milk κ-CN percentage and milk fat percentage on BTA3 and BTA14, respectively (prHOL vs. acHOL). Interestingly, the milk protein composition such as glycosylated κ-CN percentage has been shown to have implication on variations in the rate of firming and curd firmness (Robitaille et al., 1993) whereas in general, the role of the κ-CN in the stabilization of the casein micelle is commonly known (McMahon and Brown, 1984). The fact that we found these milk-related QTL only when comparing Holstein from Italy (prHOL and itHOL) with the acHOL, could be related to a greater attention of the Italian farmer to select for milk traits involved in the cheese making. Although the remaining significantly enriched QTL between acHOL and the other 2 populations were not directly linked to milk traits, they point to conformation traits indirectly important for milk production. For example, stature, withers height, and chest girth are reportedly related to feed efficiency (Dickinson et al., 1969; Yerex et al., 1988). Others limb-related conformation traits such as rear leg placement, foot angle, and feet and leg conformation are related to the length of productive life of a cow (Vollema and Groen, 1998; Peréz-Cabal and Alenda, 2002; Vacek et al., 2006; Shin et al., 2022).

Considering the comparison prHOL versus itHOL, we just found 2 significantly enriched QTL, one related to milk yield (305-d milk yield) on BTA18 and the other to somatic cell score which is a trait often selected to improve resistance to mastitis (Shook and Schutz, 1994).

Concerning the significantly enriched QTL related to milk traits highlighted on putative genomic regions detected by the EHH approaches, we found milk protein percentage, milk fat content, and milk caproic acid content on BTA10 (itHOL vs. acHOL); milk linoleic acid content and milk casein percentage on BTA18 and BTA19, respectively (prHOL vs. acHOL); and milk phosphorus content and milk α-S1-CN percentage on BTA7 (prHOL vs. itHOL).

Finally, in the Bayesian approach, the significantly enriched QTL were mainly linked to health, reproduction, and meat and carcass. As reported above, these results might stress some weaknesses in the use of a single locus approach with respect to either multilocus (windows-based) or EHH-based approaches, especially in the absence of a multicohort framework (Senczuk et al., 2020a).

In the GO and pathway enrichment analyses performed on a list of candidate genes located within significantly enriched QTL related to production traits, we found that the majority of the significant terms were related to molecular function and biological process. Among the 17 enriched GO terms, 15 were found in the comparison itHOL versus prHOL and all were associated with AA transmembrane transport, specifically alanine, glycine, and proline. The latter is a conditionally essential AA (which may not be synthesized by the organism in some conditions) and has a key role in the pathway of P53 which is involved in different mechanisms including apoptosis, responses to oxidative stress, and in the regulation of fertility (Krishnan et al., 2008, Hu et al., 2011) through the pathway of estriol (E3), which has a crucial role in female fertility (Luo et al., 2013; Otsuka and Kadokawa, 2017).

In the same comparison, the enrichment of the Kyoto Encyclopedia of Genes and Genomes pathway map00680 for *methane metabolism* is notable; in fact, even though methanogenesis is performed by the rumen microbial community, there may be an effect of the animal itself (Corral-Jara et al., 2022). Additional studies will be necessary to assess possible differences in methane emission of the studied populations.

CONCLUSIONS

In this study we showed that the 3 Holstein populations are clearly differentiated; however, both the Italian and North American Holstein showed low levels of genetic diversity indices, higher inbreeding coefficients compatible with their history of isolation, drift, and strong genomic selection. Conversely, the Holstein pop-

ulation used to produce Parmigiano Reggiano, showed higher genetic diversity and lower inbreeding values. These results suggest that also in a cosmopolitan breed, often characterized by intense genetic exchange, the different breeding goals, environmental conditions, and farm systems can produce a genetic differentiation in a very short number of generations. Finally, the selection signature analyses detected the presence of several genes within QTL regions for important production traits associated with fertility, resistance to disease but especially with milk production traits. Overall, the results of the present study indicate that the 3 Holstein populations are genetically distinguishable from each other and that our selection signature analyses can be a starting point to identify causal mutations useful in future studies and lately in more practical applications.

ACKNOWLEDGMENTS

This study was supported by "Latteco2 project, sottomisura 10.2 of the PSRNBiodiversity 2020–2023" (MIPAAF. D.M. no. 465907 del 24/09/2021, project unique code J12C21004080005). The authors have not stated any conflicts of interest.

REFERENCES

Abbasi-Moshaii, B., M. H. Moradi, T. Yin, G. Rahimi-Mianji, A. Nejati-Javaremi, and S. König. 2023. Genome-wide scan for selective sweeps identifies novel loci associated with resistance to mastitis in German Holstein cattle. J. Anim. Breed. Genet. 140:92–105. https://doi.org/10.1111/jbg.12737.

Ablondi, M., M. Malacarne, C. Cipolat-Gotet, J.-T. van Kaam, A. Sabbioni, and A. Summer. 2021. Genome-wide scan reveals genetic divergence in Italian Holstein cows bred within PDO cheese production chains. Sci. Rep. 11:12601. https://doi.org/10.1038/s41598-021-92168-1.

Ablondi, M., A. Sabbioni, G. Stocco, C. Cipolat-Gotet, C. Dadousis, J.-T. Kaam, R. Finocchiaro, and A. Summer. 2022. Genetic diversity in the Italian Holstein dairy cattle based on pedigree and SNP data prior and after genomic selection. Front. Vet. Sci. 8:773985. https://doi.org/10.3389/fvets.2021.773985.

Alexander, D. H., and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246. https://doi.org/10.1186/1471-2105-12-246.

ANAFIBJ. 2021a. ANAFIBJ Website. Accessed Mar. 25, 2022. http://www.anafibj.it.

ANAFIBJ. 2021b. Genetic Evaluation Notes : Holstein and Jersey. Accessed Jan. 12, 2022. http://server01.anafi.it/IndiciGenetici/Schede-calcolo-Indici-ING-2021.pdf.

Andersson, L., and M. Georges. 2004. Domestic-animal genomics: Deciphering the genetics of complex traits. Nat. Rev. Genet. 5:202–212. https://doi.org/10.1038/nrg1294.

Atashi, H., H. Wilmot, S. Vanderick, X. Hubin, and N. Gengler. 2022. Genome-wide association study for milk production traits in dual-purpose Belgian Blue cows. Livest. Sci. 256:104831. https://doi.org/10.1016/j.livsci.2022.104831.

Balasini, D. 2002. Zootecnia Applicata Bovini e Bufali, Edagricole – Edizioni Agricole della Calderini s.r.l. Edagricole – Edizioni Agricole della Calderini s.r.l., Bologna.

Ben-Jemaa, S., S. Mastrangelo, S. H. Lee, J. H. Lee, and M. Boussaha. 2020. Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle. Sci. Rep. 10:19466. https://doi.org/10.1038/s41598-020-76576-3.

Bertolini, F., G. Moscatelli, G. Schiavo, S. Bovo, A. Ribani, M. Ballan, M. Bonacini, M. Prandi, S. Dall'Olio, and L. Fontanesi. 2021. Signatures of selection are present in the genome of 2 close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for their coat colors. J. Anim. Breed. Genet. n/a. https://doi.org/10.1111/jbg.12659.

Bionaz, M., and J. J. Loor. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinform. Biol. Insights 5:BBI.S7003. https://doi.org/10.4137/BBI.S7003.

Breyne, K., J. Steenbrugge, K. Demeyere, C. G. Lee, J. A. Elias, W. Petzl, D. G. E. Smith, P. Germon, and E. Meyer. 2018. Immunomodulation of host chitinase 3-like 1 during a mammary pathogenic *Escherichia coli* infection. Front. Immunol. 9:1143. https://doi.org/10.3389/fimmu.2018.01143.

Cassandro, M. 2014. Genetic aspects of fertility traits in dairy cattle - review. Acta Agrar. Kaposváriensis 18:11–23. https://doi.org/10.13140/2.1.3856.9925.

Cattle QTLdb. 2021. Accessed Mar. 25, 2022. https://www.animalgenome.org/cgi-bin/QTLdb/BT/ontrait?class_ID=1.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer data sets. GigaScience 4:s13742–015–0047–8. https://doi.org/10.1186/s13742-015-0047-8.

Chen, J., E. E. Bardes, B. J. Aronow, and A. G. Jegga. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37(Web Server):W305–W311. https://doi.org/10.1093/nar/gkp427.

Chessa, S., S. Gattolin, P. Cremonesi, D. Soglia, R. Finocchiaro, J.-T. Van Kaam, M. Marusi, and G. Civati. 2020. The effect of selection on casein genetic polymorphisms and haplotypes in Italian Holstein cattle. Ital. J. Anim. Sci. 19:833–839. https://doi.org/10.1080/1828051X.2020.1802356.

Corral-Jara, K. F., Y. Ramayo-Caldas, L. Bernard, C. Martin, J. Tournayre, D. P. Morgavi, and M. Popova. 2022. An integrative metatranscriptomic analysis reveals differences in enteric methanogenesis mechanisms between cows and goats. Unpublished data. https://doi.org/10.21203/rs.3.rs-1828071/v1.

da Cruz, A. S., D. C. Silva, L. B. Minasi, L. K. de Farias Teixeira, F. M. Rodrigues, C. C. da Silva, A. S. do Carmo, M. V. G. B. da Silva, Y. T. Utsunomiya, J. F. Garcia, and A. D. da Cruz. 2021. Single-nucleotide polymorphism variations associated with specific genes putatively identified enhanced genetic predisposition for 305-day milk yield in the Girolando crossbreed. Front. Genet. 11:573344. https://doi.org/10.3389/fgene.2020.573344.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and the 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Deng, T., A. Liang, S. Liang, X. Ma, X. Lu, A. Duan, C. Pang, G. Hua, S. Liu, G. Campanile, A. Salzano, B. Gasparrini, G. Neglia, X. Liang, and L. Yang. 2019. Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. Front. Genet. 10:36. https://doi.org/10.3389/fgene.2019.00036.

Dickinson, F. N., B. T. McDaniel, and R. E. McDowell. 1969. Comparative efficiency of feed utilization during first lactation of Ayrshire, Brown Swiss, and Holstein cows. J. Dairy Sci. 52:489–497. https://doi.org/10.3168/jds.S0022-0302(69)86593-8.

Do, D. N., N. Bissonnette, P. Lacasse, F. Miglior, M. Sargolzaei, X. Zhao, and E. M. Ibeagha-Awemu. 2017. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. J. Dairy Sci. 100:1955–1970. https://doi.org/10.3168/jds.2016-11910.

Doekes, H. P., R. F. Veerkamp, P. Bijma, S. J. Hiemstra, and J. J. Windig. 2018. Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. Genet. Sel. Evol. 50:15. https://doi.org/10.1186/s12711-018-0385-y.

Du, C., T. Deng, Y. Zhou, T. Ye, Z. Zhou, S. Zhang, B. Shao, P. Wei, H. Sun, F. A. Khan, L. Yang, and G. Hua. 2019. Systematic analyses for candidate genes of milk production traits in water buffalo (*Bubalus bubalis*). Anim. Genet. 50:207–216. https://doi.org/10.1111/age.12739.

FAO. 2018. Gateway to Dairy Production and Products: Cattle. Accessed Mar. 14, 2022. https://www.fao.org/dairy-production-products/production/dairy-animals/cattle/en/.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977–993. https://doi.org/10.1534/genetics.108.092221.

Fonseca, P. A. S., A. Suárez-Vega, G. Marras, and Á. Cánovas. 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. Gigascience 9:giaa149. https://doi.org/10.1093/gigascience/giaa149.

Freitas, P. H. F., H. R. Oliveira, F. F. Silva, A. Fleming, F. Miglior, F. S. Schenkel, and L. F. Brito. 2020. Genomic analyses for predicted milk fatty acid composition throughout lactation in North American Holstein cattle. J. Dairy Sci. 103:6318–6331. https://doi.org/10.3168/jds.2019-17628.

Gautier, M., and R. Vitalis. 2012. rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28:1176–1177. https://doi.org/10.1093/bioinformatics/bts115.

Hu, W., T. Zheng, and J. Wang. 2011. Regulation of fertility by the p53 family members. Genes Cancer 2:420–430. https://doi.org/10.1177/1947601911408892.

Hu, Z.-L., C. A. Park, X.-L. Wu, and J. M. Reecy. 2013. Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 41:D871–D879. https://doi.org/10.1093/nar/gks1150.

Hu, Z.-L., C. A. Park, and J. M. Reecy. 2022. Bringing the animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. Nucleic Acids Res. 50(D1):D956–D961. https://doi.org/10.1093/nar/gkab1116.

Hulsegge, I., K. Oldenbroek, A. Bouwman, R. Veerkamp, and J. Windig. 2022. Selection and drift: A comparison between historic and recent Dutch Friesian cattle and recent Holstein Friesian using WGS data. Animals (Basel) 12:329. https://doi.org/10.3390/ani12030329.

Illa, S. K., S. Mukherjee, S. Nath, and A. Mukherjee. 2021. Genome-wide scanning for signatures of selection revealed the putative genomic regions and candidate genes controlling milk composition and coat color traits in Sahiwal cattle. Front. Genet. 12:699422. https://doi.org/10.3389/fgene.2021.699422.

Jiao, P., Y. Yuan, M. Zhang, Y. Sun, C. Wei, X. Xie, Y. Zhang, S. Wang, Z. Chen, and X. Wang. 2020. PRL/microRNA-183/IRS1 pathway regulates milk fat metabolism in cow mammary epithelial cells. Genes (Basel) 11:196. https://doi.org/10.3390/genes11020196.

Kim, E. S., J. B. Cole, H. Huson, G. R. Wiggans, C. P. Van Tassell, B. A. Crooker, G. Liu, Y. Da, and T. S. Sonstegard. 2013. Effect of artificial selection on runs of homozygosity in US Holstein cattle. PLoS One 8:e80813. https://doi.org/10.1371/journal.pone.0080813.

Klassmann, A., and M. Gautier. 2022. Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data. PLoS One 17:e0262024. https://doi.org/10.1371/journal.pone.0262024.

Klein, S. L., T. Yin, H. H. Swalve, and S. König. 2021. Single-step genomic best linear unbiased predictor genetic parameter estimations and genome-wide associations for milk fatty acid profiles, interval from calving to first insemination, and ketosis in Holstein cattle. J. Dairy Sci. 104:10921–10933. https://doi.org/10.3168/jds.2021-20416.

Kolbehdari, D., Z. Wang, J. R. Grant, B. Murdoch, A. Prasad, Z. Xiu, E. Marques, P. Stothard, and S. S. Moore. 2009. A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. J. Anim. Breed. Genet. 126:216–227. https://doi.org/10.1111/j.1439-0388.2008.00793.x.

Korkuć, P., D. Arends, K. May, S. König, and G. A. Brockmann. 2021. Genomic loci affecting milk production in German Black Pied cattle (DSN). Front. Genet. 12:640039. https://doi.org/10.3389/fgene.2021.640039.

Krishnan, N., M. B. Dickman, and D. F. Becker. 2008. Proline modulates the intracellular redox environment and protects mammalian cells against oxidative stress. Free Radic. Biol. Med. 44:671–681. https://doi.org/10.1016/j.freeradbiomed.2007.10.054.

Lin, Y., H. Lv, M. Jiang, J. Zhou, S. Song, and X. Hou. 2019. Functional analysis of the dairy cow mammary transcriptome between early lactation and mid-dry period. J. Dairy Res. 86:63–67. https://doi.org/10.1017/S0022029919000049.

Liu, X., R. T.-H. Ong, E. N. Pillai, A. M. Elzein, K. S. Small, T. G. Clark, D. P. Kwiatkowski, and Y.-Y. Teo. 2013. Detecting and characterizing genomic signatures of positive selection in global populations. Am. J. Hum. Genet. 92:866–881. https://doi.org/10.1016/j.ajhg.2013.04.021.

Luo, L., F. Li, L. Zhu, Y. Ding, and D. Deng. 2013. Electrochemical sensing platform of natural estrogens based on the poly(L-proline)-ordered mesoporous carbon composite modified glassy carbon electrode. Sens. Actuators B Chem. 187:78–83. https://doi.org/10.1016/j.snb.2012.09.056.

Ma, L., T. S. Sonstegard, J. B. Cole, C. P. VanTassell, G. R. Wiggans, B. A. Crooker, C. Tan, D. Prakapenka, G. E. Liu, and Y. Da. 2019. Genome changes due to artificial selection in US Holstein cattle. BMC Genomics 20:128. https://doi.org/10.1186/s12864-019-5459-x.

MacHugh, D. E., M. D. Shriver, R. T. Loftus, P. Cunningham, and D. G. Bradley. 1997. Microsatellite DNA variation and the evolution, domestication, and phylogeography of Taurine and Zebu cattle (*Bos taurus* and *Bos indicus*). Genetics 146:1071–1086. https://doi.org/10.1093/genetics/146.3.1071.

Maiorano, A. M., D. L. Lourenco, S. Tsuruta, A. M. T. Ospina, N. B. Stafuzza, Y. Masuda, A. E. V. Filho, J. N. D. S. G. Cyrillo, R. A. Curi, and J. A. I. D. V. Silva. 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLoS One 13:e0200694. https://doi.org/10.1371/journal.pone.0200694.

Makanjuola, B. O., F. Miglior, E. A. Abdalla, C. Maltecca, F. S. Schenkel, and C. F. Baes. 2020. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J. Dairy Sci. 103:5183–5199. https://doi.org/10.3168/jds.2019-18013.

Marete, A. G., B. Guldbrandtsen, M. S. Lund, S. Fritz, G. Sahana, and D. Boichard. 2018. A meta-analysis including pre-selected sequence variants associated with 7 traits in 3 French dairy cattle populations. Front. Genet. 9:522. https://doi.org/10.3389/fgene.2018.00522.

Marusi, M., G. Visentin, C. Cipolat-Gotet, A. Summer, A. Formigon, M. Nocetti, M. Cassandro, G. Burchiellaro, and G. Civati. 2019. Cheesemaking and sustainability index – Parmigiano Reggiano: A new breeding index for the Italian Holstein. Page 239 in Book of Abstracts ASPA 23rd Congress. Ital. J. Anim. Sci.

Mastrangelo, S., S. Ben Jemaa, E. Ciani, G. Sottile, A. Moscarelli, M. Boussaha, M. Montedoro, F. Pilla, and M. Cassandro. 2020. Genome-wide detection of signatures of selection in three Valdostana cattle populations. J. Anim. Breed. Genet. 137:609–621. https://doi.org/10.1111/jbg.12476.

Mastrangelo, S., E. Ciani, P. Ajmone Marsan, A. Bagnato, L. Battaglini, R. Bozzi, A. Carta, G. Catillo, M. Cassandro, S. Casu, R. Ciampolini, P. Crepaldi, M. D'Andrea, R. Di Gerlando, L. Fontanesi, M. Longeri, N. P. Macciotta, R. Mantovani, D. Marletta, D. Matassino, M. Mele, G. Pagnacco, C. Pieramati, B. Portolano, F. M. Sarti, M. Tolone, and F. Pilla. 2018. Conservation status and historical relatedness of Italian cattle breeds. Genet. Sel. Evol. 50:35. https://doi.org/10.1186/s12711-018-0406-x.

Mastrangelo, S., A. Criscione, G. Sottile, B. Portolano, D. Marletta, and S. Bordonaro. 2019a. Genome-wide analysis identifies potentially causative genes explaining the phenotypic variability in Pinzirita sheep. Anim. Genet. 50:189–190. https://doi.org/10.1111/age.12763.

Mastrangelo, S., G. Sottile, M. T. Sardina, A. M. Sutera, M. Tolone, R. Di Gerlando, and B. Portolano. 2019b. A combined genome-wide approach identifies a new potential candidate marker associated with the coat color sidedness in cattle. Livest. Sci. 225:91–95. https://doi.org/10.1016/j.livsci.2019.05.009.

McCafferty, D.-M., A. W. B. Craig, Y. A. Senis, and P. A. Greer. 2002. Absence of Fer protein-tyrosine kinase exacerbates leukocyte recruitment in response to endotoxin. J. Immunol. 168:4930–4935. https://doi.org/10.4049/jimmunol.168.10.4930.

McMahon, D. J., and R. J. Brown. 1984. Composition, structure, and integrity of casein micelles: A review. J. Dairy Sci. 67:499–512. https://doi.org/10.3168/jds.S0022-0302(84)81332-6.

Milanesi, M., S. Capomaccio, E. Vajana, L. Bomba, J. F. Garcia, P. Ajmone-Marsan, and L. Colli. 2017. BITE: An R package for biodiversity analyses. bioRxiv 181610. Unpublished data. https://doi.org/10.1101/181610.

Mizuno, H., G. Atwal, H. Wang, A. J. Levine, and A. Vazquez. 2010. Fine-scale detection of population-specific linkage disequilibrium using haplotype entropy in the human genome. BMC Genet. 11:27. https://doi.org/10.1186/1471-2156-11-27.

Mohammadi, A., S. Alijani, S. A. Rafat, and R. Abdollahi-Arpanahi. 2020. Genome-wide association study and pathway analysis for female fertility traits in Iranian Holstein cattle. Ann. Anim. Sci. 20:825–851. https://doi.org/10.2478/aoas-2020-0031.

Moscarelli, A., M. T. Sardina, M. Cassandro, E. Ciani, F. Pilla, G. Senczuk, B. Portolano, and S. Mastrangelo. 2021. Genome-wide assessment of diversity and differentiation between original and modern Brown cattle populations. Anim. Genet. 52:21–31. https://doi.org/10.1111/age.13019.

Nicolazzi, E. L., S. Biffani, and G. Jansen. 2013. Short communication: Imputing genotypes using PedImpute fast algorithm combining pedigree and population information. J. Dairy Sci. 96:2649–2653. https://doi.org/10.3168/jds.2012-6062.

Nizamuddin, A., K. P. Ramesha, E. Rana, S. Kasargod, S. K. Behera, M. Ashokan, and T. S. K. Prasad. 2022. Transcriptomic analysis of PBMCs reveals COX3, BCL2, ATP2B4, B4GALT1, LACTB, FASN, and CSN family genes regulating milk performance in lactating *Bos indicus* cows of Southern India. Unpublished data. https://doi.org/10.21203/rs.3.rs-1726110/v1.

Otsuka, M., and H. Kadokawa. 2017. GPR30 mediates estrone, estriol, and estradiol to suppress gonadotropin-releasing hormone-induced luteinizing hormone secretion in the anterior pituitary of heifers. J. Reprod. Dev. 63:519–525. https://doi.org/10.1262/jrd.2017-035.

Patel, J. B., and J. B. Chauhan. 2018. Computational analysis of non-synonymous single nucleotide polymorphism in the bovine cattle kappa-casein (CSN3) gene. Meta Gene 15:1–9. https://doi.org/10.1016/j.mgene.2017.10.002.

Pérez-Cabal, M. A., and R. Alenda. 2002. Genetic relationships between lifetime profit and type traits in Spanish Holstein cows. J. Dairy Sci. 85:3480–3491. https://doi.org/10.3168/jds.S0022-0302(02)74437-8.

Pimentel, E. C. G., S. Bauersachs, M. Tietze, H. Simianer, J. Tetens, G. Thaller, F. Reinhardt, E. Wolf, and S. König. 2011. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Anim. Genet. 42:251–262. https://doi.org/10.1111/j.1365-2052.2010.02148.x.

Porter, V., L. Alderson, S. J. G. Hall, and D. P. Sponenberg. 2016. Mason's World Encyclopedia of Livestock Breeds and Breeding. CABI.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. https://doi.org/10.1086/519795.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, and H. Simianer. 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Anim. Genet. https://doi.org/10.1111/j.1365-2052.2009.02016.x.

Rajawat, D., M. Panigrahi, H. Kumar, S. S. Nayak, S. Parida, B. Bhushan, G. K. Gaur, T. Dutt, and B. P. Mishra. 2022. Identification of important genomic footprints using eight different selection signature statistics in domestic cattle breeds. Gene 816:146165. https://doi.org/10.1016/j.gene.2021.146165.

Raschia, M. A., J. P. Nani, H. A. Carignano, A. F. Amadio, D. O. Maizon, and M. A. Poli. 2020. Weighted single-step genome-wide association analyses for milk traits in Holstein and Holstein x Jersey crossbred dairy cattle. Livest. Sci. 242:104294. https://doi.org/10.1016/j.livsci.2020.104294.

Robitaille, G., K. F. Ng-Kwai-Hang, and H. G. Monardes. 1993. Effect of κ-casein glycosylation on cheese yielding capacity and coagulating properties of milk. Food Res. Int. 26:365–369. https://doi.org/10.1016/0963-9969(93)90079-X.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaneld, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M. Phillippy, J. B. Cole, C. P. Van Tassell, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9:giaa021. https://doi.org/10.1093/gigascience/giaa021.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, H. C. Ackerman, S. J. Campbell, D. Altshuler, R. Cooper, D. Kwiatkowski, R. Ward, and E. S. Lander. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837. https://doi.org/10.1038/nature01140.

Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma, T. S. Mikkelsen, D. Altshuler, and E. S. Lander. 2006. Positive natural selection in the human lineage. Science 312:1614–1620. https://doi.org/10.1126/science.1124309.

Sablik, P., A. Klenowicz, M. Szewczuk, A. Olszewski, and A. Dybus. 2020. The effect of polymorphism in PGLYRP1 gene on the productivity and health traits in Holstein-Friesian cattle. Russ. J. Genet. 56:333–338. https://doi.org/10.1134/S1022795420030138.

Santiago, E., I. Novo, A. F. Pardiñas, M. Saura, J. Wang, and A. Caballero. 2020. Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. Mol. Biol. Evol. 37:3642–3653. https://doi.org/10.1093/molbev/msaa169.

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

Senczuk, G. 2023. Supplemental files. figshare. Figure. https://doi.org/10.6084/m9.figshare.22786295.v1.

Senczuk, G., L. Guerra, S. Mastrangelo, C. Campobasso, K. Zoubeyda, M. Imane, D. Marletta, S. Kusza, T. Karsli, S. B. S. Gaouar, F. Pilla, E. Ciani, and The Bovita Consortium. 2020a. Fifteen shades of Grey: Combined analysis of genome-wide SNP data in Steppe and Mediterranean Grey cattle sheds new light on the molecular basis of coat color. Genes (Basel) 11:932 https://doi.org/10.3390/genes11080932.

Senczuk, G., S. Mastrangelo, E. Ciani, L. Battaglini, F. Cendron, R. Ciampolini, P. Crepaldi, R. Mantovani, G. Bongioni, G. Pagnacco, B. Portolano, A. Rossoni, F. Pilla, and M. Cassandro. 2020b. The genetic heritage of Alpine local cattle breeds using genomic SNP data. Genet. Sel. Evol. 52:40. https://doi.org/10.1186/s12711-020-00559-1.

Shin, J., J. Lee, J. Cho, C. Dang, T. Choi, C. Do, J. Lee, and S. Lee. 2022. The estimation of genetic parameters for longevity according to lactation period using a multiple trait animal model in Korean Holstein cows. Animals (Basel) 12:701. https://doi.org/10.3390/ani12060701.

Shook, G. E., and M. M. Schutz. 1994. Selection on somatic cell score to improve resistance to mastitis in the United States. J. Dairy Sci. 77:648–658. https://doi.org/10.3168/jds.S0022-0302(94)76995-2.

Soares, R. A. N., G. Vargas, T. Duffield, F. Schenkel, and E. J. Squires. 2021. Genome-wide association study and functional analyses for

clinical and subclinical ketosis in Holstein cattle. J. Dairy Sci. 104:10076–10089. https://doi.org/10.3168/jds.2020-20101.
Tang, K., K. R. Thornton, and M. Stoneking. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 5:e171. https://doi.org/10.1371/journal.pbio.0050171.
Taye, M., W. Lee, S. Jeon, J. Yoon, T. Dessie, O. Hanotte, O. A. Mwai, S. Kemp, S. Cho, S. J. Oh, H.-K. Lee, and H. Kim. 2017. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. Mamm. Genome 28:528–541. https://doi.org/10.1007/s00335-017-9715-6.
Theunissen, B. 2012. Breeding for nobility or for production? Cultures of dairy cattle breeding in the Netherlands, 1945–1995. Isis 103:278–309. https://doi.org/10.1086/666356.
Turner, S. D. 2018. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. J. Open Source Softw. 3:731. https://doi.org/10.21105/joss.00731.
Vacek, M., M. Štípková, E. Němcová, and J. Bouška. 2006. Relationships between conformation traits and longevity of Holstein cows in the Czech Republic. Czech J. Anim. Sci. 51:327–333. https://doi.org/10.17221/3946-CJAS.
Valdez-Torres, J. M., J. A. Grado Ahuir, B. E. Castro-Valenzuela, and M. E. Burrola-Barraza. 2020. QTL analysis associated to single nucleotide polymorphisms (SNP) involved in the dairy phenotype of Holstein cattle. Rev. Mex. Cienc. Pecu. 11:1192–1207. https://doi.org/10.22319/rmcp.v11i4.5295.
Visentin, G., M. Marusi, and G. Civati. 2018. ICS-PR: Una nuova selezione per la Frisona italiana [ICS-PR: A new selection for the Italian Holstein]. Bianconero n.6 Novembre-Dicembre 2018:12–15. http://server01.anafi.it/Latteco/articoli/2018/07-BN-Nov-Dic-2018.pdf.
Vohra, V., S. Chhotaray, G. Gowane, R. Alex, A. Mukherjee, A. Verma, and S. M. Deb. 2021. Genome-wide association studies in Indian Buffalo revealed genomic regions for lactation and fertility. Front. Genet. 12:696109. https://doi.org/10.3389/fgene.2021.696109.
Vollema, A. R., and A. F. Groen. 1998. A comparison of breeding value predictors for longevity using a linear model and survival analysis. J. Dairy Sci. 81:3315–3320. https://doi.org/10.3168/jds.S0022-0302(98)75897-7.
Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370. https://doi.org/10.2307/2408641.
Wickham, H. 2011. ggplot2. Wiley Interdiscip. Rev. Comput. Stat. 3:180–185. https://doi.org/10.1002/wics.147.
Yang, Z., Z. Lian, G. Liu, M. Deng, B. Sun, Y. Guo, D. Liu, and Y. Li. 2021. Identification of genetic markers associated with milk production traits in Chinese Holstein cattle based on post genome-wide association studies. Anim. Biotechnol. 32:67–76. https://doi.org/10.1080/10495398.2019.1653901.
Yerex, R. P., C. W. Young, J. D. Donker, and G. D. Marx. 1988. Effects of selection for body size on feed efficiency and size of Holsteins. J. Dairy Sci. 71:1355–1360. https://doi.org/10.3168/jds.S0022-0302(88)79693-9.
Zhao, F., S. McParland, F. Kearney, L. Du, and D. P. Berry. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet. Sel. Evol. 47:49. https://doi.org/10.1186/s12711-015-0127-3.
Zhou, C., C. Li, W. Cai, S. Liu, H. Yin, S. Shi, Q. Zhang, and S. Zhang. 2019. Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. Front. Genet. 10:72. https://doi.org/10.3389/fgene.2019.00072.

ORCIDS

Christian Persichilli https://orcid.org/0000-0003-2664-136X
Gabriele Senczuk https://orcid.org/0000-0001-8889-2290
Salvatore Mastrangelo https://orcid.org/0000-0001-6511-1981
Maurizio Marusi https://orcid.org/0000-0002-2359-3633
Jan-Thijs van Kaam https://orcid.org/0000-0002-2592-2461
Marika Di Civita https://orcid.org/0000-0002-9412-8704
Martino Cassandro https://orcid.org/0000-0002-8709-2870
Fabio Pilla https://orcid.org/0000-0002-1781-994X

Identification of Genetic Regions Associated with Resistance to Gastrointestinal Nematodes in Comisana Sheep using a Genome-wide Association Study based on EBV Ranking

Christian Persichilli[1*], Stefano Biffani[2], Gabriele Senczuk[1], Marika Di Civita[1], Antonio Bosco[4], Silverio Grande[3], Fabio Pilla[1]

[1]University of Molise, Department of Agricultural, Environmental and Food Science, Campobasso, Italy

[2]National Council of Research, Institute for Agriculture Biology and Biotechnology, Milan, Italy

[3]National Sheep and Goat Breeders Association, Rome, Italy

[4]University of Naples Federico II, Department of Veterinary Medicine and Animal Production, CREMOPAR, Naples, NA, Italy

[*]Corresponding author: Christian Persichilli; c.persichilli@studenti.unimol.it

ABSTRACT

Gastrointestinal nematodes (GINs) have significant economic, environmental, and animal welfare implications in small ruminant farming. This study is aimed at identifying the genetic regions responsible for sheep resistance to GINs. Fecal samples were collected from 642 Comisana sheep over three years to assess Fecal Egg Counts (FEC) with the FLOTAC technique. Using pedigree data and logn(FEC+2) as phenotypes, Estimated Breeding Values (EBVs) for GIN resistance were estimated by a BLUP animal model. The EBVs in the 99.95th and 0.05th percentiles were used to identify the most and least genetically resistant individuals to GINs, later genotyped with the Illumina OvineSNP50 beadchip. Using the software PLINK, a case-control GWAS was performed. A threshold for the 0.005% most significant FDR corrected p-values was chosen. Using the R package GALLO, QTLs associated with the significant SNPs were annotated and enriched. With the ToppGenes utility, genes associated with the SNPs have been enriched for KEGG pathways using a threshold of 0.05 FDR. As a result, 18 significant SNPs involving 13 genes were identified on 12 chromosomes. Among these, many are involved in the physiology or pathology of the gastrointestinal tract (the UGT1A* family, KIF6, LOXL2, CALN1, and TWISTNB). Others play a role in adaptive processes and production traits (LOXL2, GPC6, MYT1, and SS18L1). Among the QTLs found, six are in the Health class, five of which are associated with traits related to FEC. Enrichment analysis of the found genes highlighted 11 significant pathways involved in the regulation of the immune response, drug metabolism and detoxification, and other metabolic processes. Previous research has linked some of the discovered pathways to GIN resistance, such as Ascorbate and aldarate metabolism, and the Glucuronate pathway (uronate pathway), which are involved in the regulation of immune

response during parasitic infections in sheep, and Metabolism of xenobiotics by cytochrome P450, known to be involved in the metabolism of anthelmintic drugs.

INTRODUCTION

Gastrointestinal nematodes (GIN) are one of the most serious causes of disease in domestic ruminants worldwide (McRae et al., 2014). GIN infection poses a threat to the health and welfare of livestock and is frequently linked to financial losses, primarily as a result of subclinical diseases that reduce weight gain and milk production (Charlier et al., 2020; Glover et al., 2017; Mavrot et al., 2015). The parasitic infections cause both direct and indirect economic losses in terms of sanitary costs, reduced food intake, and nutritional deficiencies, which have a negative impact on production (Rashid et al., 2019; Sanchez et al., 2004). The economic effects of GIN infection on a national scale highlight their significance and highlight how profitable and sustainable sheep production is impossible without effective control (Chagas et al., 2022). Global estimates place the cost of treating parasites alone in the tens of billions of dollars. While powerful antiparasitic medications were available to farmers in the 1980s and 1990s, parasite resistance to anthelmintics developed swiftly (Taylor et al., 2009). Anthelmintics might penetrate the environment in a number of ways, including through the excrement of animals, the disposal of unneeded medications, and runoff from treated pastures. Once in the environment, these medications may have adverse effects on organisms that are not their intended targets and may help parasites develop drug resistance. The environmental effects of sheep anthelmintics are reviewed in terms of the possibility of residues entering the environment, their toxicity, and their impact on ecosystem functioning at an adequate scale (Beynon, 2012). There is considerable variation in resistance to gastrointestinal nematodes within and between sheep breeds, which appears to be due to underlying genetic diversity (Benavides, Sonstegard, & Tassell, 2016; McRae et al., 2014). Rapid genetic progress has been shown in both research and commercial flocks through breeding plans aimed at the selection of resistant animals (McRae et al., 2014). Breeding sheep for enhanced resistance to gastrointestinal parasites is a promising strategy to limit the use of anthelmintics due to the now widespread resistance of parasites to these molecules (Aguerre et al., 2018).

In the breeding station of Asciano, owned by Assonapa (the Italian sheep and goats breeders association), are bred two “core” flocks of the dairy sheep breeds Comisana and Massese. The primary purpose of the breeding station is to produce males to be sold to selected herds with the aim of obtaining genetic improvement in the national flock (Al Kalaldeh et al., 2019; Carta et al., 2009). To date, the breeding program of the Comisana breed has taken into account only

productive and morphological traits. Genomics makes it possible to inspect the origins of complex phenotypes that are difficult to study using only a pedigree. A clear example are phenotypes linked to gastrointestinal parasite infestation (Duchemin et al., 2012).

Selection for resistance to gastrointestinal nematode infections is usually based on the use of the egg content in the feces (Faecal Egg Count, FEC) as an indirect indicator. In many breeds, FEC EBVs are already computed and are part of their breeding programs (Aguerre et al., 2018; British Texel Sheep Society, 2017; Ciappesoni et al., 2013; Else Verbeek et al., 2011; L. Li et al., 2015). Furthermore, FEC EBVs have been proven to be a valid metric to quantify genetic resistance in sheep (Weaver et al., 2023). The aim of this work is to apply genomic techniques to inspect the genetic base of resistance to parasites based on estimated breeding values for resistance to gasto-intestinal parasites in order to identify genomic regions associated with GIN resistance in Comisana dairy sheep.

MATERIALS AND METHODS

Sampling and preliminary analysis of fecal matter

A preliminary sampling was carried out in order to detect the kind of parasites that affect the flock at the breeding station of Asciano. The most represented parasites were gastrointestinal nematodes (GIN).

Following this investigation, we made six samplings in order to investigate the seasonal trend of the infestation. The animals involved in the first two samplings were chosen among the replacement, while in the third sampling, all the ewes of the flock were sampled. For the remaining samplings, we chose to take the feces from the same animals involved in the previous ones. Due to the ongoing COVID-19 epidemiological emergency's restrictions on movement, spring and autumn sampling were skipped in 2020. The sampling sessions took place at the breeding center of Asciano in March 2019, July 2019 and January 2020; July 2020; January 2021; and May 2021. Table 1 shows the number of samples collected per sampling event. Coprological samples were collected individually from the rectal ampoule, and subsequently, the Fecal Egg Count (FEC) was determined individually using the Mini-FLOTAC technique.

SAMPLE	n
1	127
2	114

3	544
4	293
5	262
6	241

Table 1: Number of samples per sampling

Dataset construction

The data from the six samplings have been merged into one dataset, and for each individual, additional information has been retrieved from the herdbook and added to the table. Furthermore, the samplings have been grouped into three sampling seasons. The variables contained in the final dataset are Animal ID, date of birth, age at the sampling date, parity order, date of sampling, season of sampling, and FEC values.

Statistical analysis of the phenotype

Given the high number of "negative samples" (FEC = 0), the distribution of the data appears to have a strong positive asymmetry. In order to normalize the distribution of the FEC, a logarithmic transformation ln(FEC+2) was performed. In Figures *1 and 2* the density plots are visible, showing the distribution of the FEC data in the two parasites before and after the logarithmic transformation.

The analyses of the transformed phenotype "ln(FEC+2)" were performed considering the age of the animal and the date of collection. Through the software R (R Core Team, 2023), descriptive statistics and analysis of variance of the phenotype have been performed.

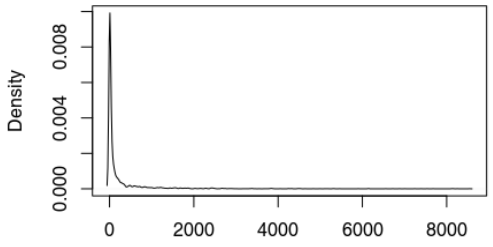


Figure 1: Distribution of the FEC

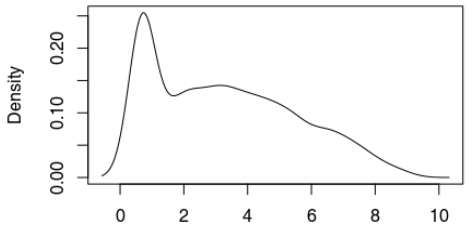


Figure 2: Distribution of logn(FEC+2)

phenotype	n	mean	sd	median	min	max	se
ln(FEC+2)	1578	3,75	2,37	3,74	0,69	9,05	0,06

Table 2: Descriptive statistics of ln(FEC+2)

Genetic parameters & Estimation of Breeding Values

Genetic parameters and estimated breeding values were calculated using the *breedR* package of the statistical programming environment R (Muñoz & Sanchez, 2020). Using the pedigree data to calculate the genetic parameters, we estimated the heritability (h2) of the phenotypic character (lnFEC+2), which indicates the severity of GIN infestation.

In a second phase, using the same package, the breeding values were calculated with the BLUP Animal Model. The random variable was the animal ID, and the fixed effects were parity order, season, and age at sampling.

Search for genotypes involved in resistance

EBVs for lnFEC(+2) were computed for all the sampled animals and for their relatives back to the third generation; as a result, the lower values indicate the best breeders for the selection of resistant animals. In order to indicate the most resistant animals to GIN with the highest EBVs, the latter have been inverted and reported on a scale of 100±5.

DNA sampling, genotyping, and quality control

The EBVs were used to rank the animals in the flock. The EBVs in the 99.95th and the 0.05th percentiles were used to identify and extract two groups of animals representing the most (31 sheep) and the least (29 sheep) genetically resistant individuals to GIN.

Consequently, biological samples have been collected from the 60 animals involved using GenoTube Livestock nasal swabs. DNA extraction was performed using the Nexttec™ "DNA Isolation Kit for Tissue and Cells" kit. The concentration of extracted DNA was assessed with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). All

animals were genotyped for 54,241 single nucleotide polymorphisms (SNPs) using the Illumina OvineSNP50K array. Using the software PLINK, the genotypes have been filtered for missingness per marker and minimum allele frequency at the standard values (--geno 0.01; --maf 0.05), and 44,524 SNPs were left.

GWAS

Using the software PLINK, a case-control GWAS (Genome-wide Association Study) was performed, considering the "less resistant" group as the case and the "most resistant" group as the control. The p-values have been adjusted by the False Discovery Rate (FDR) using the method of Benjamini & Hochberg (1995), and a threshold at the 0.005% most significant p-values was applied to identify outlying SNPs.

Enrichment

Using the R package GALLO, QTLs associated with the significant SNPs were annotated and enriched. With the ToppGenes utility, the found genes have been enriched for KEGG pathways, using a threshold of 0.05 Benjamini & Hochberg FDR.

RESULTS

Statistical analysis of phenotypes

An effect of the season on the severity of infestation is also appreciable (Figure 3). Seasonality is more marked, a peak is observed in summer. The samplings of July (2 and 4) and January (3 and 5) of the two different years have significantly different averages, an effect probably due to a different management of the pasture in the two years (Figure 4). We observed that the average and prevalence of FEC varies with age (Figure 5). This allows us to hypothesize a phenomenon of acquired immunity against GIN. ANOVA table of ln(FEC+2) against environmental variables is available in Table 3.

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**	
season	2	2457	1228.7	368.992	<2E-16	***
AGeE	1	171	171.1	51.379	1.17E-12	***
parity	6	46	7.6	2.277	0.0341	*
year	1	938	937.7	281.605	<2E-16	***

Residuals	1567	5218	3.3			
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						

Table 3: Analysis of Variance of ln(FEC+2) against season, age, parity and year of sampling

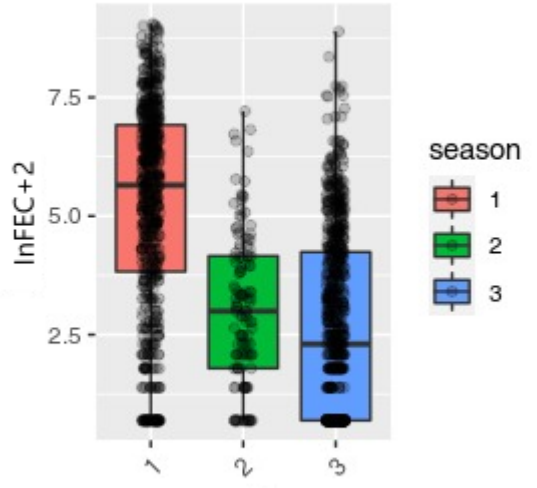


Figure 3: Boxplot of the distribution of ln(FEC+2) in the three sampling seasons

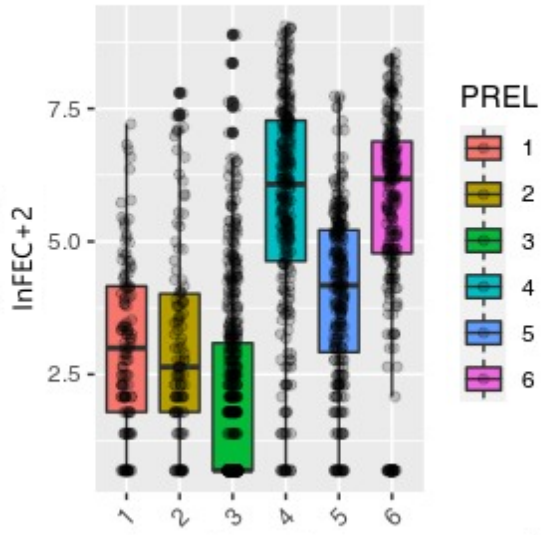


Figure 4: Boxplot of the distribution of ln(FEC+2) in the six samplings

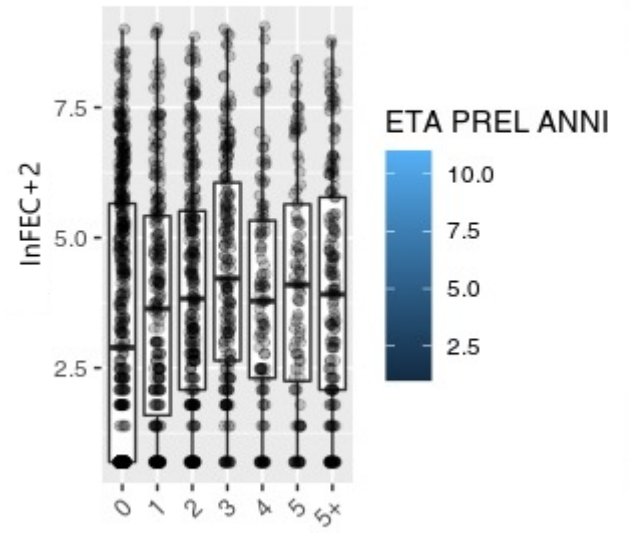


Figure 5: Distribution of ln(FEC+2) at 0, 1, 2, 3, 4, 5 and over 5 years of age of sheep

Genetic parameters

The heritability (h2) of GIN infection is 0.117. Although the heritability of the trait is low, it is still high enough to be able to calculate reliable breeding values. The values of EBV and standard error are reported in Table 4.

	min	**max**	**mean**	**median**
EBV100	83	118	100.36	100
EBV	-0.86	0.77	-0.03	-0.02
SE	0.44	0.60	0.53	0.54

Table 4: Descriptive statistics of EBVs inverted and transformed in the 100±5 (EBV100), original EBVs (EBV), and standard error (SE)

GWAS

18 significant Single Nucleotide Polymorphisms (SNPs) involving 13 genes were identified on 12 chromosomes, showing strong signals on chromosomes 1, 2, 4, 6, 13, 20 (Figure 6) that are known to contain QTLs associated parasite resistance (Benavides, Sonstegard, & Van Tassell, 2016; Diez-Tascón et al., 2002; Sayers & Sweeney, 2005). The identified SNPs fall within genes involved in the physiology or pathology of the intestinal tract including the genes of the UGT1A* family are involved in the metabolism of bilirubin and the metabolism of drugs at the

intestinal level; the genes KIF6, LOXL2 and CALN1 are involved in intestinal cancer and inflammation, immune response and in the regeneration and wounding of tissues; TWISTNB is known to have a role in gastrointestinal tract infections; LOXL2, GPC6, MYT1 and SS18L1 have a role in adaptive processes and production traits.

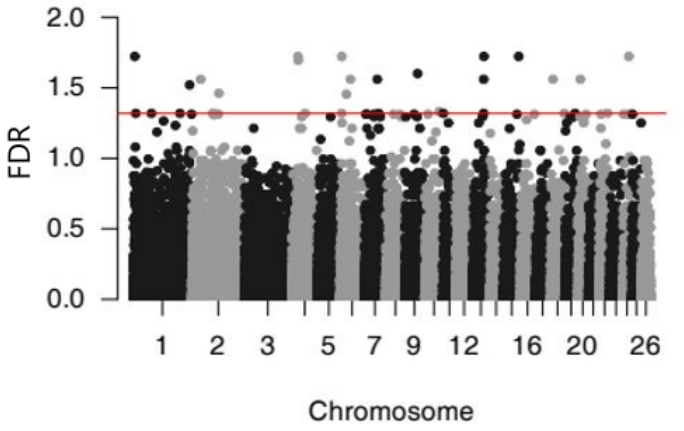


Figure 6: Manhattan plot of the GWAS analysis

Enrichment

Among the found QTLs, 6 are in the Health Class, 5 of which are associated with traits related to FEC (Table 5). Enrichment analysis of the found genes highlighted 11 significant pathways classifiable as involved in the regulation of the immune response, involved in drug metabolism and detoxification, and involved in other metabolic processes. Previous research has linked some of the discovered pathways to GIN resistance, such as Ascorbate and aldarate metabolism and Glucuronate pathway (uronate pathway), which are involved in the regulation of immune response during parasitic infections in sheep and Metabolism of xenobiotics by cytochrome P450, known to be involved in the metabolism of anthelmintic drugs.

Total_QTLs_db	**p value**	**QTL type**	**ID**
285	0.014	Health	Trichostrongylus colubriformis FEC - CHR1
416	0.043	Health	Trichostrongylus adult and larva count - CHR2
132	0.060	Health	Haemonchus contortus FEC - CHR4
240	0.071	Health	facial eczema susceptibility - CHR6
240	0.784	Health	Fecal egg count - CHR6

156	0.006	Health	Fecal egg count - CHR10

Table 5: QTLs related to Health containing outlier SNPs

ID	Name	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome	Hit in Query List
82932	Ascorbate and aldarate metabolism	8.938E-10	5	27	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
997799	Glucuronate pathway (uronate pathway)	8.938E-10	5	27	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
82929	Pentose and glucuronate interconversions	0.000000002044	5	34	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
83021	Porphyrin and chlorophyll metabolism	0.000000004662	5	42	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
83033	Drug metabolism - other enzymes	0.000000005993	5	46	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
82940	Steroid hormone biosynthesis	0.00000001656	5	58	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
83020	Retinol metabolism	0.00000002547	5	65	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
83032	Drug metabolism - cytochrome P450	0.00000003254	5	70	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
83031	Metabolism of xenobiotics by cytochrome P450	0.00000003839	5	74	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
673221	Chemical carcinogenesis	0.00000005821	5	82	UGT1A4, UGT1A1, UGT1A3, UGT1A9, UGT1A6
132956	Metabolic pathways	0.0005975	7	1272	UGT1A4, UGT1A1, UGT1A3, POLR1F, UGT1A9, DCT, UGT1A6

Table 6: Enriched KEGG pathways containing genes associated with the outlying SNPs

DISCUSSION

Gastrointestinal nematodes (GIN) are one of the most serious causes of disease in domestic ruminants worldwide. GIN infection causes both direct and indirect economic losses, with global estimates placing the cost of treating parasites alone in the tens of billions of dollars. Anthelmintics may penetrate the environment through excrement, disposal of unneeded medications, and runoff from treated pastures and may have adverse effects on organisms other than their intended targets. The environmental effects of sheep anthelmintics are reviewed in terms of residues entering the environment, their toxicity, and their impact on ecosystem functioning. Breeding sheep for enhanced resistance to gastrointestinal parasites is a promising strategy to limit the use of anthelmintics. This study aimed to identify genomic regions linked to GIN resistance in Comisana dairy sheep using FEC estimated breeding values (EBV) as a genetic resistance metric. The results of a statistical analysis of phenotypes revealed that the average and prevalence of FEC varies with age, in line with previous studies (Good et al., 2006). This may indicate an acquired resistance during the life of the animal. An effect of the season on the severity of infestation is also appreciable. Genetic parameters showed that the heritability (h2) of GIN infection is 0.117; slightly lower than values found in literature that put h2 for FEC in a range from 0.18 to 0.23 (Ngere et al., 2018), this difference may be imputed to the different GIN species found in the different geographic areas, to the small size of the sample and to the fact that not all planned samplings were undergone because of the COVID-19 restrictions. GWAS identified 18 significant Single Nucleotide Polymorphisms (SNPs) involving 13 genes on 12 chromosomes, showing strong signals on chromosomes 1, 2, 4, 6, 13, 20 that are known to contain QTLs associated parasite resistance. The identified SNPs fall within genes involved in the physiology or pathology of the intestinal tract, including the genes of the UGT1A* family known to be associated with epithelial, metabolic and innate immunity (Xiang et al., 2016), anthelmintic drugs metabolism (Bártíková et al., 2010), metabolism of dietary and environmental chemicals (Basu et al., 2004) and other metabolic processes involving the gastrointestinal tract in human and sheep (Maté et al., 2010; Peer et al., 2012; Sauer et al., 2014; Strassburg et al., 1999; Wells et al., 2004) the genes GPC6, KIF6, LOXL2, SS18L1 and CALN1 are involved in intestinal cancer and inflammation (Gan et al., 2015; Gromova et al., 2009; H. Li et al., 2014; Matsumoto et al., 2017; Tokunaga et al., 2019), immune response (Matsumoto et al., 2017; Ariel et al., 2020) and in the regeneration and wounding of tissues (Hein et al., 2001; Soulet et al., 2013; Matsumoto et al., 2017; Keil et al., 2021); while the gene TWISTNB has a role in the resistance to viral infections as Helicobacter pylori in humans (Sohn & Lee, 2011); LOXL2, GPC6, MYT1 and SS18L1 also have a role in

adaptive processes to heat tolerance (Cheruiyot et al., 2021; Passamonti et al., 2021) and production traits (Du et al., 2020; Guðmundsdóttir 1988-, 2015; Hu et al., 2022; Nayeri et al., 2017; Weber et al., 2016; Zhang et al., 2021) in livestock. The set of genes enlightened by our analyses is coherent with a supposed multifactorial resistance to parasitic infection composed of a pool of metabolic, immunological, and regenerative factors that all together protect the animal from the deleterious effects of GIN infections.

The QTL enrichment results are in line with the previous statement, the analysis found that the outlier SNPs are related with 60 QTLs, most of them related to production traits, 6 QTLs in the Health Class, 5 of which are associated with traits related to FEC. Confirming the effectiveness of the EBVs calculated in the selection of QTLs involved in GIN resistance. The pathways enrichment analysis of the found genes highlighted 11 significant KEGG pathways classifiable as involved in the regulation of the immune response, drug metabolism and detoxification, and other metabolic processes, in line with the previously discussed results. As an example ascorbate and aldarate metabolism, which has been previously associated to GIN resistance in sheep (Kumarasingha et al., 2019; Laing et al., 2013; Wang et al., 2022), this pathway is involved in the metabolism of vitamin C which has undoubtedly effects on the immune system of the animal, potentially leading to different immune response (Dede et al., 2000; Osman & I. M. Gaede, 2012). A further example is the glucuronate pathway (also known as the uronate pathway), a metabolic pathway that plays an important role in the synthesis and metabolism of various carbohydrates and other compounds in animals. While the exact role of the glucuronate pathway in gastrointestinal parasite resistance is not yet fully understood, some studies suggest that it may play a role in modulating the immune response to parasitic infections, in fact its metabolites are involved in the immune response through the activation of inflammatory cytokines (Prasanphanich et al., 2013; van Stijn et al., 2010). The pathway of metabolism of xenobiotics by cytochrome P450 includes a family of enzymes involved in the metabolism of drugs and other xenobiotics. While its exact role in gastrointestinal nematode resistance is not clear, some studies suggest that cytochrome P450 enzymes may play a role in modulating the immune response to parasitic infections (Galtier et al., 1991; Kellerová et al., 2019; Laing et al., 2015).

CONCLUSIONS

In this work, we made use of the Estimated Breeding Values (EBVs) as a metric to identify the most and least genetically resistant animals to gastrointestinal strongyles. Since most of the markers highlighted by the GWAS analysis fall within genomic regions involved in the

physiology and pathology of the gastrointestinal tract, our first outcome is that the identified regions may have a role in the resistance to gastrointestinal nematodes and other parasitic infections. Two enrichment analyses gave confirming results, clarifying which metabolic pathways involve the outlying regions. The application of this knowledge in sheep breeding may help reduce the use of anthelmintic drugs, resulting in economic and environmental benefits. Future directions would be to extend this approach to other breeds to assess the replicability of the results.

References

Aguerre, S., Jacquiet, P., Brodier, H., Bournazel, J. P., Grisez, C., Prévot, F., Michot, L., Fidelle, F., Astruc, J. M., & Moreno, C. R. (2018). Resistance to gastrointestinal nematodes in dairy sheep: Genetic variability and relevance of artificial infection of nucleus rams to select for resistant ewes on farms. *Veterinary Parasitology*, *256*, 16–23. https://doi.org/10.1016/j.vetpar.2018.04.004

Al Kalaldeh, M., Gibson, J., Duijvesteijn, N., Daetwyler, H. D., MacLeod, I., Moghaddar, N., Lee, S. H., & van der Werf, J. H. J. (2019). Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. *Genetics Selection Evolution*, *51*(1), 32. https://doi.org/10.1186/s12711-019-0476-4

Ariel, O., Gendron, D., Dudemaine, P.-L., Gévry, N., Ibeagha-Awemu, E. M., & Bissonnette, N. (2020). Transcriptome Profiling of Bovine Macrophages Infected by Mycobacterium avium spp. Paratuberculosis Depicts Foam Cell and Innate Immune Tolerance Phenotypes. *Frontiers in Immunology*, *10*. https://www.frontiersin.org/articles/10.3389/fimmu.2019.02874

Bártíková, H., Křížová, V., Štěpničková, M., Lamka, J., Kubíček, V., Skálová, L., & Szotáková, B. (2010). Activities of biotransformation enzymes and flubendazole metabolism in lambs (Ovis aries): Effect of gender and flubendazole therapy. *Pharmacological Reports*, *62*(2), 362–373. https://doi.org/10.1016/S1734-1140(10)70276-0

Basu, N. K., Ciotti, M., Hwang, M. S., Kole, L., Mitra, P. S., Cho, J. W., & Owens, I. S. (2004).

Differential and Special Properties of the Major Human UGT1-encoded Gastrointestinal UDP-glucuronosyltransferases Enhance Potential to Control Chemical Uptake*. *Journal of Biological Chemistry*, *279*(2), 1429–1441. https://doi.org/10.1074/jbc.M306439200

Benavides, M. V., Sonstegard, T. S., & Tassell, C. V. (2016). Genomic Regions Associated with Sheep Resistance to Gastrointestinal Nematodes. *Trends in Parasitology*, *32*(6), 470–480. https://doi.org/10.1016/j.pt.2016.03.007

Benavides, M. V., Sonstegard, T. S., & Van Tassell, C. (2016). Genomic Regions Associated with Sheep Resistance to Gastrointestinal Nematodes. *Trends in Parasitology*, *32*(6), 470–480. https://doi.org/10.1016/j.pt.2016.03.007

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beynon, S. A. (2012). Potential environmental consequences of administration of anthelmintics to sheep. *Veterinary Parasitology*, *189*(1), 113–124. https://doi.org/10.1016/j.vetpar.2012.03.040

British Texel Sheep Society. (2017). *Estimated Breeding Values (EBVs) – British Texel Sheep Society Ltd*. https://texel.uk/prt/breed-development-9/

Carta, A., Casu, S., & Salaris, S. (2009). Invited review: Current state of genetic improvement in dairy sheep. *Journal of Dairy Science*, *92*(12), 5814–5833. https://doi.org/10.3168/jds.2009-2479

Chagas, A. C. de S., Tupy, O., Santos, I. B. dos, & Esteves, S. N. (2022). Economic impact of gastrointestinal nematodes in Morada Nova sheep in Brazil. *Revista Brasileira de Parasitologia Veterinária*, *31*, e008722. https://doi.org/10.1590/S1984-29612022044

Charlier, J., Rinaldi, L., Musella, V., Ploeger, H. W., Chartier, C., Vineer, H. R., Hinney, B., von Samson-Himmelstjerna, G., Băcescu, B., Mickiewicz, M., Mateus, T. L.,

Martinez-Valladares, M., Quealy, S., Azaizeh, H., Sekovska, B., Akkari, H., Petkevicius, S., Hektoen, L., Höglund, J., … Claerebout, E. (2020). Initial assessment of the economic burden of major parasitic helminth infections to the ruminant livestock industry in Europe. *Preventive Veterinary Medicine*, *182*, 105103. https://doi.org/10.1016/j.prevetmed.2020.105103

Cheruiyot, E. K., Haile-Mariam, M., Cocks, B. G., MacLeod, I. M., Xiang, R., & Pryce, J. E. (2021). New loci and neuronal pathways for resilience to heat stress in cattle. *Scientific Reports*, *11*(1), Articolo 1. https://doi.org/10.1038/s41598-021-95816-8

Ciappesoni, G., Goldberg, V., & Gimeno, D. (2013). Estimates of genetic parameters for worm resistance, wool and growth traits in Merino sheep of Uruguay. *Livestock Science*, *157*(1), 65–74. https://doi.org/10.1016/j.livsci.2013.07.011

Dede, S., Değer, Y., Değer, S., & Alkan, M. (2000). Determination of the status of lipid peroxidation and antioxidants in sheep infected with certain endoparasites (Fasciola spp., Trichostrongylidae spp., Eimeria spp.). *Türkiye Parazitoloji Dergisi*, *24*(2), 190–193.

Diez-Tascón, C., MacDonald, P. A., Dodds, K. G., McEwan, J. C., & Crawford, A. M. (2002). A screen of chromosome 1 for QTL affecting nematode resistance in an ovine outcross population. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August, 2002. Session 13*, 1–4.

Du, Y., Liu, L., He, Y., Dou, T., Jia, J., & Ge, C. (2020). Endocrine and genetic factors affecting egg laying performance in chickens: A review. *British Poultry Science*, *61*(5), 538–549. https://doi.org/10.1080/00071668.2020.1758299

Duchemin, S. I., Colombani, C., Legarra, A., Baloche, G., Larroque, H., Astruc, J.-M., Barillet, F., Robert-Granié, C., & Manfredi, E. (2012). Genomic selection in the French Lacaune dairy sheep breed. *Journal of Dairy Science*, *95*(5), 2723–2733. https://doi.org/10.3168/jds.2011-4980

Else Verbeek, Egbert Kanis, Rawlynce C Bett, & Isaac S Kosgey. (2011). Optimisation of breeding schemes for litter size, lambing interval, body weight and parasite resistance for sheep in Kenya. *Livestock Research for Rural Development*, *23*, 9.

Galtier, P., Alvinerie, M., Plusquellec, Y., Tufenkji, A. E., & Houin, G. (1991). Decrease in albendazole sulphonation during experimental fascioliasis in sheep. *Xenobiotica*, *21*(7), 917–924. https://doi.org/10.3109/00498259109039531

Gan, L., Xu, M., Zhang, Y., Zhang, X., & Guo, W. (2015). Focusing on long noncoding RNA dysregulation in gastric cancer. *Tumor Biology*, *36*(1), 129–141. https://doi.org/10.1007/s13277-014-2894-9

Glover, M., Clarke, C., Nabb, L., & Schmidt, J. (2017). Anthelmintic efficacy on sheep farms in south-west England. *Veterinary Record*, *180*(15), 378–378. https://doi.org/10.1136/vr.104151

Good, B., Hanrahan, J. P., Crowley, B. A., & Mulcahy, G. (2006). Texel sheep are more resistant to natural nematode challenge than Suffolk sheep based on faecal egg count and nematode burden. *Veterinary Parasitology*, *136*(3), 317–327. https://doi.org/10.1016/j.vetpar.2005.12.001

Gromova, P., Ralea, S., Lefort, A., Libert, F., Rubin, B. P., Erneux, C., & Vanderwinden, J.-M. (2009). Kit K641E oncogene up-regulates Sprouty homolog 4 and Trophoblast glycoprotein in interstitial cells of Cajal in a murine model of gastrointestinal stromal tumours. *Journal of Cellular and Molecular Medicine*, *13*(8a), 1536–1548. https://doi.org/10.1111/j.1582-4934.2009.00768.x

Guðmundsdóttir 1988-, Ó. Ó. (2015). *Genome-wide association study of muscle traits in Icelandic sheep* [Thesis]. https://skemman.is/handle/1946/20392

Hein, S., Yamamoto, S. Y., Okazaki, K., Jourdan-Lesaux, C., Csiszar, K., & Bryant-Greenwood, G. D. (2001). Lysyl Oxidases: Expression in the Fetal Membranes and Placenta.

Placenta, *22*(1), 49–57. https://doi.org/10.1053/plac.2000.0580

Hu, L., Zhang, L., Li, Q., Liu, H., Xu, T., Zhao, N., Han, X., Xu, S., Zhao, X., & Zhang, C. (2022). Genome-wide analysis of CNVs in three populations of Tibetan sheep using whole-genome resequencing. *Frontiers in Genetics*, *13*, 971464. https://doi.org/10.3389/fgene.2022.971464

Keil, S., Gupta, M., Brand, M., & Knopf, F. (2021). Heparan sulfate proteoglycan expression in the regenerating zebrafish fin. *Developmental Dynamics*, *250*(9), 1368–1380. https://doi.org/10.1002/dvdy.321

Kellerová, P., Matoušková, P., Lamka, J., Vokřál, I., Szotáková, B., Zajíčková, M., Pasák, M., & Skálová, L. (2019). Ivermectin-induced changes in the expression of cytochromes P450 and efflux transporters in Haemonchus contortus female and male adults. *Veterinary Parasitology*, *273*, 24–31. https://doi.org/10.1016/j.vetpar.2019.07.006

Kumarasingha, R., Young, N. D., Yeo, T.-C., Lim, D. S. L., Tu, C.-L., Palombo, E. A., Shaw, J. M., Gasser, R. B., & Boag, P. R. (2019). Transcriptional alterations in Caenorhabditis elegans following exposure to an anthelmintic fraction of the plant Picria fel-terrae Lour. *Parasites & Vectors*, *12*(1), 181. https://doi.org/10.1186/s13071-019-3429-4

Laing, R., Bartley, D. J., Morrison, A. A., Rezansoff, A., Martinelli, A., Laing, S. T., & Gilleard, J. S. (2015). The cytochrome P450 family in the parasitic nematode Haemonchus contortus. *International Journal for Parasitology*, *45*(4), 243–251. https://doi.org/10.1016/j.ijpara.2014.12.001

Laing, R., Kikuchi, T., Martinelli, A., Tsai, I. J., Beech, R. N., Redman, E., Holroyd, N., Bartley, D. J., Beasley, H., Britton, C., Curran, D., Devaney, E., Gilabert, A., Hunt, M., Jackson, F., Johnston, S. L., Kryukov, I., Li, K., Morrison, A. A., … Cotton, J. A. (2013). The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. *Genome Biology*, *14*(8), R88.

https://doi.org/10.1186/gb-2013-14-8-r88

Li, H., Yu, B., Li, J., Su, L., Yan, M., Zhu, Z., & Liu, B. (2014). Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget*, *5*(8), 2318–2329. https://doi.org/10.18632/oncotarget.1913

Li, L., Swan, A., Brown, D., & Werf, J. (2015). *Australian Sheep Breeding Values for Worm Egg Count Retain Predictive Power Across Flocks in the Presence of GxE*.

Maté, L., Virkel, G., Lifschitz, A., Sallovitz, J., Ballent, M., & Lanusse, C. (2010). Phase 1 and phase 2 metabolic activities along the small intestine in adult male sheep1. *Journal of Veterinary Pharmacology and Therapeutics*, *33*(6), 537–545. https://doi.org/10.1111/j.1365-2885.2010.01177.x

Matsumoto, K., Nakai, Y., Hoshino, M., Yamazaki, K., Takioto, Y., Takadera, S., Nakagawa, T., Nishimura, R., & Kusakabe, M. (2017). Comprehensive DNA microarray expression profiles of tumors in tenascin-C-knockout mice. *Bioscience, Biotechnology, and Biochemistry*, *81*(10), 1926–1936. https://doi.org/10.1080/09168451.2017.1362975

Mavrot, F., Hertzberg, H., & Torgerson, P. (2015). Effect of gastro-intestinal nematode infection on sheep performance: A systematic review and meta-analysis. *Parasites & Vectors*, *8*(1), 557. https://doi.org/10.1186/s13071-015-1164-z

McRae, K. M., McEwan, J. C., Dodds, K. G., & Gemmell, N. J. (2014). Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genomics*, *15*(1), 637. https://doi.org/10.1186/1471-2164-15-637

Muñoz, F., & Sanchez, L. (2020). *BreedR: Statistical methods for forest genetic resources analysts* (https://doi.org/10.5281/zenodo.546023; R package version 0.12-5) [R]. Zenodo. https://github.com/famuvie/breedR

Nayeri, S., Sargolzaei, M., Abo-Ismail, M. K., Miller, S., Schenkel, F., Moore, S. S., & Stothard, P. (2017). Genome-wide association study for lactation persistency, female fertility,

longevity, and lifetime profit index traits in Holstein dairy cattle. *Journal of Dairy Science*, *100*(2), 1246–1258. https://doi.org/10.3168/jds.2016-11770

Ngere, L., Burke, J. M., Morgan, J. L. M., Miller, J. E., & Notter, D. R. (2018). Genetic parameters for fecal egg counts and their relationship with body weights in Katahdin lambs. *Journal of Animal Science*, *96*(5), 1590–1599. https://doi.org/10.1093/jas/sky064

Osman, F. A., & I. M. Gaede, H. (2012). Evaluation of the Concentrations of Products of Nitric Oxide Oxidation and some Antioxidant Vitamins in Sheep Naturally Infested with Parasites. *Assiut Veterinary Medical Journal*, *58*(134), 1–6. https://doi.org/10.21608/avmj.2012.171892

Passamonti, M. M., Somenzi, E., Barbato, M., Chillemi, G., Colli, L., Joost, S., Milanesi, M., Negrini, R., Santini, M., Vajana, E., Williams, J. L., & Ajmone-Marsan, P. (2021). The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock. *Animals*, *11*(10), Articolo 10. https://doi.org/10.3390/ani11102833

Peer, C. J., Sissung, T. M., Kim, A., Jain, L., Woo, S., Gardner, E. R., Kirkland, C. T., Troutman, S. M., English, B. C., Richardson, E. D., Federspiel, J., Venzon, D., Dahut, W., Kohn, E., Kummar, S., Yarchoan, R., Giaccone, G., Widemann, B., & Figg, W. D. (2012). Sorafenib Is an Inhibitor of UGT1A1 but Is Metabolized by UGT1A9: Implications of Genetic Variants on Pharmacokinetics and Hyperbilirubinemia. *Clinical Cancer Research*, *18*(7), 2099–2107. https://doi.org/10.1158/1078-0432.CCR-11-2484

Prasanphanich, N., Mickum, M., Heimburg-Molinaro, J., & Cummings, R. (2013). Glycoconjugates in Host-Helminth Interactions. *Frontiers in Immunology*, *4*. https://www.frontiersin.org/articles/10.3389/fimmu.2013.00240

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rashid, M., Rashid, M. I., Akbar, H., Ahmad, L., Hassan, M. A., Ashraf, K., Saeed, K., &

Gharbi, M. (2019). A systematic review on modelling approaches for economic losses studies caused by parasites and their associated diseases in cattle. *Parasitology*, *146*(2), 129–141. https://doi.org/10.1017/S0031182018001282

Sanchez, J., Dohoo, I., Carrier, J., & DesCôteaux, L. (2004). A meta-analysis of the milk-production response after anthelmintic treatment in naturally infected adult dairy cows. *Preventive Veterinary Medicine*, *63*(3), 237–256. https://doi.org/10.1016/j.prevetmed.2004.01.006

Sauer, V., Roy-Chowdhury, N., & Roy-Chowdhury, J. (2014). Monogenic Liver Diseases. In L. M. McManus & R. N. Mitchell (A c. Di), *Pathobiology of Human Disease* (pp. 1857–1865). Academic Press. https://doi.org/10.1016/B978-0-12-386456-7.04210-6

Sayers, G., & Sweeney, T. (2005). Gastrointestinal nematode infection in sheep—A review of the alternatives to anthelmintics in parasite control. *Animal Health Research Reviews*, *6*(2), 159–171. https://doi.org/10.1079/ahr2005108

Sohn, S.-H., & Lee, Y.-C. (2011). The genome-wide expression profile of gastric epithelial cells infected by naturally occurring cagA isogenic strains of Helicobacter pylori. *Environmental Toxicology and Pharmacology*, *32*(3), 382–389. https://doi.org/10.1016/j.etap.2011.08.006

Soulet, F., Kilarski, W. W., Roux-Dalvai, F., Herbert, J. M. J., Sacewicz, I., Mouton-Barbosa, E., Bicknell, R., Lalor, P., Monsarrat, B., & Bikfalvi, A. (2013). Mapping the Extracellular and Membrane Proteome Associated with the Vasculature and the Stroma in the Embryo*. *Molecular & Cellular Proteomics*, *12*(8), 2293–2312. https://doi.org/10.1074/mcp.M112.024075

Strassburg, C. P., Strassburg, A., Nguyen, N., LI, Q., Manns, M. P., & Tukey, R. H. (1999). Regulation and function of family 1 and family 2 UDP-glucuronosyltransferase genes (UGT1A, UGT2B) in human oesophagus. *Biochemical Journal*, *338*(2), 489–498.

https://doi.org/10.1042/bj3380489

Taylor, M. A., Learmount, J., Lunn, E., Morgan, C., & Craig, B. H. (2009). Multiple resistance to anthelmintics in sheep nematodes and comparison of methods used for their detection. *Small Ruminant Research*, *86*(1), 67–70. https://doi.org/10.1016/j.smallrumres.2009.09.020

Tokunaga, R., Xiu, J., Goldberg, R. M., Philip, P. A., Seeber, A., Battaglin, F., Arai, H., Lo, J. H., Puccini, A., Naseem, M., Berger, M. D., Soni, S., Zhang, W., Chen, W. S., Hwang, J. J., Shields, A. F., Marshall, J., Baba, H., Korn, W. M., & Lenz, H.-J. (2019). Gene mutations of SWI/SNF complex and molecular profile in colorectal cancer. *Journal of Clinical Oncology*, *37*(15_suppl), 3600–3600. https://doi.org/10.1200/JCO.2019.37.15_suppl.3600

van Stijn, C. M. W., van den Broek, M., Vervelde, L., Alvarez, R. A., Cummings, R. D., Tefsen, B., & Die, I. van. (2010). Vaccination-induced IgG response to Galα1–3GalNAc glycan epitopes in lambs protected against Haemonchus contortus challenge infection. *International Journal for Parasitology*, *40*(2), 215–222. https://doi.org/10.1016/j.ijpara.2009.07.009

Wang, X., Zhang, Z., Li, B., Hao, W., Yin, W., Ai, S., Han, J., Wang, R., & Duan, Z. (2022). Depicting Fecal Microbiota Characteristic in Yak, Cattle, Yak-Cattle Hybrid and Tibetan Sheep in Different Eco-Regions of Qinghai-Tibetan Plateau. *Microbiology Spectrum*, *10*(4), e00021-22. https://doi.org/10.1128/spectrum.00021-22

Weaver, A. R., Wright, D. L., Greiner, S. P., & Bowdridge, S. A. (2023). Effect of sire fecal egg count estimated breeding value on parasite resistance traits in Haemonchus contortus infected Katahdin lambs. *Small Ruminant Research*, *223*, 106970. https://doi.org/10.1016/j.smallrumres.2023.106970

Weber, K. L., Welly, B. T., Eenennaam, A. L. V., Young, A. E., Porto-Neto, L. R., Reverter, A.,

& Rincon, G. (2016). Identification of Gene Networks for Residual Feed Intake in Angus Cattle Using Genomic Prediction and RNA-seq. *PLOS ONE*, *11*(3), e0152274. https://doi.org/10.1371/journal.pone.0152274

Wells, P. G., Mackenzie, P. I., Chowdhury, J. R., Guillemette, C., Gregory, P. A., Ishii, Y., Hansen, A. J., Kessler, F. K., Kim, P. M., Chowdhury, N. R., & Ritter, J. K. (2004). Glucuronidation and the Udp-Glucuronosyltransferases in Health and Disease. *Drug Metabolism and Disposition*, *32*(3), 281–290. https://doi.org/10.1124/dmd.32.3.281

Xiang, R., Oddy, V. H., Archibald, A. L., Vercoe, P. E., & Dalrymple, B. P. (2016). Epithelial, metabolic and innate immunity transcriptomic signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues. *PeerJ*, *4*, e1762. https://doi.org/10.7717/peerj.1762

Zhang, H., Liu, A., Wang, Y., Luo, H., Yan, X., Guo, X., Li, X., Liu, L., & Su, G. (2021). Genetic Parameters and Genome-Wide Association Studies of Eight Longevity Traits Representing Either Full or Partial Lifespan in Chinese Holsteins. *Frontiers in Genetics*, *12*. https://www.frontiersin.org/articles/10.3389/fgene.2021.634986

6.2. Additional Research

The works listed in this section are not a part of my thesis, but they made a significant contribution to the growth of my knowledge, skills, and ability to train those abilities that led to the creation of this work.

Senczuk *et al. Genet Sel Evol* (2021) 53:48
https://doi.org/10.1186/s12711-021-00639-w

GSE Genetics Selection Evolution

RESEARCH ARTICLE Open Access

On the origin and diversification of Podolian cattle breeds: testing scenarios of European colonization using genome-wide SNP data

Gabriele Senczuk[1*], Salvatore Mastrangelo[2], Paolo Ajmone-Marsan[3], Zsolt Becskei[4], Paolo Colangelo[5], Licia Colli[3], Luca Ferretti[6], Taki Karsli[7], Hovirag Lancioni[8], Emiliano Lasagna[9], Donata Marletta[10], Christian Persichilli[1], Baldassare Portolano[2], Francesca M. Sarti[9], Elena Ciani[11] and Fabio Pilla[1]

Abstract

Background: During the Neolithic expansion, cattle accompanied humans and spread from their domestication centres to colonize the ancient world. In addition, European cattle occasionally intermingled with both indicine cattle and local aurochs resulting in an exclusive pattern of genetic diversity. Among the most ancient European cattle are breeds that belong to the so-called Podolian trunk, the history of which is still not well established. Here, we used genome-wide single nucleotide polymorphism (SNP) data on 806 individuals belonging to 36 breeds to reconstruct the origin and diversification of Podolian cattle and to provide a reliable scenario of the European colonization, through an approximate Bayesian computation random forest (ABC-RF) approach.

Results: Our results indicate that European Podolian cattle display higher values of genetic diversity indices than both African taurine and Asian indicine breeds. Clustering analyses show that Podolian breeds share close genomic relationships, which suggests a likely common genetic ancestry. Among the simulated and tested scenarios of the colonization of Europe from taurine cattle, the greatest support was obtained for the model assuming at least two waves of diffusion. Time estimates are in line with an early migration from the domestication centre of non-Podolian taurine breeds followed by a secondary migration of Podolian breeds. The best fitting model also suggests that the Italian Podolian breeds are the result of admixture between different genomic pools.

Conclusions: This comprehensive dataset that includes most of the autochthonous cattle breeds belonging to the so-called Podolian trunk allowed us not only to shed light onto the origin and diversification of this group of cattle, but also to gain new insights into the diffusion of European cattle. The most well-supported scenario of colonization points to two main waves of migrations: with one that occurred alongside with the Neolithic human expansion and gave rise to the non-Podolian taurine breeds, and a more recent one that favoured the diffusion of European Podolian. In this process, we highlight the importance of both the Mediterranean and Danube routes in promoting European cattle colonization. Moreover, we identified admixture as a driver of diversification in Italy, which could represent a melting pot for Podolian cattle.

Background

The tangled evolutionary history of cattle stems from two main wild aurochs subspecies, *Bos primigenius primigenius* and *Bos primigenius namadicus* which were widespread across Africa and Eurasia during Middle Pleistocene. They diverged approximately 250,000 years

*Correspondence: g.senczuk@unimol.it
[1] Department of Agricultural, Environmental and Food Sciences, University of Molise, 86100 Campobasso, Italy
Full list of author information is available at the end of the article

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

genes

MDPI

Article

Seven Shades of Grey: A Follow-Up Study on the Molecular Basis of Coat Colour in Indicine Grey Cattle Using Genome-Wide SNP Data

Gabriele Senczuk [1], Vincenzo Landi [2], Salvatore Mastrangelo [3], Christian Persichilli [1], Fabio Pilla [1] and Elena Ciani [4,*]

[1] Department of Agricultural, Environmental and Food Sciences, University of Molise, 86100 Campobasso, Italy
[2] Department of Veterinary Medicine, University of Bari "Aldo Moro", 70010 Valenzano, Italy
[3] Department of Agricultural, Food and Forest Science, University of Palermo, 90128 Palermo, Italy
[4] Department of Biosciences Biotechnologies and Biopharmaceutics, University of Bari "Aldo Moro", 70125 Bari, Italy
* Correspondence: elena.ciani@uniba.it

Abstract: Shades of grey and brown are a dominant component in mammal coat colours, representing a fundamental trait involved in a great number of processes including cryptism, sexual selection and signalling. The genetic mechanisms of the grey colouration in mammals are very complex and controlled by hundreds of genes whose effects and interactions are still largely unclear. In this study, we adopted a robust multi-cohort F_{st} outlier approach based on pairwise contrasts between seven grey indicine cattle breeds and both taurine and indicine non-grey cattle breeds in order to find genomic regions potentially related to the grey colouration. On the basis of three main drawn settings, built in order to control both the effect of the sample size and the genetic structure, we have identified some signals common to those obtained in a previous work employing only taurine cattle. In particular, using the top 1% F_{st} approach, we detected a candidate region (22.6–23.8 megabases) on chromosome 14 in which genes related to pigmentation have been already documented. In addition, when we constructed a phylogenetic tree using the significant markers identified in this study and including also the genotyping data at these loci of both the grey taurine and the extinct wild auroch, we found a topological repartition consistent with breed colour pattern rather than with the known bovine evolutionary history. Thus, on the basis of this evidence, together with the geographical distribution of the current taurine grey cattle, an ancestral indicine origin for the grey phenotype would seem to be a conceivable interpretation. In this context, a higher thermo-tolerance and less UV-induced damage of the grey phenotype might have favoured the retention of advantageous genes into the taurine genome during the post-Neolithic human-mediated cattle expansions.

Keywords: coat colour; cattle; *Bos taurus indicus*; selection signatures; F_{st}

check for updates

Citation: Senczuk, G.; Landi, V.; Mastrangelo, S.; Persichilli, C.; Pilla, F.; Ciani, E. Seven Shades of Grey: A Follow-Up Study on the Molecular Basis of Coat Colour in Indicine Grey Cattle Using Genome-Wide SNP Data. *Genes* **2022**, *13*, 1601. https://doi.org/10.3390/genes13091601

Academic Editor: Ino Curik

Received: 26 May 2022
Accepted: 31 August 2022
Published: 7 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The great majority of coat colours in mammals are shades of grey and brown as a consequence of both cryptism and concealment [1]. However, coat colour in mammals has also a relevant role for other important processes such as physiological processes (UV- and thermo-tolerance), sexual selection (mate choice) and signalling, both to conspecifics and predators [2,3].

In domestic animals, strong artificial selection has led to a wide range of coat colours, where grey and brown pigmentation still remain dominant. Up to now, the genetic mechanisms of pigmentation in mammals are still largely unknown due to the complex inheritance, consisting of more than 150 genes related to the synthesis, distribution and density of eumelanin and pheomelanin pigments and its age-dependent expression [4,5]. In livestock species, a small number of genes have been recognised as being responsible for the grey

Genetics Selection Evolution

RESEARCH ARTICLE

Open Access

A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds

Simone Ceccobelli[1*], Vincenzo Landi[2], Gabriele Senczuk[3], Salvatore Mastrangelo[4], Maria Teresa Sardina[4], Slim Ben-Jemaa[5], Christian Persichilli[3], Taki Karsli[6], Valentin-Adrian Bâlteanu[7], María Agustina Raschia[8], Mario Andrés Poli[8], Gabriel Ciappesoni[9], Farai Catherine Muchadeyi[10], Edgar Farai Dzomba[11], Nokuthula Winfred Kunene[12], Gesine Lühken[13], Tatiana Evgenievna Deniskova[14], Arsen Vladimirovich Dotsev[14], Natalia Anatolievna Zinovieva[14], Attila Zsolnai[15], István Anton[15], Szilvia Kusza[16], Nuno Carolino[17], Fátima Santos-Silva[17], Aldona Kawęcka[18], Marcin Świątek[19], Roman Niżnikowski[19], Marija Špehar[20], Gabriel Anaya[21], Antonio Granero[22], Tiago Perloiro[23], Pedro Cardoso[24], Silverio Grande[25], Beatriz López de los Santos[26], Coralie Danchin-Burge[27], Marina Pasquini[1], Amparo Martínez Martínez[28], Juan Vicente Delgado Bermejo[28], Emiliano Lasagna[29], Elena Ciani[30], Francesca Maria Sarti[29] and Fabio Pilla[3]

Abstract

Background To enhance and extend the knowledge about the global historical and phylogenetic relationships between Merino and Merino-derived breeds, 19 populations were genotyped with the OvineSNP50 BeadChip specifically for this study, while an additional 23 populations from the publicly available genotypes were retrieved. Three complementary statistical tests, Rsb (extended haplotype homozygosity between-populations), XP-EHH (cross-population extended haplotype homozygosity), and runs of homozygosity (ROH) islands were applied to identify genomic variants with potential impact on the adaptability of Merino genetic type in two contrasting climate zones.

Results The results indicate that a large part of the Merino's genetic relatedness and admixture patterns are explained by their genetic background and/or geographic origin, followed by local admixture. Multi-dimensional scaling, Neighbor-Net, Admixture, and TREEMIX analyses consistently provided evidence of the role of Australian, Rambouillet and German strains in the extensive gene introgression into the other Merino and Merino-derived breeds. The close relationship between Iberian Merinos and other South-western European breeds is consistent with the Iberian origin of the Merino genetic type, with traces from previous contributions of other Mediterranean stocks. Using Rsb and XP-EHH approaches, signatures of selection were detected spanning four genomic regions located on *Ovis aries* chromosomes (OAR) 1, 6 and 16, whereas two genomic regions on OAR6, that partially overlapped with the previous ones, were highlighted by ROH islands. Overall, the three approaches identified 106 candidate genes putatively under selection. Among them, genes related to immune response were identified via the gene interaction network.

*Correspondence:
Simone Ceccobelli
s.ceccobelli@staff.univpm.it
Full list of author information is available at the end of the article

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

In addition, several candidate genes were found, such as *LEKR1*, *LCORL*, *GHR*, *RBPJ*, *BMPR1B*, *PPARGC1A*, and *PRKAA1*, related to morphological, growth and reproductive traits, adaptive thermogenesis, and hypoxia responses.

Conclusions To the best of our knowledge, this is the first comprehensive dataset that includes most of the Merino and Merino-derived sheep breeds raised in different regions of the world. The results provide an in-depth picture of the genetic makeup of the current Merino and Merino-derived breeds, highlighting the possible selection pressures associated with the combined effect of anthropic and environmental factors. The study underlines the importance of Merino genetic types as invaluable resources of possible adaptive diversity in the context of the occurring climate changes.

Background

Sheep (*Ovis aries*) are one of the oldest livestock species, i.e. they were domesticated by humans around 10,500 years before the present (YBP) in the Fertile Crescent (South-eastern Anatolia and the Iranian Zagros Mountains) [1, 2]. Then, sheep accompanied humans in their migrations from Asia into Europe and subsequently dispersed throughout the world [3]. One of the possible routes for the spread of the first farmers into continental Europe is the Danube valley, followed by a maritime route through the Mediterranean Sea [4–6]. Sheep provide a variety of resources, including meat, milk, and wool, and they play an essential role in the global agricultural economy since the Neolithic age [7], when they were almost exclusively reared for meat consumption and their skin. Throughout the Neolithic period and across Eurasia, there is no archaeological evidence of the use of wool, which became a known practice only later, during the Bronze Age in the Near East [8]. Later, zoo-archaeological studies have shown that sheep selected for secondary products (wool and milk) were spread throughout Europe, replacing, in many areas, the primitive sheep populations [9]. During the Bronze Age the dominant colour of the wool was brown, whereas, in the Iron Age, sheep with white fleece became widespread [10]. The oldest wool textile remains date back to 1500 YBP [11, 12].

Specialisation for wool production in the early Iron Age was well documented by the Romans [10, 11]. Roman classical literature reported that the best wool sheep originated in Southern Italy (*Tarentum*) and Greece, and were later exported to other parts of the Empire, including the Iberian Peninsula [13, 14]. The development of Spanish Merino sheep for high-quality wool production dates back to the late Middle Ages [15], when they were probably used as starting genetic material for composite stocks that may well have been influenced, across centuries, by the Phoenicians', Romans', and Moors' conquests of the Iberian Peninsula [11]. In particular, a recent study by Kandoussi et al. [16], based on the analysis of the mitochondrial DNA (mtDNA) control region, provided support to the theory of a possible contribution of North African sheep to the growth of the Middle Ages Merino population. Another study [17], using genome-wide single nucleotide polymorphisms (SNPs), identified genetic patterns for intercontinental Merino populations that are compatible with a partial Italian ancestry, possibly corresponding to the documented gene flow during the Roman period. Notably, the Roman writer *Lucius Junius Moderatus Columella* [13] chronicled the complex three-way crossbreeding practice that was carried out in *Hispania Baetica* by his uncle. To improve wool quality (fine white wool), he used North African, Italian (*Tarentum*), and local sheep from the Iberian Peninsula. The resulting fine wool animals have spread over the centuries throughout the Iberian Peninsula and in other territories of the Spanish Empire. Consequently, the "Merino Breed" industry became one of the most important industries of the Iberian Peninsula and sheep were strongly protected by Royal decrees, forbidding their exportation out of the Iberian Peninsula for several centuries [18]. In the early eighteenth century, Merinos exited Spain through Kingdom agreements and smuggling. Then, Merino flocks were crossed with local breeds in many countries worldwide [19–21].

Due to centuries of selection that have led to local adaptation, sheep have thrived in a diverse range of environmental conditions along their spreading routes [22]. This is particularly evident in Merino and Merino-derived sheep, which are nowadays present in a wide variety of productive environments around the world and make them the predominant wool producing genetic type [17]. Well known for its exceptional white fleece and the abundant production of soft, fine, and curly wool, Merino sheep represent a key genetic resource thanks to their ability to face harsh climatic conditions, poor quality feeding and arid landscapes. Moreover, Merinos display an extraordinary ability to adapt and perform in an extensive production system, in a wide range of environmental conditions spanning from the Mediterranean to the Continental European and Asian climates as well as subtropical, and both warm and cold arid lands [23–25]. To date, the genetic determinism of Merino adaptability has been poorly investigated while relevant efforts have been made to identify the loci involved in wool and

7. Conclusions

The findings of this thesis contribute to our understanding of the animal genome to improve sustainability of livestock farming. The first part of the thesis focused on biodiversity, examining the genetic diversity of Pagliarola, a local endangered sheep breed, and Holstein-Friesian, a cosmopolitan cattle breed. These two populations are reared in two divergent farming systems, but for each of them, it is possible to inspect variables linked to both management and genetics, helpful to improve their sustainability. In fact, sustainability requires a balance between high yields that can be maintained even in the face of large shocks and environmental impacts. Pasture-based sheep farming in the Mediterranean basin is a significant sustainability issue, as it is strongly associated with the utilization of natural and semi-natural regions and frequently uses well-adapted autochthonous breeds. These systems are crucial for the management and protection of large farmlands in Europe, providing local communities with a variety of goods and services with a low to moderate reliance on outside resources. The key to labor profitability for sheep farming of indigenous breeds in less-favored areas is low off-farm input dependence and improved feed self-sufficiency. Diversified production improves farm flexibility and can be a tactic for dealing with uncertainty, although it may not always increase labor profitability. Conserving and utilizing genetic resources from a wide variety of animal populations is essential for low-input agricultural systems to achieve economic sustainability. On the other hand, the sustainability of dairy farms depends on multiple factors, including cow welfare, dairy farming scale and resources, access to milk markets, and dairy product portfolios. High-performing dairy herds worldwide have a lot in common, including the breeds of cattle used, the use of cutting-edge veterinary technology, dairying equipment, and record-keeping techniques. However, in many different agricultural areas, output intensity is positively correlated with greater environmental sustainability. The sustainability of the dairy sector has been enhanced by the new payment systems based on milk quality, such as the kappa-casein type, which can increase milk yield by up to 35,000 kg per 305-day lactation. It is critical for sustainable milk production for dairy cows to be placed in appropriate management and ecological niches and requires well-developed pedigree, genotype, and performance databases.

The analysis of the Pagliarola breed revealed important insights into its genetic variability, structure, relationship, and introgression with other breeds, which are crucial for its conservation and exploitation. As said above, the conservation of local breeds, such as Pagliarola, is essential for preserving rare genetic variants that may be important for addressing climate change and other environmental challenges. Furthermore, the conservation of local breeds also contributes to

ecological, economic, social, and cultural sustainability by maintaining local farming traditions and products. The results of the genetic structure analyses demonstrate that the residual population of the Pagliarola breed has a distinct and distinctive genetic background made up of two distinct genetic groups that share a common ancestor but differ in their genomic make-up, one of which has higher inbreeding coefficients and lower heterozygosity values, and the other of which is less inbred and has higher heterozygosity values. These results emphasize the importance of genomic analyses to plan conservation strategies for local and endangered breeds, as in this case, they would reveal which population could reintroduce more genetic diversity. In addition, our analyses demonstrated the relationship between Pagliarola and the Italian Merino breeds in accordance with the historical account of the traditional farming system based on transhumance. In fact, traditionally, Pagliarola sheep were left back home to provide for the needs of the families of shepherds that left the Abruzzo region for the transhumance of Merino-derived sheep (e.g., Gentile di Puglia) towards Apulia. The usefulness of this sheep was due to its ability to survive on only straw and poor pastures during the harsh winters of the Apennines. In our analyses, a divergent selection of genes linked to feed efficiency was found in Pagliarola compared to its neighboring breeds, in line with its history. Overall, the results of our analyses of the genome-wide diversity of the Pagliarola residual population help to understand how important it is to inspect genomic diversity in order to plan a breeding program and the ability of the bioinformatic tools to scan the genomes of local populations in order to underline the genomic regions involved in the adaptability to severe environmental variables. Lastly, the results of this study underline the power of genomic studies of local populations to find their connection with traditional farming practices and products.

The selection of cosmopolite breeds, such as Holstein-Friesians, is essential to obtaining more efficient animals in order to produce more product units while producing fewer emissions. The selection of cosmopolitan breeds contributes to ecological, economic, social, and cultural sustainability by adapting local farming traditions and products to the demands of the market. Local breeds conserve genetic variants that make them more suitable for a territory, while cosmopolitan breeds can be selected to have herds that are more suitable for local production than the rest of the population of the same breed. The analysis of the Holstein-Friesian breed within the Italian population selected for different breeding goals and its comparison with the North American population provided valuable information on the impact of different breeding practices on genetic diversity. The sampling method we opted for was chosen to emphasize the differences between Italian Holstein not having or having poor North American ancestry and

North American Holstein in order to ascertain if we may find genome-wide differences and try to understand the nature of these differences among the samples. The results of this study underlined differences in genetic diversity indexes within the Italian population and its higher genetic diversity compared to the North American population, stressing the importance of implementing the study of genetic diversity while programming breeding programs. In fact, the loss of genetic diversity not only reduces the possibility of obtaining factual genetic improvement but also withdraws the ability to face persistent changes in environmental variables. The detection of signatures of selection among different Holstein populations sheds light on loci that may be involved in response to different breeding goals based on diverging environmental factors such as climate, and feeding, keeping, and farming systems, or social factors such as typical productions that may influence market needs and demand. In fact, we did not only find selective sweeps in regions involved in production traits such as milk production, but also many traits known to sensibly influence the sustainability of dairy farming were revealed as having been differentially selected in the studied populations. Examples are traits linked to fertility, resistance to diseases, and methane emissions. These findings highlight the need to scan the genome of different Holstein populations to improve the sustainability of cosmopolitan breeds through genetic selection for adaptation to specific environments.

The second part of the thesis focused on identifying loci associated with resistance to parasites in Comisana, a selected dairy sheep breed local to Sicily. The analysis of FEC phenotypes and SNP data allowed the development of a model for calculating breeding values for resistance to gastrointestinal nematodes (GIN) in dairy sheep. Furthemore, comparing the most and least resistant reproductors revealed genomic areas linked to the resistance mechanism and found genes involved in metabolic pathways with differential expression. These findings provide insights into the genetic basis of resistance to parasites and can contribute to the development of breeding strategies aimed at improving resistance to GIN in sheep. Breeding more resistant sheep to GIN may reduce the use of anthelmintic drugs, helping to increase the economic sustainability of sheep farming because of the limitation of production losses and reduction of veterinary expenses, including preventive treatments. Furthermore, the reduction of the use of anthelmintic pharmaceuticals results in more environmentally sustainable farming by limiting their entry into the environment and reducing detrimental effects such as their contact with non-target species and, as a consequence, the pharmaceutical resistance of the parasites to these molecules.

In conclusion, the investigation on biodiversity and the detection of loci under selection in local and cosmopolitan breeds carried out in this study have provided valuable insights into genetic variability, adaptation to different environments, and resistance to parasites. These findings have important implications for the conservation of local breeds, the development of tailored breeding strategies, and the improvement of livestock productivity and resilience. The results of this thesis contribute to our understanding of the genetic basis of sustainability in livestock production and provide a foundation for further research in this field. It is hoped that the findings of this thesis will inform future breeding and conservation efforts to promote sustainable livestock farming practices and address the challenges facing animal production in the near future.

References

Abraham, M., & Pingali, P. (2020). Transforming Smallholder Agriculture to Achieve the SDGs. In S. Gomez y Paloma, L. Riesgo, & K. Louhichi (A c. Di), *The Role of Smallholder Farms in Food and Nutrition Security* (pp. 173–209). Springer International Publishing. https://doi.org/10.1007/978-3-030-42148-9_9

Agenda 21 | Department of Economic and Social Affairs. (s.d.). Recuperato 17 aprile 2023, da https://sdgs.un.org/publications/agenda21

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, *26*(2), 420–430. https://doi.org/10.1111/mec.13948

ANABIC. (2022). *I-BEEF*. http://www.anabic.it/i-beef/i-beef-azioni.htm#azione_02

Annan-Diab, F., & Molinari, C. (2017). Interdisciplinarity: Practical approach to advancing education for sustainability and for the Sustainable Development Goals. *The International Journal of Management Education*, *15*(2, Part B), 73–83. https://doi.org/10.1016/j.ijme.2017.03.006

Arru, B., Furesi, R., Madau, F. A., & Pulina, P. (2021). Economic performance of agritourism: An analysis of farms located in a less favoured area in Italy. *Agricultural and Food Economics*, *9*(1), 27. https://doi.org/10.1186/s40100-021-00199-z

Avise, J. C., Giblin-Davidson, C., Laerm, J., Patton, J. C., & Lansman, R. A. (1979). Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, Geomys pinetis. *Proceedings of the National Academy of Sciences*, *76*(12), 6694–6698. https://doi.org/10.1073/pnas.76.12.6694

Avise, J. C., Lansman, R. A., & Shade, R. O. (1979). THE USE OF RESTRICTION ENDONUCLEASES TO MEASURE MITOCHONDRIAL DNA SEQUENCE RELATEDNESS IN NATURAL POPULATIONS. I. POPULATION STRUCTURE AND EVOLUTION IN THE GENUS PEROMYSCUS. *Genetics*, *92*(1), 279–295.

https://doi.org/10.1093/genetics/92.1.279

Barranco, C. (2021). The human genome project. Nature Research. https://doi.org/10.1038/d42859-020-00101-9

Bayat, A. (2002). Bioinformatics. BMJ : British Medical Journal, 324(7344), 1018–1022. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1122955/

Benavides, F. J., & Guénet, J.-L. (2012). Mouse genomics. In The Laboratory Mouse (pp. 57–90). Elsevier. https://doi.org/10.1016/B978-0-12-382008-2.00004-0

Ben-Jemaa, S., Senczuk, G., Ciani, E., Ciampolini, R., Catillo, G., Boussaha, M., Pilla, F., Portolano, B., & Mastrangelo, S. (2021). Genome-wide analysis reveals selection signatures involved in meat traits and local adaptation in semi-feral maremmana cattle. Frontiers in Genetics, 12. https://www.frontiersin.org/articles/10.3389/fgene.2021.675569

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., & Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Frontiers in Genetics*, *6*. https://www.frontiersin.org/articles/10.3389/fgene.2015.00033

Blasco, A., & Toro, M. A. (2014). A short critical history of the application of genomics to animal breeding. *Livestock Science*, *166*, 4–9. https://doi.org/10.1016/j.livsci.2014.03.015

Borlaug, N. (2007). Feeding a Hungry World. *Science*, *318*(5849), 359–359. https://doi.org/10.1126/science.1151062

Brandenburg, J.-T., Clark, L., Botha, G., Panji, S., Baichoo, S., Fields, C., & Hazelhurst, S. (2022). H3AGWAS: A portable workflow for genome wide association studies. BMC Bioinformatics, 23(1), 498. https://doi.org/10.1186/s12859-022-05034-w

Brito, L. F., Bedere, N., Douhard, F., Oliveira, H. R., Arnal, M., Peñagaricano, F., Schinckel, A. P., Baes, C. F., & Miglior, F. (2021). Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. *Animal*, *15*, 100292. https://doi.org/10.1016/j.animal.2021.100292

Brodt, S., Six, J., Feenstra, G., Ingels, C., & Campbell, D. (2011). Sustainable agriculture. *Nat. Educ. Knowl*, *3*(1).

Brookes, A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177–186.

https://doi.org/10.1016/S0378-1119(99)00219-X

Broom, D. M. (2018). The scientific basis for action on animal welfare and other aspects of sustainability. In *Farming, Food and Nature*. Routledge.

Brown, W. M., & Wright, J. W. (1979). Mitochondrial DNA Analyses and the Origin and Relative Age of Parthenogenetic Lizards (Genus Cnemidophorus). *Science*, *203*(4386), 1247–1249. https://doi.org/10.1126/science.424751

Brundtland, G. H. (1987). Our Common Future—Call for Action*. *Environmental Conservation*, *14*(4), 291–294. https://doi.org/10.1017/S0376892900016805

Bruzzone, A., Iamartino, D., Blasi, M., & Pilla, F. (2003). The Pentro horse: Genetic characterization by microsatellite markers. *Italian Journal of Animal Science*, *2*(3), 223–230. https://doi.org/10.4081/ijas.2003.223

Camargo, G. M. F. de. (2018). The role of molecular genetics in livestock production. *Animal Production Science*, *59*(2), 201–206. https://doi.org/10.1071/AN18013

Campbell, B. M., Thornton, P., Zougmoré, R., van Asten, P., & Lipper, L. (2014). Sustainable intensification: What is its role in climate smart agriculture? *Current Opinion in Environmental Sustainability*, *8*, 39–43. https://doi.org/10.1016/j.cosust.2014.07.002

Canestrari, P. (2005). *GLI STRUMENTI DEL MIGLIORAMENTO GENETICO AL SERVIZIO DELL'ALLEVAMENTO BOVINO ITALIANO DA CARNE*.

Chakraborty, D., Sharma, N., Kour, S., Sodhi, S. S., Gupta, M. K., Lee, S. J., & Son, Y. O. (2022). Applications of omics technology for livestock selection and improvement. Frontiers in Genetics, 13. https://www.frontiersin.org/articles/10.3389/fgene.2022.774113

Colditz, I. G., Hine, B. C., Colditz, I. G., & Hine, B. C. (2016). Resilience in farm animals: Biology, management, breeding and implications for animal welfare. *Animal Production Science*, *56*(12), 1961–1983. https://doi.org/10.1071/AN15297

Collins, F. S., Morgan, M., & Patrinos, A. (2003a). The Human Genome Project: Lessons from Large-Scale Biology. *Science*, *300*(5617), 286–290. https://doi.org/10.1126/science.1084564

Collins, F. S., Green, E. D., Guttmacher, A. E., & Guyer, M. S. (2003b). A vision for the future of

genomics research. Nature, 422(6934), 835–847. https://doi.org/10.1038/nature01626

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12(7), 499–510. https://doi.org/10.1038/nrg3012

Demir, E., Ceccobelli, S., Bilginer, U., Pasquini, M., Attard, G., & Karsli, T. (2022). Conservation and selection of genes related to environmental adaptation in native small ruminant breeds: A review. Ruminants, 2(2), 255–270. https://doi.org/10.3390/ruminants2020017

Du Pisani, J. A. (2006). Sustainable development – historical roots of the concept. *Environmental Sciences*, *3*(2), 83–96. https://doi.org/10.1080/15693430600688831

Dumont, B., Groot, J. C. J., & Tichit, M. (2018). Review: Make ruminants green again – how can sustainable intensification and agroecology converge for a better future? *Animal*, *12*(s2), s210–s219. https://doi.org/10.1017/S1751731118001350

Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, *12*(1), Articolo 1. https://doi.org/10.1038/s41598-022-14395-4

Eusebi PG, Martinez A, Cortes O (2020) Genomic Tools for Effective Conservation of Livestock Breed Diversity. Diversity 12:8. https://doi.org/10.3390/d12010008

Farstad, W. (2018). Ethics in animal breeding. *Reproduction in Domestic Animals*, *53*(S3), 4–13. https://doi.org/10.1111/rda.13335

Feldman, M. W., Bergman, A., Pollock, D. D., & Goldstein, D. B. (1997). Microsatellite Genetic Distances With Range Constraints: Analytic Description and Problems of Estimation. *Genetics*, *145*(1), 207–216. https://doi.org/10.1093/genetics/145.1.207

Flint, A. p. f, & Woolliams, J. a. (2007). Precision animal breeding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 573–590. https://doi.org/10.1098/rstb.2007.2171

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., & Snyder, P. K.

(2005). Global Consequences of Land Use. *Science*, *309*(5734), 570–574. https://doi.org/10.1126/science.1111772

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. https://doi.org/10.1534/genetics.108.092221

Gianola, D., & Rosa, G. J. M. (2015). One Hundred Years of Statistical Developments in Animal Breeding. *Annual Review of Animal Biosciences*, *3*(1), 19–56. https://doi.org/10.1146/annurev-animal-022114-110733

Goddard, J., & Hayes, B. (2009). The challenges of applying genomics to animal breeding. Animal Genetics, 40(2), 105-112. https://doi.org/10.1111/j.1365-2052.2009.01880.x

Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J., & Hayes, B. J. (2016). Genetics of complex traits: Prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1835), 20160569. https://doi.org/10.1098/rspb.2016.0569

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science*, *327*(5967), 812–818. https://doi.org/10.1126/science.1185383

Goldfeder, R. L., Wall, D. P., Khoury, M. J., Ioannidis, J. P. A., & Ashley, E. A. (2017). Human genome sequencing at the population scale: A primer on high-throughput dna sequencing and analysis. American Journal of Epidemiology, 186(8), 1000–1009. https://doi.org/10.1093/aje/kww224

Goldstein, D. B., & Pollock, D. D. (1997). Launching Microsatellites: A Review of Mutation Processes and Methods of Phylogenetic Inference. *Journal of Heredity*, *88*(5), 335–342. https://doi.org/10.1093/oxfordjournals.jhered.a023114

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Gordillo, G., & Jerónimo, O. M. (2013). *Food Security and Sovereignty |Policy Support and Governance| Food and Agriculture Organization of the United Nations*. https://www.fao.org/policy-support/tools-and-publications/resources-details/en/c/1268249/

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

Gutiérrez-Reinoso, M. A., Aponte, P. M., & García-Herreros, M. (2023). Genomic and phenotypic udder evaluation for dairy cattle selection: A review. Animals, 13(10), 1588. https://doi.org/10.3390/ani13101588

Harris, H. (1966). Enzyme Polymorphisms in Man. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *164*(995), 298–310.

Hasan, N., Choudhary, S., Naaz, N., Sharma, N., & Laskar, R. A. (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. Journal of Genetic Engineering and Biotechnology, 19(1), 128. https://doi.org/10.1186/s43141-021-00231-1

Hassler, H. B., Probert, B., Moore, C., Lawson, E., Jackson, R. W., Russell, B. T., & Richards, V. P. (2022). Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies. Microbiome, 10(1), 104. https://doi.org/10.1186/s40168-022-01295-y

Hayden, M. J., Nguyen, T. M., Waterman, A., & Chalmers, K. J. (2008). Multiplex-Ready PCR: A new method for multiplexed SSR and SNP genotyping. *BMC Genomics*, *9*(1), 80. https://doi.org/10.1186/1471-2164-9-80

Herrero, M., Henderson, B., Havlík, P., Thornton, P. K., Conant, R. T., Smith, P., Wirsenius, S., Hristov, A. N., Gerber, P., Gill, M., Butterbach-Bahl, K., Valin, H., Garnett, T., & Stehfest, E. (2016). Greenhouse gas mitigation potentials in the livestock sector. *Nature Climate Change*, *6*(5), Articolo 5. https://doi.org/10.1038/nclimate2925

Herrero-Medrano, J. M., Megens, H.-J., Groenen, M. A., Ramis, G., Bosse, M., Pérez-Enciso,

M., & Crooijmans, R. P. (2013). Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. BMC Genetics, 14(1), 106. https://doi.org/10.1186/1471-2156-14-106

HLPE. (2020). Food security and nutrition: Building a global narrative towards 2030. *A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security*.

Hocquette, J. F., Botreau, R., Legrand, I., Polkinghorne, R., Pethick, D. W., Lherm, M., Picard, B., Doreau, M., Terlouw, E. M. C., Hocquette, J. F., Botreau, R., Legrand, I., Polkinghorne, R., Pethick, D. W., Lherm, M., Picard, B., Doreau, M., & Terlouw, E. M. C. (2014). Win–win strategies for high beef quality, consumer satisfaction, and farm efficiency, low environmental impacts and improved animal welfare. *Animal Production Science*, *54*(10), 1537–1548. https://doi.org/10.1071/AN14210

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, *41*(s1), 32–46. https://doi.org/10.1111/j.1365-2052.2010.02043.x

Hovi, M., Sundrum, A., & Thamsborg, S. M. (2003). Animal health and welfare in organic livestock production in Europe: Current state and future challenges. *Livestock Production Science*, *80*(1), 41–53. https://doi.org/10.1016/S0301-6226(02)00320-2

Iamartino, D., Bruzzone, A., Lanza, A., Blasi, M., & Pilla, F. (2005). Genetic diversity of Southern Italian goat populations assessed by microsatellite markers. *Small Ruminant Research*, *57*(2), 249–255. https://doi.org/10.1016/j.smallrumres.2004.08.003

Jamal, T., & Dredge, D. (2014). 6. Tourism and Community Development Issues. In R. Sharpley & D. J. Telfer (A c. Di), *Tourism and Development* (pp. 178–204). Multilingual Matters. https://doi.org/10.21832/9781845414740-008

John, M., Ankenbrand, M. J., Artmann, C., Freudenthal, J. A., Korte, A., & Grimm, D. G. (2022). Efficient permutation-based genome-wide association studies for normal and skewed phenotypic distributions. Bioinformatics, 38(Supplement_2), ii5–ii12. https://doi.org/10.1093/bioinformatics/btac455

Johnsson, M. (2023). Genomics in animal breeding from the perspectives of matrices and molecules. Hereditas, 160(1), 20. https://doi.org/10.1186/s41065-023-00285-w

Ke, X., Taylor, M. S., & Cardon, L. R. (2008). Singleton SNPs in the human genome and implications for genome-wide association studies. *European Journal of Human Genetics*, *16*(4), Articolo 4. https://doi.org/10.1038/sj.ejhg.5201987

Khlestkina, E. K., & Salina, E. A. (2006). SNP markers: Methods of analysis, ways of development, and comparison on an example of common wheat. *Russian Journal of Genetics*, *42*(6), 585–594. https://doi.org/10.1134/S1022795406060019

Kidd, C. V. (1992). The evolution of sustainability. *Journal of Agricultural and Environmental Ethics*, *5*(1), 1–26. https://doi.org/10.1007/BF01965413

Kozarewa, I., Ning, Z., Quail, M. A., Sanders, M. J., Berriman, M., & Turner, D. J. (2009). Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature Methods*, *6*(4), Articolo 4. https://doi.org/10.1038/nmeth.1311

Kulski, J. K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications. InTech. https://doi.org/10.5772/61964

Lange, L. (2022). Business Models, Including Higher Value Products for the New Circular, Resource-Efficient Biobased Industry. *Frontiers in Sustainability*, *3*. https://doi.org/10.3389/frsus.2022.789435

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, *92*(9), 4656–4663. https://doi.org/10.3168/jds.2009-2061

Lewontin, R. C., & Hubby, J. L. (1966). A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygosity in Natural Populations of DROSOPHILA PSEUDOOBSCURA. *Genetics*, *54*(2), 595–609.

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., & Misztal, I. (2020). Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in

BLUPF90. *Genes*, *11*(7), Articolo 7. https://doi.org/10.3390/genes11070790

Lowe, J. W. E., & Bruce, A. (2019). Genetics without genes? The centrality of genetic markers in livestock genetics and genomics. History and Philosophy of the Life Sciences, 41(4), 50. https://doi.org/10.1007/s40656-019-0290-x

MacLeod, M., Leinonen, I., Wall, E., Houdijk, J., Eory, V., Burns, J., Ahmadi, B. V., & Gómez-Barbero, M. (2019). *Impact of animal breeding on GHG emissions and farm economics*. Publications Office of the European Union.

Martínez-Castillo, R. (2016). Sustainable agricultural production systems. *Revista Tecnología En Marcha*, *29*, 70–85. https://doi.org/10.18845/tm.v29i5.2518

Martyniuk, E. (2021). Policy Effects on the Sustainability of Animal Breeding. *Sustainability*, *13*(14), Articolo 14. https://doi.org/10.3390/su13147787

Marwal, A., & Gaur, R. K. (2020). Molecular markers: Tool for genetic analysis. In Animal Biotechnology (pp. 353–372). Elsevier. https://doi.org/10.1016/B978-0-12-811710-1.00016-1

McGreevy, S. R., Rupprecht, C. D. D., Niles, D., Wiek, A., Carolan, M., Kallis, G., Kantamaturapoj, K., Mangnus, A., Jehlička, P., Taherzadeh, O., Sahakian, M., Chabay, I., Colby, A., Vivero-Pol, J.-L., Chaudhuri, R., Spiegelberg, M., Kobayashi, M., Balázs, B., Tsuchiya, K., … Tachikawa, M. (2022). Sustainable agrifood systems for a post-growth world. *Nature Sustainability*, *5*(12), Articolo 12. https://doi.org/10.1038/s41893-022-00933-5

McGuire, A. L., Gabriel, S., Tishkoff, S. A., Wonkam, A., Chakravarti, A., Furlong, E. E. M., Treutlein, B., Meissner, A., Chang, H. Y., López-Bigas, N., Segal, E., & Kim, J.-S. (2020). The road ahead in genetics and genomics. Nature Reviews Genetics, 21(10), 581–596. https://doi.org/10.1038/s41576-020-0272-6

Medema, M. H., de Rond, T., & Moore, B. S. (2021). Mining genomes to illuminate the specialized chemistry of life. Nature Reviews Genetics, 22(9), 553–571. https://doi.org/10.1038/s41576-021-00363-7

Meier, S., Korkuć, P., Arends, D., & Brockmann, G. A. (2019). DNA Sequence Variants and

Protein Haplotypes of Casein Genes in German Black Pied Cattle (DSN). *Frontiers in Genetics*, *10*. https://www.frontiersin.org/articles/10.3389/fgene.2019.01129

Mondini, L., & Pagnotta, M. A. (2015). Using molecular techniques to dissect plant genetic diversity. *In* M. R. Ahuja & S. M. Jain (A c. Di), Genetic Diversity and Erosion in Plants (Vol. 7, pp. 125–157). Springer International Publishing. https://doi.org/10.1007/978-3-319-25637-5_6

Nass, M. M. K., & Nass, S. (1963). INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS: I. Fixation and Electron Staining Reactions. *Journal of Cell Biology*, *19*(3), 593–611. https://doi.org/10.1083/jcb.19.3.593

Nass, S., & Nass, M. M. K. (1963). INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS: II. Enzymatic and Other Hydrolytic Treatments. *Journal of Cell Biology*, *19*(3), 613–629. https://doi.org/10.1083/jcb.19.3.613

Neef, A., & Neubert, D. (2011). Stakeholder participation in agricultural research projects: A conceptual framework for reflection and decision-making. *Agriculture and Human Values*, *28*(2), 179–194. https://doi.org/10.1007/s10460-010-9272-z

Ngene, M., Hayes, B., Goddard, J., & Fernando, R. (2017). Genomic selection for local breeds: Challenges and opportunities. Animal Genetics, 48(3), 327-339. https://doi.org/10.1111/age.12678

Nguyen, H. (s.d.). *Sustainable food systems: Concept and framework*.

Nicolazzi, E. L., Biffani, S., Biscarini, F., Orozco ter Wengel, P., Caprera, A., Nazzicari, N., & Stella, A. (2015). Software solutions for the livestock genomics SNP array revolution. *Animal Genetics*, *46*(4), 343–353. https://doi.org/10.1111/age.12295

Nicolazzi, E. L., Biffani, S., & Jansen, G. (2013). Short communication: Imputing genotypes using PedImpute fast algorithm combining pedigree and population information. *Journal of Dairy Science*, *96*(4), 2649–2653. https://doi.org/10.3168/jds.2012-6062

Nicolosi, A., Laganà, V. R., Di Gregorio, D., & Privitera, D. (2021). Social Farming in the Virtuous System of the Circular Economy. An Exploratory Research. *Sustainability*, *13*(2), Articolo 2. https://doi.org/10.3390/su13020989

Noble, R., & Noble, D. (2023). The Gene Delusion. In Understanding Living Systems (Understanding Life, pp. 1-13). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009277396.003

Osei-Amponsah, R., Chauhan, S. S., Leury, B. J., Cheng, L., Cullen, B., Clarke, I. J., & Dunshea, F. R. (2019). Genetic selection for thermotolerance in ruminants. Animals : an Open Access Journal from MDPI, 9(11), 948. https://doi.org/10.3390/ani9110948

Peng, W., Zheng, H., Robinson, B. E., Li, C., & Li, R. (2022). Comparing the importance of farming resource endowments and agricultural livelihood diversification for agricultural sustainability from the perspective of the food–energy–water nexus. *Journal of Cleaner Production*, *380*, 135193. https://doi.org/10.1016/j.jclepro.2022.135193

Persichilli, C., Biffani, S., Senczuk, G., Di Civita, M., Bosco, A., Grande, S., & Pilla, F. (2023b). *Identification of Genetic Regions Associated with Resistance to Gastrointestinal Nematodes in Comisana Sheep using a Genome-wide Association Study based on EBV Ranking* [Submitted for publication].

Persichilli C, Senczuk G, Mastrangelo S, et al (2023a) Exploring genome-wide differentiation and signatures of selection in Italian and North American Holstein populations. Journal of Dairy Science 106:5537–5553. https://doi.org/10.3168/jds.2022-22159

Persichilli, C., Senczuk, G., Pasqualino, L., Anzalone, D. A., Negrini, R., Ajmone Marsan, P., Colli, L., Pilla, F., & Mastrangelo, S. (2021). Genome-wide diversity of Pagliarola sheep residual population and its conservation implication. *Italian Journal of Animal Science*, *20*(1), 1695–1705. https://doi.org/10.1080/1828051X.2021.1970033

Poczai, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J. P., & Hyvönen, J. (2013). Advances in plant gene-targeted and functional markers: A review. Plant Methods, 9(1), 6. https://doi.org/10.1186/1746-4811-9-6

Quarrie, J. (1992). *Earth Summit ’92. The United Nations Conference on Environment and Development, Rio de Janeiro 1992.* UNCED. https://scholar.google.com/scholar_lookup?title=Earth+Summit+%2792.+The+United+Nations+Conference+on+Environment+and+Development%2C+Rio+de+Janeiro+1992.&a

uthor=Quarrie+J.+%28ed.%29&publication_year=1992

Raj, A., Jhariya, M. K., Khan, N., Banerjee, A., & Meena, R. S. (2021). Ecological Intensification for Sustainable Development. In M. K. Jhariya, R. S. Meena, & A. Banerjee (A c. Di), *Ecological Intensification of Natural Resources for Sustainable Agriculture* (pp. 137–170). Springer. https://doi.org/10.1007/978-981-33-4203-3_5

Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., & Rieseberg, L. H. (2018). Trends in Global Agricultural Land Use: Implications for Environmental Health and Food Security. *Annual Review of Plant Biology*, *69*(1), 789–815. https://doi.org/10.1146/annurev-arplant-042817-040256

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

Rodrik, D. (2014). The Past, Present, and Future of Economic Growth. *Challenge*, *57*(3), 5–39. https://doi.org/10.2753/0577-5132570301

Rothschild, M. F., & Plastow, G. S. (2014). Applications of genomics to improve livestock in the developing world. *Livestock Science*, *166*, 76–83. https://doi.org/10.1016/j.livsci.2014.03.020

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), Articolo 7164. https://doi.org/10.1038/nature06250

Sabeti, P. C., Walsh, E., Schaffner, S. F., Varilly, P., Fry, B., Hutcheson, H. B., Cullen, M., Mikkelsen, T. S., Roy, J., Patterson, N., Cooper, R., Reich, D., Altshuler, D., O'Brien, S., & Lander, E. S. (2005). The Case for Selection at CCR5-Δ32. *PLOS Biology*, *3*(11), e378. https://doi.org/10.1371/journal.pbio.0030378

Salgotra, R. K., & Stewart, C. N. (2020). Functional markers for precision plant breeding. International Journal of Molecular Sciences, 21(13), 4792. https://doi.org/10.3390/ijms21134792

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

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*, 104257. https://doi.org/10.1016/j.livsci.2020.104257

Schmid, M., & Bennewitz, J. (2017). Invited review: Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs. *Archives Animal Breeding*, *60*(3), 335–346. https://doi.org/10.5194/aab-60-335-2017

Segelbacher, G., Bosse, M., Burger, P., Galbusera, P., Godoy, J. A., Helsen, P., Hvilsom, C., Iacolina, L., Kahric, A., Manfrin, C., Nonic, M., Thizy, D., Tsvetkov, I., Veličković, N., Vilà, C., Wisely, S. M., & Buzan, E. (2022). New developments in the field of genomic technologies and their relevance to conservation management. Conservation Genetics, 23(2), 217–242. https://doi.org/10.1007/s10592-021-01415-5

Selman, P. (2004). Community participation in the planning and management of cultural landscapes. *Journal of Environmental Planning and Management*, *47*(3), 365–392. https://doi.org/10.1080/0964056042000216519

Shao, Z., Wang, T., Qiao, J., Zhang, Y., Huang, S., & Zeng, P. (2022). A comprehensive comparison of multilocus association methods with summary statistics in genome-wide association studies. BMC Bioinformatics, 23(1), 359. https://doi.org/10.1186/s12859-022-04897-3

Singh, S., Gautam, B., Rao, A., Tandon, G., & Kaur, S. (2018). Bioinformatics approaches for animal breeding and genetics. In G. Wadhwa, P. Shanmughavel, A. K. Singh, & J. R. Bellare (A c. Di), Current trends in Bioinformatics: An Insight (pp. 287–306). Springer. https://doi.org/10.1007/978-981-10-7483-7_17

Singh, B. D., & Singh, A. K. (2015). High-throughput snp genotyping. In B. D. Singh & A. K.

Singh, Marker-Assisted Plant Breeding: Principles and Practices (pp. 367–400). Springer India. https://doi.org/10.1007/978-81-322-2316-0_13

Sneddon, C., Howarth, R. B., & Norgaard, R. B. (2006). Sustainable development in a post-Brundtland world. *Ecological Economics*, *57*(2), 253–268. https://doi.org/10.1016/j.ecolecon.2005.04.013

Stern, S., Sonesson, U., Gunnarsson, S., Öborn, I., Kumm, K.-I., & Nybrant, T. (2005). Sustainable Development of Food Production: A Case Study on Scenarios for Pig Production. *AMBIO: A Journal of the Human Environment*, *34*(4), 402–407. https://doi.org/10.1579/0044-7447-34.4.402

Sud, A., Kinnersley, B., & Houlston, R. S. (2017). Genome-wide association studies of cancer: Current insights and future perspectives. *Nature Reviews Cancer*, *17*(11), Articolo 11. https://doi.org/10.1038/nrc.2017.82

Tait-Burkard, C., Doeschl-Wilson, A., McGrew, M. J., Archibald, A. L., Sang, H. M., Houston, R. D., Whitelaw, C. B., & Watson, M. (2018). Livestock 2.0 – genome editing for fitter, healthier, and more productive farmed animals. Genome Biology, 19(1), 204. https://doi.org/10.1186/s13059-018-1583-1

Tälle, M., Wiréhn, L., Ellström, D., Hjerpe, M., Huge-Brodin, M., Jensen, P., Lindström, T., Neset, T.-S., Wennergren, U., & Metson, G. (2019). Synergies and Trade-Offs for Sustainable Food Production in Sweden: An Integrated Approach. *Sustainability*, *11*(3), Articolo 3. https://doi.org/10.3390/su11030601

Tang, K., Thornton, K. R., & Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLOS Biology*, *5*(7), e171. https://doi.org/10.1371/journal.pbio.0050171

Thrupp, L. A. (1989). Legitimizing local knowledge: From displacement to empowerment for third world people. *Agriculture and Human Values*, *6*(3), 13–24. https://doi.org/10.1007/BF02217665

Tolone M, Sardina MT, Senczuk G, et al (2022) Genomic Tools for the Characterization of Local Animal Genetic Resources: Application in Mascaruna Goat. Animals 12:2840.

https://doi.org/10.3390/ani12202840

Tullo, E., Finzi, A., & Guarino, M. (2019). Review: Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy. *Science of The Total Environment*, *650*, 2751–2760. https://doi.org/10.1016/j.scitotenv.2018.10.018

UNCED. (1992a). Rio Declaration on Environment and Development. *Environmental Conservation*, *19*(4), 366–368. https://doi.org/10.1017/S037689290003157X

UNCED. (1992b). *United Nations Framework Convention on Climate Change*. https://unfccc.int/sites/default/files/convention_text_with_annexes_english_for_posting.pdf

UNCED. (1992c, aprile 21). *Non-legally binding authoritative statement of principles for a global consensus on the management, conservation and sustainable development of all types of forests: Note /: By the Secretary-General of the Conference.* UN Conference on Environment and Development (1992 : Rio de Janeiro, Brazil). Secretary-General. https://digitallibrary.un.org/record/144461

United Nations. (1992). Convention on biological diversity. In *Treaty Series* (Vol. 1760, p. 79). https://treaties.un.org/doc/Treaties/1992/06/19920605%2008-44%20PM/Ch_XXVII_08p.pdf

Valliant, J. C. D., Farmer, J. R., Dickinson, S. L., Bruce, A. B., & Robinson, J. M. (2017). Family as a catalyst in farms' diversifying agricultural products: A mixed methods analysis of diversified and non-diversified farms in Indiana, Michigan and Ohio. *Journal of Rural Studies*, *55*, 303–315. https://doi.org/10.1016/j.jrurstud.2017.08.017

van der Werf, J. (2013). Genomic Selection in Animal Breeding Programs. In C. Gondro, J. van der Werf, & B. Hayes (A c. Di), *Genome-Wide Association Studies and Genomic Prediction* (pp. 543–561). Humana Press. https://doi.org/10.1007/978-1-62703-447-0_26

van Marle-Köster, E., Lashmar, S. F., Retief, A., & Visser, C. (2021). Whole-genome snp characterisation provides insight for sustainable use of local south african livestock populations. Frontiers in Genetics, 12. https://doi.org/10.3389/fgene.2021.714194

Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T.,

Haudenschild, C. D., Moore, S. S., Warren, W. C., & Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, *5*(3), Articolo 3. https://doi.org/10.1038/nmeth.1185

Velten, S., Leventon, J., Jager, N., & Newig, J. (2015). What Is Sustainable Agriculture? A Systematic Review. *Sustainability*, *7*(6), Articolo 6. https://doi.org/10.3390/su7067833

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, *34*(3), 275–305. https://doi.org/10.1051/gse:2002009

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*, *23*(21), 4407–4414. https://doi.org/10.1093/nar/23.21.4407

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), Articolo 4356. https://doi.org/10.1038/171737a0

Windig, J. J., & Engelsma, K. A. (2010). Perspectives of genomics for genetic conservation of livestock. Conservation Genetics, 11(2), 635–641. https://doi.org/10.1007/s10592-009-0007-x

Yi, X., & Latch, E. K. (2022). Nonrandom missing data can bias Principal Component Analysis inference of population genetic structure. *Molecular Ecology Resources*, *22*(2), 602–611. https://doi.org/10.1111/1755-0998.13498

Youssef, N., Budd, A., & Bielawski, J. P. (2019). Introduction to genome biology and diversity. In M. Anisimova (A c. Di), Evolutionary Genomics: Statistical and Computational Methods (pp. 3–31). Springer. https://doi.org/10.1007/978-1-4939-9074-0_1

Zhao, P., Peng, C., Fang, L., Wang, Z., & Liu, G. E. (2023). Taming transposable elements in livestock and poultry: A review of their roles and applications. Genetics Selection Evolution, 55(1), 50. https://doi.org/10.1186/s12711-023-00821-2

Zhao, Y., Wang, K., Wang, W., Yin, T., Dong, W., & Xu, C. (2019). A high-throughput SNP

discovery strategy for RNA-seq data. BMC Genomics, 20(1), 160.

https://doi.org/10.1186/s12864-019-5533-4