Università degli Studi di Milano

PhD Course in Veterinary and Animal Science

Genomics of Efficiency, Resilience and Biodiversity in Holstein and local cattle breeds

Francesca Bernini

Università degli Studi di Milano

Department of Veterinary Medicine and Animal Sciences (DIVAS)

PhD Course in Veterinary and Animal Science

Class XXXVII

Genomics of Efficiency, Resilience and Biodiversity in Holstein and local cattle breeds

(AGRI-09/A: Livestock Systems, Animal Breeding and Genetics)

PhD candidate: Francesca Bernini

Supervisor: Prof. Alessandro Bagnato

PhD course coordinator: Prof. Fabrizio Ceciliani

Academic year: 2023/2024

To my big family,

for always believing in me

Table of contents

Abstract

This thesis explores the genomics of biodiversity, resilience, and efficiency in Holstein (a cosmopolitan population) and local cattle breeds, focusing on their genetic diversity, adaptation to environmental challenges and the application of genomic tools in breeding programs. The research primarily investigates genetic diversity among autochthonous cattle breeds in the Aosta Valley, particularly the Aosta Red Pied, Aosta Black Pied-Chestnut, and their subgroup Aosta Chestnut-Hérens, alongside the highly selected Holstein Friesian breed.

The study leveraged genomic information such as Single Nucleotide Polymorphism (SNP) markers to assess genetic variation across breeds and conduct genome-wide association studies. Emphasis was placed on exploring the impact of genomic variations on traits linked to disease resistance (e.g., mastitis and bovine respiratory disease) and milk production efficiency. The Holstein breed, a specialised cosmopolitan population known for its high milk production and intensive farming system, served as a reference. In contrast, local breeds like Aosta were studied for their resilience and production efficiency in harsh environmental conditions, such as alpine mountain pastures.

Key findings include the identification of Runs of Homozygosity (ROH), used to estimate inbreeding and genomic diversity and to map genomic regions under selection. The ROH analysis showed that local breeds possess unique adaptations, exhibiting lower inbreeding coefficients than the Holstein, which has faced a loss of genetic variation due to the intense selection that occurred for milk production traits in the early past. This thesis also mapped genomic regions associated with mastitis resistance in Holstein, contributing to better breeding strategies aimed at improving animal health and welfare.

Additionally, this work highlights the critical role of local breeds in maintaining biodiversity. These breeds offer a source of genetic variation essential for sustainable cattle farming and their resilience could help in facing future environmental changes. The genetic makeup of local breeds also supports cultural and economic activities in the Aosta Valley, such as the production of Fontina cheese, a Product of Designated Origin (PDO), and the *Bataille de Reines*.

Finally, this thesis proposes that coupling genomic data with phenotypic variability enhances the understanding of resilience traits in cattle. It suggests that local breeds, despite their small population size, are valuable resources in the face of global challenges such as climate change and the increasing demand for sustainable livestock systems.

1. General introduction

1.1 Cattle genomics

The animal genomics era saw its first days with two studies by Beckmann and Soller published in 1983 (Beckmann & Soller, 1983; Soller & Beckmann, 1983). Soller predicted that genes and Quantitative Trait Loci (QTL) of economically important traits in dairy cattle could be mapped and used for genetic improvement by a Marker Assisted Selection (MAS) approach (Soller, 1978). However, in the late 1970s this was still practically not possible due to the low number of known polymorphic markers. From 1989 onwards, Short Tandem Repeats (STR), also known as microsatellites, began to be identified. These markers are in high numbers throughout the genome, endow with high intraspecific variability and allowed the beginning of QTL mapping projects in cattle (Beckmann & Soller, 1990; Georges et al., 1995).

The QTL mapping registered another massive step forward with the whole genome sequencing of *Bos taurus* and the genome mapping of dozens of thousands of Single Nucleotide Polymorphism (SNP) markers.

In the ‘00s the first comparative maps, with different resolution levels, of the cattle-human genomes were released, (Band et al., 2000; Larkin et al., 2003; Wind et al., 2005). Despite the efforts, the first version of the refined and annotated sequence of the *Bos taurus* genome dates back to 2009 (Elsik et al., 2009; Zimin et al., 2009). Recently the availability of long read sequencing technology, which further improves the sequencing accuracy, allowed the release of the last revised and reassessed bovine reference genome (Eché et al., 2023).

A complete reference genome allowed the easy use of many thousands of SNP markers across the bovine genome, paving the way to a large variety of genomic-based studies on cattle populations (Oldenbroek, 2017). Since 2010 low-cost genome-wide SNP arrays allowed the genotyping of millions of individuals and revolutionized the study and management of genetic diversity as well as breeding strategies in cattle.

Before the availability of a reference genome, the additive genetic merit values of bulls were obtained by analysing the phenotypes of their descendants (*progeny testing)*. When the cattle genome sequence information became available, the Genomic Selection (GS) theorized by Meuwissen et al. in 2001 was implemented worldwide in selection programs. GS had a rapid expansion and nowadays it also incorporates the cows’ genotypic information that are widely produced at farm level (Jenko et al., 2017; Misztal et al., 2020). Recently, GS has been also proposed for innovative resilience and efficiency traits recorded by farmers themselves (Gaddis et al., 2014), such as mastitis treatments and longitudinal data obtained from the automatic milking systems, but also for the maintenance and management of autochthonous cattle populations (Jones & Wilson, 2022). Nevertheless, genomic information is also widely used to study genetic diversity among breeds (Mastrangelo et al., 2018; Senczuk et al., 2021; Strillacci et al., 2020), to identify selection signatures (Saravanan et al., 2020) and to map genes related to economically important traits (Bernini et al., 2024; Freebern et al., 2020). At present, the increased availability of female genotypes will likely open new possibilities for Genome Wide Association Studies (GWAS) to identify QTL even for low heritability and complex traits, such as mastitis resistance and efficiency (Li et al., 2019; Strillacci et al., 2023; Tiezzi et al., 2015).

1.2 Local cattle breeds and genetic diversity

The availability of high-resolution genetic maps and millions of potential genetic markers in the bovine genome sequence data have been crucial to advances in genomic selection. In the last century, many resources and studies have been focused on the genetic improvement of livestock including bovine. Primarily because cattle are still one of the main sources of protein, in the form of dairy and meat products, often allowing the usage of grasslands that are not suitable as a source of food for humans (e.g. Figure 1). Secondarily they are also particularly important in developing countries to allow a healthy and balanced diet but also economically sustain families (Herrero et al., 2013). Even though cattle farming is seen as one of the major contributors to climate change, it is an activity that minorly affects the overall Greenhouse Gases (GHG), i.e. 14.5% (https://www.fao.org/). Indeed, livestock farming and all the connected activities are some of the sectors majorly affected by environmental changes. The effects of droughts, rainfalls, floods, and high temperatures during the year affect the production of feed (Pachauri et al., 2014); but also, the stress due to heat waves can cause a reduction in productivity, welfare, and fertility of farmed animals while increasing their susceptibility to diseases, and eventually the mortality (Thornton et al., 2021). When considering the "developed countries", the effects of climate change can also have significative consequences on the Products of Designated Origin (PDO), not only by reducing their quantity but also by changing their palatability and quality. This impacts also the entire food chain commercialisation, including Large Organized Distribution (GDO) and the final consumers (Pulighe et al., 2024; Vitali et al., 2019).

In the last decade, the negative effects of climate change were registered in many countries, that started to take action to mitigate them. For example, in Europe, the Common Agricultural Policy

(CAP) 2023-2027 (https://agriculture.ec.europa.eu/cap-my-country/cap-strategic-plans/italy_en) acknowledges the climate action.

The anticipated decline in agricultural productivity caused by climate change, combined with the challenges that farmers encounter in adapting their animal management practices to it, could severely affect food security and economic sustainability. A solution to this situation is especially urgent as the demand for food continues to rise (Michalk et al., 2019; WHO, 2023).

For all the above-mentioned reasons, the current livestock production system, as it is structured nowadays, most likely will not guarantee food security for the growing world population that is expected to reach 10 billion by 2100 (Sakschewski et al., 2014). Reducing food production to only a handful of highly specialized breeds may represent a limiting factor in facing and solving these challenges (Boudalia et al., 2020) as the existence of biodiversity is one of the key points that can help us face them.

The term biodiversity encompasses the range of variation across genetic differences, biological functions, or ecosystem types (Broom et al., 2013). Local cattle breeds, as a source of genetic diversity and biodiversity, are now taken into consideration because they can help in facing future needs (Moulin & Ickowicz, 2023). They often present unique characteristics that facilitate them to adapt to different environmental and climatic conditions, implying a better response to environmental stressors and challenges (Biscarini et al., 2015). They also represent a source of immaterial socio-cultural and ecosystem benefits. Overall, local breeds are raised in pasture-based systems, an activity based on local resources that makes this practice connected with and limited to a specific ecosystem (Marsoner et al., 2018).

In the second half of the 20th century, livestock genetic diversity suffered an unprecedented decline (Bett et al., 2013) as the focus on increasing farm profitability led to the widespread use of specialized, high-yielding breeds, displacing multi-purpose local breeds. Additional contributing factors can be represented by genetic introgression, lack of public incentives and funding, but also external pressures like natural disasters, political instability, and fashion-driven trends among breeders (Hiemstra et al., 2010). Even though the problem has been well known for years, nowadays there is still a major number of local endangered breeds, especially in Europe and Caucasian countries, Figure 2.

To best manage the available biodiversity resources, is fundamental to identify and classify all the characteristics that differentiate species and breeds. For this reason, in autochthonous populations, in addition to a precise and accurate phenotype description, it's essential to frame their genetic variability and proprietary genomic features. The application of GS to improve the population and jointly manage their genomic variability is difficult in small breeds and, so far, not applied. However, researchers are currently carrying out studies to develop it (Kenny et al., 2022; Mancin et al., 2021).

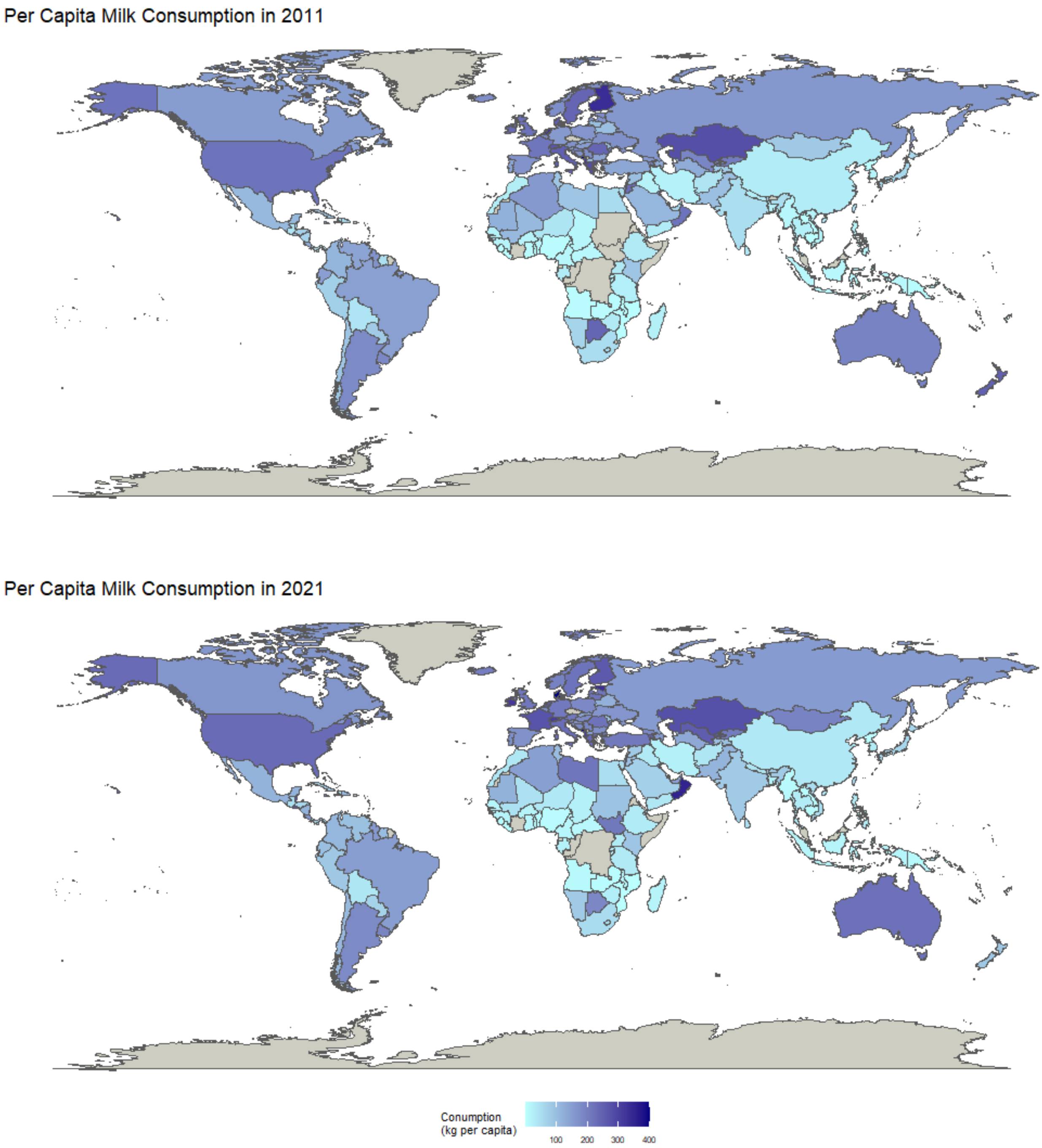


Figure 1. Average per capita milk consumption in 2011 and 2021, measured in kilograms per person per year. Including the milk equivalents of dairy products made from milk ingredients without considering the butter. Image obtained from the FAO 2023 data processed by: Ritchie et al., 2024).

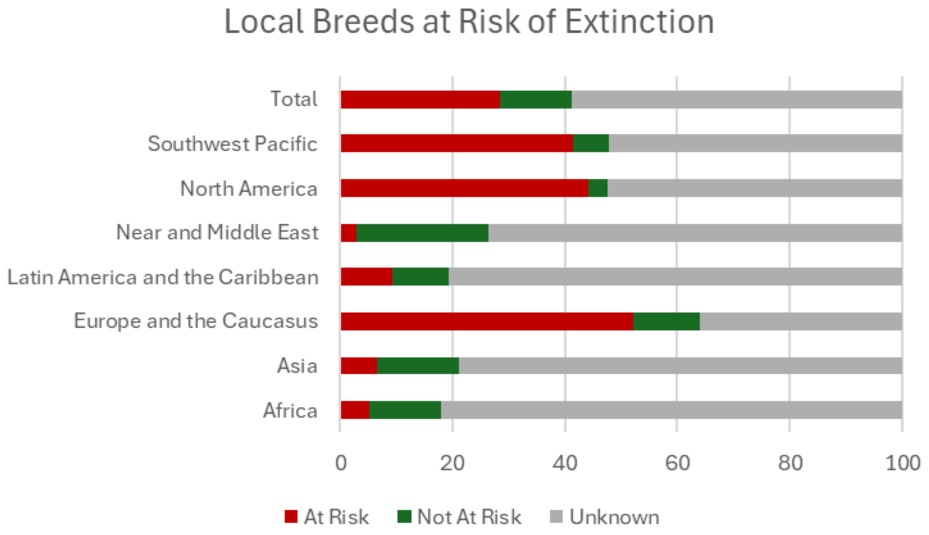


Figure 2. Proportion of local breeds, classified as being at risk, not-at risk or of unknown level of risk of extinction, data obtained from Domestic Animal Diversity Information System (DAD-IS).

1.3 Resilience and efficiency in cattle

From its introduction genetic selection in livestock was mainly addressed to increasing production, i.e., maximizing outputs in milk, meat, or eggs (Thornton, 2010). Decades of this type of selective breeding have significantly altered the energy allocation patterns in dairy cattle, leading to reduced reproductive efficiency, disease resistance, and overall resilience in highly selected breeds, such as the Holstein (Brito et al., 2021). For this reason, in the last 20 years, the selection for functional traits (i.e. robustness) has been included in breeding plans, especially in dairy cattle where the cow functionality (i.e. fertility, health and longevity) is a key factor for the farmer profitability (Miglior et al., 2017). Additionally, more recently, increasing efforts have been placed in selecting climate-resilient cattle and reducing GHG emissions (González-Recio et al., 2020; Hayes et al., 2013). The negative effects of the strong selection in cosmopolitan cattle have been partially reversed by the new breeding methods, especially regarding functional traits that are easily measured, e.g., in the Holstein milk somatic cell count - to reduce mastitis incidence and reproductive parameters - to improve fertility. However, new selection objectives, i.e. resilience, and reduction in emission, direct and affordable measures of the phenotype are still missing, limiting the possibility of selection for these traits.

Resilience can be defined as the animal’s ability to return to the normal state (regarding health, fertility, and production), after a wide range of disruptive challenges. It is worth mentioning that the Resilient Dairy Genome Project - RDGP defined resilience also as “the capacity of the animal to adapt rapidly to changing environmental conditions, without compromising its productivity, health or fertility while becoming more resource-efficient and reducing its environmental burden" (van Staaveren et al., 2024).

The common factor shared by all the various definitions is that one organism, to be defined as resilient, after facing the stressor should go back to the normal state as quickly as possible. Furthermore, resilience should not be confused with robustness which is the ability to function at optimal levels across a wide spectrum of environments. Resilience is the ability to meet the day-to-day challenges within a given environment (Colditz & Hine, 2016).

Breeding for increased disease resistance can be one of the main preventive measures to reduce the use of antibiotics in livestock farms according to a One Health approach. We all know that is impossible to breed cattle that is fully resistant to pathogens but one of the achievable goals is to select animals that can overcome infections by themselves and go back to a normal state rapidly, and so that are more resilient (Doeschl-Wilson et al., 2021). For the above cited reasons, selecting disease-resilient animals means, not only, obtaining subjects that have a higher resistance and tolerance to pathogenic infections (Knap & Doeschl-Wilson, 2020).

The term efficiency involves a plethora of different aspects that intervene in its definition and for this reason, it can be sub-divided into a wide range of aspects.

I will mainly focus on production efficiency, which is extremely important for less-industrialized regions where intensive farming cannot be practiced. Higher production efficiency is also fundamental in the context of the maintenance of local cattle breeds (Brito et al., 2021b) since they are mainly farmed in marginal areas where they contribute to maintaining the environment and landscapes, e.g. in Aosta Valley. Moreover, higher production efficiency dilutes the maintenance requirements and increases the farmers' revenues (Bauman & Capper, 2010).

In reverse, in areas where the intensified production system is higher, the focus will be placed on the reduction of the environmental impact. This is to improve the feed conversion efficiency, that is the ability of an animal to transform the energy of the diet into outputs (e.g. milk, meat, or eggs) (Bach et al., 2020).

In any case, one of the requirements for more resilient or efficient cattle selection is the knowledge of the QTLs affecting these polygenic traits. It can be achieved with genotype-to-phenotype association studies.

1.4 Studied breeds

For this project, three autochthonous and one cosmopolitan cattle breeds have been considered. The local Aosta breeds, Figure 3, are represented by the Aosta Red Pied (ARP) and the Aosta Black Pied

- Chestnut (ABC) with the subgroup of Aosta Chestnut that have registered in the pedigree an Hérens ascendant (ACH) (Bernini et al., 2023). These local breeds are extremely important for the Aosta Valley because they represent the most farmed cattle breeds in the valley, and their milk is almost entirely used to produce the Fontina cheese, a Product of Designated Origin (PDO) of the Aosta Valley. These cattle have also a really important role in the maintenance of the valley (Strillacci et al., 2020). Their farming system follows a seasonality: in the winter period the housing is inside the stables, in spring and autumn the grazing is in the valley while in summer it goes up to the mountain pastures, where they can reach up to 2500 meters above the sea level.

They also have a big cultural value for the Aosta Valley. Every year numerous tournaments take place during the grazing period across the entire region, culminating with the "final battle" in October in the Arena Coix-Noir, in Aosta. This tournament is called "*Bataille de Reines*" (Queens' fight) and consists of a series of cruelty-free fights between pregnant cows to establish the "Queen", the highest cow in the hierarchy (Sartori & Mantovani, 2010). This is not only a recognition for the farmer, that has the queen in the herd, but also it can help during the transhumance since the queen will lead all the other cows to the summer pastures.

The Aosta cattle breeds are well known for their hardiness and the ability to adapt to the harsh environmental conditions of the mountains Alps. The Aosta Red Pied breed likely traces its origins back to the spotted livestock (descendants of the *Bos frontosus*) brought by the Burgundian tribes. These tribes possibly carried the cattle from distant Sweden, across the German plains, into Western Switzerland, Central-Eastern France, and eventually into our Alpine valleys around the 5th century AD. The Aosta Black Pied and Chestnut breeds most probably descend from the *Bos brachyceros* an autochthonous bovine group that originally populated the Alps and was characterized by brachycephalism (a broad skull), a lively character, and rusticity. The lively and rather aggressive temperament of these cattle has been maintained and is shown by the Aosta breeds during the grazing season.

The Aosta Red Pied coat is deep red (intense and dark), tending to violet, with limbs and ventral regions generally white. The head is white with red ears, red spotted heads are not usually appreciated by farmers. The muzzle is pink with short yellowish horns. The size and stature are small, but the figure is harmonious. The Aosta Black Pied is very similar to the Red Pied, despite being usually more "rustic" it has a more solid and harmonious conformation. Besides the Pied coat, it generally presents white limbs and ventral regions, and a black head with white spots on the forehead. The Aosta Chestnut is instead more similar to the Swiss Hérens. The coat is uniform, generally black, or rather variable in colour from black to light brown.

The Aosta breeds are dual-purpose breeds. The milk production is appreciable when correlated to the animal size, its dietary requirements, and its breeding condition environment. In 2023, the Aosta Red Pied produced an average milk production of 3921 Kg with 3.35% fats and 3.16% proteins, and an average weight of 312 Kg at 11 months. In the same year, the Aosta Black Pied – Chestnut, less devoted to milk production, registered an average milk production of 2760Kg with 3.41% fats and 3.30% protein and a body weight of 300Kg (https://anaborava.it/rzz_duplice.php).

Figure 3. The Aosta cattle breeds during the pasture, from left to right: the Aosta Black Pied, the Aosta Chestnut, and the Aosta Red Pied.

The other breed considered in the present project is the Holstein Friesian. This is one of the most cosmopolitan dairy cattle breeds, with approximately 70 million animals worldwide. It is renowned for its high milk yield production and efficiency (Andrew et al., 2021). Originating in the Netherlands, specifically in the regions of North Holland and Friesland, the breed has been selectively bred for dairy production over the past 2000 years. Holstein Friesians are easily recognized by their distinctive black and white or red and white coat patterns, with the black and white variety being more common. These cattle are pivotal in the dairy industry as they produce large quantities of milk with an average production of 10,000 to 12,000 litres per lactation cycle (https://agriculture.canada.ca/en/sector/animal-industry/canadian-dairy-information-centre/statistics-market-information/dairy-animal-genetics/production-breed/per-cow), with fat and protein content typically around 3.6% and 3.2%, respectively. Holstein Friesians are primarily raised in intensive farming systems (Figure 4) where they thrive in temperate climates. However, their high milk yield necessitates a well-balanced and nutrient-rich diet to maintain production and health, making them more demanding in terms of feed compared to other breeds (Lawrence et al., 2016). Beyond their dairy potential, Holstein Friesians also play a significant role in genetic improvement programs worldwide, contributing to the advancement of dairy cattle genetics (Medeiros et al., 2022). Furthermore, being one of the most abundant cattle they have a pivotal role in the development of genomic selection making them one of the most studied animals breeds in the world (Miglior et al., 2017).

Figure 4. Holstein-Friesian cattle in their typical farming environment.

1.5 Context of the research

Recent advances in cattle genomics, including the widespread use of SNP arrays and Whole Genome Sequencing (WGS), have provided a more detailed view of genetic variability across cattle breeds (Strucken et al., 2021, van Breukelen et al., 2019, Makina et al., 2014, Hulsegge et al., 2022). Key genes associated with productivity, disease resistance, and environmental adaptation have been identified, yet their function and impact in local breeds remain underexplored. Local breeds exhibit greater genetic variability than cosmopolitan breeds at certain loci, suggesting the presence of unique adaptive mechanisms. However, a deeper understanding of structural polymorphisms (SNPs, Indels, CNVs) that directly influence resilience and efficiency under different environmental conditions and in different populations (e.g. Holstein in intensive farming structure, Aosta breed in extensive farming system) is still lacking. Moreover, most research has focused on production traits rather than resilience-related traits, leaving an open question about how the latter can be integrated into breeding programs.

This thesis contributes to filling this gap by investigating genomic diversity, structural variants, and the genetic basis of resilience in both local and cosmopolitan breeds.

Genetic diversity is a fundamental element for the sustainability and resilience of cattle populations, allowing adaptation to changing environmental conditions and ensuring sufficient variability for future selection (FAO, 2015). However, studies on genetic diversity have historically focused on highly selected cosmopolitan breeds, such as the Holstein, due to their economic impact. In contrast, local breeds, despite their crucial role in biodiversity conservation and adaptability, have been less characterized at the genomic level (Signer-Hasler et al., 2023). Additionally, managing genetic variability in breeding programs remains a challenge, particularly in small populations where inbreeding and genetic erosion pose significant risks (Restoux et al., 2022). Another critical aspect is the lack of integrated strategies to maintain genetic diversity in high-yielding populations without compromising economic performance. This thesis tries to address these gaps by combining genetic diversity analyses, structural variant identification, and assessments of genetic variability management in both local and cosmopolitan cattle breeds.

Moreover, local and cosmopolitan breeds have evolved under different selective pressures, leading to significant differences in their ability to adapt to harsh environmental conditions (Rovelli et al. 2020). Cosmopolitan breeds, such as Holstein, have been primarily selected for production efficiency, focusing on milk yield and growth rate (Miglior et al., 2017). This has often resulted in reduced genetic variability for traits related to disease resistance and environmental adaptability (Brito et al., 2021). In contrast, local breeds, such as the Aosta cattle, have developed in environments with limited resources and challenging climatic conditions, maintaining traits linked to resilience, including disease resistance and efficient resource utilization (Strillacci et al., 2020). However, the genetic basis of resilience and efficiency remains poorly understood, and methodologies to quantify it at the genomic level are still limited. This thesis aims to contribute to bridge this gap by comparing the structural genomic differences between local and cosmopolitan breeds and by identifying genomic variants associated with resilience and efficiency traits in different environmental contexts.

The analysis of genetic diversity and resilience in cattle breeds has been explored through complementary approaches in the different chapters of this thesis. Chapter 3 will focus on characterizing genetic diversity in both local and cosmopolitan breeds by assessing allele frequencies of key genes (i.e. monogenic controlled traits such as k-casein or Mendelian diseases) and identifying highly heterozygous genomic regions, indicative of genomic regions under selection. Additionally, Copy Number Variants (CNVs) will be examined, as they may have functional implications for productivity and resilience. Chapter 4 will investigate genome-wide association studies (GWAS) to identify loci associated with production traits, such as milk yield in Aosta breeds, mastitis susceptibility in Holstein, and disease resistance in pre-weaned dairy calves. Finally, Chapter 5 will provide general conclusions, synthesizing the findings and discussing future strategies for incorporating these insights into genetic selection programs.

References:

Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A. W., & Burghardt, T. (2021). Visual identification of individual Holstein-Friesian cattle via deep metric learning. Computers and Electronics in Agriculture, 185, 106133.

Band, M. R., Larson, J. H., Rebeiz, M., Green, C. A., Heyen, D. W., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., & Womack, J. E. (2000). An ordered comparative map of the cattle and human genomes. Genome Research, 10(9), 1359–1368.

Beckmann, J. S., & Soller, M. (1983). Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theoretical and Applied Genetics, 67, 35–43.

Beckmann, J. S., & Soller, M. (1990). Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/Technology, 8(10), 930–932.

Bernini, F., Mancin, E., Sartori, C., Mantovani, R., Vevey, M., Blanchet, V., Bagnato, A., & Strillacci, M. G. (2024). Genome Wide Association Studies for milk production traits in two autochthonous Aosta cattle breeds. Animal, 101322.

Bernini, F., Punturiero, C., Vevey, M., Blanchet, V., Milanesi, R., Delledonne, A., Bagnato, A., & Strillacci, M. G. (2023). Assessing major genes allele frequencies and the genetic diversity of the native Aosta cattle female population. Italian Journal of Animal Science, 22(1), 1008–1022.

Bett, R. C., Okeyo, M. A., Malmfors, B., Johansson, K., Agaba, M., Kugonza, D. R., Bhuiyan, A., Vercesi Filho, A. E., Mariante, A. S., & Mujibi, F. D. (2013). Cattle breeds: extinction or quasi-extant? Resources, 2(3), 335–357.

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, 33.

Boudalia, S., Ben Said, S., Tsiokos, D., Bousbia, A., Gueroui, Y., Mohamed-Brahmi, A., Smeti, S., Anastasiadou, M., & Symeon, G. (2020). BOVISOL project: breeding and management practices of indigenous bovine breeds: solutions towards a sustainable future. Sustainability, 12(23), 9891.

Brito, L. F., Bédère, N., Douhard, F., Oliveira, H. R., Arnal, M., Peñagaricano, F., Schinckel, A. P., Baes, C. F., & Miglior, F. (2021). Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. Animal, 15, 100292.

Broom, D. M., Galindo, F. A., & Murgueitio, E. (2013). Sustainable, efficient livestock production with high biodiversity and good welfare for animals. Proceedings of the Royal Society B: Biological Sciences, 280(1771), 20132025.

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.

Doeschl-Wilson, A., Knap, P. W., Opriessnig, T., & More, S. J. (2021). Livestock disease resilience: from individual to herd level. Animal, 15, 100286.

Eché, C., Iampietro, C., Birbes, C., Dréau, A., Kuchly, C., Di Franco, A., Klopp, C., Faraut, T., Djebali, S., & Castinel, A. (2023). A Bos taurus sequencing methods benchmark for assembly, haplotyping, and variant calling. Scientific Data, 10(1), 369.

Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A., Muzny, D. M., Weinstock, G. M., Adelson, D. L., Eichler, E. E., & Elnitski, L. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science, 324(5926), 522.

FAO. (2015) Coping with climate change – the roles of genetic resources for food and agriculture. Rome. ISBN 978-92-5-108441-0.

Freebern, E., Santos, D. J. A., Fang, L., Jiang, J., Parker Gaddis, K. L., Liu, G. E., VanRaden, P. M., Maltecca, C., Cole, J. B., & Ma, L. (2020). GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC Genomics, 21(1), 1–11.

Gaddis, K. L. P., Cole, J. B., Clay, J. S., & Maltecca, C. (2014). Genomic selection for producer-recorded health event data in US dairy cattle. Journal of Dairy Science, 97(5), 3190–3199.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorensen, A., Steele, M. R., & Zhao, X. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics, 139(2), 907–920.

González-Recio, O., López-Paredes, J., Ouatahar, L., Charfeddine, N., Ugarte, E., Alenda, R., & Jiménez-Montero, J. A. (2020). Mitigation of greenhouse gases in dairy cattle via genetic selection: 2. Incorporating methane emissions into the breeding goal. Journal of Dairy Science, 103(8), 7210–7221.

Hayes, B. J., Lewin, H. A., & Goddard, M. E. (2013). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. Trends in Genetics, 29(4), 206–214.

Herrero, M., Grace, D., Njuki, J., Johnson, N., Enahoro, D., Silvestri, S., & Rufino, M. C. (2013). The roles of livestock in developing countries. Animal, 7(s1), 3–18.

Hiemstra, S. J., de Haas, Y., Mäki-Tanila, A., & Gandini, G. (2010). Local cattle breeds in Europe: development of policies and strategies for self-sustaining breeds. Wageningen Academic Publishers.

Hulsegge, I., Oldenbroek, K., Bouwman, A., Veerkamp, R., & Windig, J. (2022). Selection and drift: A comparison between historic and recent Dutch Friesian cattle and recent Holstein Friesian using WGS data. Animals, 12(3), 329.

Jenko, J., Wiggans, G. R., Cooper, T. A., Eaglen, S. A. E., Luff, W. G. de L., Bichard, M., Pong-Wong, R., & Woolliams, J. A. (2017). Cow genotyping strategies for genomic selection in a small dairy cattle population. Journal of Dairy Science, 100(1), 439–452. https://doi.org/10.3168/jds.2016-11479

Jones, H. E., & Wilson, P. B. (2022). Progress and opportunities through use of genomics in animal production. Trends in Genetics, 38(12), 1228–1252.

Kenny, D., Carthy, T. R., Murphy, C. P., Sleator, R. D., Evans, R. D., & Berry, D. P. (2022). The Association between Genomic Heterozygosity and Carcass Merit in Cattle. Frontiers in Genetics, 13, 789270.

Knap, P. W., & Doeschl-Wilson, A. (2020). Why breed disease-resilient livestock, and how? Genetics Selection Evolution, 52, 1–18.

Larkin, D. M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P. A., Bachman, S., Green, C., Wright, C. L., Campos, E. J., Benson, L. D., & Edwards, J. (2003). A cattle–human comparative map built with cattle BAC-ends and human genome sequence. Genome Research, 13(8), 1966–1972.

Lawrence, D., O’Donovan, M., Boland, T. M., Lewis, E., & Kennedy, E. (2016). An examination of two concentrate allocation strategies which are based on the early lactation milk yield of autumn calving Holstein Friesian cows. Animal, 10(5), 796–804.

Li, B., Fang, L., Null, D. J., Hutchison, J. L., Connor, E. E., VanRaden, P. M., VandeHaar, M. J., Tempelman, R. J., Weigel, K. A., & Cole, J. B. (2019). High-density genome-wide association study for residual feed intake in Holstein dairy cattle. Journal of Dairy Science, 102(12), 11067–11080.

Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., MacNeil, M. D., Maiwashe, A. (2014). Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Frontiers in genetics, 5, 333.

Mancin, E., Tuliozi, B., Sartori, C., Guzzo, N., & Mantovani, R. (2021). Genomic prediction in local breeds: The rendena cattle as a case study. Animals, 11(6), 1–19. https://doi.org/10.3390/ani11061815

Marsoner, T., Vigl, L. E., Manck, F., Jaritz, G., Tappeiner, U., & Tasser, E. (2018). Indigenous livestock breeds as indicators for cultural ecosystem services: A spatial analysis within the Alpine Space. Ecological Indicators, 94, 55–63.

Mastrangelo, S., Ciani, E., Ajmone Marsan, P., Bagnato, A., Battaglini, L., Bozzi, R., Carta, A., Catillo, G., Cassandro, M., & Casu, S. (2018). Conservation status and historical relatedness of Italian cattle breeds. Genetics Selection Evolution, 50(1), 1–16.

Medeiros, I., Fernandez-Novo, A., Astiz, S., & Simões, J. (2022). Historical evolution of cattle management and herd health of dairy farms in OECD Countries. Veterinary Sciences, 9(3), 125.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4), 1819–1829.

Michalk, D. L., Kemp, D. R., Badgery, W. B., Wu, J., Zhang, Y., & Thomassin, P. J. (2019). Sustainability and future food security—A global perspective for livestock production. Land Degradation & Development, 30(5), 561–573.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., & Baes, C. F. (2017). A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. Journal of Dairy Science, 100(12), 10251–10271.

Misztal, I., Lourenco, D., & Legarra, A. (2020). Current status of genomic evaluation. Journal of Animal Science, 98(4), skaa101–skaa101.

Moulin, C.-H., & Ickowicz, A. (2023). Livestock grazing systems and sustainable development in the Mediterranean and Tropical areas: Recent knowledge on their strengths and weaknesses.

Oldenbroek, J. K. (2017). Genomic management of animal genetic diversity. Wageningen Academic Publishers.

World Health Organization (2023). The State of Food Security and Nutrition in the World 2023: Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum (Vol. 2023). Food & Agriculture Org.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., & Dasgupta, P. (2014). Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. Ipcc.

Pulighe, G., Di Fonzo, A., Gaito, M., Giuca, S., Lupia, F., Bonati, G., & De Leo, S. (2024). Climate change impact on yield and income of Italian agriculture system: a scoping review. Agricultural and Food Economics, 12(1), 23.

Restoux, G., Rognon, X., Vieaud, A., Guemene, D., Petitjean, F., Rouger, R., Brard-Fudulea, S., Lubac-Paye, S., Chiron, G., Tixier-Boichard, M. (2022). Managing genetic diversity in breeding programs of small populations: the case of French local chicken breeds. Genetic Selelection Evolution, 54(1):56.

Ritchie, H., Rosado, P., & Roser, M. (2024). Meat and dairy production. Our World in Data.

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, 104257.

Rovelli, G., Ceccobelli, S., Perini, F., Demir, E., Mastrangelo, S., Conte, G., Abeni, F., Marletta, D., Ciampolini, R., Cassandro, M., Bernabucci, U., Lasagna, E. (2020). The genetics of phenotypic plasticity in livestock in the era of climate change: a review. Italian Journal of Animal Science, 19(1), 997–1014.

Sartori, C., & Mantovani, R. (2010). Genetics of fighting ability in cattle using data from the traditional battle contest of the Valdostana breed. Journal of Animal Science, 88(10), 3206–3213. https://doi.org/10.2527/jas.2010-2899

Sakschewski, B., von Bloh, W., Huber, V., Müller, C., & Bondeau, A. (2014). Feeding 10 billion people under climate change: How large is the production gap of current agricultural systems? Ecological Modelling, 288, 103–111.

Senczuk, G., Mastrangelo, S., Ajmone-Marsan, P., Becskei, Z., Colangelo, P., Colli, L., Ferretti, L., Karsli, T., Lancioni, H., & Lasagna, E. (2021). On the origin and diversification of Podolian cattle breeds: testing scenarios of European colonization using genome-wide SNP data. Genetics Selection Evolution, 53(1), 1–16.

Signer-Hasler, H., Casanova, L., Barenco, A., Maitre, B., Bagnato, A., Vevey, M., Berger, B., Simčič, M., Boichon, D., Capitan, A., Medugorac, I., Bennewitz, J., Mészáros, G., Sölkner, J., Drögemüller, C., Flury, C. (2023). Genomic regions underlying positive selection in local, Alpine cattle breeds. Animal genetics, 54(3), 239-253.

Soller, M. (1978). The use of loci associated with quantitative effects in dairy cattle improvement. Animal Science, 27(2), 133–139.

Soller, M., & Beckmann, J. S. (1983). Genetic polymorphism in varietal identification and genetic improvement. Theoretical and Applied Genetics, 67, 25–33.

Strillacci, M. G., Punturiero, C., Milanesi, R., Bernini, F., Mason, T., & Bagnato, A. (2023). Antibiotic treatments and somatic cell count as phenotype to map QTL for mastitis susceptibility in Holstein cattle breed. Italian Journal of Animal Science, 22(1), 190–199.

Strillacci, M. G., Vevey, M., Blanchet, V., Mantovani, R., Sartori, C., & Bagnato, A. (2020). The Genomic Variation in the Aosta Cattle Breeds Raised in an Extensive Alpine Farming System. Animals, 10(12), 2385.

Strucken, E. M., Gebrehiwot, N. Z., Swaminathan, M., Joshi, S., Al Kalaldeh, M., & Gibson, J. P. (2021). Genetic diversity and effective population sizes of thirteen Indian cattle breeds. Genetics Selection Evolution, 53, 1-17.

Thornton, P. K. (2010). Livestock production: recent trends, future prospects. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), 2853–2867.

Thornton, P., Nelson, G., Mayberry, D., & Herrero, M. (2021). Increases in extreme heat stress in domesticated livestock species during the twenty-first century. Global Change Biology, 27(22), 5762–5772.

Tiezzi, F., Parker-Gaddis, K. L., Cole, J. B., Clay, J. S., & Maltecca, C. (2015). A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS One, 10(2), e0114919.

van Breukelen, A. E., Doekes, H. P., Windig, J. J., Oldenbroek, K. (2019). Characterization of genetic diversity conserved in the gene bank for Dutch cattle breeds. Diversity, *11*(12), 229.

van Staaveren, N., de Oliveira, H. R., Houlahan, K., Chud, T. C. S., Oliveira Jr, G. A., Hailemariam, D., Kistemaker, G., Miglior, F., Plastow, G., & Schenkel, F. S. (2024). The Resilient Dairy Genome Project—A general overview of methods and objectives related to feed efficiency and methane emissions. Journal of Dairy Science, 107(3), 1510–1522.

Vitali, A., Segnalini, M., Esposito, S., Lacetera, N., Nardone, A., & Bernabucci, U. (2019). The changes of climate may threat the production of Grana Padano cheese: past, recent and future scenarios. Italian Journal of Animal Science.

Wind, A. E. der, Larkin, D. M., Green, C. A., Elliott, J. S., Olmstead, C. A., Chiu, R., Schein, J. E., Marra, M. A., Womack, J. E., & Lewin, H. A. (2005). A high-resolution whole-genome cattle–human comparative map reveals details of mammalian chromosome evolution. Proceedings of the National Academy of Sciences, 102(51), 18526–18531.

Zimin, A. V, Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C. P., & Sonstegard, T. S. (2009). A whole-genome assembly of the domestic cow, Bos taurus. Genome Biology, 10, 1–10.

2. Aim and objectives

This thesis aims to leverage genomic information to explore phenotypic variability and genetic resilience to environmental challenges, integrating data from both cosmopolitan and autochthonous cattle breeds. The cost reduction in dense SNP-array genotyping makes female genotyping accessible and widely exploitable in cattle dairy farms. In the last two decades, several projects like GENOVAL (Aosta cattle breeds), GENORIP (Holstein breed), DUALBREEDING Fase 1 and 2 (dual-purpose breeds), LATTEco (Holstein and Brown Swiss breeds), and EU-Quantomics (Brown Swiss, Aosta Breed) facilitated the genotyping and recording of innovative phenotypes in cattle farms. These projects made also available Whole Genome Sequencing (WGS) data of the Aosta cattle breeds, delivering an unprecedented set of information and paving the way to new studies addressed to disclose genomic structural variation in these breeds. Comparing breeds with different selection goals and reared in distinct environments can, indeed, help to disclose i) specific genomic features related to their performances and ii) intrinsic genetic properties acquired as adaptive selection to two different environments. The two environments in here considered, are the intensive system, an artificial environment that is structured to provide maximum “comfort” to individuals, and the summer pasture, a natural environment in which the animals find severe challenging interactions.

The specific objectives of this thesis are:

a) To frame the genetic diversity of autochthonous and cosmopolitan cattle breeds, focusing on the Aosta breeds and the Holstein ones.

b) To identify genomic variants related to disease traits, with a particular focus on mastitis in the Holstein breed.

c) To identify proprietary genomic structural variants (e.g. SNP, Indel, CNV) for autochthonous cattle breeds using WGS data.

d) To identify candidate genes for resilience and efficiency traits in populations under differential selection/environments/attitude.

Modern dairy cattle populations represent a unique opportunity for genetic studies due to several factors: (i) the vast amount of genomic data generated through routine genotyping, (ii) well-documented population structures and pedigree records, and (iii) the systematic phenotyping performed within breeding programs. These characteristics make cattle an ideal model for investigating complex phenotypes, particularly in the context of resilience and adaptation to diverse farming conditions. By addressing these objectives, this thesis contributes to a deeper understanding of the genetic mechanisms underlying adaptation and resilience in cattle, providing valuable insights for the future of sustainable breeding strategies.

3. Genetic diversity of local and cosmopolitan cattle breeds

Genetic diversity is represented by all the genetic differences between and within populations and individuals (Lenstra et al., 2012). It is extremely important as it intervene in the survival of the species regulating their adaptability and evolution and, in farmed animals it is necessary to apply the selection (Woolliams & Oldenbroek, 2017). Conventionally livestock are classified in breeds, that were initially defined based on their origin and phenotype and now are differentiated also by their genomic information. Unfortunately, a specific definition of the term breed does not exist, but the Food and Agricultural Organization (FAO) defined it as “a subspecific group of domestic livestock with a common history whose members are treated in a common manner with respect to genetic management” (FAO, 2013).

Genetic diversity in populations is influenced by four key forces: mutation, migration, genetic drift, and selection (Doekes, 2020):

- Mutation is the primary source of genetic variation, introducing new mutations that are passed to offspring, when they appear in germ cells. The mutation rates are generally low, especially in cattle i.e., 1.21 × 10^-8 per bp (Harland et al., 2018), but several new mutations arise in each new generation.
- Migration impacts diversity through the exchange of genetic material between populations, facilitated in cattle by artificial insemination and practices such as the crossbreeding. Different populations (or breeds) often have distinct allele frequencies, and the direction and extent of migration affect genetic diversity within and across populations (Alcala et al., 2013).
- Genetic drift refers to random changes in allele frequency over generations, influenced by Mendelian inheritance (Masel, 2011). Smaller populations, or populations with unequal reproductive contributions, experience stronger drift, increasing the chance of allele loss by chance.
- Selection shapes diversity by favouring certain alleles, indirectly increasing genetic drift through unequal reproduction and directly reducing genetic variation by shifting allele frequencies in a specific direction (Liu et al., 2014).

Genetic diversity can be measured and characterized with different techniques based on the genomic information used, e.g. single genotypes from SNP array, WGS, sequences of mitochondrial DNA (Doekes, 2020). The genetic information used for the analysis influence the ability to identify precisely the genetic differences between breeds, the less precise being the low-density SNP array to the more precise whole genome resequencing and even better the next generation sequencing techniques (Olschewsky & Hinrichs, 2021).

References

Alcala, N., Streit, D., Goudet, J., & Vuilleumier, S. (2013). Peak and persistent excess of genetic diversity following an abrupt migration increase. Genetics, 193(3), 953–971.

Doekes, H. P. (2020). Genomic characterization and conservation of genetic diversity in cattle. Wageningen University.

Harland, C., Durkin, K., Artesi, M., Karim, L., Cambisano, N., Deckers, M., Mni, M., Mullaart, E., Coppieters, W., & Georges, M. (2018). Rate of de novo mutation in dairy cattle and potential impact of reproductive technologies. Proc World Congr Genet Appl to Livest Prod.

Lenstra, J. A., Groeneveld, L. F., Eding, H., Kantanen, J., Williams, J. L., Taberlet, P., Nicolazzi, E. L., Sölkner, J., Simianer, H., & Ciani, E. (2012). Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. Animal Genetics, 43(5), 483–502.

Liu, H., Sørensen, A. C., Meuwissen, T. H. E., & Berg, P. (2014). Allele frequency changes due to hitch-hiking in genomic selection programs. Genetics Selection Evolution, 46, 1–14.

Masel, J. (2011). Genetic drift. Current Biology, 21(20), R837–R838.

FAO. (2013). In vivo conservation of animal genetic resources. In FAO animal production and health guidelines. FAO Rome:

Olschewsky, A., & Hinrichs, D. (2021). An overview of the use of genotyping techniques for assessing genetic diversity in local farm animal breeds. Animals, 11(7), 2016.

Woolliams, J. A., & Oldenbroek, J. K. (2017). Genetic diversity issues in animal populations in the genomic era. In Genomic management of animal genetic diversity (pp. 13–47). Wageningen Academic.

3.1 Assessing major genes allele frequencies and the genetic diversity of the native Aosta cattle female population

Francesca Bernini[a], Chiara Punturiero[a], Mario Vevey[b], Veruska Blanchet[b], Raffaella Milanesi[a], Andrea Delledonne[a], Alessandro Bagnato[a] and Maria Giuseppina Strillacci[a*]

[a] *Department of Veterinary Medicine and Animal Science, Università degli Studi di Milan, Lodi, Italy;* [b] *Associazione Nazionale Bovini di Razza Valdostana, Gressan, Italy;*

Published in the Italian Journal of Animal Science (2023)
https://doi.org/10.1080/1828051X.2023.2259221

Highlights

- The autochthonous Aosta cattle breeds are free from the Mendelian inheritance disease variants found in cosmopolitan cattle.
- Allele frequencies revealed the effort in selection for variants related to cheese yield and a viable selection for variants related to meat quality.
- The use of genomic information in Aosta cattle can help maintain the biodiversity of these local breeds and their peculiarities.

Abstract

The Aosta cattle breeds have a key role in the economy of the Aosta Valley. In addition to the meat and milk production these autochthonous breeds are important for their cultural value, and for their role in the maintenance of the mountain environment. The knowledge of their genetic makeup represents a fundamental asset to managing the reproduction of the population in order to maintain the existing genetic diversity and, as a possible input, to apply genomic selection in a small population. A total of 3195 Aosta cows were genotyped with the GeneSeek Genomic Profiler ® (GGP) Bovine 100K by Neogen in the framework of the DUALBREEDING-2 project. The Aosta Black-Chestnut and Chestnut-Héren resulted to be a unique population. The Aosta female population had shorter ROH than bulls and shared ROH_islands that harbour adaptative and functional genes. Allele frequencies of major genes highlight the possibility for selection for both milk and meat quality variants and that the Aosta cattle population is free from the known Mendelian inheritance diseases found in cosmopolitan breeds (e.g. BLAD, CVM, HCD).

Keywords: Aosta cattle breeds; autochthonous cattle; genetic diversity; ROH; Mendelian traits

Introduction

The autochthonous Aosta dual-purpose cattle is raised in the northern-west alpine region of Italy, Valle d'Aosta. The Aosta Red Pied breed (ARP) most likely derives from the Northern Europe pied breeds, brought into the valley around the fifth century by the Burgundians migrations, while the Aosta Black Pied-Chestnut (ABC) with the Swiss Hérens are most likely native cattle breeds sharing a common genetic background (Vezzani 1929; Del Bo et al. 2001; Strillacci et al. 2020; Signer-Hasler et al. 2023).

According to the herd book regulations approved in 2020 (https://anaborava.it/lg_disciplinare.html), the Aosta cattle population genetic improvement program has two sections: one for the ARP and one for the ABC cattle and the ABC subjects with a Hérens ascendant, coded as Aosta Chestnut - Hérens (ACH).

Since the establishment of the National Breeders Association (Associazione Nazionale Allevatori Bovini di Razza Valdostana - ANABORAVA) in 1980 the selection criteria of Aosta cattle includes milk and meat production. In addition to production traits, farmers always paid attention to the ability of cows to adapt to the farming system based on summer grazing. Summer pasture occurs in the mountains surrounding the Aosta Valley up to 2500 m.a.s.l. that is characterised by rough climates and permanent meadows. Even if the grazing activity in such harsh environmental conditions may challenge the cows' functionality, summer pasture it is practised as it represents a great economic value for the valley preserving territories and landscapes, making them accessible to tourism and thus closely linked to the local economy and traditions.

For what concerns the productivity, in 2022 the average 305-day milk yield recorded by the Italian breeders Association (AIA) were: 3334 kg and 2459 for ABC and ACH respectively and 3781 kg for ARP (http://bollettino.aia.it/Contenuti.aspx?CD_GruppoStampe=TB&CD_Specie=C4). The milk from the Aosta cattle is mainly used to produce the Fontina cheese which is a Protected Designations of Origin (PDO), and its regulation envisages that it must be produced in the Aosta Valley, exclusively with Aosta cattle milk.

Regarding the cultural and social value of the breeds, a historical tournament, called "*Battailes des Reines*", takes place yearly in the Aosta Valley. During this folkloristic event cows challenge each other in a harmless match to establish a hierarchy, thus determining the dominant one. This traditional contest takes place in the span of time from March to October, with the final event at the "Arena Croix Noir" to proclaim the Queen of the Valley (Sartori and Mantovani 2010). Because of this tournament, the combativeness trait has been included in the official selection index of ABC and ACH since 2014, the Cheese yield, Muscularity and Combativeness Index - IRCMC (Sartori et al. 2014). Whereas the selection index for the ARP is the IRCM, based only on Cheese yield and Muscularity.

Since the introduction of SNP genotyping technology, the genomic information has been used to investigate the Aosta breeds' genetic variation (Strillacci et al. 2020; Signer-Hasler et al. 2023) as a step forward from the studies based on SSR markers (Del Bo et al. 2001).

Among the indicators of genetic variability, Runs of Homozygosity (ROH) are long portions of DNA in homozygous state that may be shared across individuals and populations (Peripolli et al. 2017), i.e., genomic regions identical by descent (IBD). ROH are investigated in populations as they may occur as a consequence of directional selection (Falchi et al. 2023) in livestock. They are generally used to investigate how the population's history, structure and demography evolved along generations (Peripolli et al. 2016).

The ROH can also be used to calculate the genomic inbreeding coefficient (F_{ROH}) which is the proportion of ROH-covered genome to the total length of the autosomal genome (Saura et al. 2015). Another commonly used inbreeding coefficient, based on the difference between the observed and expected homozygous genotypes is the F_{HOM} that can be calculated for each individual and across all markers (Keller et al. 2011).

For the conservation of animal genetic resources, in addition to a precise and accurate phenotype description, it is essential to frame the genetic variability that exists in autochthonous populations. The objectives of this study were: (i) to identify the genetic diversity among the Aosta breeds, (ii) to identify Runs of Homozygosity in Aosta cattle population; (iii) to estimate the genomic inbreeding of individuals; (iv) to determine allele frequencies at loci that influence milk and meat characteristics, mendelian diseases, and fertility haplotypes.

Materials and methods

Ethics statement

For this study the Animal Care and Welfare Committee approval was not required as genotypes were made available by ANABORAVA from their genomic database.

Sampling and genotyping

The 3195 female genotypes, obtained with the GGP Bovine 100K (GeneSeek®) by Neogen, were provided by the Associazione Nazionale Allevatori Bovini di Razza Valdostana (ANABORAVA) 929 - ABC, 158 - ACH and 2108 - ARP. The proportion of breeds in the genotypes well represents the population composition: in fact, in 2020, the Aosta breeders' association registered a total of 5,598 ABC-ACH and 13,997 ARP cows in the herd book (https://anaborava.it/lg_consistenza.html). Cows were sampled across all farms with individuals recorded in the herd book and represent as such the actual population.

The genotypes part of the ANABORAVA genomic database have been produced by the breeder association within the national project DUALBREEDING - Fase2, funded by EU EAFRD and the Italian Ministry of Agriculture. The aims of the project are to preserve cattle populations' biodiversity and to integrate the information of the DUALBREEDING project for a better genotypic and phenotypic characterisation.

The available marker dataset consisted of 89,762 SNP markers mapped on autosomes with known chromosomal positions according to the ARS-UCD1.2 bovine reference genome. This SNP dataset does not include redundant SNPs to detect the same mutation. All individuals had a call rate value > 95%.

Population Structure and Genetic Diversity Analysis

Genotypes were subjected to a quality control: SNP with a call rate lower than 95% were filtered out. To analyse the population structure and genetic variation within and among breeds the following approaches have been used:

- The principal component analysis (PCA) has been performed in SNP & Variation Suite (SVS) v8.9 (Golden Helix Inc., Bozeman, MT, USA) using only the SNP genotypes filtered for MAF >0.05; leaving a total of 84,078 SNPs for the analysis. Two analyses have been performed, one for the three breeds together and one for only the ABC and ACH breeds. The graphical representation of the results has been obtained using the R package “ggplot2” (Wickham 2011).
- The pairwise fixation index (i.e., Wright’s F-statistic - F_{ST}) was estimated using the dedicated module implemented in SVS v8.9. The F_{ST} was calculated for all possible pairs of breed combinations.
- The observed and expected heterozygosity has been obtained by PLINK v1.9 software using the “--hardy” command (Purcell et al. 2007).
- The admixture analysis has been performed with the ADMIXTURE software (Alexander et al. 2009). To correct for individual relatedness the pairwise LD pruning (using the PLINK v1.9 software) has been performed on a sliding window of 50 SNPs with a step of 5 SNPs and an r2 of 0.5. After the pruning a total of 58,836 SNPs were available for the admixture analysis. The Admixture analysis has been carried out using from 2 to 7 K and the one with the lower cross validation error (CV) has been selected (K = 2, CV = 0.58851).

Effective population size

The estimation of effective population size (Ne) was obtained using the SNePv1.1 program as described in Barbato et al. (2015). The approach is based on the known relationship between the variance in linkage disequilibrium (calculated using allele frequencies) and the effective population size. The software can estimate the historical Ne based on the relationship of the LD spread, directly calculated on the SNP genotypes, and on the recombination rate. The recombination rate between a pair of SNPs is inferred by SNePv1.1 considering the relationship between the physical and genetic distances, with a default value of 1 Mb = 1 cM.

Detection of Runs of Homozygosity

The ROH detection was carried out with the “detectRUN” package of the R software (Biscarini et al. 2018), using the “consecutiveRUNS” method with the following parameters: i) the minimum length of the ROH’s was set to 1Mb to avoid the detection of short and common ROH across the genome due to linkage disequilibrium (LD), because no LD-based pruning was performed; ii) a minimum of 30 homozygous consecutive SNPs; iii) no heterozygote nor missing genotypes were allowed; iv) a maximum gap of 1 Mb between the SNPs, to ensure that the SNP density did not affect the ROH.

The top 1% SNPs of the SNP distribution in ROH were used to identify the ROH_islands. This defines a threshold of 26% and 51% of SNPs in ROH, i.e. the proportion of individuals with uninterrupted sequences of the top 1% SNP in a ROH. The QTL annotation of the genes inside the ROH_islands, identified with the gene set (Bos taurus: Annotation Release 105) downloaded from NCBI, has been obtained using the Animal QTL database for Cattle (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index).

Inbreeding coefficients calculation

The inbreeding coefficients (F_{HOM} and F_{ROH}) were calculated for each sample as follows:

- the F_{HOM} coefficient values were calculated by the dedicated function of the SVS v8.9 software.
- the F_{ROH} inbreeding coefficient has been calculated after ROH detection as:

$$F_{ROH} = \frac{\sum_i^n L_{ROH}}{L_{AUT}}$$

where L_{ROH} is the length of the ith ROH segment, n is the number of detected ROH and L_{AUT} is the length of the autosomal genome covered by the SNPs (2,487,916,500 bp).

For the inbreeding values obtained, the minimum, maximum, mean, and standard deviation values for each breed were then obtained, to compare the results across them.

Genotypic frequencies for the traits of interest

Several SNP mapping causative loci are included in the Neogen GGP Bovine 100K SNP chip. Among them, the ones related to known cattle mendelian traits and disorders have been investigated (e. g. diseases, fertility haplotypes and milk proteins). Genotypic frequencies were obtained within breed for each locus using SVS v8.9.

Results and Discussions

Population Structure of the Aosta Breeds

Figure 1A shows the dispersion plot of the first 2 components of the PCA representing 59% of the total variability. The ARP grouped in a separate cluster, suggesting being genetically differentiated from the ABC and ACH breeds that cluster in a unique group, supporting the hypothesis of their common origin. In Figure 1B it is possible to see the graphical representation of the PCA analysis performed including only genotypes of the subjects of the two breeds ABC and ACH, that cluster overlapping one another as in Figure 1A. This result is in line with Strillacci et al. (2020) where the Aosta Black Pied (ABP) and Chestnut (CAS) were classified as 2 different breeds. To compare their results with those of this study, it is to be accounted that the ABP in their study, is here part of the ABC, while the CAS here is classified as ABC and ACH according to the ascendants, as hereinbefore described. Additionally, the comparison with the results by Signer-Hasler et al. (2023) confirmed the subdivision in two main clusters: one for the Aosta Black Pied, Chestnut, Héren (VPN, CAST and ER respectively in their study) and one for the Aosta Red Pied (VPR in their study). The sample size in this study is bigger and more evenly distributed across the Aosta cattle, reflecting the actual population-to-breed proportion. The cluster of the ARP is showing a small, separated cluster. This is particularly interesting as the sub-cluster (lower left corner of Figure 1A) is composed of daughters of only one Aosta Red Pied bull. All sires in reproduction are registered in the Herd Book and must have, as this one, registered sire and dam and must be genomically tested for their paternity and maternity. Ancestors (i.e. paternal and maternal grandsires) are in common with other sires of the analysed females. The hypothesis we may speculate is that the difference could be due to a specific variation of its genome with respect to the rest of the population.

The PCA results are supported by the F_{ST} index (Figure 1C): the difference between the ABC and ACH is 0.005, practically close to zero, meaning that the two breeds can be considered as a unique population. In fact, the value is much less than the F_{ST} obtained between ARP and the other two ones, 0.053 and 0.06 for the ABC and ACH respectively. Strillacci et al. (2020) identified similar F_{ST} values between the ARP and the other two breeds, i.e., 0.050 and 0.052 for the ABP and the CAS respectively.

The Admixture analysis identified two ancestors (K = 2). The results are shown in Figure S1, reporting the cross-validation error for K 1 to 4. For K = 2, which is the one with the lowest cross-validation error, the ABC and ACH were composed of 90% and 97% by the same population respectively, and the ARP breed results consisted of a different ancestor (94%).

Based on this evidence (low F_{ST} values and high genetic similarity), all the subsequent analyses have been developed considering the breeds grouped in ARP and ABCH, which is the union of ABC and ACH.

The Ne calculated for the ABCH and ARP is reported in Figure 2. In the most recent generation, i.e., 13th, the Ne is 301 for the ARP and 443 for the ABCH depicting the good variability present in these breeds. In the Holstein breed, Makanjuola et al. (2020) reported a Ne varying from 43 to 66 according to the input information used to calculate it, i.e. SNP genotype vs. pedigree information.

Inbreeding

The inbreeding statistics are reported in Table 1. The average F_{HOM} coefficient is close to zero in both groups showing the lowest (-0.123) and highest (0.208) values for the ARP. The expected and observed heterozygosity are similar for both breeds. Values were for the ABCH and ARP respectively 0.3761 and 0.3721 for the observed heterozygosity and 0.3763 and 0.3710 for the expected one.

F_{ROH} values calculated for each length class of ROH are very low especially when compared to values calculated with the same approach in a highly selected specialised breed, such as the Holstein (Makanjuola et al. 2020).

In their study Strillacci et al. (2020) reported very similar inbreeding coefficients to the ones obtained here: for the ARP (VPR in Strillacci et al 2020) F_{ROH} was 0.067 compared to 0.064 in this study, while the F_{HOM} was -0.003 as in this study. More recently, Signer-Hasler et al. (2023) calculated the F_{ROH} coefficient using the Illumina BovineHD bead chip for the Aosta Red Pied, Black Pied and Chestnut finding average values of 0.034, 0.029 and 0.025 respectively. The average F_{ROH} calculated in the Italian Holstein by Dadousis et al. (2022) is 0.15. It is interesting to note that, even if the ARP and the ABCH are two local breeds with a much smaller population size than the Holstein, the inbreeding coefficient is close to 1/3 with respect to the one found in the Holstein, where the intense selection is likely to be the cause of the loss of genetic variation.

The possibility of exploring at the genomic level the inbreeding in the Aosta cattle population in this study, allowed us to highlight the effectiveness of the selection programs applied for decades by farmers according to indications from the technical committee of their national breeder's association.

Runs of Homozygosity

In all the cows of the two groups, ABCH and ARP, a total of 230,948 ROH were identified (Table 2). The largest total amount of ROHs (156,711), the longest average length (2,125,789) and the highest average number (74) of ROHs were found in the ARP. The obtained results showed that the Aosta population exhibits shorter ROHs than the ones identified by Strillacci et al. (2020). This is particularly evident for the largest ROH that reaches a length up to 71,168,012 bp in the VRP (26,380,269 for ARP in this study) and 55,392,599 in the VBP (23,005,964 for ABCH in this study). Ferenčaković et al. (2013), concluded that the identification of ROHs performed with the 50K SNP array is as reliable as the one obtained with the HD SNP chip: as such different SNP chip densities should not justify differences found in this study with respect to others studies as in Strillacci et al. (2020). We may speculate that the difference in ROH found in this study with respect to Strillacci et al. (2020) is likely due to the different composition of cohorts used in the two analyses.

ROH_islands were assessed using SNP occurrences and are shown in Figure 3. A total of 6 and 4 ROH_islands were found in the ARP (threshold 51%) and in the ABCH (threshold 26%) respectively. In Table 3 the ROH_islands shared by more than 51% of cows, are reported together with the annotated genes and the associated QTL. The genes annotated in the ROH_islands identified in the ABCH breed are reported in Table 4.

In accordance with the results of the ROH statistics, this study identified smaller ROH_islands respect to the results by Strillacci et al. (2020). The only exception is the new ROH_island on BTA21 found in the ABCH breed (Table 4). This region harbours five genes that are orthologs in many species, Homo sapiens included, where it has been identified as the causative region for the Prader-Willi syndrome (Cassidy 1997). A study, conducted on eight Italian beef cattle identified a longer common ROHs on the same chromosome shared between seven of them, starting at position 83,766 and ending at different positions (Fabbri et al. 2021). Furthermore, the genes annotated in this ROH_island, MKRN3, MAGEL2, NDN and SNRPN/SNURF, have a well-known role in the epigenetic regulation of precocious puberty onset, reproductive hormones synthesis, oocytes development, and pre- or post-implantation of embryos in cattle and humans (Alves et al. 2022). Lastly, in the study of Costilla et al. (2020), NDN, SNRPN and MAGEL2 genes have been associated with the temperament of cattle. Especially for the ABCH breed, farmers select cows for their fighting aptitude and the identified ROH_island may be related to a genetic component of the combativity of the Aosta cattle population. On chromosome 5 the KITLG gene has been found in a ROH_island shared by both breeds; this gene has a QTL for the eye area pigmentation, and it has been associated with the roan phenotype in Belgian Blue and Shorthorn cattle (Seitz et al. 1999). Among the top 51% ROH_islands there are other genes that control coat colour such as the KIT gene that regulates the spotting (Fontanesi et al. 2010) and the PDGFRA gene that has been associated with white spotting (Fan et al. 2014), and also trypanotolerance in crossbred cattle (Yougbaré et al. 2021). Furthermore, the KDR gene has been associated with white pattern in Hereford cattle (Whitacre et al. 2013) and the TECRL gene associated with head pigmentation pattern (Mészáros et al. 2015), feed efficiency (de Almeida Santana et al. 2016) and fertility (Dias et al. 2017).

Among the other annotated genes in the identified ROH_islands, there are genes related to immunity, such as:

- CXCR4: that regulates the immune response (Revskij et al. 2022) and has been associated with BVDV persistently infected calves (Helal et al. 2013);
- CLDN7: that influences epithelial intestinal permeability. In intestine-specific claudin-7 knockout mice, this gene showed increased intestinal permeability and inflammation of mucosal structures (Meissner et al. 2017; Aschenbach et al. 2019);
- GPS2: that is involved in stress response and in immune system processes, particularly associated with bovine and ovine leukaemia virus response (Klener et al. 2006; Casas et al. 2020);
- TNFSF13 (also known as APRIL): that is a proliferation-inducing ligand and is a member of the BAFF system molecules that play a vital role in mature B-cell survival and in the secretion of IgA antibodies. TNFSF13 can also have an adjuvant-like effect on the immune system to enhance antigen-specific humoral immunity (Zhang et al. 2010; Mallikarjunappa et al. 2019).

Furthermore, in the identified ROH_islands, in addition to the hereinbefore listed genes related to immunity, other interesting genes related to adaptability, feed efficiency and residual feed intake have been found and are reported in Table S1. Among those, very relevant for the Aosta cattle are: the ACADVL gene, associated with the mobilisation of fatty acids during periods of low energy intake and fatty acids' beta-oxidation, resulting in a better adaptation to low feed periods (van Dorland and Bruckmaier 2013; Alaedin et al. 2021); the ACAP1 gene, involved in the cold stress response, it has been found to be an up-regulated gene in severe cold-exposed cattle (Xu et al. 2017); and the CLOCK gene that has a key role in positive modulation of the inflammatory response and regulates the circadian rhythm in mammals (Casey and Plaut 2012). All the genes found in the identified ROH_islands have been classified in Table S1 based on the available published studies.

The genes harboured in the ROH_islands of the Aosta population might be related to the ability of the Aosta breeds to adapt to the harsh farming conditions. Particularly, the presence of adaptative genes in ROH_islands (Table S1) related to fatty acids mobilisation (e.g. ACADVL), cold stress response (e.g. ACAP1), and immune response (e.g. CXCR4, CLDN7, GPS2, TNFSF13) could be related to the summer pasture practice to which these breeds are undertaken.

Genotypic and gene frequencies for milk protein variant loci

Many studies investigated milk proteins and the effects of their variants on (i) the protein concentration and composition of milk (Heck et al. 2009; Huang et al. 2012), (ii) the milk coagulation properties (Jakob and Puhan 1992; Cipolat-Gotet et al. 2018); (iii) the effect on human nutrition (Tauzin et al. 2002; Caroli et al. 2004). Milk from Aosta cattle is used mainly to produce the PDO Fontina cheese. For this reason, the Aosta breeds' selection plan pays attention to milk protein variants reporting for artificial insemination bulls their genotype for k-casein, β-casein and β-lactoglobulin.

Genotypic frequencies for the milk proteins loci, that are present on the 100K SNP chip, are reported in Table 5.

For what concerns the casein fraction of the milk proteins: the k-casein registers a higher frequency of B allele with respect to A in the ARP (i.e., B allele frequency = 0.63) while in the ABCH the A allele is the one with the largest frequency, i.e., 0.60. The A variant of the k-casein has been associated in many studies with a lower k-casein content in milk, which results in a lower cheese yield (Heck et al. 2009; Huang et al. 2012).

Comparing the Aosta population with other autochthonous breeds, Sanchez et al. (2020) found that the most frequent variant in the Abondance and Vosgienne breeds was A variant, with frequencies of 64.8% and 65.5% respectively. On the contrary the B variant was found at higher frequencies in Brown Swiss, Montbéliarde, Tarentaise and Normande cattle, 74.4%, 61.6%, 56.7% and 85.5% respectively.

The β-casein variant has been recently brought to attention for its possible impact on the milk digestibility properties (Truswell 2005; Fernández-Rico et al. 2022). Even though the relationship with digestibility and gastrointestinal sensibility of A1 milk is still unclear, some milk packaging industries are nowadays marketing A2A2 milk to consumers because of the health claims that have been stated in the latest years. Recently a study on the Holstein Friesian cattle (Chessa et al. 2020) suggested that some selection is undergoing in this breed towards an increase of A2 allele in the last decades. In Anglo-Saxon countries, where the marketing of A2A2 milk was successful, sires with A2A2 genotype were in fact preferentially selected as reproducers. The use of these bulls as reproducers may partially explain the increased frequency of A2 variant that the authors reported, i.e., from 0.49 to 0.56 in the recent population. In our study the most frequent allele of the β-casein haplotype (CSN2 gene) was the A2 allele, being 0.60 in the ARP and 0.55 in the ABCH.

In the Pinzgauer alpine population, Caroli et al. (2010) found that the A1 and A2 alleles had a frequency 0.357 and 0.572 respectively, while the B variant is the least frequent one, f(B) = 0.015.

Some studies (Brooke-Taylor et al., 2017; Sebastiani et al., 2020) affirm that the A1 variant results in an improved curd consistency and milk coagulation as well as a better micelle size but results in a lower milk digestibility because of the bioactive peptide that forms during the milk digestion. On the other hand, Heck et al. (2009) found that Holstein cows with A1A1 genotype show a reduction in protein yield respect to those with A2A2 genotype, but this variant was also associated with a higher concentration of α-S1-casein and k-casein. They also identified that the A1A2 genotype resulted, respect to the homozygous A1A1 genotype, in a smaller content of α-lactalbumin, β-lactoglobulin and α-S1-casein and a greater content of β-casein.

For the α-S1-casein locus (CSN1S1 gene) the A mutation frequency was higher in both breeds being 0.95 in the ARP and 0.80 in the ABCH. The G mutation at the α-S1-CN SNP, is a very rare mutation also in the Holstein population (Heck et al. 2009). The G mutation was also found in a low frequency in the Italian Brown population by Mancini et al. (2013). Sanchez et al. (2020) reported a detailed scheme relating missense SNP mutation and protein variants. The mutation G for the SNP on chr 6 at position 85,427,427 (ARS_UCD1.2) is coded as C variant, while the A mutation in the B one. In the study of Caroli et al. (2010) the C variant in the Pinzgauer cattle had an allelic frequency of 0.27. The Pinzgauer is an alpine cattle population and has a frequency more similar to the ABCH (i.e, 0.2) respect to the Italian Holstein i.e., 0.003 (Chessa et al. 2020) and the Italian Brown, i.e., 0.08 (Mancini et al. 2013) which is more similar to the ARP having a frequency of 0.05.

Regarding the whey proteins, representing the 20% of milk proteins (Wood et al. 2021), the 100K SNP Chip releases the information for the β-lactoglobulin locus, i.e., BLG gene. This gene is particularly important in cheese making and the most frequent allele frequency in the Aosta population refers to allele B with 0.69 and 0.73 in ABCH and ARP breeds. The β-lactoglobulin B variant has been found as the most frequent variant also in the Abondance, Brown Swiss, Tarentaise, Normande and Vosgienne cattle with frequencies of 53.5, 68.8, 68.2, 57.0 and 68.8 respectively (Sanchez et al. 2020). A study of Heck et al. (2009) affirms that selecting for the B variant of the BLG gene will result in a higher concentration of caseins, but with the same milk composition and a different whey composition; for this reason, selection for the B variant of BLG gene can improve cheese production without negatively affect the cheese making properties.

However, the complexity of the relationship among the casein variants and the milk properties for cheese production make difficult to disclose an explicit selection for an overall combination of milk protein variants to exclusively enhance renneting properties.

In the Aosta breeds at the marker for the lactoferrin - LTF gene - the B allele has a frequency of 0.73 both for the ABCH and ARP (Table 5). The polymorphism at the LTF gene has been associated with milk Somatic Cell Count (SCC); in particular, Wojdak-Maksymiec (2016) found that the heterozygote genotype AG is associated with a higher SCC in milk while the AA has the lowest value.

Genotypic and gene frequencies for meat quality loci

In the last decades, as the consumers require a good quality product, the selection of beef cattle breeds has been also oriented to meat tenderness. The availability of the Calpain, Calpastatin and Leptin genes tests made possible to identify carriers of favourable meat quality characteristics before their slaughtering.

Since Aosta breeds are double purpose cattle used to produce milk and meat, knowing the distribution of the different genotypes for the Calpain, Calpastatin and Leptin genes is important.

In Table 6 the genotypic frequencies of Calpain_316 and Calpain_4751 markers are reported. The G allele of Calpain_316 show frequencies of 0.92 and 0.93 in the ABCH and ARP breeds respectively. The most frequent genotype for Calpain_4751 is the heterozygote TC in the ABCH, while for the ARP the TT genotype outnumbers the others with a frequency of 0.546. Lisa and Di Stasio (2009) found allelic frequencies of 1 and 0.94 for the G allele at the CAPN_316 marker in the Aosta Black Pied and in the ARP respectively and for the CAPN_4751 frequencies of 0.62 and 0.61 for the T allele. The T allele frequency for the CAPN_4751 in our study is 0.56 for the ABCH and 0.74 for the ARP. The difference can be due to the effect of selection (ARP), to the analysis of Chestnut and Heren jointly, or to the smaller sample size of the study of Lisa and Di Stasio (2009).

For the Calpastatin (CAST) locus, allele frequencies varied across markers and breeds: the SNP CAST_2870 marker shows a frequency for the G allele of 0.38 (ABCH) and 0.35 (ARP); the SNP CAST_2959 marker shows a frequency for the G allele of 0.06 (ABCH) and 0.19 (ARP). Lastly, at the UoGCAST1 marker, frequencies of the G allele have a frequency of 0.65 in the ABCH and 0.53 in the ARP.

The calpain gene, firstly identified by Smith et al. (2000), works in a complex system jointly with the calpastatin one, identified by Bishop et al. (1993). Their combined effect has an important role on meat tenderness. In fact, calpains are $Ca2^+$ proteases that intervene during meat maturation processes, tenderising muscle fibres, while calpastatins are their inhibitor enzymes (Coria et al. 2018). It has been found that these proteins also influence flavour and juiciness (Casas et al. 2006) as well as colour, pH and water retention ability of meat (Reardon et al. 2010). Casas et al. (2006) found that the two homozygotes for the CAPN_4751 and 4959 markers have the most flavourful and juicy meat, with respect to the heterozygotes. Furthermore, Gill et al. (2009) reported that subjects with the CC genotype at the CAPN_316 marker show lower tenderometer values, higher hindquarter weights and tenderness scores by the taste panel, compared to the other genotypes. Many other studies identified the C allele as most favourable for meat tenderness at the markers: Calpain_316 (Page et al. 2002), Calpain_4751 (White et al. 2005) and UoGCAST1, also known as CAST_282 (Schenkel et al. 2006) (Casas et al. 2006; Gill et al. 2009; Chung et al. 2014). For the other two calpastatin gene makers (CAST_2870 and CAST_2959) identified by Cong et al. (1998) and Barendse (2002), the favourable allele is the A (Morris et al. 2006; Ribeca et al. 2009). Reardon et al. (2010) identified that at the UoGCAST1 marker, the GG genotype has a higher pH than the heterozygote and the CC. Making the Homozygote GG a good candidate for dark, firm, and dry beef (DFD) if the animals are not managed in the best way possible at the slaughterhouse.

Allele frequencies at the Leptin 2FB mutation of the leptin gene differ between the two populations: the most frequent allele is the C one found in the ARP (0.77). At the Leptin 2FB locus Schenkel et al. (2005) found that the C allele is associated with higher lean meat production with lower carcass fat. Brickell et al. (2010) found that primiparous heifers' homozygote for the C allele showed a lower incidence of calf mortality at first calving.

The quality of meat (i.e., tenderness, juiciness, flavour), with respect to the genotypes at different markers with different combinations, has been tested by several authors in meat cattle breeds, but not specifically in the Aosta population and a targeted study could then be useful for the breeds valorisation and for their selection.

Genotypic and gene frequencies for fertility haplotypes/loci and mendelian inheritance diseases loci

In the Aosta population, the haplotypes affecting cows' fertility and embryonic development found in cosmopolitan breeds (e.g., Brown Swiss Fertility Haplotypes - BHs, Holstein Fertility Haplotypes – HHs, Jerseys Fertility Haplotypes – JHs) are absent.

In Table 7 the genotypic frequencies for haplotypes and loci affecting fertility are reported. The variability identified in the Montbèliarde Haplotype 1 (MH1) could be the result of some ancestral cross between the Aosta breeds and the Montbèliarde.

The COQ9-rs109301586 marker has been associated with improved fertility and higher oocyte mitochondrial content, with positive characteristics found for the A allele; this allele can improve oocyte competence needed for supporting the following embryonic development (Ortega et al., 2017). Sugimoto et al. (2013) found that cows with the Del/Del genotype at the PKP2-988 marker had a higher conception rate than the TA/TA, for this reason the DD genotype is associated with more fertile cows than the II, in the Holstein population.

The Aosta breeds have also been tested for all the mendelian inheritance diseases markers on the 100K SNP Chip. In Table S2 only the markers with variability are reported and it is possible to see that the Aosta cattle are free from almost all the mendelian inherited diseases except for the presence of few heterozygote individuals.

For the Mulefoot_241 disease the recessive homozygous genotypes have been identified with a frequency of 0.004 and 0.001 for the ABCH and ARP, respectively.

The Infectious Bovine Keratoconjunctivitis (IBK) is a disease that has been found associated with the A/G mutation at the IBK marker on BTA8. The G mutation is the resistant allele with a substitution effect of 8.2% on reduction of pinkeye infection Kataria et al. (2011). In the Aosta population the G mutation has a frequency of 0.55 in the ABCH and 0.42 in the ARP.

Conclusions

The cost reduction in SNP genotyping favoured the genomic data production also in small local populations and in females. The Aosta breeds are part of the DUALBREEDING PSRN project envisaging a wide production of genotypes on the female population. The knowledge of the genomic characteristics of autochthonous populations is an important asset of information allowing a better understanding of their genetic background.

The autochthonous populations represent a unique source of biodiversity (Rognoni et al. 1983) and the genomic technologies made possible to accurately characterise them at the genomic level and to identify peculiarities linked to their extensive farming system and adaptation to the harsh mountain environment.

In this study the Aosta cattle female population has been analysed after the new Aosta breeds classification occurred in 2020. With this study and the one from Strillacci et al. (2020), it is possible to state that some ROH_islands are shared by bulls and cows. The genes annotated in the ROH_islands are related to efficiency, adaptability and resilience traits, including immune system. Indeed, we can suppose that the summer pasture practice favour individuals that are efficient in dealing with infections and have a prompt and fast response to stressors such as the harsh climatic conditions and the lower quantity of feed.

It is important to highlight that the Aosta breeds are free from the mendelian inheritance disorders identified in cosmopolitan breeds. The genotypic frequencies at the analysed loci well explain and highlight the selection that has occurred in these breeds, to improve milk and cheese production. Moreover, the ones related to genes influencing meat quality showed that there is possibility to improve the selection for this trait. Of particular interest are the calpain and calpastatin variants where the genotypic frequencies reflect the actual meat characteristics of the Aosta cattle. Since the favourable gene combinations are at low frequency leaving large room for improvement, it is possible to apply a gene assisted selection to improve the meat tenderness, a characteristic desired by consumers. Even though, good practices at the slaughterhouse and the use of a traditional long maturation periods can greatly improve the quality of meat; a gene assisted selection approach can be more extensively proposed and valorised since consumers are willing to pay more for good quality meat.

References

Alaedin M, Ghaffari MH, Sadri H, Meyer J, Dänicke S, Frahm J, Huber K, Grindler S, Kersten S, Rehage J. 2021. Effects of dietary l-carnitine supplementation on the response to an inflammatory challenge in mid-lactating dairy cows: Hepatic mRNA abundance of genes involved in fatty acid metabolism. J Dairy Sci. 104(10):11193–11209.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res [Internet]. 19(9):1655–1664. https://doi.org/10.1101/gr.094052.109

de Almeida Santana MH, Junior GAO, Cesar ASM, Freua MC, da Costa Gomes R, da Luz e Silva S, Leme PR, Fukumasu H, Carvalho ME, Ventura RV. 2016. Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. J Appl Genet. 57(4):495–504.

Alves AAC, da Costa RM, Fonseca LFS, Carvalheiro R, Ventura RV, de Magalhães Rosa GJ, Albuquerque LG. 2022. A Random Forest-Based Genome-Wide Scan Reveals Fertility-Related Candidate Genes and Potential Inter-Chromosomal Epistatic Regions Associated With Age at First Calving in Nellore Cattle. Front Genet. 13.

Aschenbach JR, Zebeli Q, Patra AK, Greco G, Amasheh S, Penner GB. 2019. Symposium review: The importance of the ruminal epithelial barrier for a healthy and productive cow. J Dairy Sci. 102(2):1866–1882.

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.

Barendse WJ. 2002. DNA markers for meat tenderness. Patent WO02064820.:1–88.

Biscarini F, Cozzi P, Gaspa G, Marras G. 2018. detectRUNS: an R package to detect runs of detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid homozygosity and heterozygosity in diploid genomes genomes. IBBA-CNR, PTP, Università degli Studi di Sassari, University of Guelph. 01(01):1–13. https://cran.r-project.org/package=detectRUNS

Bishop MD, Koohmaraie M, Killefer J, Kappes S. 1993. Rapid communication: restriction fragment length polymorphisms in the bovine calpastatin gene. J Anim Sci. 71(8):2277.

Del Bo L, Polli M, Longeri M, Ceriotti G, Looft C, Barre-Dirie A, Dolf G, Zanotti M, Barre-Dirie A, Dolf G, Zanotti M. 2001. Genetic diversity among some cattle breeds in the Alpine area. 118(5):317–325. https://doi.org/10.1046/j.1439-0388.2001.00306.x

Brickell JS, Pollott GE, Clempson AM, Otter N, Wathes DC. 2010. Polymorphisms in the bovine leptin gene associated with perinatal mortality in Holstein-Friesian heifers. J Dairy Sci. 93(1):340–347.

Brooke-Taylor S, Dwyer K, Woodford K, Kost N. 2017. Systematic review of the gastrointestinal effects of A1 compared with A2 β-casein. Advances in nutrition. 8(5):739–748.

Caroli A, Chessa S, Bolla P, Budelli E, Gandini GC. 2004. Genetic structure of milk protein polymorphisms and effects on milk production traits in a local dairy cattle. J Anim Breed Genet. 121:119–127. www.razzareggiana.it/

Caroli A, Rizzi R, Lühken G, Erhardt G. 2010. Milk protein genetic variation and casein haplotype structure in the Original Pinzgauer cattle. J Dairy Sci. 93(3):1260–1265.

Casas E, Ma H, Lippolis JD. 2020. Expression of viral microRNAs in serum and white blood cells of cows exposed to bovine leukemia virus. Front Vet Sci. 7:536390.

Casas E, White SN, Wheeler TL, Shackelford SD, Koohmaraie M, Riley DG, Chase Jr CC, Johnson DD, Smith TPL. 2006. Effects of calpastatin and μ-calpain markers in beef cattle on tenderness traits. J Anim Sci. 84(3):520–525.

Casey TM, Plaut K. 2012. Lactation Biology Symposium: Circadian clocks as mediators of the homeorhetic response to lactation. J Anim Sci. 90(3):744–754.

Cassidy SB. 1997. Prader-Willi syndrome. J Med Genet. 34(11):917–923.

Chessa S, Gattolin S, Cremonesi P, Soglia D, Finocchiaro R, Van Kaam J-T, Marusi M, Civati G. 2020a. The effect of selection on casein genetic polymorphisms and haplotypes in Italian Holstein cattle. Ital J Anim Sci. 19(1):833–839.

Chessa S, Gattolin S, Cremonesi P, Soglia D, Finocchiaro R, Van Kaam J-T, Marusi M, Civati G. 2020b. The effect of selection on casein genetic polymorphisms and haplotypes in Italian Holstein cattle. Ital J Anim Sci. 19(1):833–839.

Chung H, Shin S, Chung E. 2014. Effects of genetic variants for the bovine calpain gene on meat tenderness. Mol Biol Rep. 41(5):2963–2970.

Cipolat-Gotet C, Cecchinato A, Malacarne M, Bittante G, Summer A. 2018. Variations in milk protein fractions affect the efficiency of the cheese-making process. J Dairy Sci. 101(10):8788–8804.

Cong M, Thompson VF, Goll DE, Antin PB. 1998. The bovine calpastatin gene promoter and a new N-terminal region of the protein are targets for cAMP-dependent protein kinase activity. Journal of Biological Chemistry. 273(1):660–666.

Coria MS, Carranza PG, Palma GA. 2018. Calpain System in meat tenderization: A molecular approach. Rev MVZ Cordoba. 23(1):6523–6536.

Costilla R, Kemper KE, Byrne EM, Porto-Neto LR, Carvalheiro R, Purfield DC, Doyle JL, Berry DP, Moore SS, Wray NR. 2020. Genetic control of temperament traits across species: association of autism spectrum disorder risk genes with cattle temperament. Genetics Selection Evolution. 52(1):1–14.

Dadousis C, Ablondi M, Cipolat-Gotet C, van Kaam J-T, Marusi M, Cassandro M, Sabbioni A, Summer A. 2022. Genomic inbreeding coefficients using imputed genotypes: Assessing different estimators in Holstein-Friesian dairy cows. J Dairy Sci. 105(7):5926–5945. https://doi.org/10.3168/jds.2021-21125

Dias MM, Cánovas A, Mantilla-Rojas C, Riley DG, Luna-Nevarez P, Coleman SJ, Speidel SE, Enns RM, Islas-Trejo A, Medrano JF. 2017. SNP detection using RNA-sequences of candidate genes associated with puberty in cattle. Genet Mol Res. 16(1):gmr16019522.

van Dorland HA, Bruckmaier RM. 2013. Understanding diversity of hepatic metabolism and related adaptations in the early lactating dairy cow. Schweiz Arch Tierheilkd. 155(8):443–451.

Fabbri MC, Dadousis C, Tiezzi F, Maltecca C, Lozada-Soto E, Biffani S, Bozzi R. 2021. Genetic diversity and population history of eight Italian beef cattle breeds using measures of autozygosity. PLoS One. 16(10):e0248087.

Falchi L, Cesarani A, Mastrangelo S, Senczuk G, Portolano B, Pilla F, Macciotta NPP. 2023. Analysis of runs of homozygosity of cattle living in different climate zones. J Anim Sci. https://doi.org/10.1093/jas/skad061

Fan Y, Wang P, Fu W, Dong T, Qi C, Liu L, Guo G, Li C, Cui X, Zhang S. 2014. Genome-wide association study for pigmentation traits in Chinese Holstein population. Anim Genet. 45(5):740–744.

Ferenčaković M, Sölkner J, Curik I. 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genetics Selection Evolution. 45:1–9.

Fernández-Rico S, Mondragón A del C, López-Santamarina A, Cardelle-Cobas A, Regal P, Lamas A, Ibarra IS, Cepeda A, Miranda JM. 2022. A2 Milk: New Perspectives for Food Technology and Human Health. Foods. 11(16). https://doi.org/10.3390/foods11162387

Fontanesi L, Tazzoli M, Russo V, Beever J. 2010. Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus. Anim Genet. 41(3):295–303.

Gill JL, Bishop SC, McCorquodale C, Williams JL, Wiener P. 2009. Association of selected SNP with carcass and taste panel assessed meat quality traits in a commercial population of Aberdeen Angus-sired beef cattle. Genetics Selection Evolution. 41(1):1–12.

Heck JML, Schennink A, van Valenberg HJF, Bovenhuis H, Visker M, van Arendonk JAM, van Hooijdonk ACM. 2009. Effects of milk protein variants on the protein composition of bovine milk. J Dairy Sci. 92(3):1192–1202.

Helal MAY, Hoshino Y, Takagi S, Tajima M. 2013. CXC chemokine receptor type 4 and cytokine expressions in cows of a dairy herd with high prevalence of calves persistently infected with bovine viral diarrhea virus. Japanese Journal of Veterinary Research. 61(3):109–116.

Huang W, Peñagaricano F, Ahmad KR, Lucey JA, Weigel KA, Khatib H. 2012. Association between milk protein gene variants and protein composition traits in dairy cattle. J Dairy Sci. 95(1):440–449.

Jakob E, Puhan Z. 1992. Technological properties of milk as influenced by genetic polymorphism of milk proteins - A review. Int Dairy J. 2(3):157–178.

Kataria RS, Tait RG, Kumar D, Ortega MA, Rodiguez J, Reecy JM. 2011. Association of toll-like receptor four single nucleotide polymorphisms with incidence of infectious bovine keratoconjunctivitis (IBK) in cattle. Immunogenetics. 63(2):115–119.

Keller MC, Visscher PM, Goddard ME. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. https://doi.org/10.1534/genetics.111.130922

Klener P, Szynal M, Cleuter Y, Merimi M, Duvillier H, Lallemand F, Bagnis C, Griebel P, Sotiriou C, Burny A. 2006. Insights into gene expression changes impacting B-cell transformation: cross-species microarray analysis of bovine leukemia virus tax-responsive genes in ovine B cells. J Virol. 80(4):1922–1938.

Lisa C, di Stasio L. 2009. Variability of μ-Calpain and Calpastatin genes in cattle. Ital J Anim Sci. 8(sup2):99–101.

Makanjuola BO, Miglior F, Abdalla EA, Maltecca C, Schenkel FS, Baes CF. 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(6):5183–5199. https://doi.org/10.3168/jds.2019-18013

Mallikarjunappa S, Adnane M, Cormican P, Karrow NA, Meade KG. 2019. Characterization of the bovine salivary gland transcriptome associated with Mycobacterium avium subsp. paratuberculosis experimental challenge. BMC Genomics. 20(1):1–13.

Mancini G., Nicolazzi EL, Valentini A, Chillemi G, Ajmone Marsan P, Santus E, Pariset L. 2013. Association between single nucleotide polymorphisms (SNPs) and milk production traits in Italian Brown cattle. Livest Sci. 157(1):93–99. https://doi.org/10.1016/j.livsci.2013.07.008

Mancini G, Nicolazzi EL, Valentini A, Chillemi G, Ajmone Marsan P, Santus E, Pariset L. 2013. Association between single nucleotide polymorphisms (SNPs) and milk production traits in Italian Brown cattle. Livest Sci. 157(1):93–99. https://doi.org/10.1016/j.livsci.2013.07.008

Meissner S, Hagen F, Deiner C, Günzel D, Greco G, Shen Z, Aschenbach JR. 2017. Key role of short-chain fatty acids in epithelial barrier failure during ruminal acidosis. J Dairy Sci. 100(8):6662–6675.

Mészáros G, Petautschnig E, Schwarzenbacher H, Sölkner J. 2015. Genomic regions influencing coat color saturation and facial markings in Fleckvieh cattle. Anim Genet. 46(1):65–68.

Morris CA, Cullen NG, Hickey SM, Dobbie PM, Veenvliet BA, Manley TR, Pitchford WS, Kruk ZA, Bottema CDK, Wilson T. 2006. Genotypic effects of calpain 1 and calpastatin on the tenderness of cooked M. longissimus dorsi steaks from Jersey× Limousin, Angus and Hereford-cross cattle. Anim Genet. 37(4):411–414.

Ortega MS, Wohlgemuth S, Tribulo P, Siqueira LGB, Null DJ, Cole JB, da Silva M v, Hansen PJ. 2017. A single nucleotide polymorphism in COQ9 affects mitochondrial and ovarian function and fertility in Holstein cows. Biol Reprod. 96(3):652–663.

Page BT, Casas E, Heaton MP, Cullen NG, Hyndman DL, Morris CA, Crawford AM, Wheeler TL, Koohmaraie M, Keele JW. 2002. Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. J Anim Sci. 80(12):3077–3085.

Peripolli E, Munari DP, Silva M, Lima ALF, Irgang R, Baldi F. 2017. Runs of homozygosity: current knowledge and applications in livestock. Anim Genet. 48(3):255–271.

Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. 2016. Runs of homozygosity: current knowledge and applications in livestock. https://doi.org/10.1111/age.12526

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559–575. https://doi.org/10.1086/519795

Reardon W, Mullen AM, Sweeney T, Hamill RM. 2010a. Association of polymorphisms in candidate genes with colour, water-holding capacity, and composition traits in bovine M. longissimus and M. semimembranosus. Meat Sci. 86(2):270–275.

Reardon W, Mullen AM, Sweeney T, Hamill RM. 2010b. Association of polymorphisms in candidate genes with colour, water-holding capacity, and composition traits in bovine M. longissimus and M. semimembranosus. Meat Sci. 86(2):270–275.

Revskij D, Haubold S, Plinski C, Viergutz T, Tuchscherer A, Kröger-Koch C, Albrecht E, Günther J, Tröscher A, Hammon HM. 2022. Cellular detection of the chemokine receptor CXCR4 in bovine mammary glands and its distribution and regulation on bovine leukocytes. J Dairy Sci. 105(1):866–876.

Ribeca C, Bittante G, Albera A, Bonfatti V, Maretto F, Gallo L. 2009. Investigation on variability of candidate genes for meat quality traits in Piemontese cattle. Ital J Anim Sci. 8(SUPPL. 2):132–134. https://doi.org/10.4081/ijas.2009.s2.132

Rognoni G, Pagnacco G, Boselli E. 1983. Atlante etnografico delle popolazioni bovine allevate in Italia. Rome: Consiglio Nazionale delle Ricerche.

Sanchez MP, Fritz S, Patry C, Delacroix-Buchet A, Boichard D. 2020. Frequencies of milk protein variants and haplotypes estimated from genotypes of more than 1 million bulls and cows of 12 French cattle breeds. J Dairy Sci. 103(10):9124–9141.

Sartori C, Mantovani R. 2010. Genetics of fighting ability in cattle using data from the traditional battle contest of the Valdostana breed1. J Anim Sci. 88(10):3206–3213. https://doi.org/10.2527/jas.2010-2899

Sartori C, Vevey M, Mantovani R. 2014. Triplice attitudine in valdostana pezzata nera-castana: Introduzione della combattività nell'indice di selezione. Quaderno SOZOOALP. 8:133–142.

Saura M, Fernández A, Varona L, Fernández AI, de Cara MÁR, Barragán C, Villanueva B. 2015. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genetics Selection Evolution. 47(1):1. https://doi.org/10.1186/s12711-014-0081-5

Schenkel FS, Miller SP, Jiang Z, Mandell IB, Ye X, Li H, Wilton JW. 2006. Association of a single nucleotide polymorphism in the calpastatin gene with carcass and meat quality traits of beef cattle. J Anim Sci. 84(2):291–299.

Schenkel FS, Miller SP, Ye X, Moore SS, Nkrumah JD, Li C, Yu J, Mandell IB, Wilton JW, Williams JL. 2005. Association of single nucleotide polymorphisms in the leptin gene with carcass and meat quality traits of beef cattle. J Anim Sci. 83(9):2009–2020.

Sebastiani C, Arcangeli C, Ciullo M, Torricelli M, Cinti G, Fisichella S, Biagetti M. 2020. Frequencies evaluation of β-casein gene polymorphisms in dairy cows reared in Central Italy. Animals. 10(2):252.

Seitz JJ, Schmutz SM, Thue TD, Buchanan FC. 1999. A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. Mammalian Genome. 10(7):710–712.

Signer-Hasler H, Casanova L, Barenco A, Maitre B, Bagnato A, Vevey M, Berger B, Simčič M, Boichon D, Capitan A, et al. 2023. Genomic regions underlying positive selection in local, Alpine cattle breeds. Anim Genet. https://doi.org/10.1111/age.13295

Smith TPL, Casas E, Rexroad Iii CE, Kappes SM, Keele JW. 2000. Bovine CAPN1 maps to a region of BTA29 containing a quantitative trait locus for meat tenderness. J Anim Sci. 78(10):2589–2594.

Strillacci MG, Vevey M, Blanchet V, Mantovani R, Sartori C, Bagnato A. 2020a. The genomic variation in the aosta cattle breeds raised in an extensive alpine farming system. Animals. 10(12):1–18. https://doi.org/10.3390/ani10122385

Strillacci MG, Vevey M, Blanchet V, Mantovani R, Sartori C, Bagnato A. 2020b. The Genomic Variation in the Aosta Cattle Breeds Raised in an Extensive Alpine Farming System. Animals. 10(12):2385.

Sugimoto M, Sasaki S, Gotoh Y, Nakamura Y, Aoyagi Y, Kawahara T, Sugimoto Y. 2013. Genetic variants related to gap junctions and hormone secretion influence conception rates in cows. Proceedings of the National Academy of Sciences. 110(48):19495–19500.

Tauzin J, Miclo L, Gaillard J-L. 2002. Angiotensin-I-converting enzyme inhibitory peptides from tryptic hydrolysate of bovine αS2-casein. FEBS Lett. 531(2):369–374.

Truswell AS. 2005. The A2 milk case: A critical review. Eur J Clin Nutr. 59(5):623–631. https://doi.org/10.1038/sj.ejcn.1602104

Vezzani V. 1929. Razza d'Hérens o razza valdostana? Rivista di Zootecnica. 5:177–183.

Whitacre L, Decker J, Kim JW, Schnabel R, Taylor J. 2013. Kinase insert domain receptor-a candidate for the Hereford “spotted” gene. Plant and Animal Genome XXI, San Diego, CA.:P0543.

White SN, Casas E, Wheeler TL, Shackelford SD, Koohmaraie M, Riley DG, Chase Jr CC, Johnson DD, Keele JW, Smith TPL. 2005. A new single nucleotide polymorphism in CAPN1 extends the current tenderness marker test to include cattle of Bos indicus, Bos taurus, and crossbred descent. J Anim Sci. 83(9):2001–2008.

Wickham H. 2011. ggplot2. Wiley Interdiscip Rev Comput Stat. 3(2):180–185.

Wojdak-Maksymiec K. 2016. Associations between bovine lactoferrin gen polymorphism and somatic cell count in milk. Vet Med - Czech, 51(1):14-20 DOI: 10.17221/5512-VETMED

Wood EL, Christian DG, Arafat M, McColl LK, Prosser CG, Carpenter EA, Levine AS, Klockars A, Olszewski PK. 2021. Adjustment of whey: Casein ratio from 20: 80 to 60: 40 in milk formulation affects food intake and brainstem and hypothalamic neuronal activation and gene expression in laboratory mice. Foods. 10(3):658.

Xu Q, Wang YC, Liu R, Brito LF, Kang L, Yu Y, Wang DS, Wu HJ, Liu A. 2017. Differential gene expression in the peripheral blood of Chinese Sanhe cattle exposed to severe cold stress. Genet Mol Res. 16(2).

Yougbaré B, Ouédraogo D, Tapsoba ASR, Soudré A, Zoma BL, Orozco-terWengel P, Moumouni S, Ouédraogo-Koné S, Wurzinger M, Tamboura HH. 2021. Local ancestry to identify selection in response to trypanosome infection in Baoulé x Zebu crossbred cattle in Burkina Faso. Front Genet.:1606.

Zhang J-X, Ma H-W, Sang M, Hu Y-S, Liang Z-N, Ai H-X, Zhang J, Cui X-W, Zhang S-Q. 2010. Molecular structure, expression, cell and tissue distribution, immune evolution and cell proliferation of the gene encoding bovine (Bos taurus) TNFSF13 (APRIL). Dev Comp Immunol. 34(11):1199–1208.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Project name: “Le razze bovine a duplice attitudine: un modello alternativo di zootecnia eco-sostenibile – DUAL BREEDING FASE 2” – PSRN Programma di Sviluppo Rurale Nazionale 2014/2020 – Sottomisura 10.2 – PERIODO 2020-2023 – funded by the European Agricultural Fund for Rural Development.

Author contributions

FB, MGS and AB: Conceived and designed the study; FB: data analysis; AD, RM and CP: participated to data analysis; MV and VB: data curation; MV and AB project administration; MV: funding acquisition; FB, MGS and AB: wrote the original draft. All authors reviewed and approved the final version of the manuscript.

Data availability statement

The data supporting the conclusions of this manuscript are included in the Supplementary Materials.

Acknowledgments

The Associazione Nazionale Allevatori Razza Valdostana is kindly acknowledged for providing the genotypes and the data used for the analyses.

Tables

Table 1. Minimum (MIN), maximum (MAX), and average (AVG) values for inbreeding coefficients calculated for the Aosta population.

Inbreeding coefficients	**ABCH**			**ARP**		
	MIN	**MAX**	**AVG (sd)**	**MIN**	**MAX**	**AVG (sd)**
$\mathbf{F_{HOM}}$	-0.078	0.169	0.001 (0.02)	-0.123	0.208	-0.003 (0.03)
$\mathbf{F_{ROH}}^*$	0.007	0.220	0.056 (0.02)	0.005	0.250	0.064 (0.02)
$\mathbf{F_{ROH}}$ **class ≤ 2 Mbp**	0.007	0.220	0,056 (0.02)	0.005	0.250	0,064 (0,02)
$\mathbf{F_{ROH}}$ **class 2-4 Mbp**	0.001	0.172	0,030 (0.02)	0.001	0.209	0,037 (0,02)
$\mathbf{F_{ROH}}$ **class 4-8 Mbp**	0.002	0.109	0,012 (0.01)	0.002	0.160	0,016 (0.01)
$\mathbf{F_{ROH}}$ **class 8-16 Mbp**	0.003	0.049	0,006 (0.01)	0.003	0.085	0,007 (0.01)
$\mathbf{F_{ROH}}$ **class ≥ 16 Mbp**	0.006	0.009	0,007 (0)	0.006	0.016	0,008 (0)

$^*F_{ROH}$ calculated considering five Runs of Homozygosity (ROH) length (in Mbp) classes.

Table 2. Minimum (MIN), maximum (MAX), and average number (AVG) and length (in bps) of Runs of Homozygosity (ROH) identified in the Aosta population.

	ROH statistics						
	N				**Length**		
Breed	**TOTAL**	**MIN**	**MAX**	**AVG (sd)**	**MIN**	**MAX**	**AVG**
ABCH	74,237	11	179	68 (16)	1,000,004	23,005,964	2,023,899
ARP	156,711	9	183	74 (17)	1,000,000	26,380,269	2,125,789

*Average calculated per breed on total ROH length covered for each sample's genome.

Table 3. ROH_islands identified in the ARP breed, with genes mapped in the ROH and associated QTL.

Breed	Chr	from	to	Genes	QTL*
ARP	2	61057756	61281214	CXCR4	
ARP	2	61300454	61404353	DARS	
ARP	5	18259786	18430163	KITLG	KITLG: Eye area pigmentation (QTL:21151)
ARP	6	69700321	69804132	PDGFRA	
ARP	6	69826263	71305377	KIT, KDR, SRD5A3, TMEM165, CLOCK, PDCL2, NMU, EXOC1L, EXOC1	KIT: Eye area pigmentation (QTL:21154; QTL:21160), Somatic cell count (QTL:31634; QTL:31635); KDR: Bovine respiratory disease susceptibility (QTL:137210);
ARP	23	512811	1270657	KHDRBS2	KHDRBS2: Calving ease (maternal) (QTL:51867), Calving ease (QTL:51877), Dairy form (QTL:51868), Daughter pregnancy rate (QTL:51869), Foot angle (QTL:51870), Length of productive life (QTL:51874), Milk fat percentage (QTL:51871), Milk fat yield (QTL:51872), Milk protein percentage (QTL:51875), Net merit (QTL:51873), Rear leg placement - side view (QTL:51876), Somatic cell score (QTL:51878), Stillbirth (QTL:51879), Strength (QTL:51880), Teat length (QTL:51881)

*QTL (Quantitative Trait Loci) annotated in the Animal QTL Database for Cattle (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index)

Table 4. ROH_islands identified in the ABCH breed, with genes mapped in the ROH and associated QTL.

Breed	Chr	from	to	Genes	QTL*
ABCH	5	17,120,900	18,390,837	C5H12orf50, C5H12orf29, CEP290, TMTC3, KITLG	KITLG: Eye area pigmentation (QTL:21151)
ABCH	6	79,502,698	80,269,768	TECRL	
ABCH	19	26,897,393	27,496,286	ASGR2, ASGR1, DLG4, ACADVL, DVL2, PHF23, GABARAP, CTDNEP1, ELP5, CLDN7, SLC2A4, YBX2, EIF5A, GPS2, NEURL4, ACAP1, KCTD11, TMEM95, TNK1, PLSCR3, TMEM256, NLGN2, SPEM1, SPEM2, TMEM102, FGF11, CHRNB1, ZBTB4, SLC35G6, POLR2A, TNFSF12, TNFSF13, SENP3, EIF4A1, CD68, MPDU1, SOX15, FXR2, SAT2, SHBG, ATP1B2, TP53, WRAP53, EFNB3, DNAH2	ACADVL: Bovine respiratory disease susceptibility (QTL:160277); ATP1B2: 305-day milk yield (QTL:14129; QTL:14128), Average daily milk yield (QTL:14134; QTL:14132; QTL:14133), Body temperature (QTL:14130), Milk fat percentage (QTL:14126), Milk protein percentage (QTL:14127), Red blood cell Na+/K+ ATPase activity (QTL:14135; QTL:14136), Red blood cell potassium level (QTL:14131)
ABCH	21	497,917	2,006,266	MKRN3, MAGEL2, NDN, SNRPN, SNURF	

*QTL (Quantitative Trait Loci) annotated in the Animal QTL Database for Cattle (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index)

Table 5. Genotypic frequencies for milk proteins and the lactoferrin loci.

Milk traits				
Trait	**Chr**	**Genotype**	**Frequency**	
			ABCH	**ARP**
α-S1-casein	6	AA	0.644	0.903
		AG	0.314	0.094
		GG	0.042	0.002
β-casein #	6	A2A2	0,302	0,363
		A2B	0,270	0,023
		A2C	0,000	0,004
		A2E	0,000	0,000
		A2I	0,115	0,131
		A1A1	0,010	0,070
		A1A2	0,109	0,309
		A1B	0,061	0,013
		A1I	0,021	0,065
		BB	0,055	0,000
		BC	0,000	0,000
		BI	0,046	0,007
		II	0,011	0,014
β-lactoglobulin	11	AA	0.100	0.074
		AB	0.411	0.388
		BB	0.489	0.538
K-casein	6	AA	0.352	0.133
		AB	0.490	0.475
		BB	0.158	0.392
Lactoferrin*	22	GG	0.066	0.074
		AG	0.409	0.398
		AA	0.525	0.528

#The β-casein gene variants have been identified considering the genotypes at the following SNPrs, as suggested by (Chessa et al. 2020): rs433954503, rs43703013, rs43703012, rs109299401, rs43703011, rs721259074 and rs3423226649. *SNP markers on the GGP Bovine 100K.

Table 6. Genotypic frequencies at loci related to meat quality traits.

Meat traits				
Trait	**Chr**	**Genotype**	**Frequency**	
			ABCH	**ARP**
Calpain_316	29	GG	0.853	0.871
		GC	0.137	0.121
		CC	0.010	0.007
Calpain_4751	29	TT	0.307	0.546
		TC	0.502	0.385
		CC	0.190	0.069
CAST_2870*	7	GG	0.143	0.110
		AG	0.478	0.472
		AA	0.379	0.418
CAST_2959*	7	GG	0.005	0.037
		AG	0.105	0.296
		AA	0.890	0.667
Leptin 2FB	4	CC	0.340	0.588
		CT	0.489	0.361
		TT	0.171	0.051
UoGCAST1 (CAST_282)	7	GG	0.427	0.271
		GC	0.443	0.510
		CC	0.130	0.219

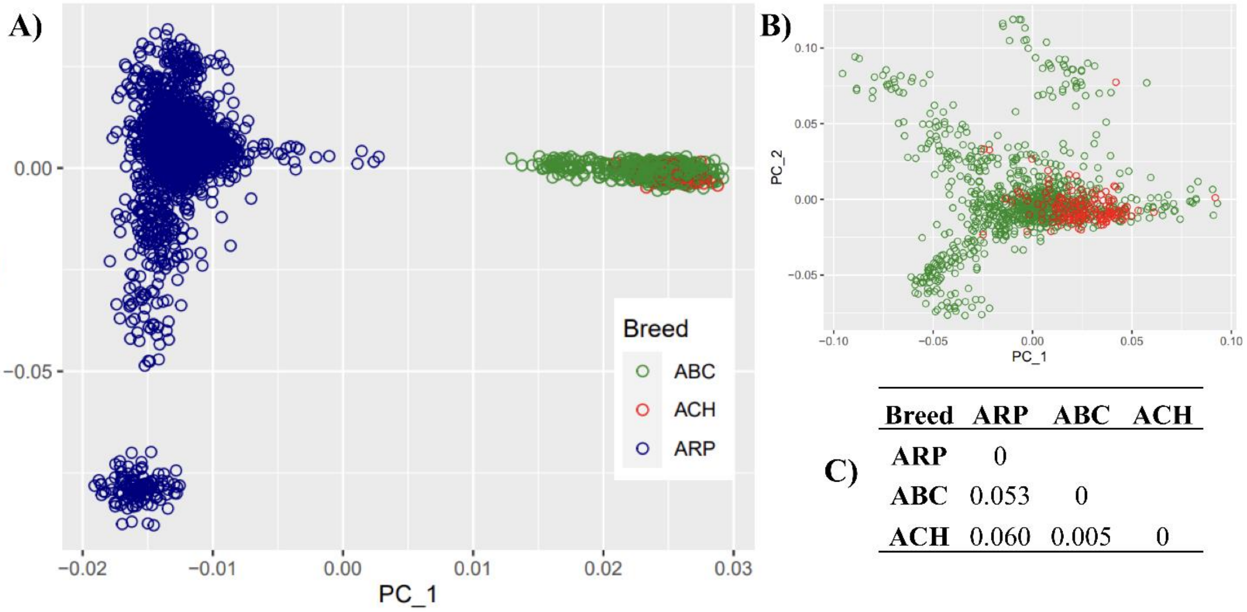
*SNP on the GGP Bovine 100K at the Calpastatin (CAST) gene.

Table 7. Genotypic frequencies for fertility traits.

Fertility				
Trait	**Chr**	**Genotype**	**Frequency**	
			ABCH	**ARP**
COQ9 rs109301586*	18	AA	0.343	0.160
		AG	0.508	0.470
		GG	0.149	0.370
MH1	19	CC	0.958	0.997
		CT	0.042	0.003
		TT	0.000	0.000
PKP2-988*	5	DD	0.065	0.161
		DI	0.357	0.469
		II	0.578	0.370
STAT3_25402*	19	GG	0.034	0.018
		GT	0.283	0.258
		TT	0.683	0.724
STAT3_19069*	19	CC	0.027	0.019
		CT	0.247	0.254
		TT	0.726	0.727
STAT5_13244* (STAT5_13516*)	19	CC(GG)	0.059	0.020
		CT(GT)	0.339	0.262
		TT(TT)	0.602	0.718
STAT5_13319*	19	AA	0.001	0.000
		AG	0.080	0.004
		GG	0.919	0.996

*SNP markers on the GGP Bovine 100K. Coenzyme Q9 (COQ9) (rs109301586); Montbéliarde Haplotype 1 (MH1); plakophilin 2 (PKP2_988). I → Insertion; D → Deletion; Signal transducer and activator of transcription (STAT).

Figures



C)

Breed	ARP	ABC	ACH
ARP	0		
ABC	0.053	0	
ACH	0.060	0.005	0

Figure 1. A) Graphical representation of Principal Component Analysis (PCA) result for the three breeds (ARP, ABC and ACH): Principal component 1 (PC_1) vs Principal Component 2 (PC_2) are plotted; **B)** Graphical representation of Principal Component Analysis (PCA) result limited to ABC and ACH: Principal component 1 (PC_1) vs Principal Component 2 (PC_2) are plotted; **C)** Aosta breeds Wright's fixation index - F_{ST}.

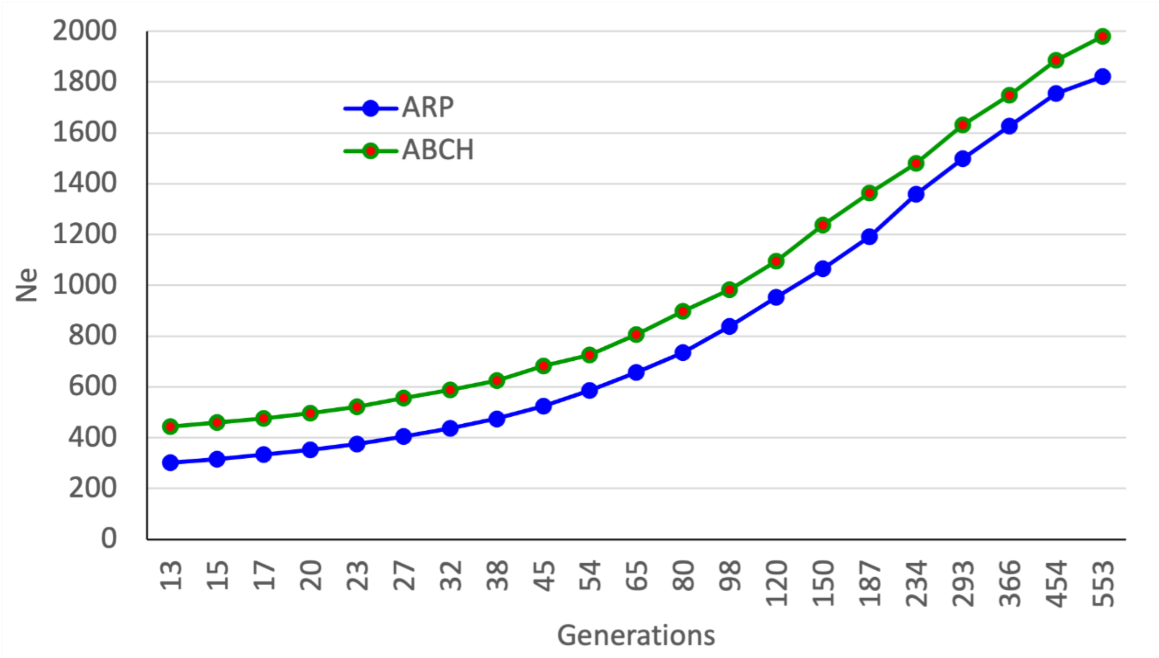


Figure 2. Effective population size (Ne) per generation ago for the ARP and ABCH breeds.

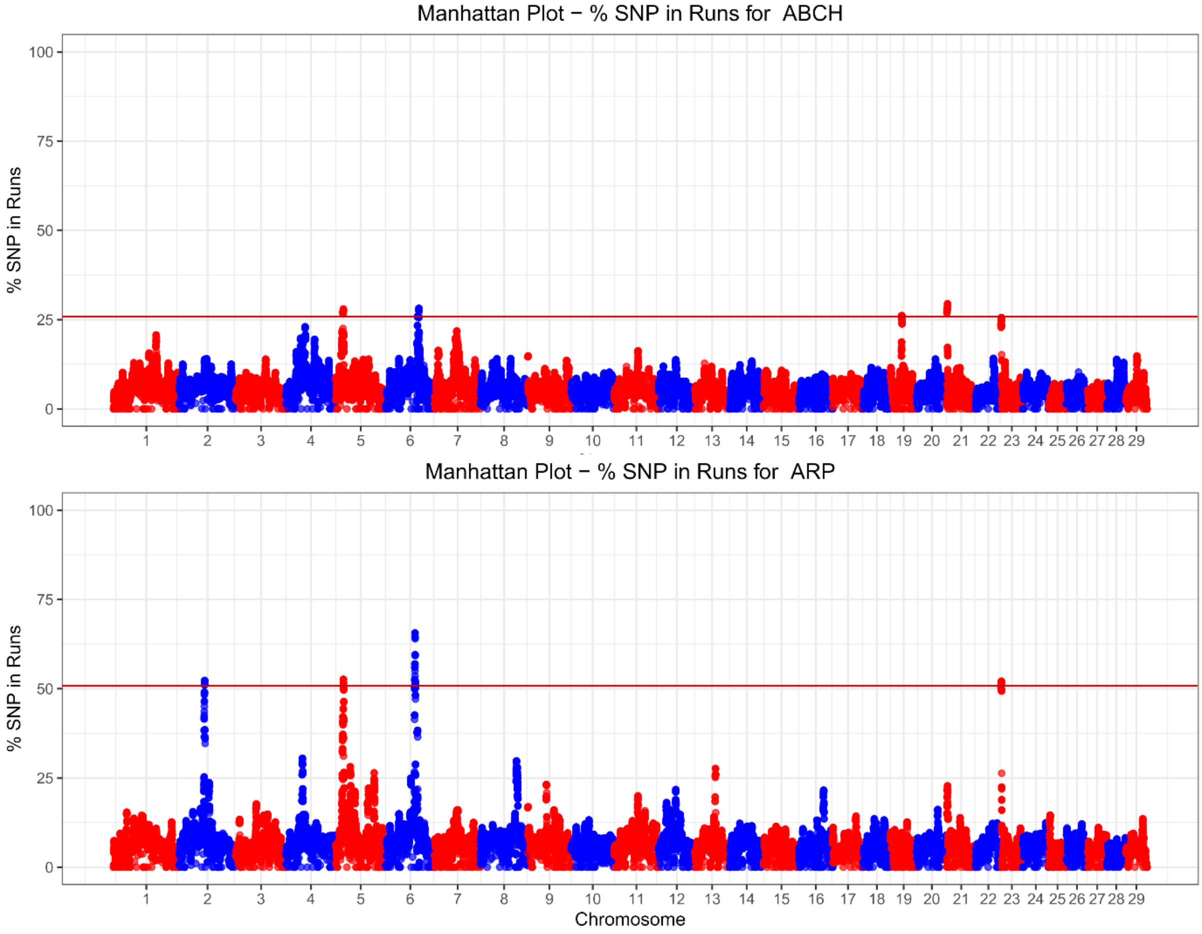


Figure 3. Manhattan plots of the SNPs incidence. Red lines threshold represents the top 1% of SNP defining the ROH_islands.

Additional information can be found at **https://doi.org/10.1080/1828051X.2023.2259221**

3.2 Genomic approach to manage genetic variability in dairy farms

Chiara Punturiero[1], Raffaella Milanesi[1], Francesca Bernini[1], Andrea Delledonne[1], Alessandro Bagnato[1] and Maria Giuseppina Strillacci[1*]

[1]*Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano, Via dell'Università 6, 26900 Lodi, Italy*

Published in the *Italian Journal of Animal Science* (2023)
https://doi.org/10.1080/1828051X.2023.2243977

Highlights

- All females of 7 herds have been genotyped with the GGP 100K SNP chip.
- Genomic information on all females can be used by farmers in the process to manage reproduction, selection and genetic herd variability.
- The availability of genomic information on the whole herd allowed to release to each farmer the genomic make-up of their herd.
- The ROH distribution together with the genotype frequencies for disease, fertility and production mendelian traits, made it possible to identify genomic regions under selection according to farmer strategies.

Abstract

In this study we investigated the genetic variability, the inbreeding and allele frequencies of monogenic traits in seven herds of Holstein breed and provided insight to farmers on the value of genomic management of reproduction in their herds. A total of 3,953 Holstein cows were sampled and genotyped with the Neogen GGP Bovine 100K SNP chip within the activities of the Regione Lombardia funded GO-PEI project 'GENOmic tool for the management of REProduction in dairy cattle and for the control of inbreeding – GENORIP'. Principal component analysis was applied for analysing the genetic variability within and among farms using the SVS software of Golden Helix. Run of Homozygosity (ROH) and the genomic inbreeding were obtained with the detectRUNS package of the R software. Genotype frequencies for mendelian disease, fertility and production traits were also obtained. A total of 458,267 ROH were identified and ROH were distributed on all autosomes with an average length of 2,703,811 bp covering 12.7% of the genome. Several genomic regions appear under selection, while a specific region on BTA4 was identified in one herd, harbouring genes mainly related to the specific selection strategy of the farmer. The FROH values obtained considering ROH greater than 16 Mb, varied from 0.004 to 0.325, with the highest FROH average value of 0.136. Among mendelian heritable diseases, the Haplotype Cholesterol Deficiency was the one with the largest proportion of carrier animals, i.e., 5.6%. A herd-tailored process to assist farmers in genomic management of reproduction was released. The ROH distribution within herd, together with the genotype frequencies for disease, fertility and production mendelian traits, suggest that similar directional selection is occurring across herds. This study released to each farmer the genomic make-up of their herd used jointly with the gEBV estimated by their national breeders' association (ANAFIBJ) for herd reproductive management.

Keywords: ROH; genomic variability; genomic inbreeding; Holstein cattle breed;

Introduction

The Holstein breed is widely recognised as the most productive dairy cattle in the world. The Holstein is the most wide-spread dairy cattle breed in Italy with 9,552 farms, 1,130,734 lactating cows and an average production of 10,710 kg of milk. In particular, in the Lombardy region located in the northern flat, which is suited to intensive dairy farming, there are 2,759 farms with a total of 566,583 animals (ANAFIBJ – National statistics 2022).

The selection index of the Holstein breed has evolved through the years, with an initial emphasis on increasing milk yield per cow, followed by a shift towards milk components and functional and health traits (Egger-Danner et al. 2015). However, as in all selection programs, the intense selection practiced over the years may led to a loss of genetic variability and to an increase in inbreeding in the population. The need to control the increase in inbreeding even in large populations under selection has been discussed for a long time (Mcdaniel 2001; Weigel 2001). In the second half of the last century, there was a motivation to introduce new molecular tools to integrate traditional phenotypic selection programs (Henderson 1975) with the use of information of loci and QTL regions that contain genes capable of influencing economically important traits in animal production (Georges et al. 1995; Andersson 2001).

In the genomics era occurred in the last decade, the paradigm of animal breeding has changed significantly. Current genotyping techniques make it possible to determine the genotype of an animal at hundreds of thousands of markers known to be associated with phenotypic variability at very low cost and use this information to select animals even without any performance available. This is the principle of genomic selection proposed by Meuwissen et al. (2001) based on the use of Single Nucleotide Polymorphisms (SNP) as markers.

The SNP are biallelic genomic markers very frequent in the genome of any individual (approximately one per 100 base pairs) (Collins et al. 1997). The SNP genotyping technology is nowadays a routine process in cattle breeding, both for males and females, producing a large number of genotyping information, allowing as such to implement efficient selection programs also for traits with low heritability (Boichard et al. 2015), and to develop comprehensive mating plans that make use of all the genomic information available to the breeder female herd.

One of the most important elements necessary to perform selection in livestock populations is the existence of genetic variability. Indeed, the occurred selection of superior animals over time has resulted in a loss of genetic diversity. This may cause a reduced response to selection and an increase of the frequency of homozygous loci (Dickerson and Hazel, 1944).

Before the advent of genomics, the study of inbreeding was based on pedigree information which, however, has limitations: 1) the value of inbreeding depends on the quality and completeness of the pedigree data (Oliehoek and Bijma 2009); 2) it does not consider the genetic variability between full siblings due to recombination during meiosis leading to an underestimation and/or overestimation of inbreeding (Hill and Weir 2011); 3) the comparison between genomic and pedigree information, showed that the frequency of misidentified bulls can be as high as 13.9% (Wiggans et al. 2012), affecting as such the inbreeding values based on this information. With the genomic selection and the development of high-density SNP arrays, it has become possible to obtain more accurate estimates of genome-wide inbreeding and relatedness (Engelsma et al. 2012). Genomic inbreeding can be calculated from a genome-wide relationship matrix (GRM) between individuals (Hayes et al. 2009), or as ratio between the length of the genome where homozygous markers form Run Of Homozygosity (ROH) (McQuillan et al. 2008) and the total genome length analysed (FROH). The length of the ROH provides also information on whether a ROH segment is the result of recent (long ROH) versus more distant (short ROH) autozygosity events (Pemberton et al. 2012). Additionally, identification of genes annotated in the ROH can provide insights on the selection occurring in the population (MacLeod et al., 2009; Purfield et al., 2012).

Due to intense selection, in dairy populations the inbreeding rates (Charlesworth and Willis, 2009) and the frequency of deleterious alleles (Ouborg et al. ,2010) have increased significantly over the years with a possible effect on productivity due to inbreeding depression (Falconer and Mackay, 1983; Keller and Waller, 2002). In Friesian cattle deleterious effects due to this phenomenon have been described on productive and functional traits (Martikainen et al., 2018; Doekes et al., 2019).

For these reasons, fostering genetic variability in the herd and controlling inbreeding is considered a priority in dairy farms and in management of cattle populations under selection. A farm-driven project funded by the Regione Lombardia in the EC EIP-AGRI Rural Development Program 2014-2020 framework is bringing genomics into the management of female replacements through the genotyping of all the animals in the herds. The project is named GENORIP: "GENOmic tool for the management of REProduction in dairy cattle and for the control of inbreeding. The project aimed to release a process to integrate the genomic management of herd reproduction and to manage inbreeding and genetic variability using dense SNP genotyping data.

This study is part of the activities of the GENORIP and aimed was to investigate the genetic variability, the genomic inbreeding and the allele frequencies of hereditary monogenic traits in the females of seven Italian Holstein large dairy cattle herds.

Materials and methods

Ethics statement

The study was approved by the OPBA of the University of Milan (Protocol number 160_2019), in accordance with the Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010, updating Directive 86/609/EEC of 1986 on the protection of animals used for scientific purposes.

Animal sampling, DNA extraction, quality control and genotyping

A total of 3,953 Holstein cows were sampled from seven different herds located in the Lombardy region of Italy. These herds were chosen for their different sizes and management practices, in order to provide some examples of dairy cattle farms in the Lombardy region. The farms differ, in fact, in their structures and available technologies (such as milking parlours vs. milking robots) and size (120 milking heads vs 600 milking cows).

Animals were sampled using ear Tissue Sampling Units (TSU) for adult individuals and bioptic ear tags for newborn calves. The collected samples were then classified in a project structured database and stored at the University of Milan tissue repository Animal Bio-Arkive (Longeri et al. 2021).

The DNA was extracted from ear tissue using the Quick-DNA™ Miniprep Kit of Zymo Research according to the manufacturer's protocol (Zymo Research Corporation). Cows were genotyped with NEOGEN's GGP Bovine 100K, consisting of approximately 100,000 SNPs, with an average SNP spacing of about 29 kbp.

All samples had a call rate value > 95%. Only SNPs located on the 29 autosomes annotated according to the ARS-UCD1.2 bovine genome assembly (n. 89,762) were considered in this study to perform all analyses. In order to avoid bias, we excluded SNPs detecting the same mutation: they were more than 600 SNPs on the autosomes.

Analysis of population structure

Principal component analysis (PCA) was used to determine genetic diversity within and among herds and has been performed using SNP & Variation Suite (SVS) v8.9 (Golden Helix Inc., Bozeman, MT, USA). The 2-D graphical visualisation of PCA results was obtained using the 'ggplot2' R library (Wickham 2016).

Genotype frequencies for health, phenotypic and productive traits

The GGP Bovine 100K chip releases a large number of SNPs genotypes associated to mendelian hereditary traits, such as genetic disorders and mutations related to phenotypic and productive traits or haplotypes linked to fertility traits. The allele and genotypic frequencies for these loci were estimated using an in-house R script.

Runs of Homozygosity Detection

Runs of Homozygosity (ROH) were obtained for each individual using the consecutive run method of the ‘detectRUNS’ library of R software (Biscarini et al., 2019). The parameters used were: (i) minimum number of 30 SNPs/ROH; (ii) a minimum length of 1 Mb for the identified ROH, to avoid the detection of short and common ROH across the genome due to Linkage Disequilibrium; (iii) a maximum distance of 1 Mb between consecutive SNPs to eliminate the bias in detection due to the density occurrence of SNPs; (iv) no missing SNPs as well as no heterozygous genotypes presence in ROH definition.

The ROH distribution per herd was evaluated separately using five classes of ROH length (<2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb and >16 Mb). Descriptive statistics relative to the total number of ROH, the ROH average number per individuals and, the average length of ROH were calculated.

The ‘detectRUNS’ library was also used to obtain: i) the graphical representations (Manhattan plots) for the percentage of occurrence of SNPs in ROH, estimated by counting the number of times that each SNP falls inside a ROH over the total number of individuals; ii) the ROH_islands, identified as peaks in Manhattan plot where SNPs are inside a ROH in more than 50% (chosen threshold) of the cows as discussed and suggested by Schiavo et al. (2021).

Gene annotation of ROH_islands and functional analyses

All ROH_islands were annotated with the genes downloaded from the NCBI online Database (NCBI Annotation Release: 106). Only genes with an official gene name were considered. Database for Annotation, Visualisation, and Integrated Discovery (DAVID) v6.8 (DAVID online Database.) was used to perform a gene ontology (GO) functional annotation and KEGG pathway analyses.

Additionally, the CattleQTLdb database (AnimalQTLdb) was used to identify – using the ‘Search by associated gene’ option – the QTL overlapping the ROH_islands.

Inbreeding coefficient

The genomic inbreeding coefficients (FROH) based on ROH were calculated for each cow as:

$$F_{ROH} = \sum L_{ROH} / L_{AUTO} \qquad (1)$$

where LROH is the total length of ROH proper of each individual genome, and LAUTO is the total genome length covered by the used SNP dataset (2,487,916,500 bp in this study). FROH were calculated for each of the five classes of ROH length (LROH) previously defined.

Results and Discussion

Table 1 reports the number of individuals sampled and genotyped in each farm, and the average gEBV for several traits released by ANAFIBJ for all genotyped females. All paternity and maternity consistency was verified based on the genome data by ANAFIBJ to solve inconsistencies due to incorrect genealogy, i.e., errors in sire or in maternal grandsire registration. The proportion of these inconsistences varied from 8% to 45%.

Population structure

A first sight of the genetic variability of the 7 herds is provided by the graphical representation of PCA shown in Figure 1. As visualised in Figure 1, within herd PCAs, cows cluster clearly in separate groups in Herd_3 (PCA1 = 15.88%, PCA2 = 13.94%), Herd_4 (PCA1 = 15.02%, PCA2 = 12.63%), Herd_6 (PCA1 = 19.67%, PCA2 = 14.44%), and to some extent also in Herd_2 (PCA1 = 16.81%, PCA2 = 16.16%) and Herd_5 (PCA1 = 13.11%, PCA2 = 12.40%), while it appears to exist more homogeneity among cows for Herd_7 (PCA1 = 14.79%, PCA2 = 13.08%) and Herd_1 (PCA1 = 15.24%, PCA2 = 12.22%). The cow clustering is expected to reflect the choices made by farmers in terms of use of sires for reproduction. More specifically it is likely that the variability shown by PCAs depends on the sire origin as system/country of selection scheme (e.g., USA, CAN, NLD, DEU, FRA, ITA) is mediated by the AI centres selling the semen to farmers. Discussion with farmers (partners of the project) on this topic disclosed a different approach in sire choice: some farmers rely on the technical advice (and semen) from a unique AI centre, some others select sires personally across all available on the market, also taking advantage of the information deriving from the mating plans offered by the farmer association. Only one, Herd_4 is selecting sires based on a herd genomic selection on females already applied for some years. All gEBVs here reported (Table 1) are based on the breeders' reproductive choices that were made without taking into account the genomic information of the females in the herd, with the exception of Herd_4.

To provide a rational for the cows' clustering in PCAs, we investigated the variability of sires used in farms as number of daughters from same sire / maternal grandsire (i.e., a bull being both sire and maternal grandsire in the same herd), within herd and among herds (Table 2).

In relation to the herd size, Herd_3 and Herd_6 use the lowest number of bulls: each sire has, in fact, an average 10.5 and 8.1 daughters, respectively. Herd_7 is the one with the largest number of sires (3.2 daughters per sire, on average) accounting for the herd size. The large size of Herd_3 somehow affects the possibility to use a large number of sires in the breeding plans, maintaining a high genetic level of the group of males: using a large number of sires to decrease the number of daughters per male, would in fact diminish the average genetic value of the reproducers. To avoid decreasing too much the genetic level of the bulls Herd_3 accept to have larger groups of daughters per sire, if compared to other herds. Nevertheless, the 5 clusters visible for Herd_3 in Figure 1 indicate that the farmer in selecting sires, in addition to the selection goal, was also paying attention to the genetic variability: the EBVs of the cows for PFT (selection index for Productivity, Functionality and Type) are in fact comparable to those of other herds, with a higher average value for milk gEBV (1,052 kg). On the contrary, the large spectrum of sires used (173 sires and 186 maternal grand sires, the largest in all herds), was widening the genetic variability, but only at a very low extent also reducing the average genetic values of used sires. It is noteworthy that Herd_7 is the one with the lowest proportion of bulls being at the same time sires and maternal grand sires of females in the herd. This indicates a fast change in sires used in reproduction and a great attention to reduce genetic inbreeding. Herd_4 has even fewer sires being also maternal grand sires, only 38%, an indication consistent with the applied selection plan of the farmer, based on genomic selection of females for several years. Herd_4 farmer is using young sires of the most recent generations as much as possible; the selection is addressed to prioritise improvement of functional traits, with particular emphasis on fertility and longevity. The mating plan based on genomic information used by Herd_4 appears very successful, both for the very good gEBVs of females, greater than all other herds for functional and production traits (Somatic Cell Count, Udder Health, Longevity and Fertility), the selection indexes (PFT, IES$, ICS-PR), and for the maintenance of genetic variability. Herd_2 is under a very successful introduction of precision farming (Automatic Milking Systems), never used genomic selection at farm female level and is systematically applying the same breeding goal in the herd for the past several years. The impact of GENORIP on this farm was positive as it allowed the farmer a fast step forward in matching the technology available in the herd with the genomic information to manage female reproduction in the herd.

When we compared the number of common bulls across herds, this can contribute to the explanation why Herd_6 is clustering separately from others: the number of sires in Herd_6, when compared to others, is a maximum 15 in common with maximum 2 other farms, while others have in common 21 to 59 sires up to 4 herds. The mating plan in Herd_6 is in fact fully relying on the technical advice of a unique AI centre with all bulls deriving from its selection program, while other herds acquire bulls from various AI centres.

Genotype frequencies for productive traits

Regarding genes linked to milk production and quality (Table 3), the AA SNP mutation of α-S1-Casein is close to 100% in all herds, while the GG SNP mutation wasn’t identified: as a key to compare to other studies the SNP allele A mutation correspond to the B variant and the G SNP mutation correspond to the C variant as usually reported (Sanchez et al. 2020). In Holstein the effect of B- and C-variants for α-S1-Casein were identified (Poulsen et al. 2013), with the B variant linked to increase in milk yield and the positive effect of C variant on curd coagulation time and curd firmness rate (Bovenhuis et al. 1992). The frequency of the B variant in indicates that the genetic potential of our herds is for milk production as it is closely to be fixed in the population.

Also for the β-Casein locus the most frequent genotype is AA in all herds (ranging from about 84% to 95%), whilst the BB genotype was found with a low genotype frequency only in Herd_1 (1.07%) and in Herd_7 (1.10%). β-Caseins show numerous genetic variants that result in different quality characteristics in milk. The most common variants are A1 and A2. A1A1 is the less frequent genotype variant of β-Casein in all herds; while the A1A2 and A2A2 vary according with herds, ranging from 37.4% to 51.7 % for A1A2 and from 28.8% to 58.0 % for A2A2. The molecular difference between the two proteins is related to a mutation resulting in an amino acid change (proline vs histidine) at position 67 of β-Casein (Ginger and Grigor 1999). The amino acid change was associated with a different gastric digestion of caseins. Indeed, during the enzymatic digestion of A1 casein, an opioid peptide (BCM-7) is released, which is not released in the digestion of A2 variants (Brooke-Taylor et al. 2017). In recent years, there has been an increased focus on the β-Casein A2 allele as some studies have suggested that the β-Casein A2 allele is better tolerated by the human population (He et al. 2017). To date, however, no relationship has been found between the consumption of cow's milk with the A1 allele for β-Casein and disease incidence. In addition, the A1 variant improves rennet coagulation properties compared to the A2 variant (Dinc et al. 2013; Ketto et al. 2017). Interest in marketing dairy products, with improved health impact, has opened the market to milk selected for its β-Casein A2 content only (Mendes et al. 2019). As Table 2 shows, it is evident that two herds, in particular Herd_5 and Herd_6, have a proportion of A2A2 genotypes >50%. These breeders, in fact, select for this genotype while the others generally have higher values for the heterozygous A1A2 genotype. The higher proportion of heterozygous genotype variants is in agreement with those reported by other authors for Holstein breed (Massella et al. 2017).

For β-Lactoglobulin, depending on the herd, the AA or AB genotypes are the most frequent genotypes: i.e., the AA genotype frequencies is higher in Herd_5 and Herd_7 (50% and 43.86%, respectively); in all other herds the higher frequency has been found for the AB genotype (ranging from 46.81% in Herd_4 to 50.95% in Herd_2). The β-Lactoglobulin is the major serum protein in cow's milk, accounting for about 50% the total amount of milk proteins and the B variant has been associated with a higher casein content, resulting in a higher cheese yield (van den Berg et al. 1992; Stasio and Mariani 2000).

The frequencies of the six K-Casein genotypes (AA, AB, AE, BB, BE, and EE) had the same pattern in all herds: the most represented genotype was AB, ranging from about 36.97% in Herd_3 to 50.45%

in Herd_5. Instead, the EE genotype has been registered in a very low number of cows (not one EE was found in Herd_2 and Herd_5) (Table 3). Studies in the literature show that in Holstein both A and B alleles are the most frequent (Prinzenberg et al. 1999; Farrell Jr et al. 2004) and the E allele the least frequent (Caroli et al. 2000). In fact, negative effects on coagulum formation during cheesemaking have been observed in milk produced by individuals carrying the E allele variant (Caroli et al. 2000; Comin et al. 2008). As Table 3 shows, heterozygous are generally the most widespread, while the BB variant, which has intermediate values, positively influences the production of cheeses, such as Parmigiano Reggiano and Grana Padano, by increasing cheese yields, as shown in the study of Mariani et al. (Mariani et al. 1976).

The heterozygous cows (AG) at marker linked to milk YellowFat feature, are still present with very low frequencies in Herd_3 (0.47%), Herd_4 (0.91%) and Herd_7 (0.36%). The AA genotypes causes a characteristic yellow colour of fat in tissues and milk, due to carotenoids depositions in adipose tissue. (Yang et al. 1992).

Finally, for the Holstein cows here analysed, at Lactoferrin locus, we found a similar distribution of the three genotype frequencies across herds, with a higher frequency of AA, mainly in Herd_4 (about 67%). AA genotype was associated with a low milk SCC values (Wojdak-Maksymiec et al. 2016).

Regarding meat traits, we observed high variability at all analysed loci (Table 3), as the Holstein breed is not selected for these traits.

Genotype frequencies for reproductive traits and disease

Table 4 shows the genotype frequencies of genes and haplotypes influencing bovine fertility. Animals bearing mutations affecting reproduction efficiency were found for all haplotypes, with different frequencies in the seven herds. The haplotype HH5 was the one with the largest number of carriers in all herds, with about 10% of females being carriers in Herd_7. Instead, high carrier haplotype frequencies were found in some herds only for specific HH. HH4 carriers are counted only for Herd_3 and Herd_7. The genotype frequencies for COQ9-rs109301586, STAT3 and 5, Leptin_2F, and PKP2_988 markers loci (all mainly involved in embryo development) are similar across all herds, counting a higher proportion of one or both homozygous genotypes, except for STAT3_25402 and STAT5_13319, for which the most frequent genotype was BB (Table 4).

In this study, the highest carrier frequency was found for HCD (haplotype cholesterol deficiency), that represents an economic loss for the farmer. Animals homozygous for this disease show the first clinical signs between 1 and 5 months of age with decreased appetite, weight loss, diarrhoea and subsequent death with a frequency in the German Friesian population of 4.2%. (Kipp et al. 2016). In our study only one animal from Herd_3 was affected. In previous study, it was shown that animals carrying HCD had significantly higher protein yields than non-carriers, but it is still unclear how HCD affects cheese yields (Cole et al. 2016).

The presence of a high proportion of carrier haplotypes/SNPs may be linked to the breeders' choice to concentrate on the selection of productive traits, which are negatively correlated with certain reproductive traits, or the use of bulls carrying some haplotype. Some of these haplotypes influence heifer conception rate, cow conception rate, milk, and protein (Cole et al. 2016). The increased consideration of these haplotypes in cattle selection criteria could lead breeders to a gradual improvement in herd fertility, reducing the losses associated with it.

Regarding genetic diseases, this study has shown the presence of carriers with low frequencies (ranging from 0.07 to 4.28, Table 3) for BLAD, GSDV and RP1 in all herds. Higher carrier frequencies were observed for Brachyspina (up to 7.22% in Herd_2).

The proportion of BLAD-carrier animals ranged from 0.07 to 0.66 which, compared with other studies such as in Brazilian Holsteins that found a carrier frequency of 5.7% (Ribeiro et al. 2000), is very low.

Avoiding carrier-to-carrier mating, and thus identifying heterozygous cows, would be a way of managing the reproduction and presence of female carriers in the herd.

Runs of Homozygosity Detection

A total of 458,267 ROH was identified in all cows of the seven herds. The count of ROH (per herd) reflects the size of herd sampling (correlation = 0.997). At the individual level, the average number of ROH ranged from 7 (Herd_1) to 251 (Herd_3), with a similar total mean ROH length close to 2.6 Mb, except for Herd_5 that had on average longer ROH, close to 3 Mb (Table 5, Figure 2A). Herd_2 showed both the lowest mean number of ROH per individual (106.8) and lowest total genome length (average value) covered by ROH (11.3%) (Table 5). Differences among cows were identified also considering the total length of the genome covered by the ROH (sum of all ROH per animal, Figure S1).

Same selection occurring across herds may have affected same regions of the genome; at genomic level some evidence may be related to the fact that across herds ROH are found in largely overlapping genomic regions among females of different herds. Over the years selection may have affected same regions where genes involved in expression of traits under selection are annotated (Zhang et al. 2015).

The ROH were found for all classes of length (Figure 2B), with shorter regions (<2 Mb), being the most frequent classes of length (about 50%), even if this proportion may be slightly overestimated according to results from (Ferenčaković et al. 2013), study however based on a 50K SNP chip and not on a 100K SNP chip array. Contrariwise, a small number of ROH longer than 16 Mb were mapped in all herds (observed frequencies ranging from 0.05% – Herd_5 to 0.30% – Herd_5) with a maximum of two ROH longer than 16Mb per individual.

Finally, ROHs were found over all chromosomes: there was no evidence of a relationship between the chromosome's length and mean ROH length, as shown in Figure 2C.

The graphs in Figure 3 shows the proportion of SNP in ROH segments across all the autosomes (Manhattan Plots) for all Herds.

ROH_islands (i.e., SNP within ROH with frequency value greater than 50% as herein before defined) were detected for the Herds 2, 4, 5, 6, and 7 and are listed in Table 6. On chr 7 two very close regions were identified for Herd_6. ROH_islands identified on chr 10 and 20 are identified in three and two herds, respectively. These two genomic regions overlap to that identified in other Holstein cows bred in Italy (Mastrangelo et al. 2018). All ROH_islands except one mapping on chr 7, harboured annotated genes (n. 68).

According to the Animal QTL database, the genes lying within the ROH_Island located on chr 10 (region shared by cows of three different herds) are mainly associated with reproduction traits (e.g., fertility traits) and morphology traits (e.g., Udder and Conformation traits). Among the genes annotated within the ROH_Island reported in Table 6, the ERBB4 and MKRN3 genes affects udder and fertility traits. In detail, ERBB4 was identified as the hub gene of the network that regulates udder growth and development and seems to affect the genes' expression that are involved in the udder involution and that promote mammary gland remodeling (Xuan et al. 2022), whereas MKRN3 controls the initiation of puberty (Abreu et al. 2015) and inhibits the reproductive axis (Abreu et al. 2020). The ROH region on chr 20 is under selection in Herd_4 and Herd_5 and includes the PELO gene. This region has been recently found under selection also in US Holstein and Jersey by (Lozada-Soto et al. 2022) who characterised the ROH in several dairy cattle populations. In their study this ROH region resulted wider respect to the one here identified.

Inbreeding coefficient

As reported in Table 7, the FROH values varied from 0.004 (Herd_1) to 0.325 (Herd_3), with an overall average value of FROH ranging from 0.113 to 0.136. These values are comparable with the genomic inbreeding calculated in the US Holstein by Lozada-Soto et al. (2022) and in Italian Holstein as reported by Dadousis et al. (2022). As shown in Figure 4, the distribution of the FROH values calculated per each class of ROH length differed among the herds, ranging from 0.113 (Herd_1) to 0.136 (Herd_4). For the two greater classes of ROH, representing the most recent genomic inbreeding, the average values (per farm) were between 0.012 and 0.023 for ROH of 8-16 Mb and 0.008 and 0.010 for ROH > 16 Mb (Table S2). We want to highlight that the maximum proportion of cows with ROH > 16 Mb was identified in Herd_5 (25.5%). In other herds this proportion is lower than 17% with a minimum value in Herd_3 of 5.7%. The inbreeding coefficients here obtained for each class of length were lower (except for the class of length <2 Mb) than calculated in Italian Holstein breed (Ablondi et al. 2022). Considering this overview, we can easily deduct that the inbreeding is taken under control in all farms in the last decades by the farmers, applying breeding strategies aimed to maintain genetic diversity among cows.

Conclusions

In the last decade, genomic selection has been very successful and rapidly adopted in the genetic improvement plans of large dairy cattle populations, such as the Italian Friesian breed. The introduction of genomic selection in selection schemes made the improvement for low heritable traits, as functional or health traits, much more efficient. In addition, the availability of SNP genotypes also on females is making it possible for farmers to customise herd breeding goals by implementing efficient selection, especially for functional and health traits, within herd and develop comprehensive mating plans that exploit all the information available to the breeder.

The use of SNP genotypes on females can be extended to optimise the herd mating plans also to manage herd genetic variability, control inbreeding at a genomic level and for specific selection for mendelian monogenic traits.

In this study, the analysis of genotypes produced by the GENORIP project provided a snapshot of the genetic variability, of the genomic inbreeding, as well as the presence of mendelian genetic variants linked to traits of interest in seven Holstein dairy herds for a total of 3,953 animals. A particularity of the project was to genotype all females present in the herds and the newborn female calves along the three years of the project duration, allowing as such to validate all the genealogical information. In addition, the availability of gEBV for all females makes it possible to evaluate the selection program adopted by farmers based on the sire side.

The knowledge of both milk properties and carriers of unfavourable traits is useful to farmers who process their milk into cheese or regarding payment according to the protein composition of the milk, and to implement mating plan, respectively.

References

Ablondi M, Sabbioni A, Stocco G, Cipolat-Gotet C, Dadousis C, Kaam J-T van, Finocchiaro R, Summer A. 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.

Abreu AP, Macedo DB, Brito VN, Kaiser UB, Latronico AC. 2015. A new pathway in the control of the initiation of puberty: the MKRN3 gene. J Mol Endocrinol. 54(3):R131.

Abreu AP, Toro CA, Song YB, Navarro VM, Bosch MA, Eren A, Liang JN, Carroll RS, Latronico AC, Rønnekleiv OK. 2020. MKRN3 inhibits the reproductive axis through actions in kisspeptin-expressing neurons. J Clin Invest. 130(8):4486–4500.

ANAFIBJ – National statistics. [accessed 2022 Oct 18]. http://www.anafi.it/it/pubblicazioni-statistiche/statistiche-nazionali

Andersson L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet. 2(2):130–138.

van den Berg G, ESCHER JTM, de Koning PJ, Bovenhuis H. 1992. Genetic polymorphism of κ-casein and β-lactoglobulin in relation to milk composition and processing properties. Nederlands melk en Zuiveltijdschrift. 46(3–4):145–168.

Biscarini F, Cozzi P, Gaspa G, Marras G. 2019. detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes genomes. R package version 0.9.5. Available at https://CRAN.R-project.org/package=detectRUNS. [place unknown].

Boichard D, Ducrocq V, Fritz S. 2015. Sustainable dairy cattle selection in the genomic era. Journal of Animal Breeding and Genetics. 132(2):135–143.

Bonfatti V, Di Martino G, Cecchinato A, Degano L, Carnier P. 2010. Effects of β-κ-casein (CSN2-CSN3) haplotypes, β-lactoglobulin (BLG) genotypes, and detailed protein composition on coagulation properties of individual milk of Simmental cows. J Dairy Sci. 93(8):3809–3817.

Bovenhuis H, Van Arendonk JAM, Korver S. 1992. Associations between milk protein polymorphisms and milk production traits. J Dairy Sci. 75(9):2549–2559.

Brooke-Taylor S, Dwyer K, Woodford K, Kost N. 2017. Systematic Review of the Gastrointestinal Effects of A1 Compared with A2 β-Casein. Advances in Nutrition [Internet]. 8(5):739–748. https://doi.org/10.3945/an.116.013953

Caroli A, Bolla P, Budelli E, Barbieri G, Leone P. 2000. Effect of κ-casein E allele on clotting aptitude of Italian Friesian milk. Zootecnica e Nutrizione Animale. 26(3):127–130.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. Nat Rev Genet. 10(11):783–796.

Cole JB, Null DJ, VanRaden PM. 2016. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. J Dairy Sci. 99(9):7274–7288.

Collins FS, Guyer MS, Chakravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. Science (1979). 278(5343):1580–1581.

Comin A, Cassandro M, Chessa S, Ojala M, Dal Zotto R, de Marchi M, Carnier P, Gallo L, Pagnacco G, Bittante G. 2008. Effects of Composite β- and κ-Casein Genotypes on Milk Coagulation, Quality, and Yield Traits in Italian Holstein Cows. J Dairy Sci. 91(10):4022–4027. https://doi.org/10.3168/jds.2007-0546

Dadousis C, Ablondi M, Cipolat-Gotet C, van Kaam J-T, Marusi M, Cassandro M, Sabbioni A, Summer A. 2022. Genomic inbreeding coefficients using imputed genotypes: Assessing different estimators in Holstein-Friesian dairy cows. J Dairy Sci. 105(7):5926–5945. https://doi.org/10.3168/jds.2021-21125

Dickerson G, Hazel LN. 1944. Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. J agric Res. 69(2):459–476.

Dinc H, Ozkan E, Koban E, Togan I. 2013. Beta-casein A1/A2, kappa-casein and beta-lactoglobulin polymorphisms in Turkish cattle breeds. Arch Anim Breed. 56(1):650–657.

Doekes HP, Veerkamp RF, Bijma P, De Jong G, Hiemstra SJ, Windig JJ. 2019. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. Genetics Selection Evolution. 51(1):1–16. https://doi.org/10.1186/s12711-019-0497-z

Egger-Danner C, Cole JB, Pryce JE, Gengler N, Heringstad B, Bradley A, Stock KF. 2015. Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal. 9(2):191–207.

Engelsma KA, Veerkamp RF, Calus MPL, Bijma P, Windig JJ. 2012. Pedigree-and marker-based methods in the estimation of genetic diversity in small groups of Holstein cattle. Journal of Animal Breeding and Genetics. 129(3):195–205.

Falconer DS, Mackay TFC. 1983. Quantitative genetics. [place unknown]: Longman London, UK.

Farrell Jr HM, Jimenez-Flores R, Bleck GT, Brown EM, Butler JE, Creamer LK, Hicks CL, Hollar CM, Ng-Kwai-Hang KF, Swaisgood HE. 2004. Nomenclature of the proteins of cows' milk—Sixth revision. J Dairy Sci. 87(6):1641–1674.

Ferenčaković, M., J. Sölkner, J., & I. Curik I. 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genetics Selection Evolution. 45:42.

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics. 139(2):907–920.

Ginger MR, Grigor MR. 1999. Comparative aspects of milk caseins. Comp Biochem Physiol B Biochem Mol Biol. 124(2):133–145. https://doi.org/10.1016/S0305-0491(99)00110-8

Hayes BJ, Visscher PM, Goddard ME. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res (Camb). 91(1):47–60.

He M, Sun J, Jiang ZQ, Yang YX. 2017. Effects of cow's milk beta-casein variants on symptoms of milk intolerance in Chinese adults: a multicentre, randomised controlled study. Nutr J. 16(1):72. https://doi.org/10.1186/s12937-017-0275-0

Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics.:423–447.

Hill WG, Weir BS. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genet Res (Camb). 93(1):47–64.

Keller LF, Waller DM. 2002. Inbreeding effects in wild populations. Trends Ecol Evol. 17(5):230–241.

Ketto IA, Knutsen TM, Øyaas J, Heringstad B, Ådnøy T, Devold TG, Skeie SB. 2017. Effects of milk protein polymorphism and composition, casein micelle size and salt distribution on the milk coagulation properties in Norwegian Red cattle. Int Dairy J. 70:55–64.

Kipp S, Segelke D, Schierenbeck S, Reinhardt F, Reents R, Wurmser C, Pausch H, Fries R, Thaller G, Tetens J, et al. 2016. Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. J Dairy Sci. 99(11):8915–8931. https://doi.org/10.3168/jds.2016-11118

Longeri ML, Zaniboni L, Cozzi MC, Milanesi R, Bagnato A. 2021. Animal Bio Arkivi: establishment of a phenotype and tissue repository for farm animals and pets at the University of Milan. In: ASPA Congress. Vol. 20. [place unknown]: Taylor & Francis; p. 136.

Lozada-Soto EA, Tiezzi F, Jiang J, Cole JB, VanRaden PM, Maltecca C. 2022. Genomic characterization of autozygosity and recent inbreeding trends in all major breeds of US dairy cattle. J Dairy Sci. 105(11):8956–8971.

MacLeod IM, Hayes BJ, Goddard ME. 2009. A novel predictor of multilocus haplotype homozygosity: comparison with existing predictors. Genet Res (Camb). 91(6):413–426.

Mariani P, Losi G, Russo V, Castagnetti GB, Grazia L, Morini D, Fossa E. 1976. [Caseification tests made with milk characterized by variants A and B of k-casein in the production of Parmigiano-Reggiano cheese]. [Paper presented at a conference]. [Italian]. Scienza e Tecnica Lattiero Casearia.

Martikainen K, Sironen A, Uimari P. 2018. Estimation of intrachromosomal inbreeding depression on female fertility using runs of homozygosity in Finnish Ayrshire cattle. J Dairy Sci. 101(12):11097–11107.

Massella E, Piva S, Giacometti F, Liuzzo G, Zambrini AV, Serraino A. 2017. Evaluation of bovine beta casein polymorphism in two dairy farms located in northern Italy. Ital J Food Saf. 6(3).

Mastrangelo S, Sardina MT, Tolone M, Di Gerlando R, Sutera AM, Fontanesi L, Portolano B. 2018. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. Animal. 12(12):2480–2488.

Mc Parland S, Kearney JF, Rath M, Berry DP. 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. J Dairy Sci. 90(9):4411–4419.

Mcdaniel BT. 2001. Uncontrolled Inbreeding. J Dairy Sci. 84:E185–E186. https://doi.org/10.3168/jds.S0022-0302(01)70214-7

McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, et al. 2008. Runs of homozygosity in European populations. Am J Hum Genet. 83(3):359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

Mendes MO, de Morais MF, Rodrigues JF. 2019. A2A2 milk: Brazilian consumers' opinions and effect on sensory characteristics of Petit Suisse and Minas cheeses. LWT. 108:207–213.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819–1829.

Oliehoek PA, Bijma P. 2009. Effects of pedigree errors on the efficiency of conservation decisions. Genetics Selection Evolution. 41(1):9.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. 2010. Conservation genetics in transition to conservation genomics. Trends in genetics. 26(4):177–187.

Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. 2012. Genomic patterns of homozygosity in worldwide human populations. The American Journal of Human Genetics. 91(2):275–292.

Poulsen NA, Bertelsen HP, Jensen HB, Gustavsson F, Glantz M, Månsson HL, Andrén A, Paulsson M, Bendixen C, Buitenhuis AJ. 2013. The occurrence of noncoagulating milk and the association of bovine milk coagulation properties with genetic variants of the caseins in 3 Scandinavian dairy breeds. J Dairy Sci. 96(8):4830–4842.

Prinzenberg E, Krause I, Erhardt G. 1999. SSCP analysis at the bovine CSN3 locus discriminates six alleles corresponding to known protein variants (A, B, C, E, F, G) and three new DNA polymorphisms (H, I, A1). Anim Biotechnol. 10(1–2):49–62.

Purfield DC, Berry DP, McParland S, Bradley DG. 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13(1):70.

Ribeiro LA, Baron EE, Martinez ML, Coutinho LL. 2000. PCR screening and allele frequency estimation of bovine leukocyte adhesion deficiency in Holstein and Gir cattle in Brazil. Genet Mol Biol. 23:831–834.

Sanchez M-P, Fritz S, Patry C, Delacroix-Buchet A, Boichard D. 2020. Frequencies of milk protein variants and haplotypes estimated from genotypes of more than 1 million bulls and cows of 12 French cattle breeds. J Dairy Sci. 103(10):9124–9141.

Schiavo G, Bovo S, Muñoz M, Ribani A, Alves E, Araújo JP, Bozzi R, Čandek-Potokar M, Charneca R, Fernandez AI. 2021. Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds. Anim Genet. 52(2):155–170.

Stasio L di, Mariani P. 2000. The role of protein polymorphism in the genetic improvement of milk production. Zootecnica e Nutrizione Animale. 26(3):69–90.

Weigel KA. 2001. Controlling inbreeding in modern breeding programs. J Dairy Sci. 84:E177–E184.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag, editor. New York; [accessed 2021 May 2]. https://ggplot2.tidyverse.org

Wiggans GR, Cooper TA, VanRaden PM, Olson KM, Tooker ME. 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. J Dairy Sci. 95(3):1552–1558.

Xuan R, Chao T, Zhao X, Wang A, Chu Y, Li Q, Zhao Y, Ji Z, Wang J. 2022. Transcriptome profiling of the nonlactating mammary glands of dairy goats reveals the molecular genetic mechanism of mammary cell remodeling. J Dairy Sci. 105(6):5238–5260.

Yang A, Larsen T, Tume R. 1992. Carotenoid and retinol concentrations in serum, adipose tissue and liver and carotenoid transport in sheep, goats and cattle. Aust J Agric Res. 43(8):1809. https://doi.org/10.1071/AR9921809

Zhang Q, Guldbrandtsen B, Bosse M, Lund MS, Sahana G. 2015. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics. 16(1):1–16. https://doi.org/10.1186/s12864-015-1715-x

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This project was founded by EAFRD Rural Development Program 2014-2020, Management Autority Regione Lombardia - OP. 16.1.01 Project ID n. 201801062430 – ‘Operational Group EIP AGRI’ https://ec.europa.eu/eip/agriculture/en/eip-agriprojects/projects/operational-groups.

Author contributions

CP: collected the samples; MGS and AB: Conceived and designed the study; RM: performed the experiments, CP and MGS: data analysis; AD and FB: participated to data analysis; CP, MGS and AB: wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

Data availability statement

The data supporting the conclusions of this manuscript are included in the Supplementary Materials.

Tables

Table 1. Sampling and Milk destination of each herd considered, together with descriptive statistics (minimum, maximum and mean) of gEBV (genomic EBV) values calculated by ANAFBJ.

HERD_N.	**HERD_1**	**HERD_2**	**HERD_3**	**HERD_4**	**HERD_5**	**HERD_6**	**HERD_7**
N. samples	455	263	1494	329	220	638	554
GPFT	**1559/4474 (3045.87)**	**1373/3875 (2689.16)**	**1726/4543 (3436.17)**	**1726/4543 (3436.2)**	**1760/4346 (3093.47)**	**1608/4072 (3002.92)**	**1847/4324 (3181.58)**
IES€	**-387/1228 (388.59)**	**-451/944 (257)**	**-345/875 (288.62)**	**-199/1266 (668.22)**	**-310/1124 (416.91)**	**-447/1007 (312.27)**	**-256/1083 (421.58)**
ICS-PR	**-990/1469 (108)**	**-1117/870 (-69.05)**	**-1235/888 (-55.85)**	**-684/1411 (519.04)**	**-824/1510 (168.35)**	**-1122/1135 (70.77)**	**-749/1386 (207.93)**
Milk (Kg)	-1168/2514 (605.81)	-896/1872 (424.01)	-1211/2437 (725.55)	-870/2055 (751.66)	-1265/2060 (574.73)	-1144/2021 (580.94)	-964/2061 (634.58)
Fat (kg)	-46/107 (32.78)	-48/99 (12.27)	-64/110 (25.12)	-44/122 (45.6)	-45/116 (31.69)	-55/94 (26.62)	-33/114 (30.05)
Protein (Kg)	-30/85 (27.77)	-33/70 (16.52)	-27/84 (29.54)	-41/83 (36.28)	-37/78 (29.44)	-36/75 (23.06)	-22/86 (30.07)
Fat (%)	-0.57/0.84 (0.09)	-0.53/0.7 (-0.04)	-0.58/0.65 (-0.02)	-0.72/0.8 (0.16)	-0.53/0.66 (0.09)	-0.61/0.82 (0.05)	-0.61/0.76 (0.06)
Protein (%)	-0.32/0.39 (0.06)	-0.19/0.32 (0.02)	-0.29/0.34 (0.04)	-0.23/0.46 (0.09)	-0.18/0.39 (0.09)	-0.39/0.41 (0.03)	-0.28/0.36 (0.07)
Feed efficiency	94/106 (99.61)	94/105 (98.92)	94/108 (101.33)	90/104 (98.52)	94/105 (99.39)	89/104 (97.23)	88/106 (99.02)
Type	-1.15/2.07 (0.62)	-1.61/1.63 (0.11)	-1.67/2.27 (0.16)	-1.52/1.73 (0.51)	-0.91/1.93 (0.59)	-1.29/2.75 (1.17)	-0.56/3.01 (0.95)
ICM	-1.91/2.82 (0.79)	-2.38/2.89 (0.28)	-2.22/2.82 (0.18)	-2.24/3.23 (0.89)	-1.67/2.77 (0.78)	-1.95/3.78 (1.47)	-1.15/4.32 (1.3)
Feet and leg	-2.74/3.7 (0.44)	-3.26/2.64 (-0.11)	-3.51/2.81 (-0.18)	-2.91/3.69 (0.27)	-2.48/3.48 (0.3)	-2.07/3.83 (0.72)	-2.82/3.77 (0.89)
Somatic cell	92/113 (102.5)	93/112 (102.14)	89/115 (100.84)	91/121 (104.39)	90/116 (101.86)	91/116 (103.34)	88/120 (103.41)
Udder health	92/113 (102.38)	92/113 (102.34)	86/115 (100.83)	91/118 (103.91)	89/115 (102.58)	90/116 (102.55)	91/118 (103.21)
Longevity	91/118 (105.24)	94/117 (104.97)	91/117 (102.96)	93/121 (110.23)	95/118 (105.66)	94/117 (105.24)	94/118 (105.42)
Fertility	88/114 (102.06)	93/113 (101.92)	85/116 (100.37)	89/116 (105.03)	94/116 (102.54)	90/112 (101.01)	91/113 (102.4)
BCS	86/114 (98.18)	88/113 (99.65)	84/114 (98.81)	87/116 (99.28)	78/109 (97.28)	82/109 (96)	78/113 (97.67)

GPFT: Production, Functionality and Type genomic index; IES€: Indice economico salute (Health Economic Index); ICS-PR: Indice Caseificazione e Sostenibilità-Parmigiano Reggiano (Cheesemaking and Sustainability Index – Parmigiano Reggiano); ICM: Indice Composto Mammella (Udder Composite Index); BCS: Body Condition Score.

Table 2. Number of sire and maternal grandsires used as reproducers in each herd and shared among them.

Herd_n	N. Sires	N. min -max of daughters per sire in the herd (average)	N. maternal grandsires	N. sires also maternal grand sire* (%)	N. of common sires (%) in N. of Herds (min - max)	N. of daughters of common sires in other herds (average)	N. daughters of common sires across all herds (average)
		Values within herds			Values among herds		
Herd_1	101	1-16 (4.5)	121	75 (74)	59 (58.0) – (1-4)	1-81 (7.2)	2-85 (14.9)
Herd_2	50	1-35 (5.3)	67	34 (68)	23 (46.0) – (1-3)	1-31 (7.8)	3-52 (17.9)
Herd_3	142	1-81 (10.5)	159	104 (73)	38 (26.8) – (1-4)	1-21 (4.4)	3-85 (22.4)
Herd_4	53	1-30 (6.2)	82	20 (38)	22 (41.0) – (1-4)	1-63 (7.6)	2-65 (18.5)
Herd_5	56	1-19 (3.9)	72	41 (73)	21 (37.5) – (1-4)	1-48 (7.8)	2-53 (16.8)
Herd_6	79	1-117 (8.1)	98	59 (75)	15 (18.8) – (1-2)	1-16 (4.4)	2-31 (11.3)
Herd_7	173	1-14 (3.2)	186	90 (52)	56 (32.4) – (1-4)	1-81 (8.0)	2-85 (14.6)

*N. of sires that are maternal grandsire of another female. Proportion calculated as (n. sires also maternal grandsires) / (n. sires)

Table 3. Genotype frequencies for productive traits (i.e., milk protein and meat variant of monogenic related loci) in each herd.

Milk trait related loci								
Gene	**Genotypes/** *mutation* **at SNP level**	**Herd_1**	**Herd_2**	**Herd_3**	**Herd_4**	**Herd_5**	**Herd_6**	**Herd_7**
α-S1-Casein	BB (*AA*)	98.90	100	99.87	100	99.50	100	99.80
	BC (*AG*)	1.10	-	0.13	-	0.50	-	0.20
	CC (*GG*)	-	-	-	-	-	-	-
β-Casein BB	AA	93.41	84.03	96.79	93.92	95.90	93.26	93.50
	AB	6.37	15.97	3.21	6.08	4.10	6.74	6.30
	BB	1.07	-	-	-	-	-	1.10
β-Casein variants *	A1A1	13.41	20.15	12.58	15.20	6.82	4.70	13.92
	A1A2	44.62	51.71	48.33	51.37	39.09	37.77	48.64
	A2A2	41.98	28.14	39.09	33.43	54.09	57.52	37.43
β-Lactoglobulin	AA	35.82	22.81	33.22	40.73	50.00	37.80	43.86
	AB	51.21	50.95	50.03	46.81	40.00	47.72	42.42
	BB	12.97	26.24	16.74	12.46	10.00	14.49	13.72
K-Casein	AA	17.58	36.12	29.74	14.29	16.82	10.03	18.95
	AB	42.86	39.16	36.97	39.82	50.45	48.12	39.35
	AE	10.99	6.46	11.52	6.99	5.45	5.17	10.83
	BB	17.14	12.93	12.12	26.75	25.00	29.00	19.49
	BE	9.89	5.32	8.51	10.33	2.27	7.21	10.65
	EE	1.54	-	1.14	1.82	-	0.47	0.72
Lactoferrin	AA (*AA*)	54.73	33.84	50.94	67.48	42.73	46.39	55.78
	AB (*AG*)	39.12	53.61	40.70	28.27	48.18	44.36	38.45
	BB (*GG*)	6.15	12.55	8.37	4.26	9.09	9.25	5.78
Yellow Fat	*GG*	100	100	99.53	99.09	100	100	99.64
	AG	-	-	0.47	0.91	-	-	0.36
	AA	-	-	-	-	-	-	-
Meat traits related loci								
Calpain_316	*CC*	36.92	19.77	28.67	34.95	34.09	26.33	37.36
	CG	49.45	51.33	51.11	46.20	46.36	50.47	47.83
	GG	13.63	28.90	20.23	18.84	19.55	23.20	14.80
Calpain_4751	*CC*	52.97	56.65	53.68	51.67	61.82	56.43	56.32
	CT	41.10	36.50	39.02	39.82	34.55	38.24	38.99
	TT	5.93	6.84	7.30	8.51	3.64	5.33	4.69
CAST_2870	AA (*AA*)	20.66	9.89	19.08	22.19	22.37	28.68	14.08
	AB (*AG*)	49.89	44.11	50.67	49.85	47.95	49.06	45.85
	BB (*GG*)	29.45	46.01	30.25	27.96	29.68	22.26	40.07
CAST_2959	AA (*AA*)	34.73	50.57	43.11	42.55	39.27	27.74	46.57
	AB (*AG*)	48.57	41.44	44.91	43.47	45.66	49.06	43.68
	BB (*GG*)	16.70	7.98	11.98	13.98	15.07	23.04	9.75
UoGCAST1	CC	27.21	35.88	35.37	29.18	28.90	23.38	37.64
	CG	49.12	51.53	47.11	48.63	48.17	47.87	48.18
	GG	23.67	12.60	17.52	22.19	22.94	28.75	14.18

*Genotype frequencies at marker CSN2_7 (the same frequencies were obtained for CSN2_X14711_8101 locus)

Table 4. Genotype frequencies calculated for genes (haplotype) affecting fertility traits and genetic diseases in each herd.

Fertility related traits loci								
Gene haplotype (HH)*	**Genotypes at SNP level**	**Herd_1**	**Herd_2**	**Herd_3**	**Herd_4**	**Herd_5**	**Herd_6**	**Herd_7**
COQ9	AA (AA)	24.84	27.76	16.80	24.01	17.27	10.50	19.89
rs109301586	AB (AG)	45.05	52.85	48.46	51.37	50.91	47.02	45.21
	BB (GG)	30.11	19.39	34.74	24.62	31.82	42.48	34.90
HH1	Free	96.26	100	97.59	100	93.18	97.34	93.50
	Carrier	3.74	-	2.41	-	6.82	2.66	6.50
HH3	Free	98.24	92.78	93.17	96.96	97.73	99.37	94.58
	Carrier	1.76	7.22	6.83	3.04	2.27	0.63	5.42
HH4	Free	100	100	96.72	100	100	100	99.09
	Carrier	-	-	3.28	-	-	-	0.91
HH5	Free	93.61	94.30	93.37	92.40	94.55	91.22	89.71
	Carrier	6.39	5.70	6.63	7.60	5.45	8.78	10.29
HH6	AA (AA)	99.12	99.62	97.39	99.39	93.64	100	99.10
	AB (AG)	0.88	0.38	2.61	0.61	6.36	-	0.90
HH7 CENPU	AA (DD)	-	-	-	-	-	-	-
	AB (DI)	0.20	0.40	0.40	-	-	0.20	0.40
	BB (II)	99.80	99.60	99.60	100	100	99.80	99.60
Leptin_2FB	CC	50.77	39.54	33.27	36.17	37.27	41.07	38.45
	CT	40.44	47.91	48.59	52.58	42.73	47.18	46.39
	TT	8.79	12.55	18.14	11.25	20.00	11.76	15.16
STAT3_19069	AA (TT)	17.36	23.95	16.33	14.29	15.91	15.99	13.18
	AB (TC)	48.13	52.85	48.39	49.54	52.73	48.75	44.95
	BB (CC)	34.51	23.19	35.27	36.17	31.36	35.27	41.88
STAT3_25402	AA (TT)	7.69	8.37	7.76	3.34	4.09	7.05	5.60
	AB (TG)	37.36	44.11	37.62	34.65	29.55	45.61	35.74
	BB (GG)	54.95	47.53	54.62	62.01	66.36	47.34	58.66
STAT5_13244	AA (TT)	14.07	14.83	18.14	7.29	7.27	19.75	12.45
	AB (TC)	47.25	50.19	47.86	44.68	41.82	54.55	45.85
	BB (CC)	38.68	34.98	34.00	48.02	50.91	25.71	41.70
STAT5_13319	AA (TT)	2.20	3.42	2.01	2.43	5.45	1.25	1.62
	AB (TG)	25.33	33.08	23.90	31.61	35.91	18.50	21.12
	BB (GG)	72.47	63.50	74.10	65.96	58.64	80.25	77.26
PKP2_988	AA (I/I)	41.32	31.68	37.51	31.00	45.91	57.77	43.14
	AB (I/D)	44.84	51.15	48.96	52.89	46.82	35.95	45.85
	BB (D/D)	13.85	17.18	13.53	16.11	7.27	6.28	11.01
Mendellian heritable disease loci								
BLAD	Free	99.34	100	99.93	100	100	99.84	100
	Carrier	0.66	-	0.07	-	-	0.16	-
	Affected	-	-	-	-	-	-	-
Brachyspina	Free	97.58	92.78	97.59	98.78	100	96.55	99.64
	Carrier	2.42	7.22	2.41	1.22	-	3.45	0.36
	Affected	-	-	-	-	-	-	-
GSDV	Free	99.78	100	99.8	100	100	100	100
	Carrier	0.22	-	0.2	-	-	-	-
	Affected	-	-	-	-	-	-	-
HCD	Free	90.99	98.10	92.97	98.48	92.73	92.95	94.22
	Carrier	9.01	1.90	6.96	1.52	7.27	7.05	5.78
	Affected	-	-	1	-	-	-	-
RP1	AA (I/I)	-	-	-	-	-	-	-
	AB (I/D)	1.76	0.38	4.28	-	0.50	-	-
	BB (D/D)	98.24	99.62	95.72	100	99.50	100	100

*Coenzyme Q9 (COQ9) (G→A; rs109301586); Holstein Haplotype 1 (HH1); Holstein Haplotype 3 (HH3); Holstein Haplotype 4 (HH4); Holstein Haplotype 5 (HH5); Holstein Haplotype 6 (HH6); Holstein Haplotype 7 (HH7 CENPU); Leptin_2FB; Brachyspina; Bovine Leukocyte Adhesion Deficiency (BLAD); Holstein Cholesterol Disorder (HCD); Glycogen storage disease type V (GSDV);

Progressive retinal degeneration (RP1); Signal transducer and activator of transcription (STAT); plakophilin 2 (PKP2_988). I=insertion; D=Deletion.

Table 5. Runs of Homozygosity (ROH) descriptive statistics.

Herd	Total ROH	Min-Max (mean) n. ROH	Min ROH length	Max ROH length	Mean ROH length	Mean coverage[1] (%)[2]
Herd_1	53,513	7-208 (117.6)	1,000,172	33,248,436	2,659,912	312,834,920 (12.6)
Herd_2	28,094	57-174 (106.8)	1,000,172	23,591,055	2,630,000	280,940,043 (11.3)
Herd_3	164,064	26-251 (109.8)	1,000,014	45,522,015	2,685,399	294,89,7811 (11.8)
Herd_4	41,637	20-188 (126.6)	1,000,172	26,347,829	2,677,836	338,896,839 (13.6)
Herd_5	24,913	64-172 (113.2)	1,000,172	33,218,333	2,967,912	336,089,051 (13.5)
Herd_6	79,827	47-208 (125.1)	1,000,124	26,861,540	2,586,512	323,626,185 (13.0)
Herd_7	66,219	53-187 (119.5)	1,000,039	29,616,394	2,719,103	325,011,350 (13.1)

[1] Mean Coverage= average calculated by Herds on total ROH length cover each sample's genome; [2] Proportion of genome covered by ROH (Mean Coverage) on 2,487,916,500 bp (Genome length covered by the 89,762 SNP). ROH lengths are expressed in base pairs (bps).

Table 6. Runs of Homozygosity (ROH)_Islands detected and shared in at least 50% of cows together with the annotated genes annotated and associated traits.

Herd_n	**chrom**	**nSNP**	**From**	**To**	**Genes**	**Cattle QTL database**
Herd_5	2	7	100503013	100668079	ERBB4	
Herd_4	4	29	54647672	55649011	GPR85, BMT2, TMEM168	
Herd_6	7	33	94623898	95679696	FAM81B, TTC37, ARSK, GPR150, RFESD, SPATA9, RHOBTB3, GLRX, ELL2	
Herd_6	7	3	95832919	95890433	-	
Herd_4	10	58	31539600	32997552	C10H15orf41*, MEIS2	Bone quality, Feet and leg conformation, Foot angle
Herd_6	10	77	36261135	38520889	RAD51*, RMDN3, GCHFR, DNAJC17, C10H15orf62, ZFYVE19, PPP1R14D, SPINT1, RHOV, *VPS18*, ***DLL4, CHAC1, INO80, EXD1, CHP1, OIP5, NUSAP1*, NDUFAF1, RTF1*, ITPKA, RPAP1*, TYRO3, MGA, MAPKBP1*, JMJD7, PLA2G4B, SPTBN5, EHD4***, *PLA2G4E, PLA2G4D, PLA2G4F*, VPS39, TMEM87A, GANC*, CAPN3*, ZNF106, SNAP23, LRRC57, HAUS2, STARD9, CDAN1, TTBK2, UBR1*, TMEM62, CCNDBP1, EPB42*	Body depth, Bovine respiratory disease susceptibility, Calving ease, Dairy form, Daughter pregnancy rate, Feet and leg conformation, First service conception, Inseminations per conception, Length of productive life, Milk protein percentage, Net merit, PTA type, Quality grade, Rear leg placement - rear view, Rump width, Shear force, Somatic cell score, Stature, Stillbirth, Strength, Teat length, Teat placement, Teat placement – rear, Udder attachment, Udder cleft, Udder depth, Udder height
Herd_*7*	*10*	*72*	*36420875*	*38520889*		
Herd_2	**10**	**28**	**36459819**	**37304540**		
Herd_4	20	42	26292282	27490668	PELO	Milk protein percentage
Herd_5	20	30	26591866	27490668		
Herd_6	21	46	497917	2416354	MKRN3, MAGEL2, NDN, SNRPN, SNURF, UBE3A	

*Genes associated with traits reported in '**Cattle QTL Database**' column.

Figures

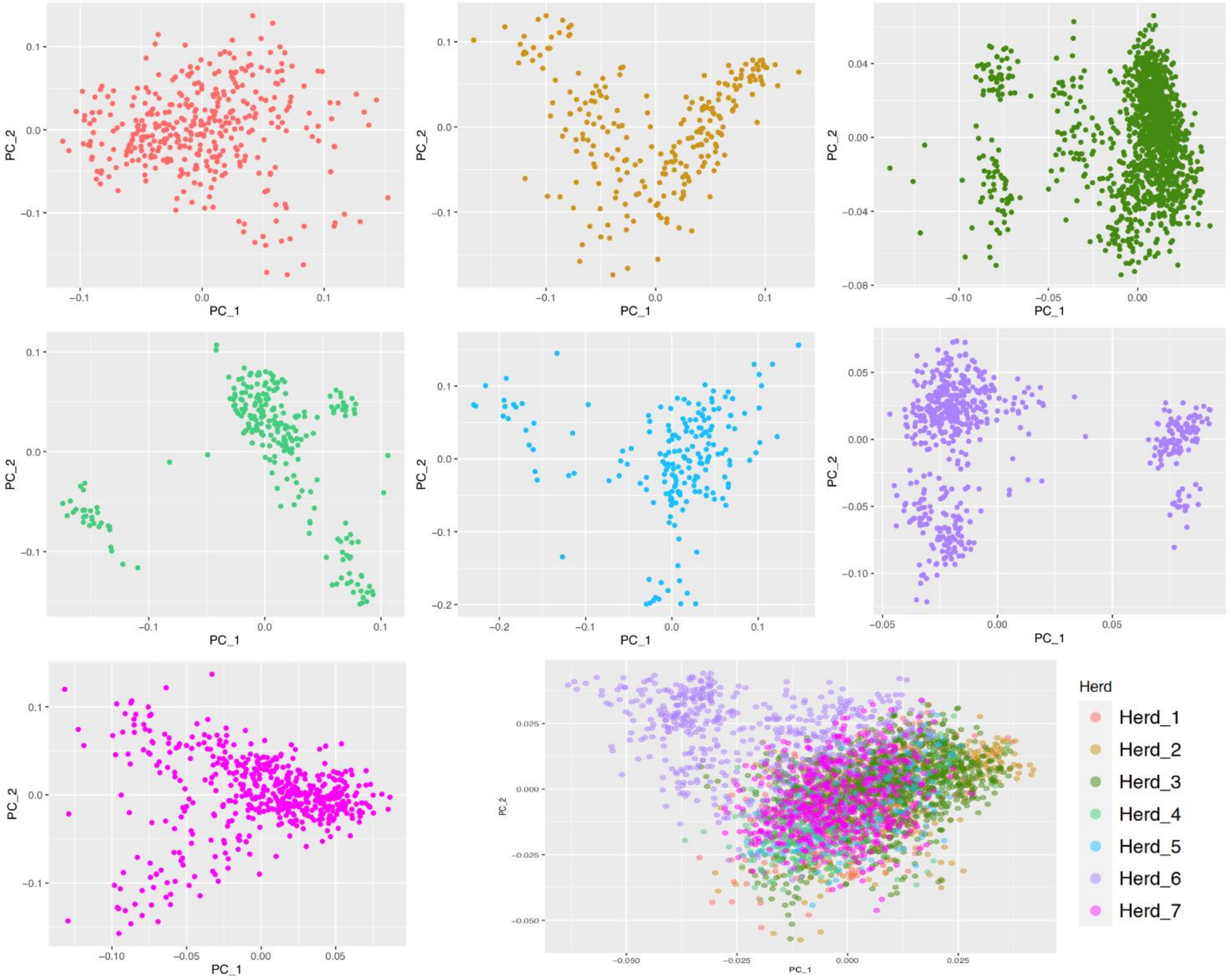


Figure 1. Graphical representation of PCA results both for each herd and for all individual together.

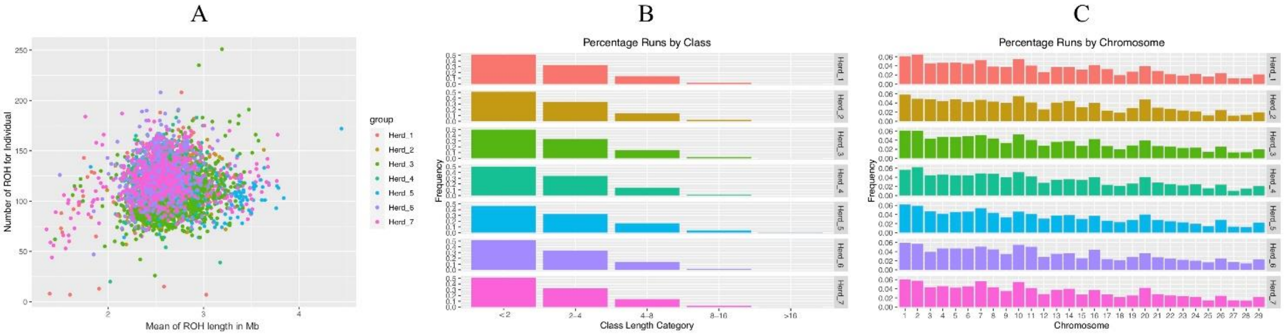


Figure 2. Graphical representation of ROH statistics per herd: (A) relationship between number and averaged total length (Mb) of ROH proper of each cow; (B) frequencies of ROH for each class of length together with details on the > 16 Mb ROH class of length.

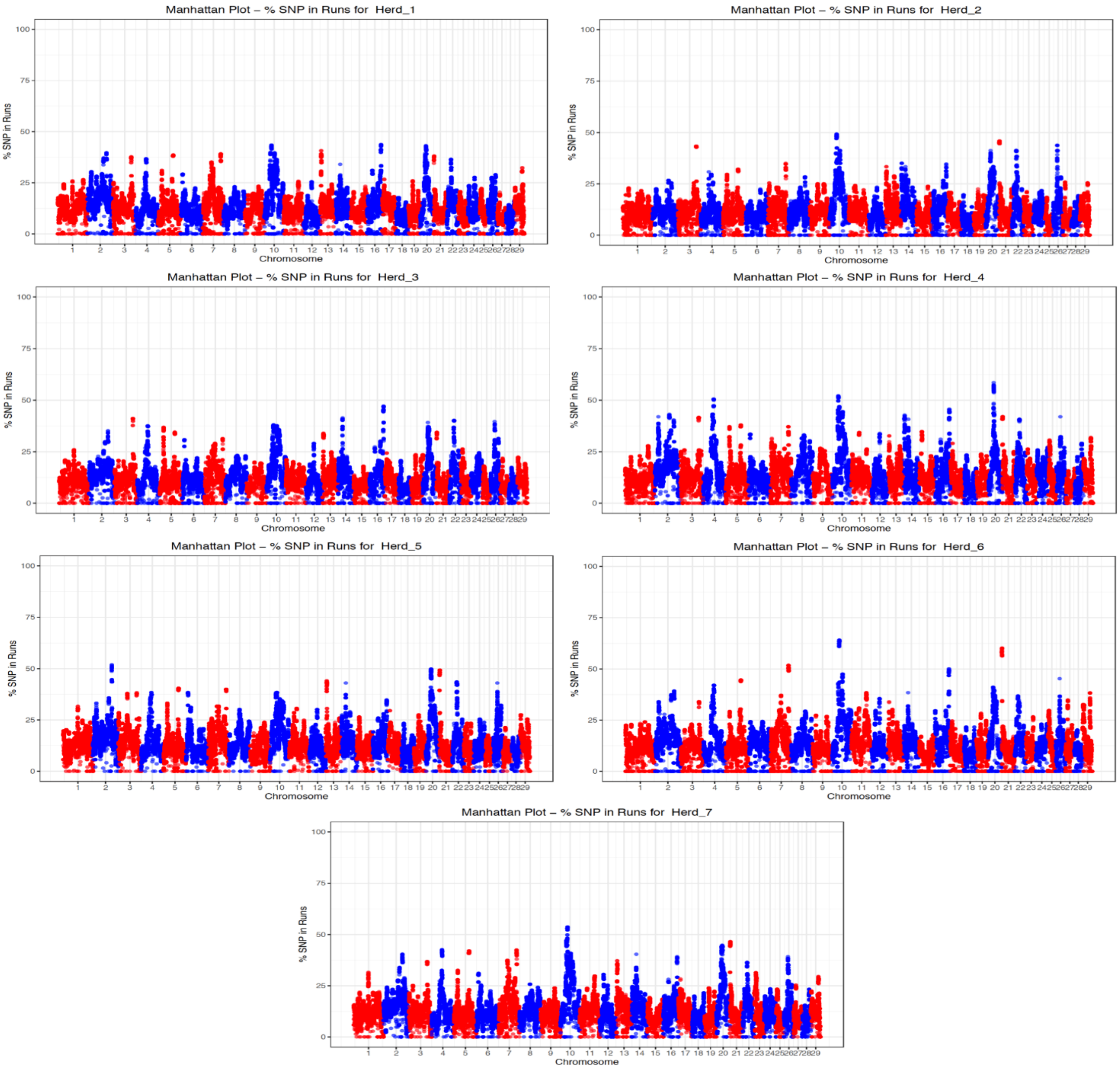
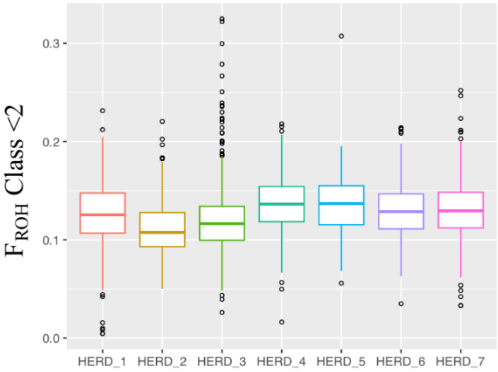


Figure 3. Manhattan plot of the proportion of SNPs in identified Runs of Homozygosity (ROH), along all the autosomes, for all analysed Herds.

Table 7. Descriptive statistics Total F_{ROH} inbreeding coefficient.

Herd_n	**F_{ROH} Min**	**F_{ROH} Max**	**F_{ROH} Mean**	**F_{ROH} SD**
Herd_1	0.004	0.231	0.126	0.032
Herd_2	0.050	0.220	0.113	0.027
Herd_3	0.026	0.325	0.119	0.029
Herd_4	0.016	0.218	0.136	0.029
Herd_5	0.056	0.307	0.135	0.028
Herd_6	0.035	0.214	0.130	0.027
Herd_7	0.033	0.252	0.131	0.030



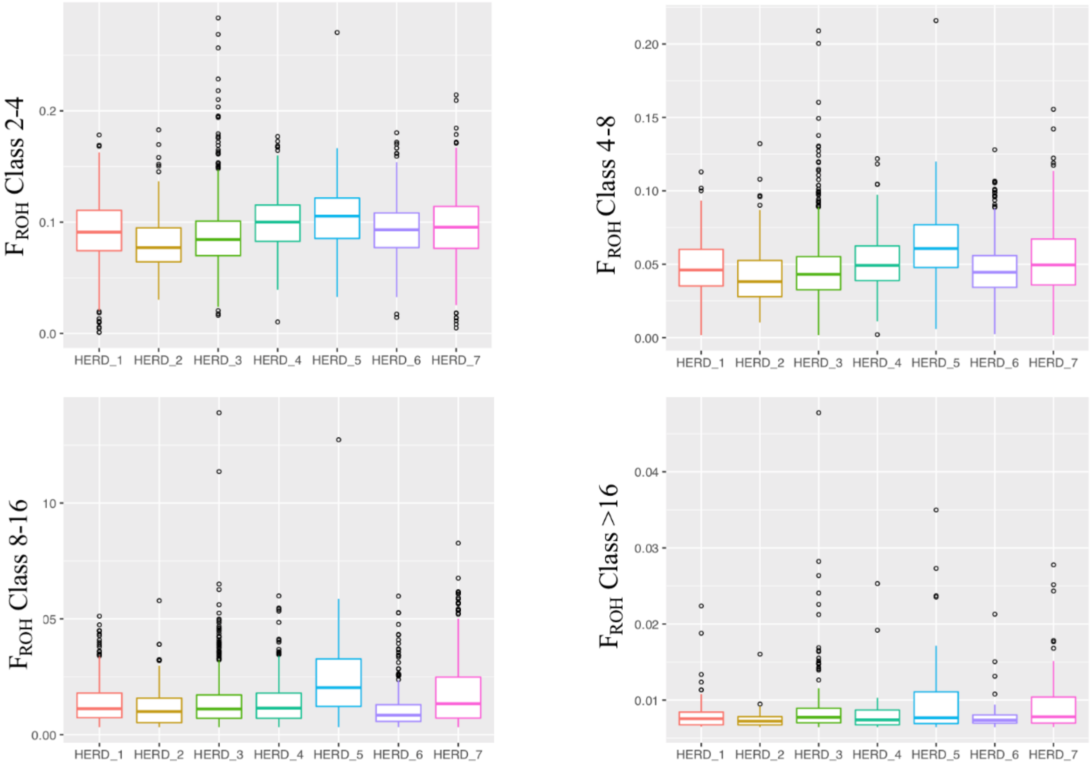


Figure 4. Descriptive statistics of total F_{ROH} (Table 7) and graphical representations (Boxplots) of F_{ROH} calculated in concordance with the five Runs of Homozygosity (ROH) classes of length.

Additional information can be found at https://doi.org/10.1080/1828051X.2023.2243977

3.3 Copy number variant scan in more than four thousand Holstein cows bred in Lombardy, Italy

Andrea Delledonne[1], Chiara Punturiero[1], Carlotta Ferrari[1], Francesca Bernini[1], Raffaella Milanesi[1], Alessandro Bagnato[1], Maria G. Strillacci[1*]

[1]*Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano, Via dell'Università 6, 26900 Lodi, Italy.*

Published in *PloS One* (2024)
https://doi.org/10.1371/journal.pone.0303044

Abstract

Copy Number Variants (CNV) are modifications affecting the genome sequence of DNA, for instance, they can be duplications or deletions of a considerable number of base pairs (i.e., greater than 1000 bp and up to millions of bp). Their impact on the variation of the phenotypic traits has been widely demonstrated. In addition, CNVs are a class of markers useful to identify the genetic biodiversity among populations related to adaptation to the environment.

The aim of this study was to detect CNVs in more than four thousand Holstein cows, using information derived by a genotyping done with the GGP (GeneSeek Genomic Profiler) bovine 100K SNP chip. To detect CNV the SVS 8.9 software was used, then CNV regions (CNVRs) were detected. A total of 123,814 CNVs (4,150 non redundant) were called and aggregated into 1,397 CNVRs. The PCA results obtained using the CNVs information, showed that there is some variability among animals. For many genes annotated within the CNVRs, the role in immune response is well known, as well as their association with important and economic traits object of selection in Holstein, such as milk production and quality, udder conformation and body morphology. Comparison with reference revealed unique CNVRs of the Holstein breed, and others in common with Jersey and Brown.

The information regarding CNVs represents a valuable resource to understand how this class of markers may improve the accuracy in prediction of genomic value, nowadays solely based on SNPs markers.

Introduction

For millennia, humans have established a profound relationship with cattle domesticating them to exploite their resources, obtain food as milk, and meat, and meet various needs [1]. Since the 20th century, the selection to improve production traits in animal species, such as the Holstein cattle breed, represents a fundamental step in the development of modern animal husbandry. The Holstein breed, nowadays worldwide recognized for its milk production, has undergone a strong selection effort aimed at improving milk yield, quality, and in the last two decades in enhancing overall functionality and health [2]. In recent years, the evolution of nanotechnology made available the SNP genotyping platforms that made possible the genomic selection revolution in cattle breeding theorized by Meuwissen et al. [3]. The utilization of SNP chips in genotyping has proven to be a potent tool in animal selection, empowering breeders to make well-informed decisions based on the collective genetic information [4]. SNP genotyping data also enable the detection of Copy Number Variants (CNVs) through the computation of the Log R Ratio (LRR) and B Allele Frequency (BAF). LRR represents a normalized measure of the total signal intensity for two alleles of a SNP, and the B allele frequency (BAF), the one measuring the allelic intensity ratio at marker level [5]. The LRR and BAF facilitate the assessment of CNV status (loss vs gain, LRR; homozygote vs heterozygote, BAF). CNVs represent a category of genomic structural variants recognized to influence phenotypic diversity through the deletion (loss status) or duplication (gain status) of DNA segments, potentially affecting gene structure and regulating expression [6,7]. These variations typically range in size from 1 kilobase (kb) to 5 mega bases (Mb) [8].

The functional impact of CNVs has been studied across various animal species, highlighting their role in influencing a range of phenotypic traits [9–13]. The fact that CNVs affect a multitude of traits across different animal species underlines their role also in adaptive responses to various environmental conditions [14–17]. In several studies on Holstein cattle, CNVRs have been identified to impact economically important traits as milk production, residual feed intake, fertility and somatic cell score [18–21].

Although CNVRs cover a small part of bovine genome length (about 2-10%), as reported by [22], these structural variants can be integrated with SNP information in genomic prediction, offering new insights to explain complex traits and understand the proportion of missing heritability not explained by SNP.

Thus, taking into account all information related to Copy Number Variations (CNVs), the objectives of this study were to examine a substantial population comprising 4,282 Holstein cows from seven distinct farms in Italy, with the purpose of map CNVs across the autosomal genome. Additionally, within the more frequent CNVRs, the goal encompassed the annotation of genes and of quantitative trait loci (QTL) associated with relevant traits in this breed. To validate our findings, we conducted a comparative analysis both within and across different cattle breeds, drawing on insights from prior research studies.

Materials and methods

Animal sampling, genotyping and ethics statement

All cows of 7 herds of the Lombardy region were genotyped with the Illumina GGP Bovine 100K (GeneSeek®) from 2019 to 2023 for a total of 4,282 individuals. These 7 herds are representative of the possible farming systems and selection objectives of Holstien farmers: they in fact spans from a small family run farm (110 cows in lactation) with historically low selection, to a large farm with Automatic Milking System and with more than 3 decades of directional selection to improve production and functionality (about 550 lactating cows) and a medium size farm producing Parmigiano Reggiano cheese and thus, requiring specific nutritional practices (no silage) and selection for milk quality. Log R Ratio (LRR) available from the SNP chip processing were used to map CNVs. The quality assessment of LRR and the mapping of CNVs was performed with the Golden Helix Inc. SVS 8.9 software (SVS).

The sampling of individual was approved by the OPBA (i.e., Animal Welfare Organisation) of the University of Milan (Protocol number 160_2019), by Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010, updating Directive 86/609/EEC on the protection of animals used for scientific purposes.

Quality control of genotyping data

The quality assessment of LRR values was performed considering the Derivative Log Ratio Spread (DLRS) as described by Pinto et al. [23] and the GC Wave Factor (GCWF) [24], both affecting signal intensity and possible cause of bias in CNVs mapping. A total of 47 samples were excluded due to their high DLRS values, while other 135 samples were excluded because of the elevated GCWF values. The detection of CNVs was then conducted on a dataset of 4,100 samples.

CNVs and CNVRs detection

CNVs detection was obtained on autosomes with SNPs mapped on the ARS UCD1.2 assembly reference genome. The detection was performed using the Copy Number Analysis Module (CNAM) of SVS by means of the univariate analysis based on LRR values. Default parameters for CNVs calling in CNAM were set as follows: i) a maximum of 100 segments per 10,000 markers; ii) a minimum of 3 markers per segment; iii) 2000 permutations per pair with a p-value cut-off of 0.005.

To identify animals with outliers CNVs frequencies and length, their distributions were analysed using QQ plots (R routine in ggplot2 library [25]). Outliers were identified as samples having CNV length greater than 7.5 Mbp. After the identification and exclusion of the individuals considered outliers (3,809 subjects were left), the individual frequency of gain and loss in relation to each sample mean CNVs length was plotted with the ggplot2 library of R.

Using the Bedtools -mergeBed command [26], CNVs that overlapped by at least one bp and were shared by a minimum of two animals were combined to generate CNV regions (CNVRs). Then, CNVRs were classified as gain, loss, or complex if comprising both deletions (loss) and duplications (gain). A CNV found in a single individual was classified as a singleton CNVR.

To be representative, only CNVRs shared by at least 2% of the population were selected for descriptive statistics as well as for downstream analyses.

The R package HandyCNV [27] was used to visualized the physical distribution of CNVRs on autosomes.

Genes and QTL annotations

The genes list with official “gene name ID” was downloaded from NCBI online Database. Genes were then annotated within the detected CNVRs using the Bedtools “-intersectBed” command [26], while the QTL associated with the genes found in the CNVRs were identified thanks to the cattle QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/BT/search) by gene name, using the “Search by associated gene” option of QTLdb.

The Cytoscape plugin ClueGo was used to identify potential biological connections among candidate genes identified in the CNVRs [28,29]. The network construction relied on information from GO and KEGG database. This analysis utilized the bovine databases integrated into the ClueGO app. Only connections with a p-value lower than 0.05 were considered.

Diversity at the population level

To study the diversity within the breed we recoded CNVs defining a CNVR for each cow as follows: i) '1' for loss state; ii) '0' for normal state; iii) '2' for gain state. We used the Past 4.03 software to perform a principal component analysis (PCA).

Comparison with results from the Literature

Our identified CNVRs were compared with the results reported in recent literature studies using the HandyCNV library of R-Studio software (compare_cnvr() function).

As reported in Table 3, two distinct comparisons were performed in order to validate Holstein specific CNVRs (comparison within breed), and to identify genomic regions shared by different breeds (comparison among breeds), i.e. Jersey (JER) and Brown Swiss (BSW). For studies with CNVRs using a different genome assembly from ARS-UCD1.2, the positions were remapped using the UCSC Lift Genome Annotations tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). A graphical visualization of overlapped CNVRs was realized through a Venn diagram built using an online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Results

CNVs and CNVRs detections

According to the number of CNVs per cow and their total length (sum of each CNV length), 291 samples were identified as outliers and subsequently removed to avoid the introduction of possible false positive CNVs; the final dataset comprising 123,814 CNVs was obtained in 3,809 cows (S1 Table); with a total of 4,150 non-redundant CNVs.

As reported in Table 1, CNVs have a maximum, minimum, and average length of 1,860,579, 1,005 and 86,166 bp, respectively. The frequency of loss CNVs doubles the frequency of gain CNVs and the mean length of losses (90,439.4) is longer than the mean length of gains (77,785.5).

Fig 1A shows the different distribution of gain and loss CNVs according to the relationship between the CNV mean length and their number per samples. Furthermore, as shown in Fig 1B, the majority of CNVs falls into the first three classes of length. Over 30,000 loss state CNVs exhibited a length below 0.05, falling in the first length class. Conversely, the majority of gain CNVs had a length ranging between 0.05 and 1 Mb. The longest CNVs were low representated for both of CNV states.

The 123,814 CNVs were aggregated into 1,397 CNVRs (Table 2 and S2 Table), covering 9.18% (228 Mbp) of the total autosomal length (2,489 Mbp). After removing singletons and CNVRs shared by less than 2% of the population, 267 CNVRs remained (CNVRs_2% in Table 2 and S2 Table): 76 in gain state, 129 in loss and 62 categorized as complex. CNVs in CNVR_2% are listed in the S2 Table. These latter CNVRs cover 2.92% of the autosomal genome length and their physical distribution on autosomes is shown according to their states in Fig 2. Values (%) on this graph represent the genomic proportion covered by CNVRs with respect to each chromosome length. CNVRs on chromosomes 12, 18 and 23 covered more than 5% of chromosomal length, 9.5%, 7.4% and 5.1% respectively, while all other chromosomes were impacted by a lower proportion of CNVRs. The CNVRs shared by the largest number of cows were on BTA 10 at 22,676,353 bp (n. 3,528 cows, loss) and on BTA 2 at 93,926,090 (n. 3,107 cows, loss). Instead, CNVRs shared by the lowest number of cows, i.e. 76 animals, were found in gain state within chromosome 20 (at 66,818,777 bp).

S1 Fig shows the genome-wide distribution of the 267 CNVRs across the chromosomes together with the mean CNVRs coverage length. The maximum number of CNVRs are on BTA 1 and BTA 9. The mean CNVRs length is not uniform along all chromosomes, and the maximum mean CNVR length was on BTA 12 (717,015.8 bp).

Principal component analysis results (Fig 3A and 3B) depict the genetic variability in the 3,809 cows analyzed, according to the presence or absence of CNVs in the identified CNVRs, considering their state. Each point in the scatter plots represents an individual animal, coloured as unique population (Fig 3A) or taking into account the herd from which it was sampled (Fig 3B).

Gene content and annotation

A total of 996 genes were annotated within 194 Holstein CNVRs (72.6% of the CNVRS_2%).Their functional classification, according to the David database, is reported in the S3 Table (recognized gene IDs = 942).

In S2 Fig (ClueGo network) it's possible to observe the presence of five macro-groups of genes associated with the following categories: troponin complex, sensory perception of smell, nervous system process, tuberculosis, and MHC class II protein complex. The KEGG pathway comprising the majority of genes is the one connected to tuberculosis, the same result has been obtained with David analysis.

After consulting the Cattle QTLdb, 142 genes were associated with a total of 122 different “Trait Name”, grouped into 24 “Trait Types” corresponding to 6 “Trait Classes” (Exterior, Healthy, Meat and Carcass, Milk, Production, and Reproduction Traits), in concordance with the database nomenclature (Fig 4). As Fig. 4 shows, the most of traits associated with the genes annotated in the CNVRs are related to the phenotypes for which the Holstein population has been selected for years.

Comparison with references

CNVRs here identified were compared with those identified in three other Holstein populations (comparison within breed) and in two different breeds (comparison among breed; one dairy cattle – Jersey; one dual-purpose ¬cattle – Brown Swiss) (Table 3 and Fig 5 and S2 Fig and S4 Table). As reported in Table 3, the minimum and the maximum number of overlapping regions were 7 and 27, respectively.

ehe 48 CNVRs resulted overlapping regions (S4 Table) included 32 regions identified in others Holstein samples, i.e. CNVRs mapped in at least two studies (shared_HOL) as shown in Fig 5A. When the comparison was performed with the JER and BSW cattle, the 32 shared_HOL regions in Fig 5B, resulted in 11 Holstein proprietary CNVRs and 4 ones found in all breeds. As in Fig 5B, the BSW breed shared the largest number of overlapping regions CNVRs. The total overlapping CNVR length was similar for those studies in which CNVs were identified with the same software (< 2 Mb – PennCNV and > 10 Mb – SVS, Table 3).

Discussion

In the literature there are several studies investigating genetic variability of Holstein’ population using SNPs, and to increase knowledge on this breed, a large set of Italian Holstein cows has been inhere analyzed through CNVs detection. CNVs, a class of structural variation, can inform about population variability and are known to occur in the genome in response to environmental stressors, including positive selection, as a consequence of farming strategies [33].

This study, based on a medium density SNP chip, i.e. the Illumina GGP Bovine 100K, allowed the identification of a high number of CNVs in a substantial number of Holstein cows. The number of CNVs per sample (32, on average), is relatively higher compared to studies that rely on non-dense SNP chips, but lower compared to studies that rely on dense SNP chips or use sequences to call CNVs [34–36]. As reported in the majority of CNV mapping studies performed with Illumina SNP chips, the number of deletions calls was approximately 1.98 more recurrent than duplications [21,31,37]. The mean length of deletion calls inhere (90,439.4 bp) is bigger than the mean length of found gains (77,785.5 bp). Interestingly Lee et al. [31], using the Illumina BovineHD BeadChip, found that duplications are longer than deletions.

Overlapping CNVs resulted in 1,397 CNVRs covering 9.18% of the cattle genome. This value is much higher than the ones reported in the literature for Holsteins, which range from 0.5% to 2.8% [31,38], but in line with the coverage found by Butty et al. [30], depending on the density of the SNP chip and the detection algorithm used [30,39]. When CNV regions shared by at least 2% of the population were selected, the percentage of genome covered by CNVRs decreased (2.9% of the autosomal genome length, Fig 2), a value similar to those reported by other authors [31,38].

As shown in S1 Fig, CNVRs are not uniformly distributed on the autosomes, and the distribution of CNVRs according to their length class (Fig 1C) shows that the majority are short to medium in length and only a few are observed in the long classes, consistently with previous findings [31].

To visualize the genomic variability related to CNVs detected in our study population, we performed a Principal Component Analysis and the results in Fig 3A, at first glance, show that all animals are spread in the graph without any clustering tendency.

The homogeneous grouping in this study appears to be related to the fact that all the cows, although bred on different farms, undergone similar intensive farming system. Nevertheess, the genetic selection performed by the farmers seems to produce an effect: when the grouping animals by herd (Fig. 3B) a slight clustering can be observed, mainly for animals in Herd_6 (magenta colour). In Herd_6, mating plans have been based on bulls from a unique AI center for years, while all other herds use sires from different semen providers [40]. When the gain/loss ratio was calculated in each herd to explain our findings, it was equal to 0.40 in Herd_6 (this value corresponds to a loss/gain ratio = 2.40) and up to 0.49 in all the others herds (maximum value was 0.70 in Herd_5; loss/gain ratio = 1.41). The lower proportion of gain CNVs found in Herd_6 may be linked to the highest number of daughters for sire in Herd_6, with a reduction of variability in specific genomic regions. The lower number of common bulls across all herds (as reported by Punturiero et al. [40]) can explain the cows' distribution of Herd_6 respect to the ones belong to all other farms. In Herd_5, the number of daughters per sire is one of the lower.

Gene content and annotation

According to the David database (S3 Table), the genes annotated within the CNVRs were classified in 91 Go-Terms. The KEGG pathway analysis revealed that among the genes under analysis 56 are mainly represented in the pathway of immune system, namely, in the classes "Tuberculosis" and "Staphylococcus aureus infection", and in the pathway of thermogenesis. Disease resistance (or susceptibility) is a complex trait and interestingly it could be affected by genomic variations, as found by different authors reporting a substantial immune gene enhancement within CNV regions [21,41–43].

The network constructed with ClueGO (S2 Fig) aligns with the results found with the David analysis. It's possible to see genes connected to different GO categories linked to nervous system, troponin complex, sensory perception of smell, nervous system process, together with the KEGG category of susceptibility to tuberclosis. Some genes are connected with more than one category, for example BOLA genes.

Variation in gene copy number leads to phenotypic variation among animals. After consulting the AnimalQTLdb for cattle we grouped the QTLs in 24 trait types. As listed in the S3 Table and showed in Fig 4, the most common trait type is milk composition, for which 102 QTL were found. This result is in line with the expectations, being the animals part of commercial farms that sell milk for the dairy industry. Milk composition, together with udder conformation, fertility, and growth (more representative trait types in Fig 4) are all object traits of selection for high-productive breeds, such as the Holstein.

Noteworthy CNVRs and comparison with references

Nine CNVRs resulted over-represented due to a high number (> 2,000) of CNV defining these regions: 4 CNVRs do not harbor genes, and most of them are in loss state. The only duplication region is the cnvr_234 identified on BTA 25 (in 2,400 cows) (S2 Table). In this CNVR, map the EEF2K and POLR3E genes that are involved in the cellular response to oxidative stress [44] and the host innate immune defense against viruses [45], respectively. Even for the genes mapped in the cnvr_024 (3,107 cows) on BTA 2 (PARD3B, NRP2) a role in immune response was reported [46,47]. Finally, the cnvr_069 located on BTA 7 (2,156 cows) overlaps the CNVR20 (complex state) identified by [30]. This region harbors five genes belonging to the family 2 of olfactory receptor genes (OR). CNVs are frequently found within OR genes and this variability may contribute to individual or breed-specific differences in olfactory capacity [48], which is also associated with feed intake and efficiency [49]. This aligns with the findings in our research; indeed, conducting gene ontology analysis with ClueGO (S2 Fig) yielded results for 35 genes in a copy number variation state linked with the following functional categories: sensory perception of smell, detection of stimulus involved in sensory perception, detection of chemical stimulus involved in sensory perception, olfactory perception activity and sensory perception of chemical stimulus. Nonetheless, these results only contribute to a small portion of our understanding given the size and complexity of this gene family comprising more than 1,000 known OR genes.

Regarding the comparison with references, as reported in Table 4, among the 267 CNVRs, 11 overlapped with the ones identified only in Holstein populations and 4 in all the considered breeds (Holstein, Jersey, and Brown). It is important to note that the size of the CNVRs identified in this study decrease after comparison (we reported only regions perfectly overlapping). This is particular evident for cnvr_225, splitted in two small regions as listed in Table 4. The entire cnvr_225 harbour genes belonging to the BOLA family, a well known gene implicated in host immune response. In the cnvr_133, located on BTA 13 (both in loss and complex states, according to breeds, see S4 Table), lied the SIRPB1 gene, also involved in the immune response [50].

Across the identified CNVRs proper of the Holstein cows, a wider variability in the regions state can be observed, more than 70% are in fact in complex state. Only 4 CNVRs harbour genes. Among them, cnvr_137 contains genes such as LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, TSNARE1, and ARC associated to clinical mastitis in US Holstein dairy cows [51]. While the cnvr_245 includes the BNIP3 gene, that plays a critical role in inducting autophagy during heat stress and was associated with the immune response phenotype [52]. The same region partially overlaps the CNVR_1549_P (the region comprising the JAKMIP3, DPYSL4, STK32C, LRRC27, PWWP2B) resulted associated with clinical mastitis in Mexican Holstein Cattle [21].

Conclusions

The study provides novel insights into CNVs mapped within the Italian Holstein cows. To date, this is the only study that conducted a CNV analysis on such a large number of animals within this breed. Based on CNVs, the Principal Component Analysis (PCA) revealed a homogeneous distribution of cows, indicating a shared effect of the intensive farming system on these animals. The slight clustering observed among cows from the same farm implies that genetic selection may influence CNV distribution, underscoring the potential impact of selective breeding practices.

The functional analysis of genes annotated in the more common CNVRs revealed biological mechanism related to immune resistance to infection and adaptability. QTL linked with the main traits object of directional selection overlapped with many CNVRs here identified. Genes involved in immune response and defense against oxidative stress were identified within CNVRs, suggesting that genetic variability could affect the animals' ability to respond to environmental stressors.

The analysis of CNVs not only provides an additional dimension of genetic information, but also represents a valuable resource to optimise (new prespective) genomic selection in a more complete and accurate way

Acknowledgments

The authors gratefully acknowledge the farmers.

References

[1]Frantz LAF, Bradley DG, Larson G, Orlando L. Animal domestication in the era of ancient genomics. Nat Rev Genet 2020;21:449–60.

[2]Egger-Danner C, Cole JB, Pryce JE, Gengler N, Heringstad B, Bradley A, et al. Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal 2015;9:191–207.

[3]Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics 2001;157:1819–29. https://doi.org/10.1093/GENETICS/157.4.1819.

[4]Wiggans GR, Cole JB, Hubbard SM, Sonstegard TS. Genomic selection in dairy cattle: the USDA experience. Annu Rev Anim Biosci 2017;5:309–27.

[5]Wang K, Bucan M. Copy number variation detection via high-density SNP genotyping. Cold Spring Harb Protoc 2008;2008:pdb-top46.

[6]Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazlsy C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315:848–53. https://doi.org/10.1126/SCIENCE.1136678.

[7]Margareto J, Leis O, Larrarte E, Pomposo IC, Garibi JM, Lafuente JV. DNA copy number variation and gene expression analyses reveal the implication of specific oncogenes and genes in GBM. Cancer Invest 2009;27:541–8. https://doi.org/10.1080/07357900802563044.

[8]Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. Nature 2011;470:59–65. https://doi.org/10.1038/nature09708.

[9]Wright D, Boije H, Meadows JRS, Bed'Hom B, Gourichon D, Vieaud A, et al. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS Genet 2009;5:e1000512.

[10] Olsson M, Meadows JRS, Truve K, Pielberg GR, Puppo F, Mauceli E, et al. A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. PLoS Genet 2011;7:e1001332.

[11] Kropatsch R, Dekomien G, Akkad DA, Gerding WM, Petrasch-Parwez E, Young ND, et al. SOX9 duplication linked to intersex in deer. PLoS One 2013;8:e73734.

[12] Venhoranta H, Pausch H, Wysocki M, Szczerbal I, Hänninen R, Taponen J, et al. Ectopic KIT copy number variation underlies impaired migration of primordial germ cells associated with gonadal hypoplasia in cattle (Bos taurus). PLoS One 2013;8:e75659.

[13] Awasthi Mishra N, Drögemüller C, Jagannathan V, Keller I, Wüthrich D, Bruggmann R, et al. A structural variant in the 5'-flanking region of the TWIST2 gene affects melanocyte development in belted cattle. PLoS One 2017;12:e0180170.

[14] Pierce MD, Dzama K, Muchadeyi FC. Genetic diversity of seven cattle breeds inferred using copy number variations. Front Genet 2018;9:163.

[15] Xu L, Hou Y, Bickhart DM, Zhou Y, Hay EHA, Song J, et al. Population-genetic properties of differentiated copy number variations in cattle. Sci Rep 2016;6:23161.

[16] Arendt M, Cairns KM, Ballard JWO, Savolainen P, Axelsson E. Diet adaptation in dog reflects spread of prehistoric agriculture. Heredity (Edinb) 2016;117:301–6. https://doi.org/10.1038/hdy.2016.48.

[17] Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet 2007;39:1256–60. https://doi.org/10.1038/ng2123.

[18] Hou Y, Bickhart DM, Chung H, Hutchison JL, Norman HD, Connor EE, et al. Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. Funct Integr Genomics 2012;12:717–23.

[19] Glick G, Shirak A, Seroussi E, Zeron Y, Ezra E, Weller JI, et al. Fine Mapping of a QTL for Fertility on BTA7 and Its Association With a CNV in the Israeli Holsteins. G3: Genes| Genomes| Genetics 2011;1:65–74.

[20] Xu L, Cole JB, Bickhart DM, Hou Y, Song J, VanRaden PM, et al. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics 2014;15:1–10.

[21] Durán Aguilar M, Román Ponce SI, Ruiz López FJ, González Padilla E, Vásquez Peláez CG, Bagnato A, et al. Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. J Anim Breed Genet 2017;134:49–59. https://doi.org/10.1111/JBG.12238.

[22] Hay EHA, Utsunomiya YT, Xu L, Zhou Y, Neves HHR, Carvalheiro R, et al. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. BMC Genomics 2018;19:1–8.

[23] Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat Biotechnol 2011;29:512–20. https://doi.org/10.1038/nbt.1852.

[24] Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res 2008;36:e126–e126. https://doi.org/10.1093/nar/gkn556.

[25] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: 2016.

[26] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–842. https://doi.org/10.1093/bioinformatics/btq033.

[27] Zhou J, Liu L, Lopdell TJ, Garrick DJ, Shi Y. HandyCNV: Standardized Summary, Annotation, Comparison, and Visualization of Copy Number Variant, Copy Number Variation Region, and Runs of Homozygosity. Front Genet 2021;12:731355. https://doi.org/10.3389/FGENE.2021.731355/BIBTEX.

[28] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

[29] Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009;25:1091–3.

[30] Butty AM, Chud TCS, Miglior F, Schenkel FS, Kommadath A, Krivushin K, et al. High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data. Scientific Reports 2020 10:1 2020;10:1–13. https://doi.org/10.1038/s41598-020-64680-3.

[31] Lee Y-L, Bosse M, Mullaart E, Groenen MAM, Veerkamp RF, Bouwman AC. Functional and population genetic features of copy number variations in two dairy cattle populations. BMC Genomics 2020;21:89. https://doi.org/10.1186/s12864-020-6496-1.

[32] Prinsen RTMM, Rossoni A, Gredler B, Bieber A, Bagnato A, Strillacci MG. A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. Livest Sci 2017;202. https://doi.org/10.1016/j.livsci.2017.05.011.

[33] Strillacci MG, Gorla E, Cozzi MC, Vevey M, Genova F, Scienski K, et al. A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations. PLoS One 2018;13:1–18. https://doi.org/10.1371/journal.pone.0204669.

[34] Ahmad SF, Singh A, Panda S, Malla WA, Kumar A, Dutt T. Genome-wide elucidation of CNV regions and their association with production and reproduction traits in composite Vrindavani cattle. Gene 2022;830:146510.

[35] Sassi N Ben, González-Recio Ó, de Paz-Del Río R, Rodríguez-Ramilo ST, Fernández AI. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. J Dairy Sci 2016;99:6371–80.

[36] Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim E, et al. Genomic characteristics of cattle copy number variations. BMC Genomics 2011;12:1–11.

[37] Prinsen RTMM, Strillacci MG, Schiavini F, Santus E, Rossoni A, Maurer V, et al. A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle. Livest Sci 2016;191. https://doi.org/10.1016/j.livsci.2016.08.006.

[38] Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, et al. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. BMC Genomics 2013;14:1–10. https://doi.org/10.1186/1471-2164-14-131/TABLES/2.

[39] Xu L, Hou Y, Bickhart DM, Song J, Liu GE. Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data. Microarrays 2013;2:171–85.

[40] Punturiero C, Milanesi R, Bernini F, Delledonne A, Bagnato A, Strillacci MG. Genomic approach to manage genetic variability in dairy farms. Ital J Anim Sci 2023;22:769–83. https://doi.org/10.1080/1828051X.2023.2243977.

[41] Suchocki T, Szyda J. Genome-wide association study for semen production traits in Holstein-Friesian bulls. J Dairy Sci 2015;98:5774–80. https://doi.org/10.3168/jds.2014-8951.

[42] Szyda J, Mielczarek M, Frąszczak M, Minozzi G, Williams JL, Wojdak-Maksymiec K. The genetic background of clinical mastitis in Holstein-Friesian cattle. Animal 2019;13:2156–63.

[43] Lee Y-L, Takeda H, Costa Monteiro Moreira G, Karim L, Mullaart E, Coppieters W, et al. A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle. PLoS Genet 2021;17:e1009331.

[44] Sanchez M, Lin Y, Yang C-C, McQuary P, Campos AR, Blanc PA, et al. Cross talk between eIF2α and eEF2 phosphorylation pathways optimizes translational arrest in response to oxidative stress. IScience 2019;20:466–80.

[45] Ramanathan A, Weintraub M, Orlovetskie N, Serruya R, Mani D, Marcu O, et al. A mutation in POLR3E impairs antiviral immune response and RNA polymerase III. Proceedings of the National Academy of Sciences 2020;117:22113–21.

[46] Schramek H, Sarközi R, Lauterberg C, Kronbichler A, Pirklbauer M, Albrecht R, et al. Neuropilin-1 and neuropilin-2 are differentially expressed in human proteinuric nephropathies and cytokine-stimulated proximal tubular cells. Laboratory Investigation 2009;89:1304–16.

[47] Raphaka K, Matika O, Sánchez-Molano E, Mrode R, Coffey MP, Riggio V, et al. Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle. BMC Genet 2017;18:1–10.

[48] Hasin Y, Olender T, Khen M, Gonzaga-Jauregui C, Kim PM, Urban AE, et al. High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. PLoS Genet 2008;4:e1000249.

[49] Connor EE, Zhou Y, Liu GE. The essence of appetite: does olfactory receptor variation play a role? J Anim Sci 2018;96:1551–8.

[50] Van Beek EM, Cochrane F, Barclay AN, van den Berg TK. Signal regulatory proteins in the immune system. The Journal of Immunology 2005;175:7781–7.

[51] Tiezzi F, Parker-Gaddis KL, Cole JB, Clay JS, Maltecca C. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS One 2015;10:e0114919.

[52] Livernois AM, Mallard BA, Cartwright SL, Cánovas A. Heat stress and immune response phenotype affect DNA methylation in blood mononuclear cells from Holstein dairy cows. Sci Rep 2021;11:11371.

Tables

Table 1. Descriptive statistics of identified CNVs.

N.	N. CNVs	N. Gain	N. Loss	Loss/Gain	Min-Max CNV per ID (mean)	Min-Max (mean) length*	Min-Max (mean) coverage per ID*
3,809	123,814	41,556	82,258	1.98	13-51 (32.5)	1-1,860 (86)	947-7,483 (2,792)

*Value expressed in Mbp.

Table 2. Descriptive statistics of identified CNVRs.

CNVRs	Tot n. CNVRs	Tot n. Singleton	CNVRs State			CNVRs length		
			Loss	Gain	Complex	Min	Max	Mean
CNVRs	1,397	329	714	513	170	1,005	2,286,232	163,678
CNVRs_2%	267	-	129	76	62	1,716	2,286,232	272,307

Table 3. Comparison with literature.

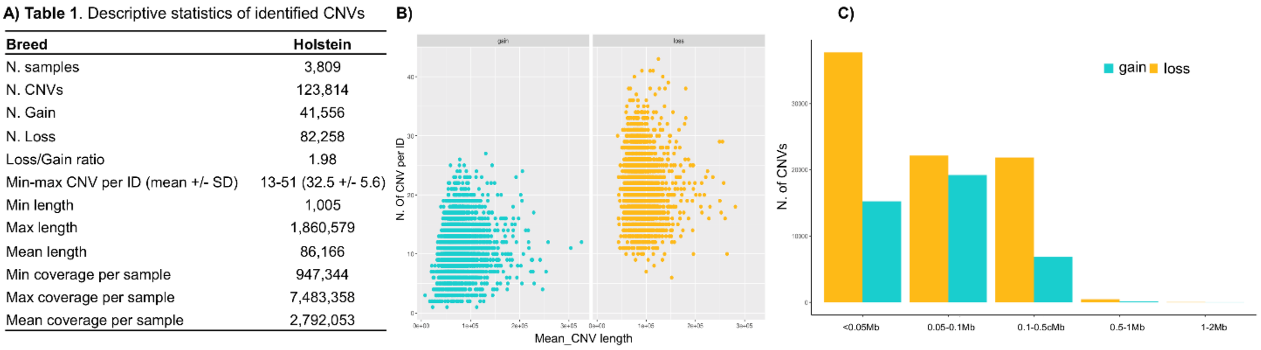
Platform	Software	Breed (N. of IDs)	Reference Genome	N. CNVRs[a]	N. overlapped CNVRs[b]	Overlapping length (bp)	Ref
Comparison within breed							
Illumina HD; 50K; GGP150K	PennCNV	Holstein (96)	ARS-UCD1.2	36	7 (2.6%; 19.4%;)	1,239,370	[30]
Ilumina HD	PennCNV	Holstein (315)	ARS-UCD1.2	135	14 (5.2%; 10.4%)	1,374,082	[31]
Illumina HD	CNAM (SVS)	Holstein (242)	UMD3.1 remapped	112	23 (8.6%; 20.5%)	12,017,083	[21]
Comparison among breeds with different aptitude							
Illumina HD	PennCNV	Jersey (107)	ARS-UCD1.2	142	15 (5.6%; 10.5%)	1,915,749	[31]
Illumina HD	CNAM (SVS)	Brown Swiss (1,116)	UMD3.1 remapped	233	27 (8.6%; 11.6%)	10,651,728	[32]

[a] When remapped, this number refers to the CNVRs resulting after the positions remapping; [b] Proportion of overlapping: calculated as n. overlapped CNVRs/n. CNVR in this study; n. overlapped CNVRs/n. CNVR in other study.

Table 4. CNVRs in common between our study and the ones found in Holstein and in different cattle breeds.

CNVR_ID this study	**Chr**	**Start**	**End**	**State**	**Genes**	**QTL**
Common CNVRs (HOL, JER, BSW)						
cnvr_075	8	76336567	76348332	loss		
cnvr_121	12	71701903	71765886	complex		
cnvr_133	13	53463194	53511604	complex	SIRPB1	SIRPB1: Milk protein percentage (QTL: 174904)
cnvr_225	23	25953514	26064642	complex		
cnvr_225	23	26113327	26350925	complex		
Only HOL CNVRs						
cnvr_034	3	119662571	119718948	loss	COPS9, OTOS	
cnvr_035	4	182210	217902	loss		
cnvr_055	6	37695352	37736960	complex		
cnvr_058	6	51884459	52200066	complex		
cnvr_068	7	17374656	17409367	complex		
cnvr_072	7	81385397	81392696	loss		
cnvr_137	14	1645654	2064157	complex	LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, TSNARE1, ARC, ADGRB1, JRK	LY6D: Milk fat percentage (QTL:33308; 166962; 161706), Milk protein percentage (QTL:161824)
cnvr_176	18	27856333	28303561	complex		
cnvr_181	18	57234258	57254890	complex		
cnvr_216	21	69040055	69788216	complex	C21H14orf180, TMEM179, INF2, ADSSL1, SIVA1, AKT1, ZBTB42, CEP170B, PLD4, AHNAK2, CLBA1, CDCA4, GPR132, JAG2, NUDT14, BRF1, BTBD6, PACS2, TEX22, MTA1, CRIP2, CRIP1, TEDC1, TMEM121	AKT1: Bovine respiratory disease susceptibility (QTL: 160320; 160321); BRF1: Conception rate (QTL: 123998)
cnvr_245	26	50598736	51990348	complex	KNDC1, ADGRA1, CFAP46, NKX6-2, INPP5A, BNIP3, JAKMIP3, DPYSL4, STK32C, LRRC27, PWWP2B	

Figures



A) Table 1. Descriptive statistics of identified CNVs

Breed	Holstein
N. samples	3,809
N. CNVs	123,814
N. Gain	41,556
N. Loss	82,258
Loss/Gain ratio	1.98
Min-max CNV per ID (mean +/- SD)	13-51 (32.5 +/- 5.6)
Min length	1,005
Max length	1,860,579
Mean length	86,166
Min coverage per sample	947,344
Max coverage per sample	7,483,358
Mean coverage per sample	2,792,053

Fig 1. Summary of statistic for detected CNVs. A) Relationship between number and mean total length (bp) of CNVs identified in each sample by state (gain vs loss); B) Number of CNVs for five classes of length.

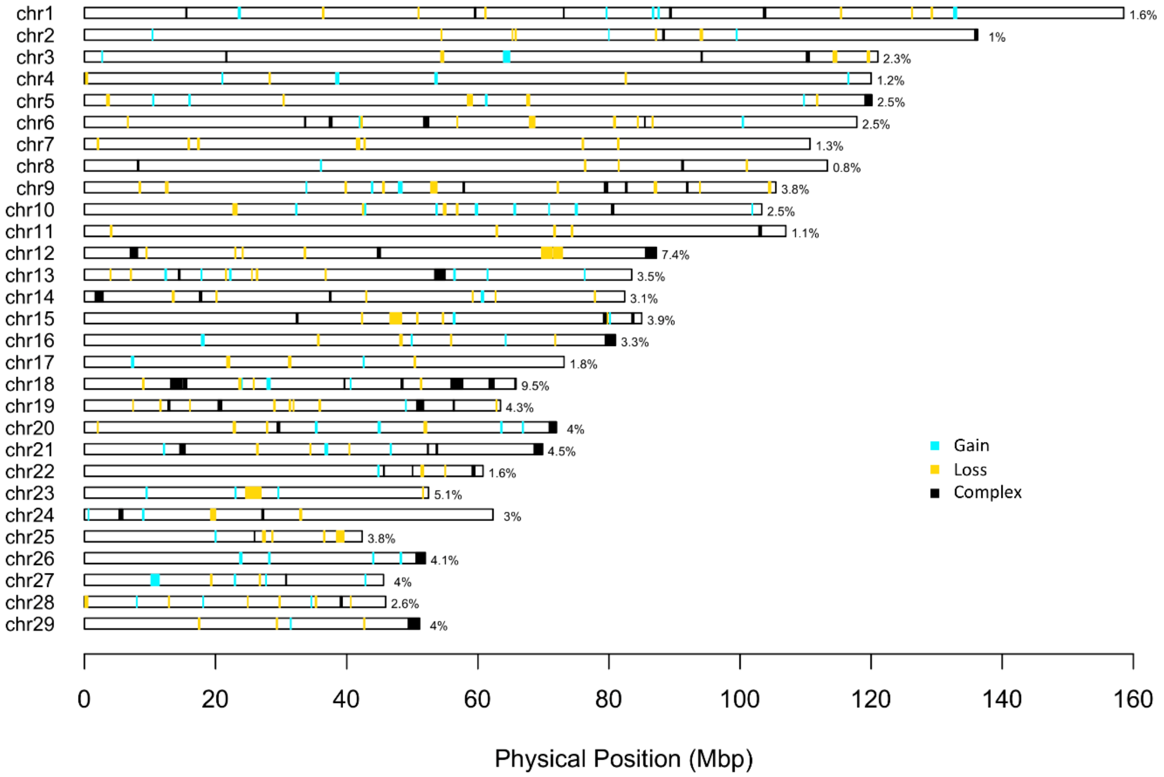


Fig 2. Physical distribution of the Copy Number Variants Regions (CNVRs) according to states (complex, gain and loss) on the *Bos taurus* ARS-UCD 1.2 assembly. Plotted CNVRs are those shared by at least 2% of individuals. Percentage values refer to the genomic proportion covered by CNVRs respect to the BTA length.

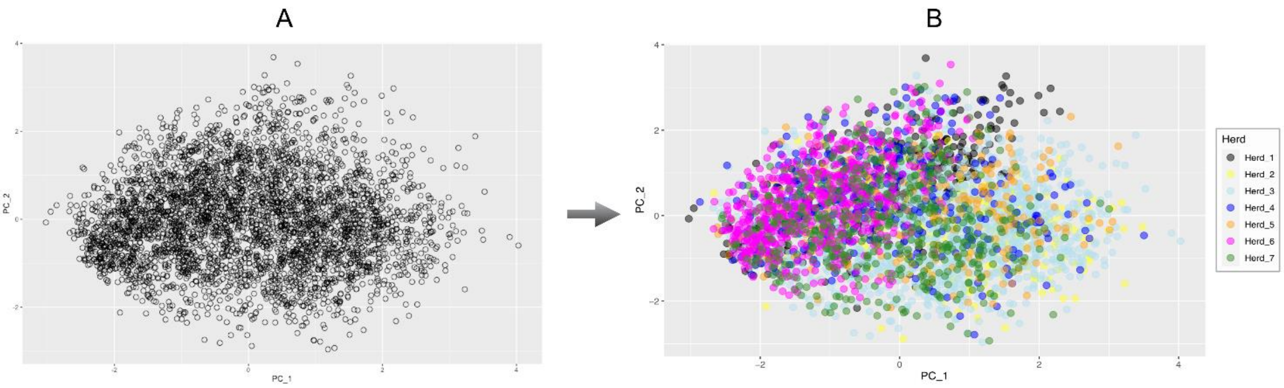


Fig 3. Principal component analysis results. A) Samples are coloured in black as unique Holstein breed; B) Samples are coloured according to the herds in which the cows were sampled.

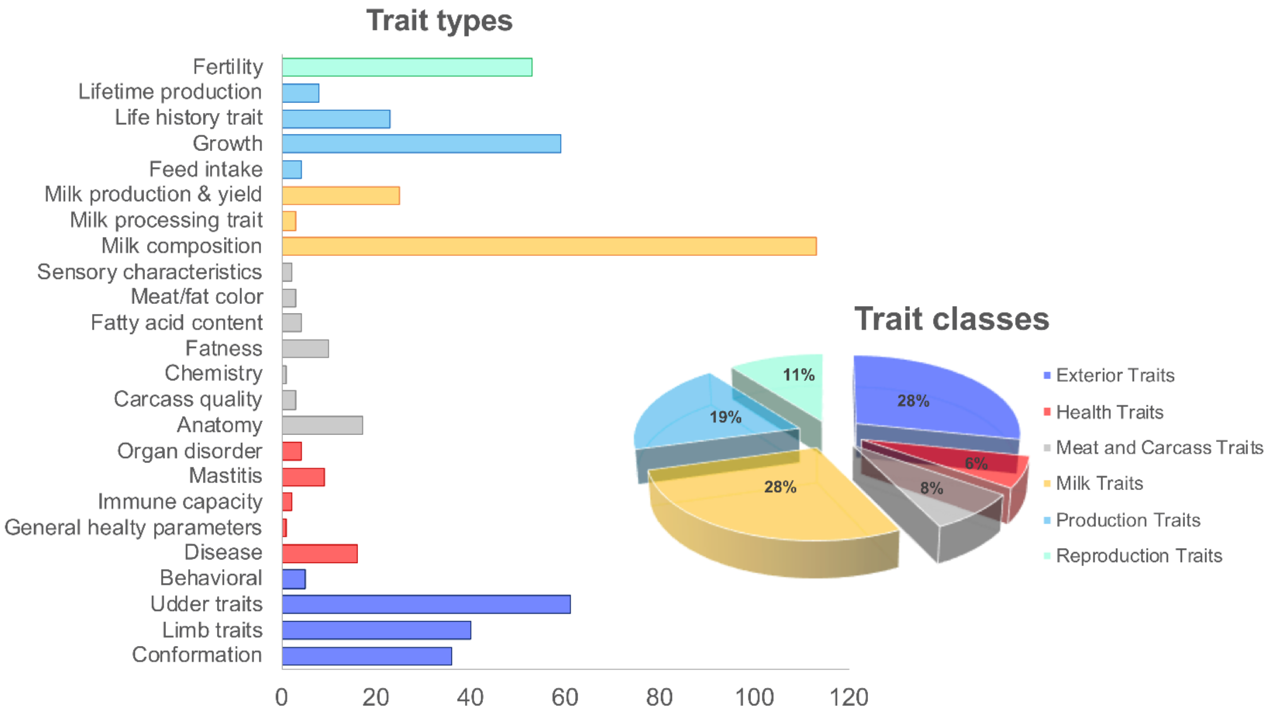


Fig 4. **Graphical representation of QTL_terms (Trait types and Trait classes) associated with genes mapped in CNVRs.** Colours of Trait types corresponded to the ones in Trait classes.

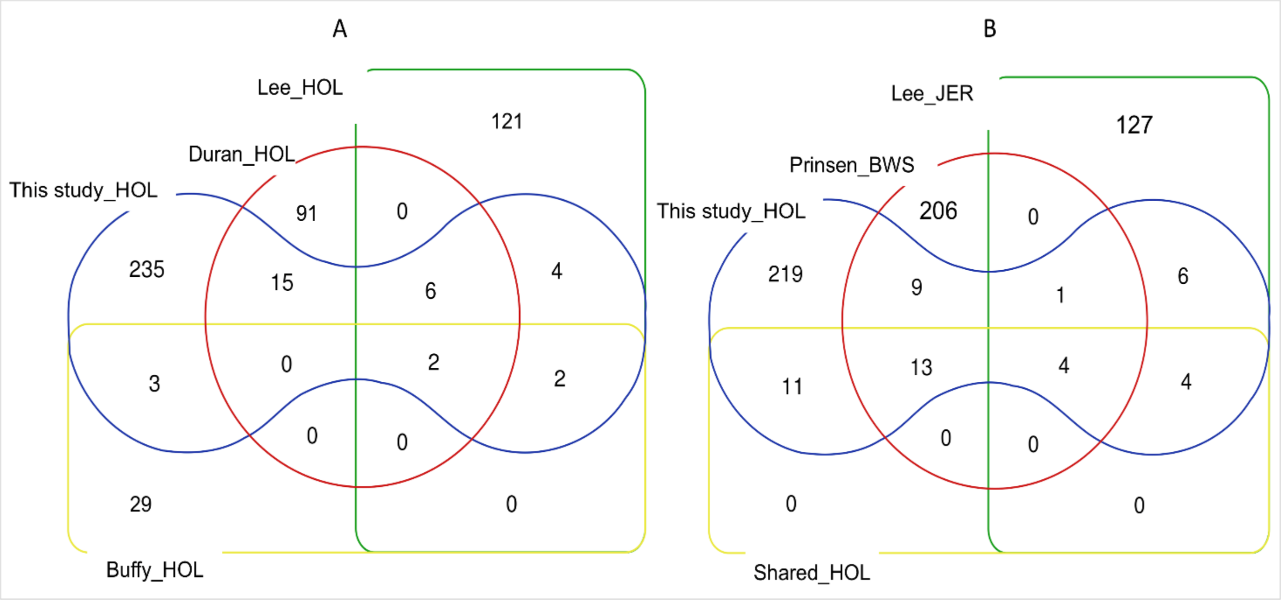


Fig 5. Comparison of CNVRs identified in different Holstein populations (A) and in others two breeds (B). Shared_HOL are those CNVRs (n.32) identified in at least two studies (part A of this Fig).

Supporting information can be found https://doi.org/10.1371/journal.pone.0303044

S1 Table. CNVs identified in Holstein breed.

S2 Table. List of the total CNVR (sheet_1); CNVRs identified in at least 2% of cows (sheet_2), ad list of CNV defining CNVRs identified in at least 2% of cows.

S3 Table. Gene functional annotation from David database.

S4 Table. CNVRs comparison with references.

S1 Fig. Graphical representation of CNVRs number and mean CNVR coverage length on autosomes.

S2 Fig. ClueGo network of genes annotated in CNVRs identified in at least 2% of cows.

3.4 Investigation of heterozygosity rich regions in autochthonous and cosmopolitan cattle breeds

Bernini Francesca[1], Punturiero Chiara[1], Ferrari Carlotta[1], Vevey Mario[2], Blanchet Veruska[2], Bagnato Alessandro[1] and Strillacci M. Giuseppina[1]

[1] *Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano, Via dell'Università 6, 26900 Lodi, Italy;*

[2] *Associazione Nazionale Bovini di Razza Valdostana, Fraz. Favret, 5, 11020 Gressan, Italy*

Ongoing research

Abstract

Heterozygosity-Rich Regions (HRRs) are contiguous genome segments characterized by heterozygous genotypes. The cosmopolitan Holstein breed, specialized in dairy production, is managed under intensive farming systems and has undergone strong directional selection to enhance milk yield. In contrast, the autochthonous cattle breeds of the Aosta Valley, Red Pied, and Black Pied-Chestnut, are raised in extensive farming systems and have experienced adaptive selection to overcome the challenging mountain environment. Holstein and Aosta populations have two significantly different schemes and strategies of selection: the hypothesis is that selection could have acted differently on the formation of HRR in each breed.

For this study, 4,755 Holstein and 4,246 Aosta (Red Pied, Black Pied, and Chestnut) cows have been genotyped with NEOGEN's GGP Bovine 100K SNP chip. Additionally, four cows per breed have been whole genome sequenced. The analysis to identify HRR has been performed using the detectRUNS package for the R software. A total of 79,013 and 47,824 HRR have been identified in the Holsteins and Aosta breeds sampled cows, respectively. Four and five HRR_islands were detected in the Aosta Black Pied – Chestnut and un the Aosta Red Pied breeds respectively of which three (on BTA 2, 20, and 28) were shared. In Holstein HRR_islands were located on BTA 1, 3, 8, and 9, none of them mapping in the same regions where HRR mapped in the Aosta's.

In the detected HRR_islands, genes coding for membrane proteins and receptors, DNA transcription, translation, and methylation factors, microtubule organization factors, and RNA polymerase activators (e.g. TERT, REEP3, ZRANB2, and CEP72) were annotated. These results show evident genomic differences between Holstein and Aosta breeds that can be hypothesised due to the different selection.

Introduction

Differently from ROH (uninterrupted long runs of homozygous genotypes), ROHet cannot similarly be defined as long stretches of heterozygous genotypes, but as Heterozygosity Rich Region (and thus better identified by the acronym HRR; Ferenčaković et al., 2016; Marras et al., 2018). Again, differently from ROH, that have been studied for many years and whose precise biological origin has been clarified (Ceballos et al., 2018; Curik et al., 2014; Peripolli et al., 2017), a defined biological meaning of HRR is still missing.

Recently high-density SNP chip genotyping data, available in some livestock populations, have been used to map HRR. Williams et al. introduced the concept of HRR in 2016, where particularly heterozygosity-dense regions have been identified in the Chillingham breed genome. In this breed, SNP loci have been seen to tend to cluster into ROHet/HRR and a possible balancing selection has been suggested (Williams et al., 2016). When most of the population shares those regions, they can be defined as HRR_islands (Selli et al., 2021). Apart from this feature, other parameters that should be used to define HRR are still missing.

Recent research on HRR aimed to investigate their potential association with high genetic variability and balancing selection (Ruan et al., 2022; Selli et al., 2021; Szmatoła et al., 2023, Falchi et al., 2024). Authors carrying on studies on HRR are proposing to use them to identify regions where genes under balancing selection possibly map.

Balancing selection is a type of selection that acts in two different ways:

- The heterozygote advantage. It happens when the heterozygous genotype generates a higher fitness than the two homozygous genotypes. A classic example is the “malaria hypothesis” where the β-globin gene, which in homozygous state causes sickle-cell anaemia in humans, but is maintained in heterozygous form to confer resistance to malaria (Charlesworth & Langley, 2013)
- The frequency-dependent selection. It happens when in different generations one homozygote is advantageous compared to the other and vice versa. One example can be the scale eating fish. In these fishes, the mouth is on a different side in different generations. This mechanism allows the maintenance of an equal number of subjects with the mouth in both directions in the population (Hori, 1993).

In the present study, the cattle breeds analysed are the local Aosta cattle breeds, represented by two sub-breeds: the Aosta Red Pied (ARP) and the Aosta Black Pied-Chestnut with the subgroup of Chestnut's subjects that have an Hérens ascendant (ABCH), and the Holstein (HOL) breed. It is well known that the cosmopolitan and specialized Holstein cattle breed, raised in intensive farming systems, has been strongly directionally selected, to increase milk yield. On the contrary, the two autochthonous Aosta breeds are under a low intensive selection scheme and being raised in an extensive farming system are most likely have been subjected also to an adaptive selection, to cope with the harsh mountain environmental conditions. In this study the hypothesis is that the two characterizing farming environments, i.e., the summer pasture and the intensive farming, could have acted differently on the shaping of the HRR in the analysed breeds.

For these reasons, the objective of the present study was to identify HRR_islands in local and cosmopolitan cattle breeds and to further investigate their biological meaning using Whole Genome Sequencing (WGS) data.

Materials and methods

Genotype data were collected from 4,755 HOL of the GENORIP project (Punturiero et al., 2023), and from 1,361 ABCH and 2,885 ARP of the National Breeders Association of the Aosta Breed - ANABORAVA genomic database, adding up to a total of 9,001 animals. All the genotype data have produced using the GGP Bovine 100K (GeneSeek®) at the Neogen service laboratory (UK).To further inspect the biological mining of HRR a preliminary analysis of four Illumina paired-end short reads WGS per breed has been used to perform the Tajima's D test and evaluate the nucleotide diversity. The WGS of ARP and ABCH were available from the DUALBREEDIG project, the fasq files were produced and aligned by Novogene who also performed the quality control. The WGS of HOL (samples: SAMEA10833765, SAMEA5415483, SAMEA10833766, and SAMEA10833769) were mined out of the ENA database.

Both genotype data and WGS were mapped to the ARS-UCD1.2 reference genome.

Genotype data analysis

Principal component analysis (PCA) has been carried out with plink on the filtered data (CR>0.95 and MAF>0.05) and the results have been plotted with ggplot2 using R studio. The effective population size has been obtained using the LD based method implemented in the SNeP software (Barbato et al., 2015) after pruning the dataset with plink using the sliding window of 50 SNPs method with a step of 5 SNPs and an r^2 of 0.5. Using the Runs of Homozygosity detection parameters (Bernini et al., 2023) the identified ROH were used to obtain the F_{ROH} inbreeding coefficient, based on the total ROH length compared to the total genome length. The consecutive HRR have been identified with the dectectRUNS R package (Biscarini et al., 2018) using a minimum length of $5x10^5$ and more than 20 SNP to define an HRR, 2 opposite and 1 missing genotype allowed, and a maximum gap of 10^6 bp, according to Mulim et al. (2022). After the identification of the HRR, the HRR_islands were defined by the top 1% of SNP appearing in an HRR. The gene annotation of the regions identified as HRR_island was done using the ARS-UCD1.2 reference map, release 108.

WGS variant calling and analysis

Aosta cattle breeds WGS alignments were performed by Novogene (UK), while Holsteins' sequences were downloaded from the ENA database. For both breeds the variant calling was performed using bcftools (Danecek et al., 2021) mpileup command with the options "-C 50 -d 1000 -a AD, DP, SP". The called variants were then filtered keeping only the SNPs with a depth of the variate position >4 and a mapping quality >20. On the final dataset, the Tajima's D test was performed using vcftools (Danecek et al., 2011) in windows of 10^4 bp to identify the presence of balancing selection in the HRR_islands. The average Tajima's D in each window was obtained using R studio and to verify if the HRR_islands are associated with a higher Tajima's D value a t-test was used.

Results and Discussion

The PCA, Figure 1, representing 81.76% of the total variability, shows a clear and separated clustering of the three analysed breeds. The estimated average FROH inbreeding coefficient for the analysed breeds were 0.05, 0.06, and 0.13 for the ABCH, ARP, and HOL, respectively. Furthermore, from the effective population size evaluation, the Aosta breeds showed a higher Ne (443 for the ABCH and 301 for the ARP) compared to the Holstein (i.e. 107 at 13 generations ago). This indicates a higher genetic diversity within the two dual-purpose breeds than within the Holstein.

A total of 126,837 HRRs have been identified in the 9,001 subjects with the consecutive method. The highest average number of HRR per individual has been registered in HOL (16.62±4.01) while the two Aosta breeds have a similar average number (11.28±3.62 and 11.25±3.78 in ABCH and ARP, respectively). The average length of the identified HRR ranged from 0.66 to 0.67 Mbp in the Aosta and HOL cattle, respectively. Manhattan plots of the identified HRR for each breed are displayed in Figure 2 (A: ABCH, B: ARP, and C: HOL).

The HRR_islands identified for each breed are reported in Table 1. ABCH showed four HRR_islands (on BTA2, 20, 23, and 28), while ARP had five (on BTA 2, 3, 20, and 28), three of which were detected also in ABCH.

The first shared region, on BTA2, harbours the GPR39 - G protein-coupled receptor 39 gene that has a hypothesised fundamental role in gastrointestinal and metabolic function in bovines (Yamamoto et al., 2009). In ABCH this region is ~250kbp longer and includes the solute carrier family 35 member F5 gene (*SLC35F5*).

In the second shared region, on BTA20, and longer in ARP, harbours *LPCAT1* - lysophosphatidylcholine acyltransferase 1 gene, Solute carrier family 6 member 3 (*SLC6A3*), the *CLPTMI* like gene and *TERT* - telomerase reverse transcriptase gene. The transgenic expression of the *TERT* gene extends telomere length and enhances telomerase activity in somatic and embryonic bovine cells (Garrels et al., 2012). This gene also enhanced the bovine blastocysts' development competence (Xu et al., 2021). On the same chromosome in a private ARP large HRR_island *SLC6A18*, *SLC6A19, SLC12A7, NKD2, TRIP13, BRD9, ZDHHC11, TPPP, CEP72* genes are annotated. Out of these genes is worth mentioning *NKD2* - NKD inhibitor of WNT signaling pathway 2, *TRIP13* - thyroid hormone receptor interactor 13, TPPP tubulin polymerization promoting protein and *CEP72* - centrosomal protein 72.

The third and last shared region maps on BTA28 and harbours two genes: *NRBF2* - nuclear receptor binding factor 2 and the *JMJD1C* - jumonji domain containing 1C. The latter's expression is important for oocytes and early embryo development in bovines (Li et al., 2015). Furthermore, an additional annotated gene, *REEP3* - receptor accessory protein 3, maps in this BTA28 region, into HRR_island 4, private of ABCH.

In HOL breed a total of four HRR islands have been identified with a 19% threshold representing the top 1% of the SNPs in a HRR distribution.

In the first HRR_island region, 755,293bp long, on BTA1, no genes were annotated.

The second and third were shorter, but a total of three genes in both HRR2 and 3 were identified. Into HRR_2 (BTA3) map *NEGR1* - neuronal growth regulator 1, *ZRANB2* - zinc finger RANBP2-type containing 2, and *PTGER3* - prostaglandin E receptor 3 genes. This latter gene has been shown to undergo alternative splicing in the C-terminal peptides and its isoforms can activate different cascades inducing a change in the receptor activity in mice (Namba et al., 1993). The HOL_HRR_3 includes also three genes: *PRUNE2* - prune homolog 2 with BCH domain, *FOXB2* - forkhead box B2, and *VPS13A* - vacuolar protein sorting 13 homolog A. Out of these three genes, *PRUNE2* gene expression was up-regulated in the liver of Nelore steers that had a lower feed efficiency (Lima et al., 2022).

In the fourth region, HRR_4, the shortest identified in HOL (341477bp, on BTA9) five genes were annotated: *REV3-like* DNA directed polymerase zeta catalytic subunit, *MFSD4B* - major facilitator superfamily domain containing 4B, the SLC16A10 - solute carrier family 16 member 10, RPF2 - ribosome production factor 2 homolog and *GTF3C6* - general transcription factor IIIC subunit 6.

All the identified genes may potentially have played a special role in the survival of the bovine species. For this reason, we wanted to further analyse the identified HRR_islands regions to investigate the possibility of them being kept under balancing selection. To do so we called the variants from WGS data and on the filtered dataset of called variants we performed the Tajima’s D test and calculated the nucleotide diversity (π).

The results for the Tajma’s D test and π cannot be discussed with an interpretation at a population level as for such a discussion a larger number of sequences representing the populations should be used. However, just as a proof of concept, we will consider here the obtained results as if they were a representative sample. On the four WGS per breed the average values of the genome wide Tajma’s D test are very different for each breed, ranging from 0.09 for the ARP to 0.17 for the ABCH and to 0.31 for the HOL. The t-test showed a significant difference between the D values of the HRR_islands with respect to the rest of the genome for each breed. The average D value within each region and the significance level of the t-test for each breed is displayed in the boxplot in Figure 3. The Tajima's D test is a statistical test used to detect deviations from the neutral theory of evolution, which postulates that most mutations are neither advantageous nor disadvantageous and are primarily shaped by genetic drift. The interpretation of this test is based on the value of D where: D < 0 suggests an excess of rare alleles (recent mutations) compared to expectations under neutrality (e.g. population expansion, and purifying selection); while D > 0 suggests an excess of intermediate-frequency alleles (balancing selection, and population contraction; Tajima, 1989).

The results obtained show that on average each breed results in a high variation state but in the HRR_islands the Tajima's D values are significantly higher than the values identified for the rest of the genome. However, further inspecting each region clearly shows that in our sample, not all the HRR_islands have a positive Tajima's D test (Table 2).

Conclusions

The preliminary results of this analysis show that the Aosta cattle breeds do not share HRR_islands with the HOL, while the Aosta breeds have three regions in common on BTA2, 20 and 28. The genes annotated in the HRR_islands encode for membrane proteins and receptors, DNA transcription, translation and methylation factors, microtubule organization factors and RNA polymerase activators.

The fact that in the Aosta breeds an average lower number if HRR per individual has been identified with respect to the HOL population can be explained by the genotyping technology used. In fact, the major limit in HRR detection in autochthonous cattle breeds using commercial DNA SNPchip is that the Chips are optimized for capturing the SNP variability mainly in cosmopolitan breeds. This is an important technical limit, that must be considered, when using these tools for HRR mapping especially for those breeds that have not been included in the building panel of the SNP chip. The alternative and more correct approach should be than to use WGS to correctly identify HRR in different breeds and to understand the reason why genes in these regions are in a heterozygous state. Using WGS available in the Aosta population on a larger sampling will be the next step aimed at refining and improving the identification of such structural genomic regions and exploring their gene variants.

References

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in Genetics, 6, 109. https://doi.org/10.3389/fgene.2015.00109

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). detectRUNS: an R package to detect runs of detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid homozygosity and heterozygosity in diploid genomes genomes. IBBA-CNR, PTP, Università Degli Studi Di Sassari, University of Guelph, 01(01), 1–13. https://cran.r-project.org/package=detectRUNS

Charlesworth, B., & Langley, C. H. (2013). Population genetics. Encyclopedia of Biodiversity, 2nd Edition. Elsevier Science, 182–198.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics, 19(4), 220.

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

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. Gigascience, 10(2), giab008.

Falchi, L., Cesarani, A., Criscione, A., Hidalgo, J., Garcia, A., Mastrangelo, S., & Macciotta, N. P. P. (2024). Effect of genotyping density on the detection of runs of homozygosity and heterozygosity in cattle. Journal of Animal Science, skae147.

Ferenčaković, M., Banadinović, M., Mercvajler, M., Khayat Zadeh, N., Mészáros, G., Cubric-Curik, V., Curik, I., & Söòkner, J. (2016). Mapping of heterozygosity rich regions in Austrian pinzgauer cattle. Acta Agric Slov, 5, S41-4.

Garrels, W., Kues, W. B., Herrmann, D., Holler, S., Baulain, U., & Niemann, H. (2012). Ectopic expression of human telomerase RNA component results in increased telomerase activity and elongated telomeres in bovine blastocysts. Biology of Reproduction, 87(4), 91–95.

Hori, M. (1993). Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. Science, 260(5105), 216–219.

Li, C. H., Gao, Y., Wang, S., Xu, F. F., Dai, L. S., Jiang, H., Yu, X. F., Chen, C. Z., Yuan, B., & Zhang, J. B. (2015). Expression pattern of JMJD1C in oocytes and its impact on early embryonic development. Genet Mol Res, 14(4), 18249–18258.

Lima, A. O., Malheiros, J. M., Afonso, J., Petrini, J., Coutinho, L. L., da Silva Diniz, W. J., Bressani, F. A., Tizioto, P. C., De Oliveira, P. S. N., & Ribeiro, J. A. S. (2022). Prune homolog 2 with BCH domain (PRUNE2) gene expression is associated with feed efficiency-related traits in Nelore steers. Mammalian Genome, 33(4), 629–641.

Marras, G., Wood, B. J., Makanjuola, B., Malchiodi, F., Peeters, K., van As, P., Baes, C. F., & Biscarini, F. (2018). Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (Meleagris gallopavo) population. 11th World Congress on Genetics Applied to Livestock Production, 763.

Mulim, H. A., Brito, L. F., Pinto, L. F. B., Ferraz, J. B. S., Grigoletto, L., Silva, M. R., & Pedrosa, V. B. (2022). Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. BMC Genomics, 23(1), 209.

Namba, T., Sugimoto, Y., Negishi, M., Irie, A., Ushikubi, F., Kakizuka, A., Ito, S., Ichikawa, A., & Narumiya, S. (1993). Alternative splicing of C-terminal tail of prostaglandin E receptor subtype EP3 determines G-protein specificity. Nature, 365(6442), 166–170.

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

Punturiero, C., Milanesi, R., Bernini, F., Delledonne, A., Bagnato, A., & Strillacci, M. G. (2023). Genomic approach to manage genetic variability in dairy farms. Italian Journal of Animal Science, 22(1), 769–783. https://doi.org/10.1080/1828051X.2023.2243977

Radwan, J., Babik, W., Kaufman, J., Lenz, T. L., & Winternitz, J. (2020). Advances in the evolutionary understanding of MHC polymorphism. Trends in Genetics, 36(4), 298–311.

Ruan, D., Yang, J., Zhuang, Z., Ding, R., Huang, J., Quan, J., Gu, T., Hong, L., Zheng, E., & Li, Z. (2022). Assessment of heterozygosity and genome-wide analysis of heterozygosity regions in two Duroc pig populations. Frontiers in Genetics, 12, 812456.

Selli, A., Ventura, R. V, Fonseca, P. A. S., Buzanskas, M. E., Andrietta, L. T., Balieiro, J. C. C., & Brito, L. F. (2021). Detection and visualization of heterozygosity-rich regions and runs of homozygosity in worldwide sheep populations. Animals, 11(9), 2696.

Szmatoła, T., Gurgul, A., Jasielczuk, I., & Ropka-Molik, K. (2023). Comprehensive analysis of runs of homozygosity and heterozygosity of Holstein cattle on the basis of medium and high density SNP panels and large population sample. Annals of Animal Science.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123(3), 585–595.

Williams, J. L., Hall, S. J. G., Del Corvo, M., Ballingall, K. T., Colli, L., Ajmone Marsan, P., & Biscarini, F. (2016). Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. Animal Genetics, 47(1), 19–27.

Xu, L., Idrees, M., Joo, M.-D., Sidrat, T., Wei, Y., Song, S.-H., Lee, K.-L., & Kong, I.-K. (2021). Constitutive expression of TERT enhances β-Klotho expression and improves age-related deterioration in early bovine embryos. International Journal of Molecular Sciences, 22(10), 5327.

Yamamoto, I., Kimura, N., Arai, T., & Tanaka, M. (2009). cDNA cloning and mRNA expression of bovine GPR39. Journal of Veterinary Medical Science, 71(5), 641–644.

Tables

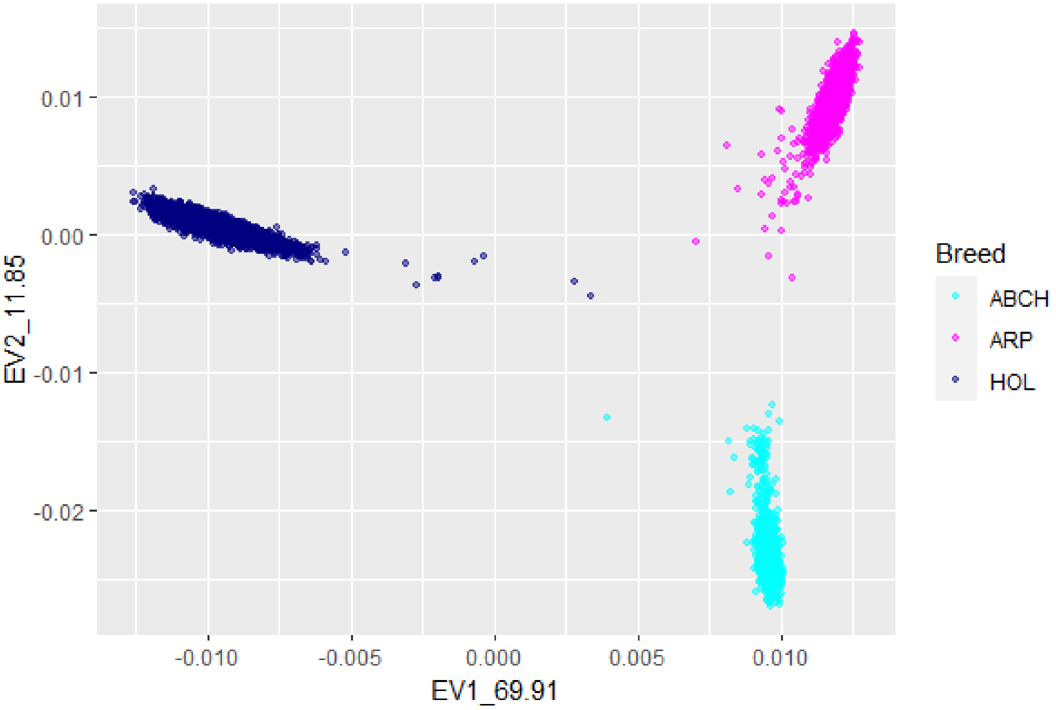
Table 1. List of HRR_islands for each breed, in bold the shared regions, with the annotated genes and QTL.

HRR_island	**Chr**	**From**	**To**	**Lenght_bp**	**Genes**	**QTL**
ABCH						
HRR_1	2	64,977,766	65,496,830	519064	**GPR39**, SLC35F5	
HRR_2	20	70,907,342	71,153,153	245811	**LPCAT1, SLC6A3, CLPTM1L, TERT**	SLC6A3: susceptibility to BRD (QTL:160304)
HRR_3	23	1,020,504	1,889,633	869129		
HRR_4	28	19,213,609	19,677,184	463575	**NRBF2, JMJD1C,** REEP3	REEP3: Residual Feed Intake (QTL:14672)
ARP						
HRR_1	2	65,126,586	65,232,327	105741	**GPR39**	
HRR_2	3	76,049,091	76,569,112	520021		
HRR_3	20	50,081,360	50,660,858	579498		
HRR_4	20	70,883,892	71,502,380	618488	**LPCAT1, SLC6A3, CLPTM1L, TERT,** SLC6A18, SLC6A19, SLC12A7, NKD2, TRIP13, BRD9, ZDHHC11, TPPP, CEP72	SLC6A3: susceptibility to BRD (QTL:160304); CEP72: Longissimus muscle area (QTL:157119)
HRR_5	28	19,213,609	19,582,759	369150	**NRBF2, JMJD1C**	
HOL						
HRR_1	1	102,391,934	103,147,227	755293		
HRR_2	3	73,527,504	74,246,790	719286	NEGR1, ZRANB2, PTGER3	NEGR1: Dairy capacity composite index (QTL:261878)
HRR_3	8	52,932,839	53,453,281	520442	PRUNE2, FOXB2, VPS13A	PRUNE2: body weight (QTL:224433)
HRR_4	9	39,036,234	39,377,711	341477	REV3L, MFSD4B, SLC16A10, RPF2, GTF3C6	

Table 2. Tajima's D test values for each HRR_island and for the regions not in a HRR_island for each breed

Tajima D values		
ABCH_1	2	1.19
ABCH_2	20	0.39
ABCH_3	23	0.72
ABCH_4	28	0.04
NO_HRR	GW	0.17
ARP_1	2	1.00
ARP_2	3	1.14
ARP_3	20	-0.01
ARP_4	28	-0.15
NO_HRR	GW	0.09
HOL_1	1	1.14
HOL_2	3	0.61
HOL_3	8	1.13
HOL_4	9	-1.09
NO_HRR	GW	0.31

Figures



.

Figure 1. PCA representing 81.76% of the total variation

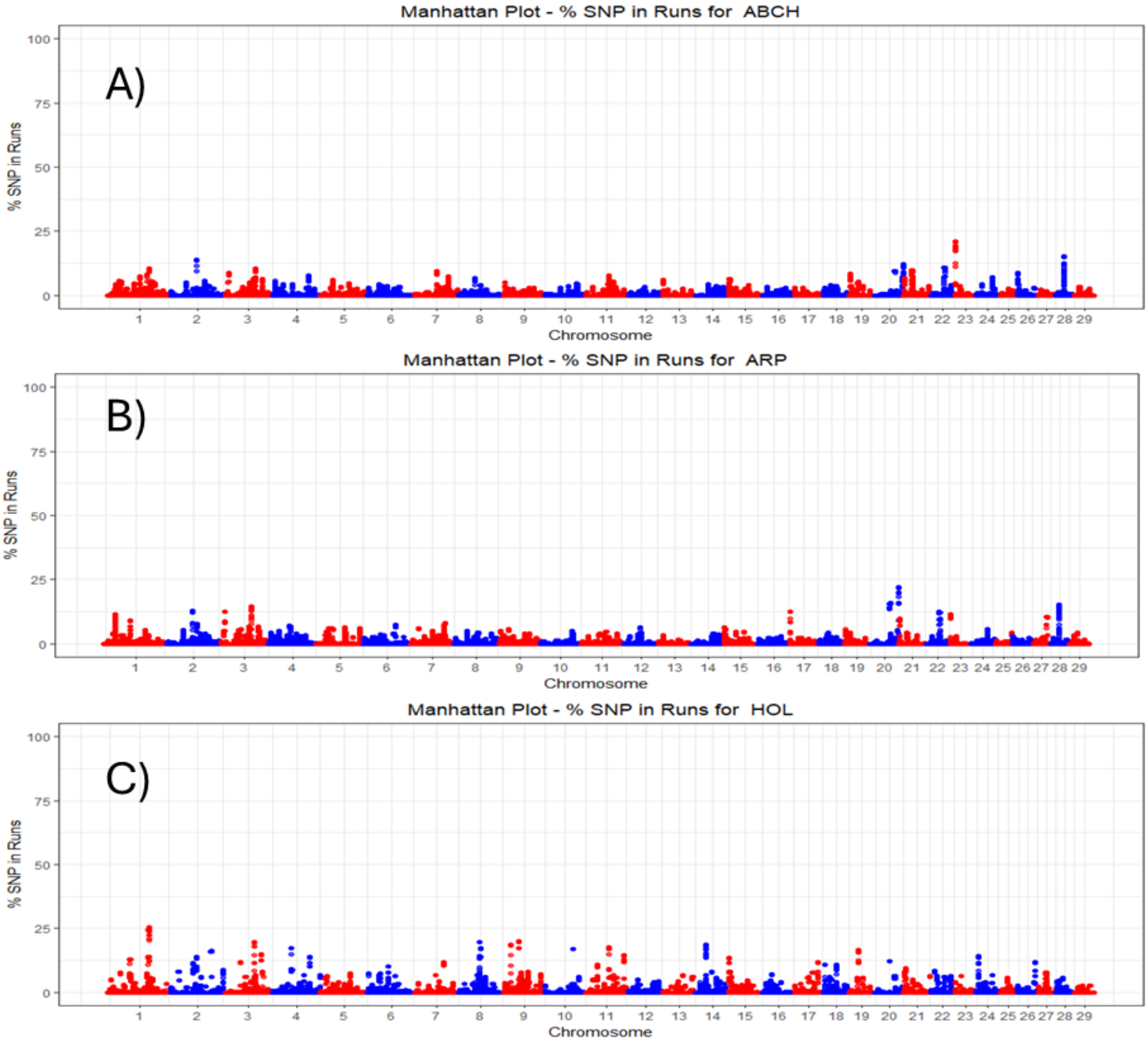


Figure 2. Manhattan plots of the identified HRR for each breed.

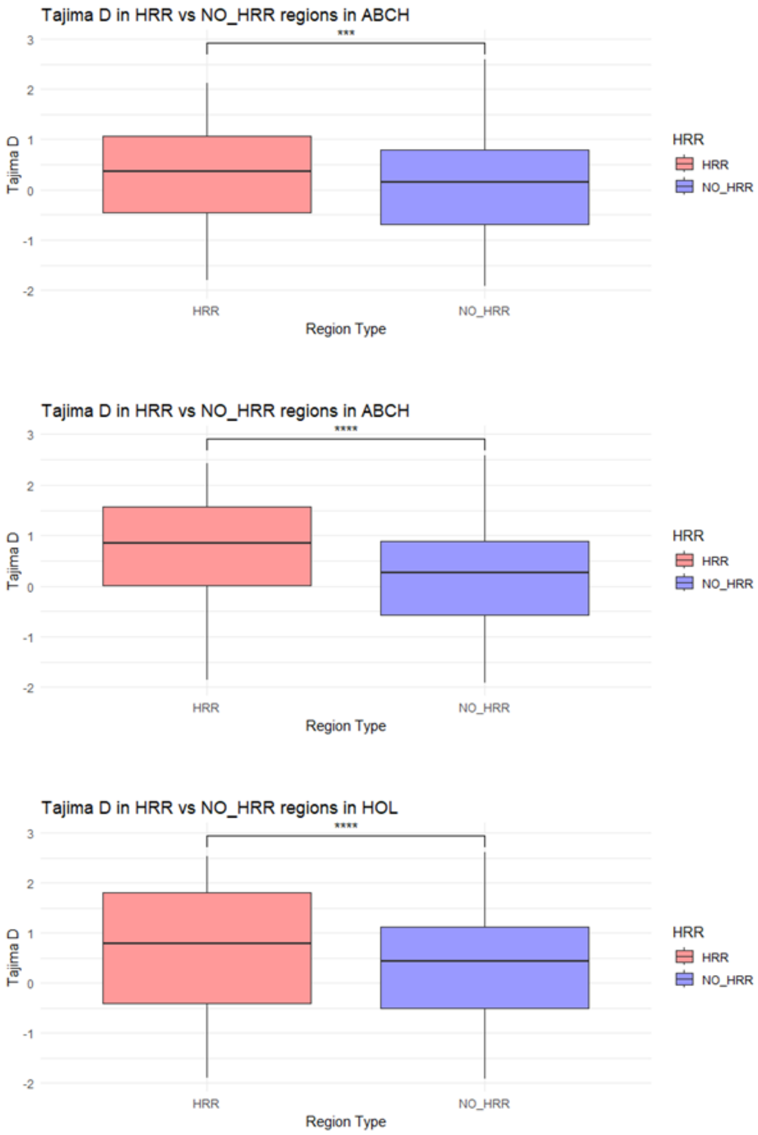


Figure 3. Boxplots representing the average Tajima's D test within HRR_islands. Asterisks indicate the level of statistical significance identified with the t-test.

4. Genome Wide Association Studies

Genome-Wide Association Studies (GWAS) have revolutionized the field of genetics by providing a powerful tool to uncover the genetic basis of complex traits. By systematically scanning the genome for associations between genetic variants and observable phenotypes, GWAS have become indispensable in human, plant, and animal research (Mai et al., 2010; Ogura & Busch, 2015, Pryce et al., 2010; The Wellcome Trust Case Control Consortium, 2007). In livestock, particularly cattle, GWAS had a profound impact on breeding programs by identifying the genomic regions responsible for traits of economic importance (Sahana et al., 2023).

Prior to the development of high-density SNP arrays, genetic association studies were conducted using microsatellite markers (Lipkin et al., 1998). But the advent of SNP chips enabled researchers to examine millions of genetic variants across the genome, marking a major breakthrough. In cattle, this technology has been widely adopted to identify Quantitative Trait Loci (QTL) that influence traits such as milk production, growth, and disease resistance (Fontanesi et al., 2014; Mancin et al., 2021; Olsen et al., 2011).

Beyond the identification of economically important traits, GWAS has also played a pivotal role in uncovering the genetic origins of diseases, not only in humans but also in livestock populations (Charlier et al., 2008; Freebern et al., 2020; Marioni et al., 2018). As animal health and welfare become increasingly important for sustainable livestock production, GWAS helps address challenges related to genetic disorders, leading to improved breeding strategies and the management of genetic diversity.

Despite its many successes, GWAS faces limitations, particularly in studying complex polygenic traits, where multiple genes contribute to the observed variation. Nonetheless, advances in whole-genome sequencing and bioinformatics are continuing to expand the potential of GWAS, ensuring its relevance in modern animal breeding and genetics research.

References

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J. S., Cambisano, N., Carta, E., Dardano, S., Dive, M., & Fasquelle, C. (2008). Highly effective SNP-based association mapping and management of recessive defects in livestock. Nature Genetics, 40(4), 449–454.

Fontanesi, L., Calò, D. G., Galimberti, G., Negrini, R., Marino, R., Nardone, A., Ajmone-Marsan, P., & Russo, V. (2014). A candidate gene association study for nine economically important traits in Italian Holstein cattle. Animal Genetics, 45(4), 576–580.

Freebern, E., Santos, D. J. A., Fang, L., Jiang, J., Parker Gaddis, K. L., Liu, G. E., VanRaden, P. M., Maltecca, C., Cole, J. B., & Ma, L. (2020). GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC Genomics, 21, 1–11.

Lipkin, E., Mosig, M. O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A., & Soller, M. (1998). Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. Genetics, 149(3), 1557–1567.

Mai, M. D., Sahana, G., Christiansen, F. B., & Guldbrandtsen, B. (2010). A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. Journal of Animal Science, 88(11), 3522–3528.

Mancin, E., Tuliozi, B., Pegolo, S., Sartori, C., & Mantovani, R. (2021). Genome Wide Association Study of Beef Traits in Local Alpine Breed Reveals the Diversity of the Pathways Involved and the Role of Time Stratification. Frontiers in Genetics, 12, 746665.

Marioni, R. E., Harris, S. E., Zhang, Q., McRae, A. F., Hagenaars, S. P., Hill, W. D., Davies, G., Ritchie, C. W., Gale, C. R., & Starr, J. M. (2018). GWAS on family history of Alzheimer’s disease. Translational Psychiatry, 8(1), 99.

Ogura, T., & Busch, W. (2015). From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. Current Opinion in Plant Biology, 23, 98–108.

Olsen, H. G., Hayes, B. J., Kent, M. P., Nome, T., Svendsen, M., Larsgard, A. G., & Lien, S. (2011). Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Animal Genetics, 42(5), 466–474.

Pryce, J. E., Bolormaa, S., Chamberlain, A. J., Bowman, P. J., Savin, K., Goddard, M. E., & Hayes, B. J. (2010). A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. Journal of Dairy Science, 93(7), 3331–3345.

Sahana, G., Cai, Z., Sanchez, M. P., Bouwman, A. C., & Boichard, D. (2023). Invited review: Good practices in genome-wide association studies to identify candidate sequence variants in dairy cattle. Journal of Dairy Science.

The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447(7145), 661–678. https://doi.org/10.1038/nature05911

4.1 Genome-wide association studies for milk production traits in two autochthonous Aosta cattle breeds

Francesca Bernini [a,*], Enrico Mancin [b], Cristina Sartori [b], Roberto Mantovani [b], Mario Vevey[c], Veruska Blanchet [c], Alessandro Bagnato [a], Maria Giuseppina Strillacci [a]

[a] *Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano, Via dell'Università 6, 26900, Lodi, Italy*

[b] *Department of Agronomy, Food, Natural Resources, Animals and Environment, Università degli Studi di Padova, Viale dell'Università 16, 35020, Legnaro, Italy*

[c] *Associazione Nazionale Bovini di Razza Valdostana, Frazione Favret 5, 11020, Gressan, Italy*

Published in *Animal* (2024)
https://doi.org/10.1016/j.animal.2024.101322

Abstract

Genome-wide association studies (GWASs) are used to identify quantitative trait loci for phenotypic traits of interest. The use of multi-locus mixed models allows to correct for population stratification and account for long-range linkage disequilibrium. In this study, GWASs were conducted to identify the genetic bases of milk production (milk yield, protein and fat composition, and yield) in two autochthonous dual-purpose cattle breeds from the Aosta Valley. Using either the breeding values or the deregressed proofs, common significative single nucleotide polymorphisms have been identified for milk yield, protein percentage, and fat percentage. Two major quantitative trait loci regions have been identified on the chromosomes 5 and 14 for the fat percentage, harbouring the MGST1, CYHR1, VPS28, and CPSF1 genes. For the protein percentage, a candidate region has been identified on BTA 6; in this region the CSN1S1, CSN2, HSTN, CSN3, and RUFY3 genes are annotated. Most of the identified genes have already been associated with milk composition in other studies on cosmopolitan and local cattle. These results show that the genes involved in milk composition quantitative traits in the Aosta cattle are common also in other cattle breeds and they can be further investigated with the use of whole genome sequencing data.

Keywords: Cattle, Genomics, Productive traits, Local breeds, Quantitative Traits Loci.

Implications

This genome-wide association study is a first step towards the development of genomic selection models for the autochthonous Aosta cattle breeds. Implementing genomic selection in these cattle represents an important step forward in the efficiency of selection, maintenance of their genetic variability, and most importantly, preservation of their hardiness. Despite further analyses are still needed to develop genomic selection for these breeds, this study highlighted that Aosta cattle breeds share quantitative trait loci with other cosmopolitan breeds.

Introduction

Genome-wide association studies have been used for years to identify quantitative trait loci for phenotypic traits of interest in cosmopolitan cattle breeds (Chen et al., 2022; Raven et al., 2014). At present, the increased availability of genotypes collected directly on female cattle, opens new possibilities for genome-wide association studies to disclose quantitative trait loci for low heritability and innovative traits (Pedrosa et al., 2023; Strillacci et al., 2023) and to perform genome-wide association studies on small autochthonous cattle, as already occurred in many local breeds (Korkuć et al., 2021; Mancin et al., 2022). In cattle, genome-wide association studies have been used for decades to identify quantitative trait loci for complex traits (Glantz et al., 2012) and more recently for innovative traits, e.g., heat tolerance, methane emissions, and feed efficiency (Manzanilla-Pech et al., 2021; Nguyen et al., 2017). To date, most genome-wide association studies have been carried out within single breeds, however multi-breed genome-wide association studies may lead to increased power and precision (van den Berg et al., 2016). Multi-breed genome-wide association studies can enhance the statistical power, reduce the likelihood of false-positive associations, and improve the mapping resolution of genetic variants (Bouwman et al., 2018). This approach captures a broader spectrum of genetic diversity, allowing for the identification of both universal and breed-specific genetic markers associated with economically important traits such as milk production.

However, in dairy cattle, only a fraction of individuals in a population are genotyped, and not all genotyped animals possess the phenotypes. Therefore, the most straightforward approach appeared to use Estimated Breeding Value (EBV) as pseudo-phenotypes to include animals with genotypes but no phenotypes in genome-wide association studies. In fact, EBVs are already adjusted for environmental factors and readily available from the routine evaluations.

Nonetheless, due to the random nature of EBVs, they tend to shrink towards zero when their reliability declines. This can lead to spurious associations between Single Nucleotide Polymorphism (SNP) and EBVs, increasing the likelihood of type-I errors (Sahana et al., 2023). In a small, heterogeneous population such as the Aosta breeds, this may result in two main drawbacks: (i) the effects of SNPs would be underestimated due to lower accuracy compared to more cosmopolitan breeds, and (ii) different cohorts of animals might experience varying degrees of reliability values for EBVs, leading to different levels of shrinkage. To address this problem, the most common approach is to deregress the EBV by their accuracy and/or remove redundant information such as parent average effects, obtaining the Deregressed Proofs (DRPs) (Garrick et al., 2009).

Aosta cattle have been recently classified into two breeds: the Aosta Red Pied (ARP) and the Aosta Black Pied - Chestnut with its subgroup of Chestnuts with Heréns ascendant (ABCH); the latter ones are two strains of the same population with different evolutionary history and only recently grouped as a unique breed by the Herd book based on their genomic similarity (Strillacci et al., 2020). These two breeds have different selection programmes: the ARP is selected for meat and milk production, while the ABCH is also selected for combativity. Like many other local cattle, Aosta breeds are particularly important for their own region, not only for their production of milk and meat, but also for the cultural value and the maintenance of the mountain landscapes and environment (Strillacci et al., 2020).

The milk of Aosta cattle is almost entirely used to produce the Protected Designation of Origin Fontina cheese. These products own a disciplinary rule that specifies that the production, processing, and preparation process must take place in a specific geographical location of the production zone. In the case of Fontina cheese, the disciplinary requires that all the milk used for the cheese production is obtained by Aosta cattle, and the production and ageing processes can only take place in the Aosta Valley.

The selection process of Aosta cattle breeds is based on performance testing for meat traits, and sires of cows (young bulls) and sires of bulls (proven bulls) are also tested for milk traits (Pagnacco et al., 1989). The same authors proposed a mating plan to maintain as much as possible the genetic variability in the population while addressing the relaxed selection programmee, which has been adopted in the population for decades. A routine milk data collection is performed every 4-5 weeks by the expert technicians of the Italian Farmers Associations and involves all the farms registered to the National Breeders Association for Aosta cattle breeds, representing the majority of the Aosta cattle breeders. The estimated breeding values for productive traits are then calculated twice per year for the entire population using a repeatability animal model accounting for pedigree information, similar to the one described in the pilot study of Mazza et al. (2016). The implementation of genomic selection would represent an important step forward in making the efficiency of selection for milk and meat traits higher, maintaining the genetic variability, and, even more importantly, keeping the hardiness of the Aosta cattle. Recently, many females have been genotyped with SNP arrays thanks to the funding of the European Agricultural Fund for Rural Development, and this information can now be used to identify quantitative trait loci for traits of interest and to apply the genomic estimation of EBV. The aim of this study was to provide a first insight into the genomic identification of quantitative trait loci for milk production traits, i.e. milk yield - MY, protein yield - PY, fat yield - FY, protein content - PP, and fat content – FP, and annotating genes in the regions identified with SNP markers.

Material and methods

Sampling and genotyping

For this study, 4 247 female genotypes were provided by the National Breeders Association for Aosta cattle breeds. Sample distribution for each Aosta breed was as follows: 1 361 - ABCH and 2 886 - ARP. The initial dataset consisted of 89 762 SNP markers, obtained with the GGP Bovine 100 K SNP chip (GeneSeek®) by Neogen. These SNPs were mapped according to the ARS-UCD1.2 bovine reference genome (GCA_002263795.2), on the bovine autosomes.

The samples used for the analysis had a call rate higher than 95%. The SNPs were subject to quality control, and only the ones with (i) call rate ≥ 0.95, (ii) minor allele frequency > 0.01, and (iii) Hardy-Weinberg equilibrium $P > 1e\text{-}6$ were kept. The pruned dataset was composed of 78 194 SNPs.

Statistical analysis for obtaining the estimated breeding values and deregression

Variance components for MY, PY, FY, PP and FP were obtained using the repeatability test day model currently used in routine genetic evaluations for productive traits. Variances were estimated on datasets of 42 716 individuals with records and 65 081 in the pedigree for ARP, and 22 799 with records and 35 914 in the pedigree for ABCH provided by the National Breeders Association for Aosta cattle breeds. The model included the fixed effects of herd, lactation number, gestation class, age at parity class within lactation, and month at parity class within lactation. The last two effects were covaried by the days in milk expressed as third-order Legendre polynomials. The herd-test day, the permanent environment, and the additive genetic component were included as random effects. The analyses were run under a Bayesian framework using a Gibbs sampling algorithm implemented within the software GIBBS3f90 of the BLUPF90 software family (Misztal et al., 2014). The EBVs were thus obtained from the analysis as individual additive genetic values. For 2 228 animals (521 ABCH and 1 707 ARP) part of the initial dataset, the DRPs were calculated according to Garrick et al. (2009). Only animals with reliability greater than 0.25 and only informative animals (with a reliability higher than the sum of the reliability of the two parents divided by four) were kept.

Genome-wide association analysis

The Mixed Model GWA analysis was performed with the software SNP & Variation Suite v8.9.1 by Golden Helix ®, and the data of ARP and ABCH were used together. The Efficient Mixed Model Association eXpedited - EMMAX algorithm was used considering the additive genetic model. The identity-by-state matrix was included to account for the relatedness of the subjects sampled, and the breed was considered as fixed effect (Kang et al., 2010). After the analysis, the False Discovery Rate and Bonferroni correction thresholds were set at 5% genome-wide to correct for multiple testing.

Gene annotation

For all the significant SNPs, over the 5% false discovery rate threshold, the rsID was assigned based on the SNP position using the Ensembl Database (McLaren et al., 2016). Once the rsID was obtained, the gene annotation was performed using the Variant Effect Predictor - VEP tool by Ensembl (Hubbard et al., 2002).

Results

The additive genetic variance and heritability of target traits (Table 1) are greater in ABCH than in ARP for all the traits considered. Target traits showed a moderate heritability, ranging from 0.12 and 0.13 or FY and FP in ARP, to 0.28 for MY in ABCH. Also, the average EBV and DRP values and SD were different between the two breeds (Table 1). In particular, the variability of MY, FY, and PY was smaller in the ARP with respect to the ABCH for both EBV and DRP. The results of the genome-wide association studies are shown in the Manhattan plots of Fig. 1 for the EBV and DRP of MY, PP, and FP. Significant SNPs associated to quantitative trait loci were found in several chromosomes and are reported in Tables 2-4 with their genomic classification and gene annotation.

When considering the Bonferroni threshold, a total of (i) 22 and 14 SNPs were significantly associated with EBV_FP and DRP_FP, respectively (Table 2), (ii) 21 and 30 SNPs were significantly associated with EBV_PP and DRP_PP, respectively (Table 3), and (iii) only two SNPs were significantly associated with EBV_MY (Table 4). For FY and PY, no significant SNPs have been found, either using the EBV or DRP, probably due to a lower heritability compared to the FP and PP. The Manhattan plots for these two traits are reported in Supplementary Figure S1.

Discussion

The genetic parameters in Table 1 show a greater genetic variability for ABCH compared to the ARP. This may be explained as this group, recently arranged but administratively considered as the same breed, includes both the Aosta Black Pied and the Aosta Chestnut strains (Strillacci et al., 2020). Heritability values in ARP were previously estimated by Mazza et al. (2016) and in the ABCH by Sartori et al. (2020). The ARP values of 0.198, 0.132, and 0.169 were respectively estimated for MY, FY, and PY, while for the same traits, they were 0.227, 0.129 and 0.167 in the ABCH.

Fat percentage trait

Six quantitative trait loci regions were significantly associated with FP. The region identified on chromosome 5, at about 93 Mbp, is defined by 21 SNPs, with rs211210569 and rs210744919 being the rsID numbers with the highest significance (Fig. 1, Table 2). This region harbours the microsomal glutathione S-transferase 1 - MGST1 gene, which is involved in glutathione transport (GO:0034635), cellular detoxification processes (GO:0098869), and cellular response to lipid hydroperoxide (GO:0071449), an oxygenated product of polyunsaturated fatty acids, suggesting a potential role in lipid metabolism (Jayawardana et al., 2023). MGST1 is a quantitative trait loci in numerous association studies performed on different breeds (e.g., Holstein, Braunvieh, Fleckvieh, Montbéliarde, and Normande) exploring the genetic basis of milk composition, especially for fat content (Sanchez et al., 2017; Tribout et al., 2020). Research in this field has aimed to decipher the complex genetic factors influencing milk composition, considering the economic and nutritional importance of milk and dairy products. Understanding the genetic determinants, including the potential contribution of MGST1, could have implications for livestock breeding programmes and the dairy industry, ultimately influencing milk quality and its nutritional value.

Recently, Korkuć et al. (2023) conducted multiple genome-wide association studies on Whole Genome Sequencing data for milk production traits in German Black Pied cattle. They identified significative SNPs for FP in the MGST1 gene, speculating that this gene might contribute to milk fat via the regulation of energy and/or fatty acids to produce milk fat in the mammary gland. Cruz et al. 2019 identified two important regions for milk fatty acids groups, one on chromosome 5 harbouring the MGST1 gene, that was confirmed also when fitting the effect of the diacylglycerol O-acyltransferase 1 - DGAT1 gene, and one on the chromosome 14 on the cysteine and histidine-rich 1- CYHR1 gene, which was also identified in our study. Integrating genome-wide and RNA sequencing information, Littlejohn et al. (2016), suggested a role for MGST1 as a detoxification enzyme whose impact on milk lipid synthesis or secretion is still unknown.

The quantitative trait loci region identified on chromosome 14 (Table 2) harbours eight genes, previously reported in association with milk traits. In particular, the CYHR1 gene has been associated with milk fat yield and content in numerous studies considering different breeds (H. R. Oliveira et al., 2019; Pedrosa et al., 2021). Nevertheless, its role is still unclear due to the possible effect of the nearby DGAT1 gene that is well known to be a locus affecting milk fat content in cattle (Kühn et al., 2004). In our study, the linkage disequilibrium squared correlation statistics suggests no linkage between each of the eight genes reported in Table 2 and the DGAT1, as shown in Supplementary Figure S2.

A meta-analysis on Holstein cattle from different countries (i.e., Australia, Canada, China, France, Germany, Ireland, and Italy) (Bakhshalizadeh et al., 2021) reported two significant SNPs for FP also found in our study, the rs17870736 and rs134432442. They are mapping within the VPS28 subunit of ESCRT-I (VPS28) and the cleavage and polyadenylation specific factor 1 (CPSF1) genes respectively, and in linkage with one SNP (rs109968515, about 2.5 Kbp apart from our significant rs137727465) that annotated close to the CYHR1 gene.

The rs134432442 SNP is a missense variant (ACC/ATC codon) causing a change of the Threonine (amino acid position: 403) with an Isoleucine. Other previous studies found on chromosome 14 the same significant quantitative trait loci region harbouring the CYHR1 and VPS28 genes to explain the variability of fat content and fatty acids in milk in the studied populations (Iung et al., 2019; Jiang et al., 2019). On the same chromosome, the present study found other interesting SNPs, as follows:

i. three SNPs (identified only for EBV) were annotated in intron positions of the Maestro Heat Like Repeat Family Member 1 (MROH1) gene that has been already associated with different milk traits (Iung et al., 2019; Jiang et al., 2019; Tribout et al., 2020) ;

ii. two SNPs (rs41256919 and rs110929299) that map in the MAF1 homolog, a negative regulator of RNA polymerase III (MAF1), a gene that has been associated with all five milk production traits and milk cholesterol content (Jiang et al., 2019; D. Wang et al., 2019);

iii. the rs136792973 SNP, annotated in the glycosylphosphatidylinositol anchor attachment 1 – (GPAA1) gene, which was already associated with protein yield (Pedrosa et al., 2021) and milk production (Raschia et al., 2020). However, Massender et al. (2023) found it to be associated with FY and FP in Canadian dairy goats using 305-day lactation milk production records as phenotypes.

iv. three SNPs located in the Plectin (PLEC) gene, which showed associations with milk fat percentage in numerous studies (Su et al., 2023; D. Wang et al., 2019; P. Wang et al., 2022). The Plectin gene has a pleiotropic effect on most milk production traits, explaining the results found here with the SNPs significant for FP (Bekele et al., 2023; Z. Yang et al., 2021).

One gene in the region identified on chromosome 19, which to the best of our knowledge was never associated with FP or milk production traits, was the Casein Kinase 1 delta - CSNK1D. This gene showed an increased expression in adipose tissue and mammary gland in postpartum cows (M. Wang et al., 2015); it was also identified in a genome-wide association studies related to the occurrence of clinical ketosis in first parity dairy cows (Soares et al., 2021). Cruz et al. (2019) found a significant region for milk short-chain fatty acids on chromosome 19 encoding the LOC101909618, now identified as the Tubulin Folding Cofactor D – (TBCD) gene, but it is ~830 kb far from the CSNK1D gene.

Three genes did not overpass the Bonferroni threshold, i.e. C11H9orf50, OR10AG83 and CCDC57, and only the last one, was found to be associated with FP and other milk traits. This gene is the Coiled-Coil Domain Containing 57 (CCDC57), here mapped by the rs41922153 and rs135528222 SNPs, which has also been associated with FY and FP as well as carcass fatty acids composition (Bouwman et al., 2014; Jiang et al., 2019; Tribout et al., 2020). The other two genes identified (OR10AG83 and C11H9orf50) have never been associated with milk production traits. However, the olfactory receptor family 10 subfamily AG member 83 (OR10AG83) was associated with milk citrate. Milk citrate is a potential early biomarker for negative energy balance in dairy cows, and for this reason, we may speculate that this gene could be somehow involved in metabolism regulation, affecting, as a consequence, milk fat content (Y. Chen et al., 2023).

Protein percentage trait

For PP, different significant regions have been identified for both EBV and DRP. A first quantitative trait loci region, defined by 13 SNPs, is located on chromosome 5. This region harbours five genes previously found as associated with milk traits in different cattle breeds, such as the TBC1 domain family member 22A (TBC1D22A) and cadherin EGF LAG seven-pass G-type receptor 1 (CELSR1), associated with PP (Tribout et al., 2020), the fibulin 1 (FBLN1) with PP and PY (Raven et al., 2014), and the tetratricopeptide repeat domain 38 (TTC38) with milk eicosapentaenoic acid content (Ibeagha-Awemu et al., 2016).

A second quantitative trait loci region, defined by 33 SNPs (about 1.4 Mbp long), was identified on chromosome 6. This region harbours five genes, three of which are part of the casein's family (the αs1-casein – CSN1S1, the β-casein – CSN2, and the k-casein gene – CSN3) (Table 3). As recently reported by Bernini et al. (2023), these three genes are polymorphic in the Aosta cattle breeds and show different allele frequencies for the two breeds (i.e., for the k-casein the B allele has a frequency of 0.40 for the ABCH and 0.63 for the ARP while for the αs1-casein the A allele has frequencies of 0.80 and 0.95, respectively). The most significant SNP (rs109193501) is annotated within the αs1-casein gene (CSN1S1) and is an intronic mutation. As in this study, Kemper et al. (2016) found an association between the rs109193501 and PP concentration in the milk of Jersey cows. For the same SNP, other authors found an association with PP but also with FP in three different cattle breeds, the Braunvieh, the Fleckvieh, and the Holstein (Pausch, Emmerling, et al., 2017; Pausch, MacLeod, et al., 2017). As discussed by Kuss et al. (2005) and Korkuć et al. (2023), polymorphisms of the regulatory region of the gene have a role in modulating transcription levels that impact the production of αs1-casein, which has an exerting influence not only on milk protein but also on milk fat content and overall milk properties.

The rs109299401 SNP was here found in association with PP variability. This SNP is a missense variant (ATG/CTG) mapping in the exon 7 of the β-casein gene (CSN2) determining a substitution of the Methionine (amino acid position: 143) with a Leucine responsible for the I/H2 variants of the β-casein (Chessa et al., 2020). The rs109299401 SNP marker was previously associated with PP and PY in two studies carried out in Holstein cattle, one from Fontanesi et al. (2014) in which the major allele (T) was associated with a reduction of PY, PP, and higher milk somatic cell count and the other one from Viale et al. (2017) in which the association study was done considering the minor allele that resulted in an increase of 0.056% of the PP.

The three significative SNPs in the CSN3 gene include: (i) two missense variants, the rs43703015 (ATC/ACC), causing a substitution Ile157Thr, and the rs43703016 (GCT/GAT), causing a substitution Ala169Asp, both mapped in exon 4 of the k-casein gene (CSN3) encoding the variants A and B of this gene (Farrell Jr et al., 2004); (ii) a synonymous variant, the rs110014544. Numerous studies focused on the k-casein gene because of its well-known effects in the cheese production process (Viale et al., 2017). Generally, the allele A correlates with reduced protein content and increased milk yield, while allele B is linked to high protein content, better milk quality but lower milk production (Caroli et al., 2004; Schopen et al., 2011). Schopen et al. (2011) identified in Holstein-Friesian cows a significant association with milk k-casein content, PP, β-lactoglobulin content, and casein index at the rs43703016 SNP.

In the quantitative trait loci region on chromosome 6, two more genes were in common between the results obtained with the DRP and EBV. The first one is the histatherin - HSTN gene which is near a regulatory element that affects the expression of the β-casein gene (Pegolo et al., 2021). In many studies, the HSTN gene has been found associated with PY, PP, and αs1-casein and β-casein concentration in milk (Jiang et al. 2019; Tribout et al. 2020; Pegolo et al. 2021). Even the genes mapping on chromosome 10 (death-associated protein kinase 2 - DAPK2) and chromosome 13 (oxysterol binding protein-like 2 - OSBPL2) were already associated with milk PP as well as with milk traits (NRDE2, necessary for RNA interference, domain containing - NRDE2, chromosome 10) (Jiang et al., 2019).

The Aosta cattle national breeder association reports the individual genotype at the k-casein, β-casein and β-lactoglobulin genes in the bull's catalogue. This information allowed the selection of the different casein variants for cheese yield, increasing as such also the allele frequencies of the B allele at the β-lactoglobulin, reported to be 0.57 and 0.68 in the 1980s by (Merlin and Di Stasio, 1982) to current frequencies of 0.73 and 0.69 for the ARP and ABCH respectively (Bernini et al., 2023). Indeed, the favourable alleles for the protein content and cheese-making proprieties are also the most present in the population (e.g., for the β-lactoglobulin 49 and 54% of subjects with BB genotype for the ABCH and ARP breed respectively, and at the k-casein gene the B allele has the highest frequency in the ARP breed with the absence of the E allele in the whole population) (Bernini et al., 2023). The differences in the genotypic frequencies may reflect the diverse evolutionary history of the two breeds, showing different origins and bred for a slightly different purpose: ARP is mainly bred for milk yield, to produce Fontina cheese, whereas the leading interest for ABCH breeders is the fighting (Sartori et al., 2020).

Milk yield trait

Two quantitative trait loci regions were associated with MY_EBV but not for the MY_DRP dependent variable, probably because of the bigger sample size of the EBV sample with compared to the one of the DRP. Most likely, the reason why this is occurring for MY and not for PP or FP is related to the average lower genetic variance of the phenotype MY (8.32 and 9.72 for the ABCH and ARP, respectively) compared to the other two traits: FP (43.57 for the ABCH and 58.62 for the ARP) and PP (20.5 in the ABCH and 22.56 in the ARP). The significant SNP rs110434046 has been associated with the Daughter Pregnancy Rate in a study of Liang et al. (2023) on a sample of a million

Holstein cows. In a study based on the same Holstein cow's dataset, the same SNP (rs110434046) resulted in an epistatic effect with the rs109421300 SNP in the DGAT1 gene (Prakapenka et al. 2024).

Conclusion

The results of this study show that even if the majority of the quantitative trait loci identified in this study have been previously associated with milk production traits, the Aosta population owns a peculiar genetic structure that differentiates them from the cosmopolitan specialized breeds intensively selected for milk yield (Signer-Hasler et al., 2023). The in population studied here has been selected with a much lower intensity for milk traits with respect to specialised dairy breeds such as the Holstein one. Being a double-purpose population with a strong aptitude for pasture in harsh mountain environments the selection programme is in fact oriented to several breeding objectives: (i) to improve the milk yield and its physical and chemical characteristics as a function of cheese yield and renneting properties; (ii) to improve the amount of beef produced and the estimated carcass quality; (iii) to maintain the genetic variability; (iv) to maintain the hardiness and longevity; these latter two are intrinsic characteristics of the population. In addition to the fact that the selection is both for meat and milk traits, the mating scheme is oriented to make the gene flow of males used in reproduction as homogeneous as possible in the female population, to maintain the genetic variability as large as possible. The ongoing breeding and selection scheme has been active for decades making the population strongly homogenous in its genomic makeup, a condition that is possibly affecting the identification of region containing quantitative trait loci under segregation for milk traits. For some of the novel identified quantitative trait loci regions, the analysis of sequence data may further explore the genomic variation here found, to determine the presence of proprietary polymorphism of the Aosta cattle with respect to other cosmopolitan and specialised populations.

Ethics approval

Not Applicable.

Data and model availability statement

The data supporting the findings of this study are available within the article and its Supplementary Materials. The raw genetic datasets generated during the current study are available from the corresponding author upon reasonable request. The data were not deposited in an official repository.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) did not use any AI and AI-assisted technologies.

Author ORCIDs

Francesca Bernini: https://orcid.org/0000-0001-8430-6813
Enrico Mancin: https://orcid.org/0000-0002-4130-4639
Cristina Sartori: https://orcid.org/0000-0002-2091-2280
Mario Vevey: N/A
Veruska Blanchet: N/A
Roberto Mantovani: https://orcid.org/0000-0002-1919-0559
Alessandro Bagnato: https://orcid.org/0000-0003-2087-1287
Maria Giuseppina Strillacci: https://orcid.org/0000-0002-8972-1730

CRediT authorship contribution statement

F. Bernini: Writing– review & editing, Writing– original draft, Investigation, Formal analysis, Conceptualization. E. Mancin: Writ ing– review & editing, Formal analysis. C. Sartori: Writing review & editing, Formal analysis. R. Mantovani: Writing– review & editing, Resources, Data curation. M. Vevey: Resources, Funding acquisition. V. Blanchet: Resources, Data curation. A. Bagnato: Writing– review & editing, Writing– original draft, Project admin istration, Investigation, Conceptualization. M.G. Strillacci: Writing– review & editing, Writing– original draft, Supervision.

Declaration of interest

None.

Acknowledgements

None

Financial support statement

Project name: “Le razze bovine a duplice attitudine: un modello alternativo di zootecnia eco-sostenibile – DUAL BREEDING FASE 2” – PSRN Programma di Sviluppo Rurale Nazionale 2014/2020 – Sottomisura 10.2 – PERIODO 2020-2023 – funded by the Ministero dell’Agricoltura della Sovranità Alimentare e delle Foreste with economic support of European Agricultural Fund for Rural Development. Grant n.: CUP_J72C21000650005. Project name: Prin 2022 "GENOmic Breeding VALue Estimation in a Native Alpine Cattle Breed" grant CUP_G53D23004040006 - Finanziato dall'Unione europea - Next Generation EU.

References

Bakhshalizadeh, S., Zerehdaran, S., & Javadmanesh, A. (2021). Meta-analysis of genome-wide association studies and gene networks analysis for milk production traits in Holstein cows. Livestock Science, 250, 104605.

Bekele, R., Taye, M., Abebe, G., & Meseret, S. (2023). Genomic Regions and Candidate Genes Associated with Milk Production Traits in Holstein and Its Crossbred Cattle: A Review. International Journal of Genomics, 2023, 8497453. https://doi.org/10.1155/2023/8497453

Bernini, F., Punturiero, C., Vevey, M., Blanchet, V., Milanesi, R., Delledonne, A., Bagnato, A., & Strillacci, M. G. (2023). Assessing major genes allele frequencies and the genetic diversity of the native Aosta cattle female population. Italian Journal of Animal Science, 22(1), 1008–1022.

Bouwman, A. C., Visker, M. H. P. W., van Arendonk, J. M., & Bovenhuis, H. (2014). Fine mapping of a quantitative trait locus for bovine milk fat composition on Bos taurus autosome 19. Journal of Dairy Science, 97, 1139–1149.

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., Sahana, G., Govignon-Gion, A., Boitard, S., & Dolezal, M. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics, 50, 362–367.

Caroli, A., Chessa, S., Bolla, P., Budelli, E., & Gandini, G. C. (2004). Genetic structure of milk protein polymorphisms and effects on milk production traits in a local dairy cattle. Journal of Animal Breeding and Genetics, 121, 119–127.

Chen, Y., Hu, H., Atashi, H., Grelet, C., Wijnrocx, K., Lemal, P., Gengler, N., 2023. Genetic analysis of milk citrate predicted by milk mid-infrared spectra of Holstein cows in early lactation. Journal of Dairy Science 107, 3047–3061 10. 3168/jds.2023-23903.

Chen, S.-Y., Schenkel, F. S., Melo, A. L. P., Oliveira, H. R., Pedrosa, V. B., Araujo, A. C., Melka, M. G., & Brito, L. F. (2022). Identifying pleiotropic variants and candidate genes for fertility and reproduction traits in Holstein cattle via association studies based on imputed whole-genome sequence genotypes. BMC Genomics, 23, 331.

Chessa, S., Gattolin, S., Cremonesi, P., Soglia, D., Finocchiaro, R., Van Kaam, J.-T., Marusi, M., & Civati, G. (2020). The effect of selection on casein genetic polymorphisms and haplotypes in Italian Holstein cattle. Italian Journal of Animal Science, 19, 833–839.

Cruz, V. A. R., Oliveira, H. R., Brito, L. F., Fleming, A., Larmer, S., Miglior, F., & Schenkel, F. S. (2019). Genome-wide association study for milk fatty acids in Holstein cattle accounting for the DGAT1 gene effect. Animals (Basel) 9: 997.

Farrell Jr, H. M., Jimenez-Flores, R., Bleck, G. T., Brown, E. M., Butler, J. E., Creamer, L. K., Hicks, C. L., Hollar, C. M., Ng-Kwai-Hang, K. F., & Swaisgood, H. E. (2004). Nomenclature of the proteins of cows' milk—Sixth revision. Journal of Dairy Science, 87, 1641–1674.

Fontanesi, L., Calò, D. G., Galimberti, G., Negrini, R., Marino, R., Nardone, A., Ajmone-Marsan, P., & Russo, V. (2014). A candidate gene association study for nine economically important traits in Italian Holstein cattle. Animal Genetics, 45, 576–580.

Garrick, D. J., Taylor, J. F., & Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution, 41, 1–8.

Glantz, M., Månsson, H. L., Paulsson, M., & Stålhammar, H. (2012). Genomic selection in relation to bovine milk composition and processability. Journal of Dairy Research, 79, 53–59.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., & Down, T. (2002). The Ensembl genome database project. Nucleic Acids Research, 30, 38–41.

Ibeagha-Awemu, E. M., Peters, S. O., Akwanji, K. A., Imumorin, I. G., & Zhao, X. (2016). High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. Scientific Reports, 6, 31109.

Iung, L. H. S., Petrini, J., Ramírez-Díaz, J., Salvian, M., Rovadoscki, G. A., Pilonetto, F., Dauria, B. D., Machado, P. F., Coutinho, L. L., & Wiggans, G. R. (2019). Genome-wide association study for milk production traits in a Brazilian Holstein population. Journal of Dairy Science, 102, 5305–5314.

Jayawardana, J., Lopez-Villalobos, N., McNaughton, L. R., & Hickson, R. E. (2023). Genomic regions associated with milk composition and fertility traits in spring-calved dairy cows in New Zealand. Genes, 14, 860.

Jiang, J., Ma, L., Prakapenka, D., VanRaden, P. M., Cole, J. B., & Da, Y. (2019). A large-scale genome-wide association study in US Holstein cattle. Frontiers in Genetics, 10, 442321.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nature Genetics, 42(4), 348–354.

Kemper, K. E., Littlejohn, M. D., Lopdell, T., Hayes, B. J., Bennett, L. E., Williams, R. P., Xu, X. Q., Visscher, P. M., Carrick, M. J., & Goddard, M. E. (2016). Leveraging genetically simple traits to identify small-effect variants for complex phenotypes. BMC Genomics, 17, 1–9.

Korkuć, P., Arends, D., May, K., & König, S. (2021). Genomic loci affecting milk production in German Black Pied cattle (DSN). Frontiers in Genetics, 12, 640039.

Korkuć, P., Neumann, G. B., Hesse, D., Arends, D., Reißmann, M., Rahmatalla, S., May, K., Wolf, M. J., König, S., & Brockmann, G. A. (2023). Whole-Genome Sequencing Data Reveal New Loci Affecting Milk Production in German Black Pied Cattle (DSN). Genes, 14, 581.

Kühn, C., Thaller, G., Winter, A., Bininda-Emonds, O. R. P., Kaupe, B., Erhardt, G., Bennewitz, J., Schwerin, M., & Fries, R. (2004). Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. Genetics, 167, 1873–1881.

Kuss, A. W., Gogol, J., Bartenschlager, H., & Geldermann, H. (2005). Polymorphic AP-1 binding site in bovine CSN1S1 shows quantitative differences in protein binding associated with milk protein expression. Journal of Dairy Science, 88, 2246–2252.

Liang, Z., Prakapenka, D., VanRaden, P. M., Jiang, J., Ma, L., & Da, Y. (2023). A million-cow genome-wide association study of three fertility traits in US Holstein cows. International Journal of Molecular Sciences, 24, 10496.

Littlejohn, M. D., Tiplady, K., Fink, T. A., Lehnert, K., Lopdell, T., Johnson, T., Couldrey, C., Keehan, M., Sherlock, R. G., & Harland, C. (2016). Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. Scientific Reports, 6, 25376.

Mancin, E., Tuliozi, B., Pegolo, S., Sartori, C., & Mantovani, R. (2022). Genome wide association study of beef traits in local Alpine breed reveals the diversity of the pathways involved and the role of time stratification. Frontiers in Genetics, 12, 746665.

Manzanilla-Pech, C. I. V, Løvendahl, P., Gordo, D. M., Difford, G. F., Pryce, J. E., Schenkel, F., Wegmann, S., Miglior, F., Chud, T. C., & Moate, P. J. (2021). Breeding for reduced methane emission and feed-efficient Holstein cows: An international response. Journal of Dairy Science, 104, 8983–9001.

Massender, E., Oliveira, H. R., Brito, L. F., Maignel, L., Jafarikia, M., Baes, C. F., Sullivan, B., & Schenkel, F. S. (2023). Genome-wide association study for milk production and conformation traits in Canadian Alpine and Saanen dairy goats. Journal of Dairy Science, 106, 1168–1189.

Mazza, S., Guzzo, N., Sartori, C., & Mantovani, R. (2016). Genetic correlations between type and test-day milk yield in small dual-purpose cattle populations: The Aosta Red Pied breed as a case study. Journal of Dairy Science, 99, 8127–8136.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biology, 17, 1–14.

Merlin, P., & Di Stasio, L. (1982). Study on milk proteins loci in some decreasing Italian cattle breeds. Annales de Génétique et de Sélection Animale, 14, 17–28.

Misztal, I., Wang, H., Aguilar, I., Legarra, A., Tsuruta, S., Lourenco, D. A. L., Fragomeni, B., Zhang, X., Muir, W., & Cheng, H. H. (2014). GWAS using ssGBLUP. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, 17–22.

Nguyen, T. T. T., Bowman, P. J., Haile-Mariam, M., Nieuwhof, G. J., Hayes, B. J., & Pryce, J. E. (2017). Implementation of a breeding value for heat tolerance in Australian dairy cattle. Journal of Dairy Science, 100, 7362–7367.

Oliveira, H. R., Cant, J. P., Brito, L. F., Feitosa, F. L. B., Chud, T. C. S., Fonseca, P. A. S., Jamrozik, J., Silva, F. F., Lourenco, D. A. L., & Schenkel, F. S. (2019). Genome-wide association for milk production traits and somatic cell score in different lactation stages of Ayrshire, Holstein, and Jersey dairy cattle. Journal of Dairy Science, 102, 8159–8174.

Pagnacco, G., Gandini, G. C., Bagnato, A., Miglior, F., & Caroli, A. (1989). Genetic gain and conservation in a small alpine cattle breed. J. Anim. Breed. Genet, 106, 351–357.

Pausch, H., Emmerling, R., Gredler-Grandl, B., Fries, R., Daetwyler, H. D., & Goddard, M. E. (2017). Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. BMC Genomics, 18, 1–11.

Pausch, H., MacLeod, I. M., Fries, R., Emmerling, R., Bowman, P. J., Daetwyler, H. D., & Goddard, M. E. (2017). Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. Genetics Selection Evolution, 49, 1–14.

Pedrosa, V. B., Schenkel, F. S., Chen, S.-Y., Oliveira, H. R., Casey, T. M., Melka, M. G., & Brito, L. F. (2021). Genomewide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. Genes, 12, 1830.

Pedrosa, V. B., Boerman, J. P., Gloria, L. S., Chen, S.-Y., Montes, M. E., Doucette, J. S., & Brito, L. F. (2023). Genomic-based genetic parameters for milkability traits derived from automatic milking systems in North American Holstein cattle. Journal of Dairy Science, 106, 2613–2629.

Pegolo, S., Yu, H., Morota, G., Bisutti, V., Rosa, G. J. M., Bittante, G., & Cecchinato, A. (2021). Structural equation modeling for unraveling the multivariate genomic architecture of milk proteins in dairy cattle. Journal of Dairy Science, 104, 5705–5718.

Prakapenka, D., Liang, Z., Zaabza, H. B., VanRaden, P. M., Van Tassell, C. P., & Da, Y. (2024). A million-cow validation of a chromosome 14 region interacting with all chromosomes for fat percentage in US Holstein cows. International Journal of Molecular Sciences, 25, 674.

Raschia, M. A., Nani, J. P., Carignano, H. A., Amadio, A. F., Maizon, D. O., & Poli, M. A. (2020). Weighted single-step genome-wide association analyses for milk traits in Holstein and Holstein x Jersey crossbred dairy cattle. Livestock Science, 242, 104294.

Raven, L.-A., Cocks, B. G., & Hayes, B. J. (2014). Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics, 15, 1–14.

Sahana, G., Cai, Z., Sanchez, M. P., Bouwman, A. C., & Boichard, D. (2023). Invited review: Good practices in genome-wide association studies to identify candidate sequence variants in dairy cattle. Journal of Dairy Science, 106, 5218–5241. https://doi.org/10.3168/jds.2022-22694

Sanchez, M.-P., Govignon-Gion, A., Croiseau, P., Fritz, S., Hozé, C., Miranda, G., Martin, P., Barbat-Leterrier, A., Letaïef, R., & Rocha, D. (2017). Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. Genetics Selection Evolution, 49, 1–16.

Sartori, C., Guzzo, N., & Mantovani, R. (2020). Genetic correlations of fighting ability with somatic cells and longevity in cattle. Animal, 14, 13–21. https://doi.org/DOI: 10.1017/S175173111900168X

Schopen, G. C. B., Visker, M., Koks, P. D., Mullaart, E., Van Arendonk, J. A. M., & Bovenhuis, H. (2011). Whole-genome association study for milk protein composition in dairy cattle. Journal of Dairy Science, 94, 3148–3158.

Signer-Hasler, H., Casanova, L., Barenco, A., Maitre, B., Bagnato, A., Vevey, M., Berger, B., Simčič, M., Boichon, D., & Capitan, A. (2023). Genomic regions underlying positive selection in local, Alpine cattle breeds. Animal Genetics, 54, 239–253.

Soares, R. A. N., Vargas, G., Duffield, T., Schenkel, F., & Squires, E. J. (2021). Genome-wide association study and functional analyses for clinical and subclinical ketosis in Holstein cattle. Journal of Dairy Science, 104, 10076–10089.

Strillacci, M. G., Vevey, M., Blanchet, V., Mantovani, R., Sartori, C., & Bagnato, A. (2020). The Genomic Variation in the Aosta Cattle Breeds Raised in an Extensive Alpine Farming System. Animals, 10, 2385.

Strillacci, M. G., Punturiero, C., Milanesi, R., Bernini, F., Mason, T., & Bagnato, A. (2023). Antibiotic treatments and somatic cell count as phenotype to map QTL for mastitis susceptibility in Holstein cattle breed. Italian Journal of Animal Science, 22, 190–199.

Su, M., Lin, X., Xiao, Z., She, Y., Deng, M., Liu, G., Sun, B., Guo, Y., Liu, D., & Li, Y. (2023). Genome-Wide Association Study of Lactation Traits in Chinese Holstein Cows in Southern China. Animals, 13, 2545.

Tribout, T., Croiseau, P., Lefebvre, R., Barbat, A., Boussaha, M., Fritz, S., Boichard, D., Hoze, C., & Sanchez, M.-P. (2020). Confirmed effects of candidate variants for milk production, udder health, and udder morphology in dairy cattle. Genetics Selection Evolution, 52, 1–26.

van den Berg, I., Boichard, D., & Lund, M. S. (2016). Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds. Journal of Dairy Science, 99, 8932–8945.

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

Wang, P., Li, X., Zhu, Y., Wei, J., Zhang, C., Kong, Q., Nie, X., Zhang, Q., & Wang, Z. (2022). Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows. Frontiers in Veterinary Science, 9, 932034.

Wang, D., Ning, C., Liu, J.-F., Zhang, Q., & Jiang, L. (2019). Replication of genome-wide association studies for milk production traits in Chinese Holstein by an efficient rotated linear mixed model. Journal of Dairy Science, 102, 2378–2383.

Wang, M., Zhou, Z., Khan, M. J., Gao, J., & Loor, J. J. (2015). Clock circadian regulator (CLOCK) gene network expression patterns in bovine adipose, liver, and mammary gland at 3 time points during the transition from pregnancy into lactation. Journal of Dairy Science, 98, 4601–4612.

Yang, Z., Lian, Z., Liu, G., Deng, M., Sun, B., Guo, Y., Liu, D., & Li, Y. (2021). Identification of genetic markers associated with milk production traits in Chinese Holstein cattle based on post genome-wide association studies. Animal Biotechnology, 32, 67–76.

Tables

Table 1.

Average (SD) values for each trait of the additive genetic variance (σ_a^2), heritability (h^2), EBV and DRP, for the Aosta cattle breeds.

Item	**ABCH**				**ARP**			
	σ_a^2	**h^2**	**EBV**	**DRP**	**σ_a^2**	**h^2**	**EBV**	**DRP**
n*	291 060	291 060	1 361	521	718 966	718 966	2 886	1 707
MY	9.72 (0.42)	0.28 (0.01)	-4.81 (340.42)	1.68 (1.75)	8.32 (0.29)	0.22 (0.01)	-12.23 (274.32)	3.01 (1.32)
FP	0.06 (0.002)	0.16 (0.01)	0.03 (0.14)	0.52 (0.23)	0.04 (0.001)	0.13 (0.003)	0.04 (0.14)	0.44 (0.20)
PP	0.02 (0.001)	0.27 (0.01)	0.02 (0.1)	0.05 (0.15)	0.02 (0.000)	0.25 (0.01)	0.01 (0.11)	0.52 (0.15)
FY	0.01 (0.001)	0.20 (0.01)	0.56 (11.03)	0.06 (0.06)	0.01 (0.000)	0.12 (0.005)	1.26 (9.13)	0.09 (0.05)
PY	0.01 (0.000)	0.25 (0.01)	0.33 (10.5)	0.06 (0.05)	0.01 (0.000)	0.18 (0.01)	0.04 (7.85)	0.10 (0.04)

*n refers to the number of observations used for the evaluation of σ_a^2 and h^2, and the number of samples analysed for each breed for each trait EBV and DRP.

Abbreviations: ABCH=Aosta Black Pied-Chestnut, ARP=Aosta Red Pied, MY=milk yield, FP=fat percentage, PP=protein percentage, FY=fat yield, PP=protein yield, EBV=estimated breeding value, DRP=deregressed proof

Table 2.

List of SNPs above the Bonferroni (underlined) and false discovery rate 0.05 thresholds for both EBVs and DRP for FP trait, identified in the Aosta cattle breeds. The Table reports the name of the SNP markers, the rsID number, the chromosome, the position in base pairs (bp), the P-values for the EBV and DPR associated with the marker, the gene in which the SNP lays, and the position of the marker respect to the gene.

Marker	RS_SNP_ID	Chr	Position (bp)	P-Value EBV*	P-Value DRP*	Gene	Position
Hapmap47387-BTA-72195	rs41591555	4	106144190	2.18E-05			Intergenic
BovineHD0500026451	rs135438063	5	92753124	2.94E-06			Intergenic
BTA-37834-no-rs	rs109957658	5	92758201	5.77E-06	1.43E-05		Intergenic
BovineHD0500026553	rs110579160	5	93109175	1.96E-05	7.06E-07		Intergenic
BovineHD0500026624	rs134155693	5	93378265	2.41E-06			Intergenic
BovineHD0500026635	rs133160309	5	93413676	5.59E-06			Intergenic
Hapmap32415-BTA-74556	rs41602750	5	93441439	3.04E-06			Intergenic
DB-335-seq-rs109307833	rs109307833	5	93450091	6.44E-07			Intergenic
BovineHD0500026649	rs132674836	5	93470430	2.70E-06	3.84E-08		Intergenic
BovineHD0500026655	rs133517677	5	93503991	3.30E-06	3.44E-06	*MGST1*	Intronic
BovineHD0500026662	rs134637616	5	93515983	5.52E-11	1.09E-09	*MGST1*	Intronic
DB-337-seq-rs211210569	rs211210569	5	93516066	1.57E-16	2.97E-13	*MGST1*	Intronic
BovineHD0500026664	rs137705840	5	93517967	2.29E-05		*MGST1*	Intronic
DB-339-seq-rs210744919	rs210744919	5	93520138	2.01E-16	1.98E-14	*MGST1*	Intronic
DB-340-seq-rs208014256	rs208014256	5	93520616	4.12E-06		*MGST1*	5’ UTR variant
chr5_93950333	rs209210458	5	93520661	1.23E-06	8.89E-06	*MGST1*	5’ UTR variant
chr5_93950346	rs210155966	5	93520674	3.13E-06		*MGST1*	5’ UTR variant
BovineHD0500026666	rs133918820	5	93521394	4.12E-06			Intergenic
BovineHD0500026668	rs135807129	5	93524134	3.90E-10	9.88E-09		Intergenic
DB-341-seq-rs209288972	rs209288972	5	93525079	1.39E-14	3.73E-12		Intergenic
BovineHD0500034417	rs109812511	5	117310460	1.51E-05			Intergenic
BovineHD0500034461	rs134294234	5	117440580	5.90E-06			Intergenic
ARS-BFGL-NGS-23468	rs109206555	11	100011096		8.28E-06	*C11H9orf50*	Intronic
ARS-BFGL-NGS-57332	rs109338243	13	16659292		9.72E-06		Intergenic
14-1322168-C-A-rs208813903	rs208813903	14	160524		1.66E-05	*OR10AG83*	Missense variant (C/A)
BovineHD1400000152	rs110508680	14	255765	1.43E-09	3.74E-08		Intergenic
ARS-BFGL-NGS-57820	rs109146371	14	465742	3.76E-07	1.04E-06		Intergenic
Chr14_1653693	rs110984572	14	468124	1.51E-12	1.07E-09		Intergenic
BovineHD1400000204	rs137727465	14	487527	4.71E-14	2.41E-11	*CYHR1*	Intronic
BovineHD1400000206	rs137472016	14	494621	1.60E-13	1.27E-10		Intergenic
ARS-BFGL-NGS-94706	rs17870736	14	511247	8.42E-18	1.52E-15	*VPS28*	Intronic
Chr14_1699016	rs136784996	14	513203	4.00E-18	3.43E-14		Intergenic
UFL-rs134432442	rs134432442	14	550784	5.44E-15	1.20E-15	*CPSF1*	Missense variant (C/T)
Chr14_1757935	rs211309638	14	572120	1.25E-21	3.89E-19		Intergenic
BovineHD1400000239	rs133299034	14	663029	2.65E-06		*MROH1*	Intronic
BovineHD1400000241	rs110966735	14	669738	3.67E-07		*MROH1*	Intronic
BovineHD1400000246	rs137787931	14	688317	1.48E-05		*MROH1*	Intronic
Hapmap52798-ss46526455	rs41256919	14	731230	5.81E-07	2.99E-06	*MAF1*	Synonymous variant (T/C)
BovineHD1400000256	rs110929299	14	751534	5.39E-07	1.89E-06	*MAF1*	Intronic
BovineHD1400000271	rs136792973	14	810116	1.85E-07	9.10E-07	*GPAA1*	Intronic
UA-IFASA-6878	rs41629750	14	810863	9.37E-08			Intergenic
BovineHD1400000282	rs136051530	14	859251	8.04E-07	2.51E-06	*PLEC*	Intronic
BovineHD1400000287	rs109662548	14	883732	6.65E-06		*PLEC*	Intronic
BovineHD1400000288	rs135270011	14	891340	8.81E-09	1.32E-06	*PLEC*	Synonymous variant (T/C)
BovineHD1900014321	rs41922195	19	50607587	1.40E-08	7.33E-06	*CSNK1D*	Intronic
BovineHD1900014337	rs41922153	19	50666822	1.99E-06	9.00E-06	*CCDC57*	Intronic
BovineHD1900014340	rs135528222	19	50674342	1.67E-05		*CCDC57*	Intronic

Abbreviations: FP=fat percentage, SNP=single nucleotide polymorphism, Chr=chromosome, EBV=estimated breeding value, DRP=deregressed proof, UTR = untranslated region

* P-values have been reported only for the significative SNPs in each category.

Table 3.

List of SNPs above the Bonferroni (underlined) and false discovery rate 0.05 thresholds for both EBVs and DRP for PP trait, identified in the Aosta cattle breeds. See Table 2 for further details.

Marker	RS_SNP_ID	Chr	Position (bp)	P-Value EBV*	P-Value DRP*	Gene	Position
BovineHD0500033518	rs3423212215	5	114916479	3.27E-06	1.93E-05		Intergenic
ARS-BFGL-NGS-69589	rs109415265	5	115890490	2.51E-05		*FBLN1*	Intronic
ARS-BFGL-NGS-18620	rs110810286	5	116521777		3.73E-06	*CDPF1*	Intronic
BovineHD0500034166	rs135679475	5	116591924	2.72E-06	7.88E-06	*TTC38*	Intronic
BovineHD0500034225	rs109155800	5	116760590	1.35E-05	2.31E-05	*CELSR1*	Intronic
ARS-BFGL-NGS-6245	rs41593908	5	117072309	2.62E-05	5.20E-06	*TBC1D22A*	Intronic
BovineHD4100004172	rs41593907	5	117075478	5.47E-07	5.70E-09	*TBC1D22A*	Intronic
ARS-BFGL-NGS-14632	rs110117542	5	117391029	1.39E-07	2.66E-06		Intergenic
BovineHD0500034461	rs134294234	5	117440580	7.23E-07	7.39E-06		Intergenic
BovineHD0500034507	rs109252331	5	117532054	8.11E-09	1.36E-07		Intergenic
BovineHD0500034524	rs110310756	5	117568705	5.85E-06			Intergenic
BovineHD0500035205	rs3423206779	5	118938470	3.78E-06			Intergenic
BovineHD0500035210	rs3423206754	5	118949961	1.78E-05			Intergenic
BovineHD0600021245	rs132642659	6	74839211	2.02E-05			Intergenic
ARS-BFGL-NGS-27958	rs110239739	6	82968804	3.20E-05			Intergenic
BovineHD0600023788	rs133627704	6	84939092	8.04E-07	4.95E-08		Intergenic
Hapmap25708-BTC-043671	rs110063049	6	85383787	2.05E-14	1.64E-12		Intergenic
DB-429-seq-rs109193501	rs109193501	6	85424759	1.95E-26	3.82E-20	*CSN1S1*	Intronic
Hapmap33451-BTC-060559	rs110914422	6	85446151	5.17E-10	4.55E-09		Intergenic
CSN2_4	rs109299401	6	85451221	4.95E-14	3.01E-12	*CSN2*	Missense variant (T/G)
chr6_87188128	rs108993011	6	85457804	1.20E-11	1.93E-11		Intergenic
chr6_87202566	rs384705370	6	85470165	1.58E-12	4.77E-11	*HSTN*	3' UTR variant
chr6_87202599	rs378595205	6	85470198	3.71E-14	5.28E-12	*HSTN*	Splice region variant (G/A)
BovineHD0600023888	rs136049155	6	85471455	6.41E-07	3.73E-07	*HSTN*	Intronic
chr6_87204247	rs382297554	6	85471846	1.81E-14	3.01E-12	*HSTN*	3' UTR variant
chr6_87204311	rs386014273	6	85471910	5.27E-11	5.34E-10	*HSTN*	3' UTR variant
chr6_87204315	rs379649542	6	85471914	1.81E-14	3.01E-12	*HSTN*	3' UTR variant
chr6_87204358	rs449985830	6	85471957	3.94E-14	7.99E-12	*HSTN*	3' UTR variant
chr6_87204403	rs385251021	6	85472002	1.81E-14	3.01E-12	*HSTN*	3' UTR variant
chr6_87204870	rs382158121	6	85472469	9.92E-14	7.19E-12		Intergenic
chr6_87204878	rs383383092	6	85472477	4.59E-14	3.55E-12		Intergenic
chr6_87205080	rs381311750	6	85472679	1.84E-14	1.89E-12		Intergenic
chr6_87205162	rs384622341	6	85472761	2.05E-14	3.01E-12		Intergenic
chr6_87205336	rs382862058	6	85472936	2.92E-13	2.77E-11		Intergenic
chr6_87205349	rs385917248	6	85472949	2.84E-14	3.01E-12		Intergenic
BovineHD4100005320	rs133035102	6	85578801		6.11E-06		Intergenic
DB-434-seq-rs43703015	rs43703015	6	85656736	9.84E-06	3.82E-20	*CSN3*	Missense variant (T/C)
CSN3_AY380228_13104_1	rs43703016	6	85656772	9.67E-06	5.20E-09	*CSN3*	Missense variant (C/A)
CSN3_AY380228_13165	rs110014544	6	85656833	1.13E-05	7.29E-09	*CSN3*	Synonymous variant (G/A)
Hapmap52348-rs29024684	rs29024684	6	85662466	3.59E-06	4.29E-09		Intergenic
BovineHD0600023926	rs110312754	6	85686637	1.07E-10	3.59E-14		Intergenic
BTA-115149-no-rs	rs109581772	6	85784380	4.55E-06	1.98E-07		Intergenic
BovineHD0600023955	rs137754062	6	85817621		1.40E-05		Intergenic
ARS-BFGL-NGS-24522	rs110064541	6	86151977	1.72E-06	8.19E-08	*RUFY3*	Intronic
BovineHD0600024093	rs110091883	6	86380145	2.07E-06	5.38E-07		Intergenic
BovineHD0600024315	rs109452259	6	87068809		1.45E-05		Intergenic
BovineHD1000013824	rs136032713	10	46034552	2.78E-06		*DAPK2*	Intronic
BovineHD1000030067	rs136238611	10	101887955	1.28E-05		*NRDE2*	Intronic
ARS-BFGL-NGS-116624	rs41696761	13	54955140	3.28E-05		*OSBPL2*	Intronic
ARS-BFGL-NGS-107234	rs110249976	15	52381562	9.22E-06		*FCHSD2*	Synonymous variant (C/T)
BovineHD1500015413	rs135702946	15	52598144	9.22E-06		*FCHSD2*	Intronic
ARS-BFGL-NGS-4613	rs110428369	15	53384296	1.62E-05		*PAAF1*	Intronic
BovineHD1600000346	rs110355156	16	1543848	9.96E-07			Intergenic
BovineHD1600000400	rs109375222	16	1771270	5.74E-07			Intergenic
BovineHD1900017689	rs41931384	19	61035099		2.09E-05		Intergenic
BovineHD2200013363	rs110741058	22	45910129	1.76E-05			Intergenic
DB-1451-seq-rs384691767	rs384691767	29	9510570	1.78E-05			Intergenic

Abbreviations: PP=protein percentage, SNP=single nucleotide polymorphism, Chr=chromosome, EBV=estimated breeding value, DRP=deregressed proof, UTR = untranslated region

* P-values have been reported only for the significative SNPs in each category.

Table 4.

List of SNPs above the Bonferroni (underlined) and false discovery rate 0.05 thresholds for DRP for MY trait, identified in the Aosta cattle breeds.

Marker	RS_SNP_ID	Chr	Position (bp)	P-Value EBV*	P-Value DRP	Gene	Position
BovineHD0600024355	rs110434046	6	87184768	1.17E-07	-		Intergenic
BovineHD0600024357	rs137712965	6	87187812	8.14E-08	-		Intergenic
BovineHD0600024093	rs110091883	6	86380145	3.50E-06	-		Intergenic
BovineHD0800031775	rs109292185	8	104696362	3.17E-06	-		Intergenic

Abbreviations: MY=milk yield, SNP=single nucleotide polymorphism, Chr=chromosome, EBV=estimated breeding value, DRP=deregressed proof.

[1] See Table 2 for further details.

* P-values have been reported only for the significative SNPs in each category.

Additional information can be found at https://doi.org/10.1016/j.animal.2024.101322

Figures

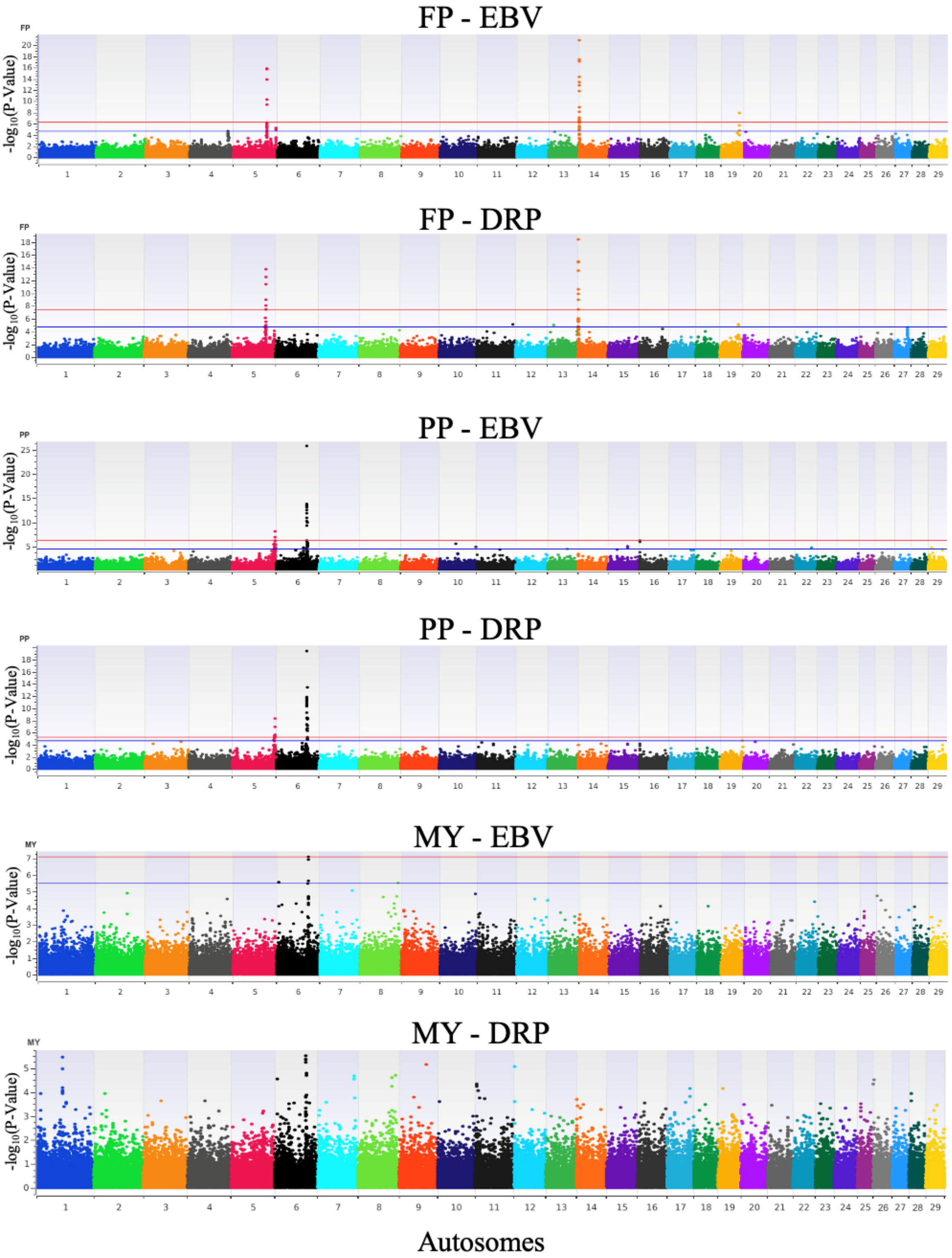


Fig. 1. Manhattan plots of the GWAS result for the FP, PP and for MY in the Aosta cattle breeds. Red and blue lines represent the Bonferroni and false discovery rate thresholds (both set at 5% genome-wide). Abbreviations: MY=milk yield, FP=fat percentage, PP=protein percentage, GWASs=genome wide association studies; EBV=estimated breeding value, DRP=deregressed proof.

4.2 Antibiotic treatments and somatic cell count as phenotype to map QTL for mastitis susceptibility in Holstein cattle breed

Maria Giuseppina Strillacci [a], Chiara Punturiero [a], Raffaella Milanesi [a], Francesca Bernini [a], Tiziano Mason [a], and Alessandro Bagnato [a]

[a] *Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano, Lodi, Italy*

Published in the *Italian Journal of Animal Science* (2023)
https://doi.org/10.1080/1828051X.2023.2175498

Abstract

Mastitis is one of the most significant diseases affecting dairy cattle profitability. A genome wide association study (GWAS) was performed using the selective genotyping approach to identify Quantitative Trait Loci (QTL) associated with mastitis susceptibility. According to the received antibiotic treatments in all their productive life, 52 lactating cows have been classified in resistant (no therapy and low SCC) and susceptible (more than one treatment and high SCC) to mastitis. Genotyping of animals was performed using the NEOGEN's GGP Bovine 100K SNP chip and QTL were identified comparing the SNPs allelic frequencies between the two groups. 26 SNPs related to mastitis susceptibility were identified in 9 chromosomes. The use of treatment data, coupled with SCC from milk recording, improved cow's classification accuracy in resistant and susceptible individuals. For this reason, the mandatory recording of treatments active from January 2022, could be a new source of information to improve the genetic selection for mastitis resistance.

Keywords: Mastitis, antibiotic treatment, genome wide association study, SCC, selective genotyping approach

Introduction

Mastitis is one of the most significant diseases affecting dairy cattle herds' profitability and is normally treated with antibiotic therapy. As reported by Mitchell et al. (1998) the most common use of antibiotic drugs in dairy cattle farming is to treat mastitis infection. In fact, clinical mastitis in lactating cows is mainly treated with intramammary administration of antibiotic while in severe cases of mastitis antibiotics are also used parenterally (Merle et al. 2013; Oliveira and Ruegg 2014; Gomes and Henriques 2016).

The blanket therapy method to reduce the prevalence of intramammary infections, foresee that all the animals of a herd are treated with antibiotic during the dry period. This practice led to an increase in bacteria antimicrobial resistance (Kabera et al. 2021). The Regulation (EU) 2019/6 on veterinary medicinal products, entered in to force in January 2022, promotes a more conscious use of antibiotic products (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0006&from=EN). To reduce the use of antibiotics, in the recent years, the selective dry method has been implemented in dairy farms. According to this approach only cows that show mastitis symptoms during lactation are treated.

Immune response to mastitis is a complex trait under the control of multiple genes (Naserkheil et al. 2022). With the availability of high throughput genotyping technologies, genome-wide association studies (GWAS) have become a useful tool for fine-scale quantitative trait loci (QTL) mapping

(Hayes et al. 2009). The discovery of genomic regions that influence quantitative (complex) traits may provide an opportunity to gather new information about gene function, and lead to a better understanding of how they interact to impact physiology and immune response to infectious diseases (Sharifi et al. 2020). In the recent past, many studies identified QTLs affecting mastitis susceptibility in dairy cattle, using genetically correlated measures such as Somatic Cell Count (SCC) and its logarithmic transformation Somatic Cell Score (SCS) (Sahana et al. 2014; Welderufael et al. 2018; Meier et al. 2020).

Given the complexity of this trait, genetic selection to increase immune resistance to mastitis in dairy cattle populations has been developed for decades using, as indicator, milk SCC (Thompson-Crispi et al. 2014; Moretti et al. 2021).

Selective DNA pooling (Darvasi and Soller 1994) is a recognised method used by several authors in GWAS analyses to map QTL either in case vs control studies or applied to complex traits (Strillacci et al. 2014; Lipkin et al. 2016; Peletto et al. 2017; Kurz et al. 2019). This approach is based on the theoretical demonstration that the most information deriving from a GWAS depends on the marker allelic frequencies of the best and worst individuals according to the phenotype distribution in the population: with a maximum of 25% per tail, the less and more extreme is the proportion of considered individuals, the more powerful is the statistical analysis (Darvasi and Soller 1992). With the selective genotyping and DNA pooling approach, it’s then possible to genotype only individuals showing the most extreme values in an extended sampling, thus reducing the costs of genomic analysis and maintaining high statistical power.

A key factor in a GWAS study to disclose QTL associated to mastitis resistance is to accurately classify cows in resistant and susceptible. In this study, to achieve an accurate classification of resistant and susceptible cows to mastitis, we proposed an innovative selection criterion based on multiple information on the same individual along all its productive life. The criterion here used has been based on the antibiotic treatments for mastitis across all the lactations of each cow, coupled with the available SCC data from routine milk tests registered during cows' productive life. We performed then a GWAS using a selective genotyping of individuals and the DNA pooling design approach and statistics to identify QTLs associated with mastitis susceptibility in Holstein dairy cows.

Materials and Methods

Ethical issue

All animal procedures performed in this study were carried out in accordance with the guidelines of the care and use of laboratory animals established by the Italian and UE laws (D. Lgs n. 2014/26, 2010/63/UE); the project operative procedures on animals were approved by the Animal Welfare Committee of the Università degli Studi di Milano (OPBA) protocol number 160_2019.

Sampling and genotyping

For this study, data from a total of 170 adult lactating Holstein cows from a single herd with at least 2 completed lactations were used. All the antibiotic treatments for mastitis done on each animal recorded by the owner, a professional veterinarian, as well as the SCC data from routine milk recording scheme were available for each cow since 2014 (the oldest cow was of 7th parity).

The use of data of only one herd allows to eliminate any environmental difference among management. Additionally, the owner is a professional veterinarian capable as such to clearly identify mastitis cases to be treated.

Animals were classified as susceptible and resistant using as primary criterion the antibiotic treatments provided to animals for mastitis: i) resistant cows didn't receive any treatment along their productive life; ii) susceptible cows received at least one treatment.

In addition, the information from milk SCC has been considered. Milk SCC values were from the official milking recording by ARAL, the Lombardy region farmers association (data available for each cow according to the AT4 ICAR system for all lactations). All the registered data for SCC for all the lactations of each cow were investigated to identify peaks of SCC also simultaneously to an antibiotic treatment. This information has been used to assign a Case Score (CS) as follows: CS=1 for any SCC test higher than 1 million cells/ml; CS=0.5 for any SCC value between 400,000 and 1 million cells/ml; CS=0 for SCC values below 400,000 cells/mL. The cut-off of 400,000 cells/mL has been chosen because representative of clinical mastitis events according to (Beaudeau et al. 1997).

More to less susceptible cows were then ranked according to number of treatments received (from zero to 9) and, within treatments CS (from zero to 6) (see Table S1 to visualise the CS values attributed to each cow).

There were 26 cows with no treatment and CS equal zero. Consequently 26 susceptible cows were identified according to number of treatments and CS. These cows represented the 15% best and worst cows, in the normal distribution of the population's phenotypes, respect to the trait "resistance to mastitis". The individual EBVs for SCC were explored to compare consistency with additive genetic values.

The 26 resistant and 26 susceptible cows (Table S1) were then sampled using a Tissue Sampling Unit to collect ear tissue. DNA was obtained using the Quick-DNA™ Miniprep Kit of Zymo Research (Zymo Research Corporation). SNP genotyping was performed on each individual with the NEOGEN's GGP Bovine 100K. Only SNPs located on the 29 autosomes, according to the bovine genome assembly ARS-UCD1.2, were considered in this study.

Statistical analysis

According to the selective DNA pooling design, both the resistant (R) and the susceptible (S) cows were randomly divided in two subgroups (R1 and R2 and S1 and S2, respectively), each of them representing a biological replicate. For each of the 4 subgroups (R1, R2, S1 and S2), the allele frequency at each marker was obtained. The frequency of alleles in each subgroup was obtained from the individual genotyping.

The frequencies were then analysed as in the selective DNA pooling design proposed by (Darvasi and Soller 1994). Here, instead of pooling before the genotyping, we have the individual genotyping data to calculate the allele frequency, reducing as such the possible error in DNA pooling and allele frequency estimate from pools and maintaining the power and the properties of the experimental design and analysis.

The multiple markers test

The statistical analysis was performed using the A allele frequencies at each marker, one of the two possible alleles (A or B) in a genotype. The frequencies were obtained separately for each subgroup of the tails (R1, R2, S1, S2), using the “genotype statistics by marker” function of Golden Helix's SVS software (SNP and Variation Suite v8.9 module, Golden Helix Inc., Bozeman, MT, USA).

We performed the GWAS analysis comparing the average A allele frequency at each SNP of resistant individuals with those of the susceptible ones, using an in-house script of the R software (version 4.0.5). Monomorphic markers were filtered out because non informative. Additionally, those SNPs lying in the top 5% of the absolute value of allele difference between replicates (i.e., R1 vs R2 and S1 vs S2) were also excluded from the analysis. After this editing out of the 89,784 SNPs on the autosomes, 58,285 SNPs were available for the association analysis.

A single-marker test for the marker-trait association was used, and the p-value for each marker was calculated as follows:

$$\mathrm{Ztest} = \mathrm{Dtest}/[\mathrm{SD}(\mathrm{Dnull})], \qquad (1)$$

where: Dtest is the difference of the A allele frequencies means among tails, and Dnull is the difference of the A allele frequencies means within tails.

The association analysis results were then visualised through a Manhattan plot made with the R package ‘qqman’ (Turner 2014). Both the Bonferroni and the FDR genome wide correction were used to set the 5% significance thresholds.

Gene annotation and functional analyses

The multi-species SNPchiMp v.3 database (Nicolazzi et al. 2015) was used to convert the Illumina SNP name of significant markers in the SNP rsID code (Reference SNP cluster ID).

The Variant Effect Predictor (VEP) tool of Ensembl database (https://www.ensembl.org/Tools/VEP) allowed to annotate the significant SNPs through the rsID codes according to the *Bos taurus* genome assembly ARS-UCD1.2 (Annotation Release: 106). The gene annotation analysis was performed using the Database for Annotation, Visualization and Integrated Discovery – DAVID (https://david.ncifcrf.gov/home.jsp) with default parameters.

Results and Discussions

One of the main weaknesses in QTL analysis for mastitis is the availability of accurate phenotype to cluster animals in susceptible and resistant. Genetic improvement has been performed based on SCC from milk recording tests. Nevertheless, the sole use of SCC data from milk recording to identify mastitis cases is suboptimal: a mastitis case can in fact happen within two subsequent milk recording tests occurring every 4 weeks in the AT4 system.

In this study, to classify in an accurate manner susceptible and resistant animals to mastitis, we used veterinary treatments provided to each cow along their complete productive life. To minimise false positive (and negatives) we considered only cows with at least 2 completed lactations. The resistant in particular (Table S1) were cows with at least 3 lactations without any treatment. To make the classification more robust we coupled the treatment with the SCC. The expectation was towards high SCC in correspondence of treatment cases. This was always the case in the 26 individuals selected in this study as susceptible, except one. This individual received 2 treatments in 3 lactations: the date of treatments was in between two subsequent milk recording tests for SCC.

All resistant individuals had CS equal zero and the average EBV for SCC was 106.5, with all EBVs greater than 100, value representing the population’s genetic basis. The susceptible had an average EBV for SCC of 99.2, but several of them had positive value (i.e., greater than 100). A possible interpretation is that the genetic evaluation is very effective in identifying resistant to mastitis individuals, while the susceptible ones may be misidentified due to the AT4 method. For this reason, the classification usually done in susceptible vs resistant animals based on EBV may be suboptimal in GWAS study aimed to map QTL.

In Figure 1 the scheme of the sampling and of the experimental design is summarised according to the criteria here used to cluster individuals in susceptible and resistant, i.e., case to control study.

GWAS results

The Manhattan plot in Figure 2 shows the *-log10(p)* of the GWAS analysis with the Bonferroni (red line) and the FDR (blue line) genome wide thresholds, both set at 0.05 significance value. The 3 SNPs over the red threshold are located in chromosomes 16 and 28.

We considered as QTL the region of about 400,000 base pair located upstream and downstream (± 200,000 bp) each significant SNP (Table S2). In Table 1, we reported the significant SNP and the genes falling within these regions. Twenty-six SNPs (annotated on 9 chromosomes) resulted located above the 5% FDR genome wide threshold (in bold, the ones over the 5% Bonferroni threshold). Among these SNPs, 16 were intragenic, all located in intronic positions. Only 10 SNPs lied between genes (intergenic). Gene enrichment results for the genes listed in Table 1 (± 200,000 bp) are reported in Table S3.

As shown in Figure 3 and in concordance with available literature, 21 of the genes listed in the Table 1 were already associated with immune response and mastitis resistance (n = 12), with adipogenesis and feed efficiency traits (n = 5) and with productive and reproductive traits (n = 4).

Genes involved in mastitis and in immune response

Among the genes associated with immune response and with mastitis, the *PTK2B* – Protein Tyrosine Kinase 2 Beta – gene plays a role in the regulation of humoral immune response and is required for macrophage polarisation and migration towards sites of inflammation[1]. This gene resulted associated with clinical mastitis in Holstein Chinese cattle via GWAS and through a Post-transcriptional Analysis (Yang et al., 2019). The same authors found that the PTK2B gene was down-regulated in peripheral blood leukocytes of cows affected by clinical mastitis as well as in vitro lipopolysaccharides (*E. coli*) –stimulated bovine mammary epithelial cells.

Other genes were found differentially expressed in case-control studies focused on mastitis resistance: the EPHX2 gene – Epoxide Hydrolase 2 – (Li et al., 2019), the PLXNA2 gene – Plexin A2 – (Asselstine, 2021), the DGL2 gene – Discs Large MAGUK Scaffold Protein 2 – (D. Wang et al., 2020), and the LYST gene – Lysosomal Trafficking Regulator – (Naserkheil et al., 2022). This last gene encodes for a protein involved in the transport of molecules into and from lysosomes and from lysosome-related organelles: loss of its function lead to the development of abnormally large lysosomes, and when the immune cells are involved, these altered lysosomes could interfere with the immune response to pathogens (Westphal et al., 2017).

---

[1] https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTK2B

Naderi et al. (2018) applying a random forest statistical approach, identified the GAS1 – Growth Arrest Specific 1 – as candidate gene involved in clinical mastitis resistance; the role of GAS1 in v morphology and physiology mechanisms proper of the udder has been suggested by (Jaggi et al., 1996). Finally, the PTPRZ1 – Protein Tyrosine Phosphatase Receptor Type Z1 – is one of the genes mapped closely to genomic regions found to be statistically associated with somatic cell content in dairy cows (Chen et al., 2015).

The remaining genes reported in Figure 3 are all involved in immune response: the LGALS8 (Galectin-8) gene encodes for a membrane receptor with the ability to activate intracellular bacterial killing and it would seem to have an important role in neutrophils migration into target tissues (Lopreiato et al., 2020). The AGK (Acylglycerol Kinase) gene is indispensable for the metabolic reprogramming of CD8+ T cell and for their function in immune responses (Hu et al., 2019); the TRIM35 (Tripartite motif TRIM35) gene has an important role on innate immunity against viral infections (Sun et al., 2020). Finally, also for PRKCQ Protein kinase C theta (PKC-θ) and PPP2R2A (Protein Phosphatase 2 Regulatory Subunit Balpha) genes an immunity role has been reported, as indicated in human studies (Anel et al., 2012; Prince et al., 2020; Li et al., 2021).

Genes involved in adipogenesis and feed efficiency traits

Food represents the energy input that the body uses for physiological or to cope with pathological processes, and for this reason feed efficiency is considered one of the most important characteristics in animal husbandry (Friggens et al., 2013). Over the years, several authors investigated a possible link between a prompt immune response and high feed efficiency in different species (Vigors et al., 2016; Zerjal et al., 2021). Among the genes identified in this study as possible candidate ones for mastitis resistance, two have been previously associated with feed efficiency traits: ETS1 – ETS Proto-Oncogene 1, Transcription Factor – (de Lima et al., 2020) and IPO11 – Importin 11. This latter is part of a group of genes harbouring SNPs that are part of a patented marker panel to select animals for feed efficiency[2]. IPO11 is also a nearby gene associated with SCS (Somatic Cell Score) in Holstein cattle (Welderufael et al., 2018).

Moreover, the PCP4 gene (Purkinje cell protein 4) has been identified as one of the major genes regulating feed efficiency in beef cattle (Machugh et al., 2019) and Amaral et al. (2019) found that the DSCAM – DS cell adhesion molecule – regulates the Residual Feed Intake in pigs. The homolog of this gene (DSCAM) in insects and crustaceans have an immunity function regulating their immune response to pathogens (Ng and Kurtz, 2020); even though the results of this study make us speculate

---

[2] https://patents.google.com/patent/US20210195876A1/en

that this gene could have an immunity role in the bovine species, there are no evidence of this function in the vertebrates.

The EBF2 (B-cell factor 2) gene promotes differentiation of brown adipocytes and controls its adaptive cold response (Angueira et al., 2020). Several studies have shown that adipocytes have immunological functions capable of employing and activating immune cells: adipocyte in fact, would appear to be an antigen-presenting cell (APC) expressing MHC class I and II molecules (Song and Deng, 2020).

Genes involved in productive and reproductive traits

Over the years, selection for increased production in specialised dairy breeds, such as the Holstein one, determined a correlated genetic response causing a reduction in reproductive efficiency and in the ability to cope with diseases and infections, such as mastitis.

In this research two candidate genes for mastitis resistance (RSU1 - Ras Suppressor Protein 1; B3GALNT2 - β-1,3-N-acetylgalactosaminyltransferase 2) were previously reported as associated with production milk yield (Farhadian et al., 2018; Poulsen et al., 2019). The CACNB2 (Calcium Voltage-Gated Channel Auxiliary Subunit Beta 2) gene here found associated with mastitis resistance was previously found associated with reproductive traits (Sammad et al., 2022) as well as the WEE2 (oocyte meiosis inhibiting kinase, also known as WEE1B) gene that has been associated with oocyte development in donkeys (F.-L. Zhang et al., 2022) and fertilisation failure in pigs and humans (Shimaoka et al. 2009; Hanna et al. 2020).

The existence of a clear genetic correlation between mastitis with production and fertility may explain why the genes here found as candidate ones for mastitis resistance were previously reported for other traits. In fact, according to Martin et al. (2019), the genetic correlation values between age at first insemination and clinical mastitis or considering SCS indirect measure are -0.04 and -0.24, respectively. The same authors estimated a more negative correlation value between first service to conception and clinical mastitis (-0.41). Also, between SCS and milk yield the genetic correlation values found in literature result negative (Haile-Mariam et al., 2001; Samoré et al., 2011).

Conclusions

In this study we used the data of antibiotic treatments for mastitis, recorded by the herd veterinarian on 170 Holstein cows for 7 years. These data were coupled with longitudinal recording of SCC allowing to rank individuals from the most susceptible (largest number of treatments along lactations and SCC peaks (recorded during the milk recording tests) to the most resistant (no treatments in at least 3 lactations, and no SCC peaks). The cows used in the GWAS analysis are then assumed to represent the extreme 15% more susceptible and more resistant of the underlying distribution of mastitis resistance.

Some of the genomic regions containing QTL identified in the present study were confirmed in previous studies that considered different phenotypes to map genetic basis of mastitis resistance: among them clinical mastitis, SCC and SCS. This suggests that the method used to classify the individuals with the medical treatment records, could improve the identification of QTL regions related to immune resistance to mastitis.

The new regulations on veterinary treatments make mandatory the recording of antibiotic prescription for each individuals treatment. The integration of this new database with routine milk recording data for SCC may represent then a new possibility to improve the genetic selection for mastitis resistance.

References

Amaral AJ, Bressan MC, Almeida J, Bettencourt C, Moreira O, Sá J, Gama-Carvalho M, Bessa R, Gama LT. 2019. Combining genome-wide association analyses and gene interaction networks to reveal new genes associated with carcass traits, meat quality and fatty acid profiles in pigs. Livest Sci. 220:180–189. https://doi.org/10.1016/J.LIVSCI.2018.12.025

Anel A, Aguiló JI, Catalán E, Garaude J, Rathore MG, Pardo J, Villalba M. 2012. Protein kinase C-θ (PKC-θ) in natural killer cell function and anti-tumor immunity. Front Immunol. 3:187.

Angueira AR, Shapira SN, Ishibashi J, Sampat S, Sostre-Colón J, Emmett MJ, Titchenell PM, Lazar MA, Lim H-W, Seale P. 2020. Early B cell factor activity controls developmental and adaptive thermogenic gene programming in adipocytes. Cell Rep. 30(9):2869–2878.

Asselstine V. 2021. Functional regulatory elements influencing mastitis in Holstein dairy cows.

Beaudeau F, Seegers H, Fourichon C, Hortet P. 1997. Association between milk somatic cell counts up to 400000 cells/ml and clinical mastitis in dairy cows. In: 4 Rencontres autour des recherches sur les ruminants, Paris (France), 4-5 Dec 1997. Paris (France): Institut de l'Elevage.

Chen X, Cheng Z, Zhang S, Werling D, Wathes DC. 2015. Combining genome wide association studies and differential gene expression data analyses identifies candidate genes affecting mastitis caused by two different pathogens in the dairy cow. Open J Anim Sci. 5(4):358–393.

Darvasi A, Soller M. 1992. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theoretical and applied genetics. 85(2):353–359.

Darvasi A, Soller M. 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics. 138(4):1365–1373.

de Lima AO, Koltes JE, Diniz WJS, de Oliveira PSN, Cesar ASM, Tizioto PC, Afonso J, de Souza MM, Petrini J, Rocha MIP, et al. 2020. Potential biomarkers for feed efficiency related traits in nelore cattle identified by co-expression network and integrative genomics analyses. Front Genet. 11:189.

Farhadian M, Rafat SA, Hasanpur K, Ebrahimi M, Ebrahimie E. 2018. Cross-species meta-analysis of transcriptomic data in combination with supervised machine learning models identifies the common gene signature of lactation process. Front Genet. 9:235.

Friggens NC, Brun-Lafleur L, Faverdin P, Sauvant D, Martin O. 2013. Advances in predicting nutrient partitioning in the dairy cow: recognizing the central role of genotype and its expression through time. animal. 7(s1):89–101.

Gomes F, Henriques M. 2016. Control of bovine mastitis: old and recent therapeutic approaches. Curr Microbiol. 72(4):377–382.

Haile-Mariam M, Bowman PJ, Goddard ME. 2001. Genetic and environmental correlations between test-day somatic cell count and milk yield traits. Livest Prod Sci. 73(1):1–13.

Hanna CB, Mudaliar D, John K, Allen CL, Sun L, Hawkinson JE, Schönbrunn E, Georg GI, Jensen JT. 2020. Development of WEE2 kinase inhibitors as novel non-hormonal female contraceptives that target meiosis †. Biol Reprod. 2020(2):368–377. https://doi.org/10.1093/biolre/ioaa097

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci. 92(2):433–443. https://doi.org/10.3168/jds.2008-1646

Hu Z, Qu G, Yu X, Jiang H, Teng X-L, Ding L, Hu Q, Guo X, Zhou Y, Wang F. 2019. Acylglycerol kinase maintains metabolic state and immune responses of CD8+ T cells. Cell Metab. 30(2):290–302.

Jaggi R, Marti A, Guo KE, Feng Z, Friis RR. 1996. Regulation of a physiological apoptosis: mouse mammary involution. J Dairy Sci. 79(6):1074–1084.

Kabera F, Roy J-P, Afifi M, Godden S, Stryhn H, Sanchez J, Dufour S. 2021. Comparing blanket vs. selective dry cow treatment approaches for elimination and prevention of intramammary infections during the dry period: A systematic review and meta-analysis. Front Vet Sci.:604.

Kurz JP, Yang Z, Weiss RB, Wilson DJ, Rood KA, Liu GE, Wang Z. 2019. A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. Immunogenetics. 71(1). https://doi.org/10.1007/s00251-018-1088-9

Li T, Gao J, Zhao X, Ma Y. 2019. Digital gene expression analyses of mammary glands from meat ewes naturally infected with clinical mastitis. R Soc Open Sci. 6(7):181604.

Li Z, Li Y, Wang X, Yang Q. 2021. PPP2R2B downregulation is associated with immune evasion and predicts poor clinical outcomes in triple-negative breast cancer. Cancer Cell Int. 21(1):13. https://doi.org/10.1186/s12935-020-01707-9

De Lima AO, Koltes JE, Diniz WJS, De Oliveira PSN, Cesar ASM, Tizioto PC, Afonso J, de Souza MM, Petrini J, Rocha MIP. 2020. Potential biomarkers for feed efficiency-related traits in nelore cattle identified by co-expression network and integrative genomics analyses. Front Genet. 11:189.

Lipkin E, Strillacci MG, Eitam H, Yishay M, Schiavini F, Soller M, Bagnato A, Shabtay A. 2016. The use of Kosher phenotyping for mapping QTL affecting susceptibility to Bovine respiratory disease. PLoS One. 11(4). https://doi.org/10.1371/journal.pone.0153423

Lopreiato V, Minuti A, Morittu VM, Britti D, Piccioli-Cappelli F, Loor JJ, Trevisi E. 2020. Inflammation, migration, and cell-cell interaction-related gene network expression in leukocytes is enhanced in Simmental compared with Holstein dairy cows after calving. J Dairy Sci. 103(2):1908–1913.

Machugh DE, Giuffra E, Alexandre PA, Naval-Sanchez M, Porto-Neto LR, Bento J, Ferraz S, Reverter A, Fukumasu H. 2019. Systems Biology Reveals NR2F6 and TGFB1 as Key Regulators of Feed Efficiency in Beef Cattle [Internet]. https://doi.org/10.3389/fgene.2019.00230

Martin P, Baes C, Houlahan K, Richardson CM, Jamrozik J, Miglior F. 2019. Genetic correlations among selected traits in Canadian Holsteins. Can J Anim Sci. 99(4):693–704.

Meier S, Arends D, Korkuć P, Neumann GB, Brockmann GA. 2020. A genome-wide association study for clinical mastitis in the dual-purpose German Black Pied cattle breed. J Dairy Sci. 103(11):10289–10298. https://doi.org/10.3168/jds.2020-18209

Merle R, Mollenhauer Y, Hajek P, Robanus M, Hegger-Gravenhorst C, Honscha W, Kaesbohrer A, Kreienbrock L. 2013. Monitoring of antibiotic consumption in cattle on agricultural farms. Berl Munch Tierarztl Wochenschr. 126(7–8):318–325.

Mitchell JM, Griffiths MW, McEwen SA, McNab WB, Yee AJ. 1998. Antimicrobial drug residues in milk and meat: causes, concerns, prevalence, regulations, tests, and test performance. J Food Prot. 61(6):742–756.

Moretti R, Soglia D, Chessa S, Sartore S, Finocchiaro R, Rasero R, Sacchi P. 2021. Identification of SNPs associated with somatic cell score in candidate genes in Italian Holstein Friesian bulls. Animals. 11(2):366.

Naderi S, Bohlouli M, Yin T, König S. 2018. Genomic breeding values, SNP effects and gene identification for disease traits in cow training sets. Anim Genet. 49(3):178–192.

Naserkheil M, Ghafouri F, Zakizadeh S, Pirany N, Manzari Z, Ghorbani S, Banabazi MH, Bakhtiarizadeh MR, Huq M, Park MN. 2022. Multi-Omics Integration and Network Analysis Reveal Potential Hub Genes and Genetic Mechanisms Regulating Bovine Mastitis. Curr Issues Mol Biol. 44(1):309–328.

Ng TH, Kurtz J. 2020. Dscam in immunity: A question of diversity in insects and crustaceans. Dev Comp Immunol. 105:103539. https://doi.org/10.1016/J.DCI.2019.103539

Nicolazzi EL, Caprera A, Nazzicari N, Cozzi P, Strozzi F, Lawley C, Pirani A, Soans C, Brew F, Jorjani H, et al. 2015. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. BMC Genomics [Internet]. 16(1):283. https://doi.org/10.1186/s12864-015-1497-1

Oliveira L, Ruegg PL. 2014. Treatments of clinical mastitis occurring in cows on 51 large dairy herds in Wisconsin. J Dairy Sci. 97(9):5426–5436.

Peletto S, Strillacci MG, Capucchio MT, Biasibetti E, Modesto P, Acutis PL, Bagnato A. 2017. Genetic basis of Lipomatous Myopathy in Piedmontese beef cattle. Livest Sci. 206:9–16.

Poulsen NA, Robinson RC, Barile D, Larsen LB, Buitenhuis B. 2019. A genome-wide association study reveals specific transferases as candidate loci for bovine milk oligosaccharides synthesis. BMC Genomics. 20(1):404. https://doi.org/10.1186/s12864-019-5786-y

Prince C, Mitchell RE, Richardson TG. 2020. An integrative multi-omics analysis of 16 autoimmune diseases and cancer outcomes highlights immune-cell regulatory mechanisms and shared genetic architecture. medRxiv.

Sahana G, Guldbrandtsen B, Thomsen B, Holm LE, Panitz F, Brøndum RF, Bendixen C, Lund MS. 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. J Dairy Sci. 97(11):7258–7275.

Sammad A, Zhang H, Shi R, Dong Y, Luo H, Chen Z, Liu L, Guo G, Liu A, Wang Y. 2022. A Post-GWAS Functional Analysis Confirming Effects of Three BTA13 Genes CACNB2, SLC39A12, and ZEB1 on Dairy Cattle Reproduction. Front Genet 13: 882951 doi: 103389/fgene.

Samoré AB, Román-Ponce SI, Vacirca F, Frigo E, Canavesi F, Bagnato A, Maltecca C. 2011. Bimodality and the genetics of milk flow traits in the Italian Holstein-Friesian breed. J Dairy Sci. 94(8):4081–4089.

Sharifi S, Pakdel A, Jahanbakhsh J, Aryan Y, Mahdavi A, Ebrahimie E. 2020. Molecular mechanisms of resistance to bovine mastitis. Livest Sci. 239:104068.

Shimaoka T, Nishimura T, Kano K, Naito K. 2009. Cell Cycle Critical effect of pigWee1B on the regulation of meiotic resumption in porcine immature oocytes. Cell Cycle [Internet]. 8(15):2375–2384. https://doi.org/10.4161/cc.8.15.9073

Song J, Deng T. 2020. The adipocyte and adaptive immunity. Front Immunol. 11:593058.

Strillacci MG, Frigo E, Schiavini F, Samoré AB, Canavesi F, Vevey M, Cozzi MC, Soller M, Lipkin E, Bagnato A. 2014. Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. BMC Genet. 15(1):1–10.

Sun N, Jiang L, Ye M, Wang Y, Wang G, Wan X, Zhao Y, Wen X, Liang L, Ma S. 2020. TRIM35 mediates protection against influenza infection by activating TRAF3 and degrading viral PB2. Protein Cell. 11(12):894–914.

Thompson-Crispi K, Atalla H, Miglior F, Mallard BA. 2014. Bovine mastitis: frontiers in immunogenetics. Front Immunol. 5:493.

Turner SD. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. Biorxiv.:5165.

Vigors S, O'Doherty J V, Kelly AK, O'Shea CJ, Sweeney T. 2016. The effect of divergence in feed efficiency on the intestinal microbiota and the intestinal immune response in both unchallenged and lipopolysaccharide challenged ileal and colonic explants. PLoS One. 11(2):e0148145.

Wang D, Liu L, Augustino S, Duan T, Hall TJ, MacHugh DE, Dou J, Zhang Y, Wang Y, Yu Y. 2020. Identification of novel molecular markers of mastitis caused by Staphylococcus aureus using gene expression profiling in two consecutive generations of Chinese Holstein dairy cattle. J Anim Sci Biotechnol. 11(1):1–17.

Welderufael BG, Løvendahl P, De Koning D-J, Janss LLG, Fikse WF. 2018. Genome-wide association study for susceptibility to and recoverability from mastitis in Danish Holstein cows. Front Genet. 9:141.

Westphal A, Cheng W, Yu J, Grassl G, Krautkrämer M, Holst O, Föger N, Lee K-H. 2017. Lysosomal trafficking regulator Lyst links membrane trafficking to toll-like receptor–mediated inflammatory responses. Journal of Experimental Medicine. 214(1):227–244.

Yang F, Chen F, Li L, Yan L, Badri T, Lv C, Yu D, Zhang M, Jang X, Li J. 2019. Three novel players: PTK2B, SYK, and TNFRSF21 were identified to be involved in the regulation of bovine mastitis susceptibility via GWAS and post-transcriptional analysis. Front Immunol.:1579.

Zerjal T, Härtle S, Gourichon D, Guillory V, Bruneau N, Laloë D, Pinard-van der Laan M-H, Trapp S, Bed'hom B, Quéré P. 2021. Assessment of trade-offs between feed efficiency, growth-related traits, and immune activity in experimental lines of layer chickens. Genetics Selection Evolution. 53(1):44. https://doi.org/10.1186/s12711-021-00636-z

Zhang F-L, Zhang S-E, Sun Y-J, Wang J-J, Shen W. 2022. Comparative transcriptomics uncover the uniqueness of oocyte development in the donkey. Front Genet.:85.

Disclosure statement

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

Author contributions

MGS and AB: Conceived and designed the study; CP: participated in data interpretation and revised the manuscript; RM: performed the experiments, MGS and FB: data analysis; TM: collected all data and sampled animals; MGS and AB: wrote and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding

This project was founded by EAFRD Rural Development Program 2014-2020, Management Autority Regione Lombardia - OP. 16.1.01 Project ID n. 201801062430 – “Operational Group EIP AGRI” https://ec.europa.eu/eip/agriculture/en/eip-agriprojects/projects/operational-groups.

Data availability statement

The data are still under embargo as part of a project still active.

Tables

Table 1. Genome Wide Association Study (GWAS) results: SNPs over the Bonferroni (bold) and False Discovery Rate (FDR) genome wide thresholds (0.05 significance value) together with candidate genes.

QTL n	rs SNP_ID	SNP	Chr	bp	P	-log10 (P)	Gene*	Tot. N. of SNPs in ± 200,000 bp from significant SNP**	Upstream gene	Downstream gene
1	rs110002854	BovineHD0100040526	1	140133491	7.79E-06	5.11	DSCAM	13	PCP4	
2	rs41255939	BovineHD0100040576	1	140200349	7.79E-06	5.11	DSCAM	14	PCP4	
3	rs109241831	BovineHD0400020205	4	72505962	2.37E-06	5.63	intergenic	8	ADAM22, SRI, STEAP4	
4	rs43410876	BovineHD0400024205	4	86508557	2.37E-06	5.63	PTPRZ1	9		AASS
5	rs136624841	BovineHD0400024691	4	88411333	1.83E-07	6.74	intergenic	6	SPAM1	TMEM229A
6	rs110017402	ARS-BFGL-NGS-15450	4	104990332	7.79E-06	5.11	WEE2	12	AGK, KIAA1147	SSBP1, TAS2R3, TAS2R4, TAS2R5, PRSS37, OR9A4
7	rs110740218	BovineHD0400033015	4	113478441	7.79E-06	5.11	intergenic	15	TMEM176B, TMEM176A, AOC1	KCNH2, NOS3, ATG9B, ABCB8
8	rs109301545	BovineHD0800022352	8	73389421	1.93E-06	5.71	intergenic	10	EBF2	PPP2R2A
9	rs41257363	Hapmap57092-ss46526915	8	74384468	2.37E-06	5.63	TRIM35	7	STMN4	PTK2B, CHRNA2
10	rs136747884	BovineHD0800022668	8	74425816	2.37E-06	5.63	PTK2B	6	STMN4, TRIM35	CHRNA2, EPHX2
11	rs41668712	Hapmap38552-BTA-99266	8	80228616	7.79E-06	5.11	intergenic	12	GAS1	
12	rs134625158	BovineHD1300004878	13	17015829	7.79E-06	5.11	PRKCQ	13		
13	rs136055233	BovineHD1300004890	13	17044866	7.79E-06	5.11	PRKCQ	14		
14	rs110714783	BovineHD1300009067	13	30887849	7.79E-06	5.11	intergenic	5		PTER, C1QL3, RSU1
15	rs41683068	BovineHD1300009702	13	32964826	7.79E-06	5.11	CACNB2	8		NSUN6, EPC1
16	rs29026933	Hapmap52547-rs29026933	13	33163943	2.37E-06	5.63	EPC1	8	CACNB2	KIF5B
17	rs41568633	BovineHD4100012751	16	74734877	7.79E-06	5.11	intergenic	12		
18	**rs110512949**	**BovineHD1600022337**	**16**	**75276029**	**2.03E-11**	**10.69**	**PLXNA2**	**11**		**CD34**
19	rs29017498	BovineHD4100014559	20	16909623	2.37E-06	5.63	IPO11	7		DIMT1, KIF2A
20	rs137039912	BovineHD2300010969	23	38130211	7.79E-06	5.11	intergenic	6	ID4	

21	**rs43681636**	**BovineHD2800001676**	**28**	**5559644**	**1.11E-08**	**7.95**	**intergenic**	**10**		**MAP10**
22	rs133891814	BovineHD2800002454	28	8195446	7.79E-06	5.11	B3GALNT2	12	GGPS1, TBCE	GNG4, LYST
23	rs133968650	BovineHD2800002457	28	8199021	7.79E-06	5.11	B3GALNT2	12	GGPS1, TBCE	GNG4, LYST
24	**rs137468382**	**BovineHD2800002736**	**28**	**8997685**	**1.06E-10**	**9.97**	**intergenic**	**4**	**GPR137B, ERO1B**	**EDARADD, LGALS8**
25	rs133507315	BovineHD2900003304	29	11356071	6.78E-07	6.17	DLG2	8		
26	rs137442517	BovineHD2900009666	29	31826445	2.37E-06	5.63	ETS1	9		

*Genes where significative SNPs are located. Intergenic = SNPs located in a genomic region between two genes. **List of these SNPs are reported in Table S2.

Figures

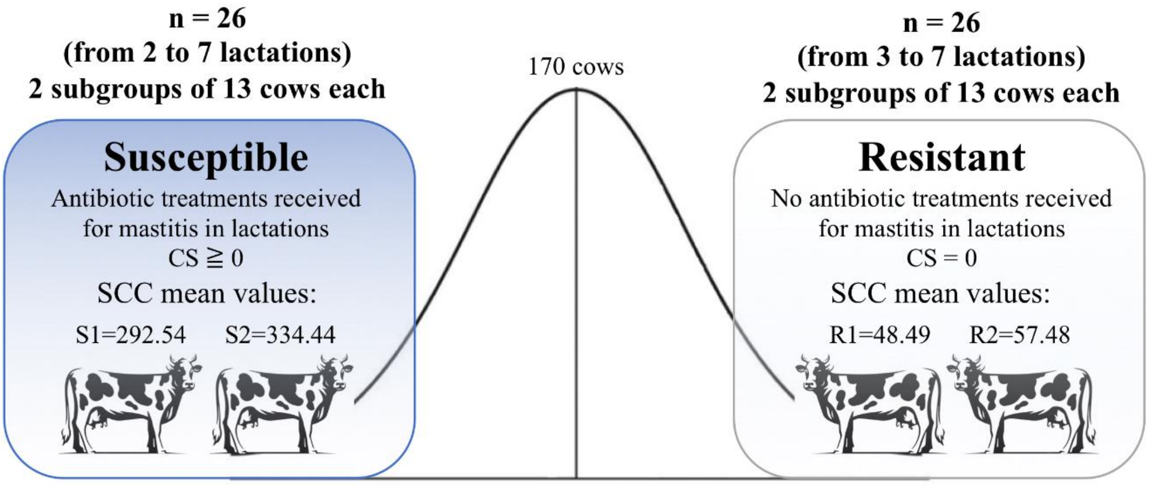


Figure 1. Experimental design scheme: Susceptible (S1 and S2) and Resistant (R1 and R2) cows in a Selective genotyping and pooling approach.

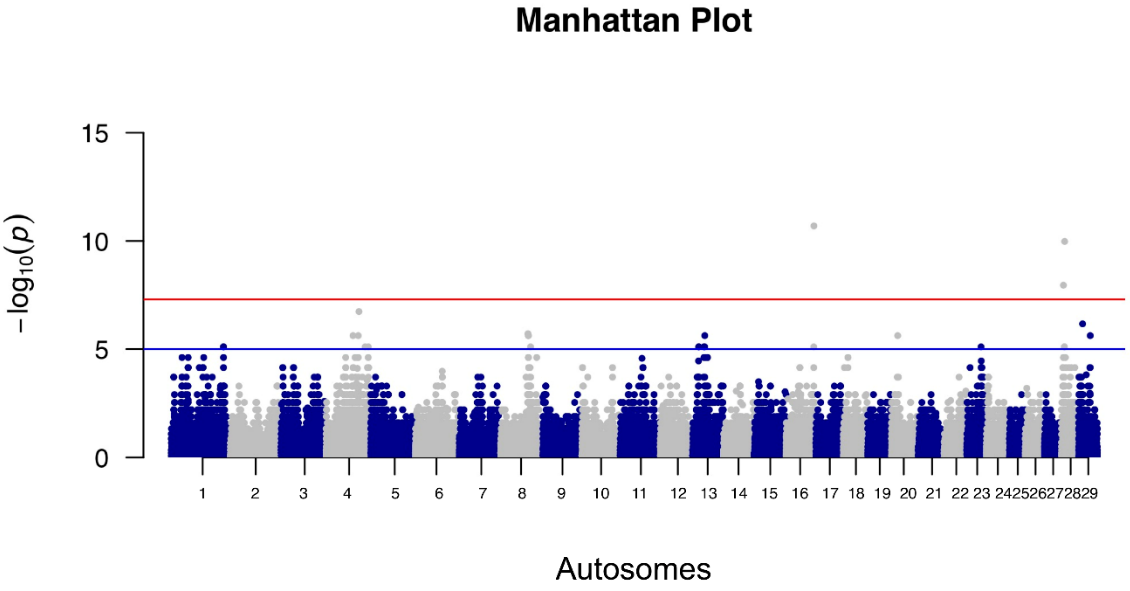


Figure 2. Graphical representation (Manhattan plot) of Genome Wide Association Study (GWAS) results. Horizontal lines represent the Bonferroni (red line) and the False Discovery Rate (FDR) (blue line) genome wide thresholds, both set at 0.05 significance value.

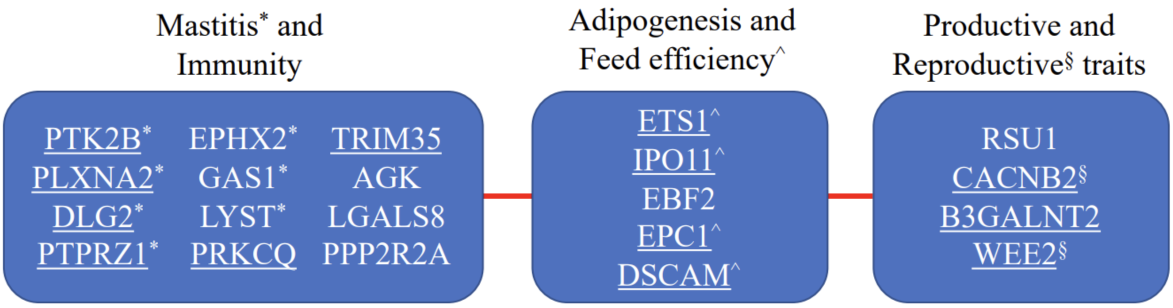


Figure 3. Scheme of gene functions: underlined genes were those in which significant SNPs mapped (intragenic SNPs – all in intronic positions); the others were nearby-genes (Table 1).

Additional information can be found at https://doi.org/10.1080/1828051X.2023.2175498

4.3 Genomic analysis of bovine respiratory disease resistance in pre-weaned dairy calves diagnosed by a combination of clinical signs and thoracic ultrasonography

Maria G. Strillacci [1#], Vincenzo Ferrulli [1,#,] Francesca Bernini [1], Davide Pravettoni[1], Alessandro Bagnato[1], Ilaria Martucci[1], Antonio Boccardo[1]

[1] *Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano, Via dell'Università 6, 26900 Lodi, Italy.*
[#] *contributed equally*

Submitted to *PloS One*

Abstract

Bovine respiratory disease (BRD) poses a significant risk of morbidity and mortality in pre-weaned dairy calves. Research indicates that this multifactorial disorder can be attributed to the involvement of various pathogens. Currently, there is little information from genome-wide association studies (GWAS) for BRD resistance in young calves lack a classification of BRD based on objective measures and clssifications of the desease. In this study, we moved a step forward in phenotyping BRD by coupling two diagnostic tests, the thoracic ultrasonography (TUS) and Wisconsin respiratory score (WISC), in order to classify susceptible and resistant animals to BRD. A total of 240 individuals were scored for BRD using TUS and WISC. A GWAS was performed using a selective genptyping approach to identify Quantitative Trait Loci (QTL) for BRD resistance. A total of 47 calves classified as BRD resistant (TUS ≤ 1 / WISC ≤ 4) and 47 as BRD susceptible (TUS = 5 / any WISC) were genotyped with the Neogen's GGP Bovine 100K SNP chip. QTL were then identified comparing the SNPs allelic frequencies between the two groups. A total of 28 QTL regions (QTLRs) were defined according to significative SNPs, 141 genes were annotated in the defined QTLRs. These genes were functionally classified into 4 main categories, i.e., i) regulation of systemic arterial blood pressure, ii) fertility, iii) immune function, and iv) filament cytoskeleton. Furthermore, 61 out of 141 genes identified here, can be considered promising candidate genes since they were already associated with BRD resistance in published GWAS studies in dairy cattle. The *ASB9, BMX, EPSTI1, and OLFM4* were identified in 4 of the 6 cosidered studies. This study paves the way for further research to mine the genome for resistance to respiratory diseases such as BRD in cattle, utilizing an accurate classification process.

Introduction

Bovine respiratory disease (BRD) is a common medical condition that affects cattle and causes significant economic losses worldwide [1]. It is also one of the main reasons for using antimicrobials in cattle [2]. Despite the scientific community's extensive focus on BRD in recent decades, in pre-weaned dairy calves' morbidity and mortality are still estimated at 12% and 14%, respectively, in the United States [3,4].

The occurrence of BRD in calves is likely attributed to several significant factors related to this disease. Bovine respiratory disease is, in fact, a multifactorial disorder with known predisposal factors and environmental components, host-pathogen interaction, and external stressors. The immune system plays an important role in influencing the individual immune response, development, and progression of the disease throughout the individual's life [5].

A significant challenge in managing BRD resistance is the lack of a field gold-standard scoring diagnostic tool for discriminating affected vs. non-affected individuals, both in pre-weaning calves and older individuals. Recent studies have compared thoracic ultrasonography (TUS) with clinical signs to enhance the diagnosis of BRD. The findings indicate that TUS is an accurate method for the in vivo assessment of lung lesions during BRD-related suppurative bronchopneumonia episodes, the most common lower respiratory tract infection in young dairy calves. Thoracic ultrasonography TUS has demonstrated good inter-rater agreement [6] and satisfactory accuracy in detecting active bronchopneumonia [7,8]. Longitudinal studies focusing on lung lesions assessed by TUS in preweaned dairy calves have shown that farm management practices, including the types of rearing, affected the dynamics of the morbidity and the severity of the disease [9–12].

To our knowledge, genome-wide association studies (GWAS) for BRD in calves using lung lesions detected by TUS combined with a clinical scoring system was recently studied in US Holstein [13]. Similarly, a recent study reported nineteen differentially expressed genes in peripheral leukocytes of pre-weaned Holstein calves with respiratory disease diagnosed with TUS and clinical signs were reported [14]. Another GWAS for BRD resistance based on objective measures was performed by Lipkin et al. (2018) [15], who recorded lung lesions at the slaughterhouse and identified QTL related to BRD resistance. Furthermore, other findings support the hypothesis that a genetic component is involved in individual immune response and, the estimated heritability for different pathogen susceptibility ranges from 0.04 to 0.24 [13,16–20].

In many studies, the evaluation of the BRD phenotype relies on subjective assessments of clinical signs or drug treatments reported by breeders. However, these assessment methods may not specifically address lower airway disease, the primary pathological challenge associated with BRD. A classification system that combines TUS and clinical signs can help mitigate the limitations of existing field data, mainly when performed in several herds, excluding a confounding effect between genetic and environmental factors [21]. Additionally, GWAS based on a selective genotyping approach allows the creation of a case-control design by selecting the two extreme tails of the recorded phenotype [22]. This experimental design, in association with the comparison of the allele frequencies of markers in the two groups [23], showed its effectiveness in the discovery of quantitative trait locus (QTL) in many studies [24–26] and was applied previously by Lipkin et al. (2018) specifically on BRD resistance [15]. The objectives of this study were i) to use TUS scoring system and clinical signs to assess BRD in multiple herds as an effective method for collecting reliable phenotypic records of resistant and susceptible female calves under field conditions and ii) to use the obtained information in a selective genotyping GWAS to identify the QTL related to this disease.

Materials and methods

Animal sampling, genotyping and ethics statement

For this study, the clinical information of 240 Holstein female calves from 10 commercial dairy farms (minimum and maximum number of pre-weaning calves per herd was 10 and 43, respectively) located in northern Italy, were available from previous studies [27,28]. Farms were selected for a history of cough in pre-weaned Holstein heifers housed in multiple pens with automatic calf feeders, with no registered antimicrobial or anti-inflammatory treatments in the 15 days before the study. Preweaned calves were randomly subjected to a systematic bilateral TUS using ventral landmarks described by [29]. The assessment was performed on the right side at intercostal space [ICS] 10-1 and the left at ICS 10-2 using a portable ultrasound unit equipped with a 7.5 MHz linear transducer (Esaote MyLab Five Vet, Esaote S.p.A., Genova, Italy). Data on lung lesions (type of lesions or depth in cm for consolidation) from each intercostal space were manually transferred into an Excel spreadsheet and simultaneously evaluated. The severity of the disease was examined and assessed by the mass of lung tissue involved according to the methods described by Ollivett et al., (2016) [30]. Thoracic ultrasonography has been, therefore, assigned on a 0-to-5-point scale. Briefly: TUS=0 indicated normal aerated lung parenchyma (with none or few comet-tail artifacts); TUS=1 indicated diffuse comet-tail artifacts without consolidation; TUS=2 consisted of lobular or patchy pneumonia, a pulmonary consolidation of >1 cm depth among normal aerated lung parenchyma; TUS=3 consisted of a full-thickness consolidation of one lung lobe; TUS=4 lobar pneumonia affected two lobe were evaluated individually. Each enrolled calf underwent clinical evaluation using the Wisconsin Respiratory Score (WISC), where 0 to 3 points were assigned for each of the following categories (i.e., 0 = normal, 3 = severely abnormal): i) rectal temperature; ii) nasal discharge; iii) ocular discharge; iv) cough; v) ear position. The eye and ear scores were considered collectively, and the higher of the two was used. A total WISC was then assigned to each calf as the sum of all categories, i.e., values can range from 0 to 12.

To perform the association study with the selective genotyping approach, the TUS and WISC scores (BRD score) were considered. Out of the 240 scored individuals, calves with a TUS score of 5 were considered susceptible to BRD (S-BRD), and the ones with a TUS score of 0 or 1 coupled with a total WISC score ≤ 4 without cough were classified resistant to BRD (R-BRD). S1 Table lists the scores of the 47 S-BRD and 47 R-BRD calves.

Residual blood samples, collected for clinical screening by Boccardo et al. (2023, 2024) [27,28] in their research, were stored at -80°C and then made available for this study, approved by the ethical committee of the University of Milan (approval number 2/16, February 15, 2016). In addition, the study was also approved by the Institutional Animal Care and Use Committee of the University of Milan (approval number 104/2020, January 15, 2020).

DNA from the residual blood of selected calves (47 samples of S-BRD and 47 samples of R-BRD representing the top and bottom 19% of the distribution of the scored animals that can be classifiable as a case vs control study) was extracted using the Quick-DNATM Miniprep Kit of Zymo Research (Zymo Research Corporation). After the DNA quality check (DNA integrity, purity and concentration were verified by standard procedures; i.e., 1% Agarose Gel and NanoQuant Infinite m200 - Tecan), and its dilution to a 50 ng/ul, the genotyping was performed with the Neogen's GGP Bovine 100K (GeneSeek®) SNP array by an external laboratory. SNP coordinates were mapped according to the bovine genome assembly ARS-UCD1.2.

Statistical analysis

To undertake a GWAS for BRD, we performed a selective genotyping [22] of animals with high versus low BRD scores (R-BRD and S-BRD animals), i.e., the top and bottom 19% of combined TUS and WISC scoring for BRD. Then, we combined the genotype data from the groups of animals at each extreme (R-BRD vs. S-BRD) and analyzed the genotype frequencies as if we had undertaken DNA pooling [23].

Genotypes of both selected R-BRD and S-BRD cows were randomly divided into two biological replicates for each of the two groups (R1-BRD and R2-BRD, S1-BRD and S2-BRD). The allele frequency at a single marker level was obtained for each replicate using the 'genotype statistics by marker' function of Golden Helix's SVS software (SNP and Variation Suite v8.9 module, Golden Helix Inc. Bozeman, MT, USA). Monomorphic and duplicated markers were filtered out because they were non-informative. Additionally, using an in-house R-script, the SNPs lying in the top 5% of the absolute value of allele difference between replicates (i.e., R1-BRD vs. R2-BRD, S1-BRD vs. S2-BRD) were also excluded from the analysis using an in-house R-script. After these edits, 68,256 SNPs (on the autosomes, BTAs) and 2,656 SNPs (on the chromosome X, BTX) were available for the association analyses.

Two separate genome association analyses, the first GWAS using autosomal SNPs and the second one using the polymorphism on chromosome X, were performed using an in-house script written in R as reported in Strillacci et al. (2023) [26]. Briefly, the p-value for each marker was calculated as follows:

$$\mathrm{Ztest} = \frac{\mathrm{Dtest}}{\mathrm{SD(Dnull)}}$$

Dtest is the difference of the allele frequencies among tails, and Dnull is the difference of the allele frequencies within tails. The GWAS results were visualized through a Manhattan plot with the R package 'qqman' [32]. The Bonferroni genome-wide correction was used to set the 5% significance threshold.

Gene Annotation

For all the SNPs over the 5% Bonferroni threshold the SNP's rsID code (Reference SNP cluster ID) of each Illumina SNP marker name was obtained using the multi-species SNPchiMp v.3 database [31]. The Variant Effect Predictor (VEP) tool of the Ensembl database [32] allowed the annotation of genes corresponding to the significant SNPs through the rsID codes according to the Bos taurus genome assembly ARS-UCD1.2 (Annotation Release: 106).

QTL regions (QTLRs) are here defined as the genome centered over single significant SNPs, or around more consecutive ones (from 2 to 5 SNPs in our study) when the distances of the consecutive SNPs mapped on the same chromosome were less than 1 Mbp, the gap between the positions of the first and last SNPs was calculated, and the difference respect to 1 Mbp was equally accounted in upstream (from the first SNP) and downstream (from the last SNP) positions, to cover the total length of 1 Mbp (these new positions defined the QTLR boundaries).

The list of genes annotated within these defined QTLRs was downloaded from the NCBI database using the tool integrated into the Genome Data Viewer (https://www.ncbi.nlm.nih.gov/gdv?org=bos-taurus). The functional enrichment analysis was then conducted using the ClueGo and Genemania tools of Cytoscape software [33–35].

Inbreeding coefficients

In order to investigate the possible role of inbreeding on the calf's capability to cope with an infection disease, for each anima, two different inbreeding coefficients (F_{HOM}, based on the excess in the observed number of homozygous genotypes [36] and F_{ROH}, based on the runs of homozygosity (ROH), that are continuous genomic homozygous segments inherited from common ancestors [37]) were both calculated using the SVS software. ROHs were identified setting the following parameters: i) a minimum number of 30 SNPs/ROH; (ii) a minimum ROH length of 1Mb, to avoid the identification of short and common ROH due to Linkage Disequilibrium; (iii) no missing SNPs and no heterozygous genotypes presence in ROH definition. F_{ROH} for all samples were calculated as:

$$\boldsymbol{F_{ROH}} = \frac{\sum_i^n \boldsymbol{L_{ROH}}}{\boldsymbol{L_{AUT}}}$$

where L_{ROH} is the length of the i^{th} ROH segment, n is the number of detected ROH and L_{AUT} is the length of the autosomal genome covered by the SNPs used in the GWAS.

Results

The results of WISC, TUS, age, and farm number of the S-BRD and R-BRD are included in the S1 Table. Fig 1 shows the two Manhattan plots of the genome-wide association analyses performed on BTAs (A) and BTX (B), with the Bonferroni (red line) threshold set at 0.05.

Twenty-seven QTLRs, defined by 45 significant SNPs, were identified on 11 autosomes, and only one QTLR was defined on BTX by 2 significant SNPs, as reported in Table 1. For 19 SNPs the p-value was greater than 1.40E-07, as being located over the Bonferroni multiple testing correction threshold of 1%. The highest number of QTLRs was identified on BTA 9 (n. 8). The significant markers were annotated as intronic markers (n. 21), 3_prime_UTR variants (n. 1), and intergenic SNPs (n. 21).

A total of 141 annotated genes (protein-coding genes) were found, including 4 miRNA (Table 1). Furthermore, 53 and 59 genes were distributed respectively in upstream and downstream positions respect to significant SNPs, 20 genes were the ones in which significant SNPs mapped, and 9 genes covered all (e.g., PLCB1 and FSTL5) or part (e.g., PIK3C2G, FILIP1, MEI4, SYNJ2, RYR2, OPCML, and NTM) of the QTLR 1 Mbp length region.

Out of the 141 genes, 62 were available and annotated with the Cytoscape software and subsequently included in the gene networks shown in Fig 2 (only networks with at least 3 terms were displayed). All these genes were significantly (P≤ 0.05) belonging to 58 GO Terms (n. 55) and KEGG pathways (n. 3), and 57 (92%) appeared in more than one term (S2 Table), mainly involved in immune response, fertility, and blood pressure regulation processes.

Inbreeding coefficients

The descriptive statistics for FHOM and FROH coefficients are reported in Table 2. Both coefficients resulted in slightly lower values in R-BRD than those calculated for S-BRD calves.

Discussion

The key factor in disclosing QTL associated with resistance or susceptibility to a specific disease is the correct classification based on a chosen phenotype. In our study, calves were identified as affected or non-affected by BRD using a quantitative scoring system derived from the combination of TUS, a noninvasive method for the visualization of lung lesions with high diagnostic accuracy [38,39], and clinical signs scored with WISC. The TUS scoring method is recognized to have a high sensitivity and specificity [40,41] and allows the classification of calves in all 5 TUS scoring across the herds here sampled. It is worth mentioning that the classification obtained for BRD resistance is objective and can be proposed as a field classification method to score pre-weaning calves.

We assumed that the R-BRD and S-BRD calves underwent the same management factors and environmental conditions: within each herd, they were farmed in the same pens, closely located, and, therefore, assumed to be equally exposed to BRD pathogens. It is additionally to be noticed that they were similar in age and had at least 15 days of minimum exposure to pathogens in the same pen. Additionally, the 10 herds here used were sampled among a much wider number of farms routinely visited by the Veterinary Mobile Clinic of our Department.

As calves were scored between February 2021 and February 2022, we can identify those individuals that are still in production after three years, among those genotyped. The survival rate of calves (now heifers or first calving cows) was verified by consulting the ANAFIBJ website (Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana), which allows access to information on the animals genotyped and registered in the herd book using the identification number.

The survival rate of female calves was notably higher for R-BRD calves (80%) compared to the 56% for the S-BRD ones. While this survival rate is not directly linked to BRD, it does raise the possibility of BRD's influence on the female population in the herd. The literature reports that among the risk factors involved in BRD mortality there is an inbreeding level, possibly related to inbreeding depression (ID) [42,43]. We have calculated here the inbreeding based on genomic information, i.e., F_{HOM} and F_{ROH}. Even if both coefficients resulted slightly lower in

R-BRD calves, the differences in F_{HOM} and F_{ROH} cannot explain by per se the diverse capability of an animal to cope with this multi pathogens disease.

In this study, a total of 141 significant SNPs were associated with BRD susceptibility vs. resistance. The most significant SNP, the rs109072042, was found in QTLR_18 where 5 other significant SNPs mapped in intronic positions of the PLCB1 gene on BTA13. This gene seems to play a role in inhibiting endothelial inflammation, and as reported by Gohari et al. (2023) [44] the expression of PLCB1 is positively correlated with chronic obstructive pulmonary disease. On BTA9, more significant SNPs (n. 15) were mapped, and three were located within the HIVEP2 gene. Schumann et al. (2020) reported that this gene influenced regulatory T-cell function in suppressing inflammation [45]. On the same BTA, some significant SNPs were annotated in intronic position, one in the FILIP1 gene, involved in skeletal muscle cell differentiation [46], three in the FSTL5 gene, implicated in activation specific protein of the caspase pathway mediating the apoptosis process [47], and one in the EPSTI1 gene, a modulator of macrophage activation with a suggested role in immunotherapies against inflammatory diseases [48]. The functional gene annotation, conducted with the ClueGO tool, revealed that the candidate genes annotated in the 28 QTLRs can be grouped into four main categories: regulation of systemic arterial blood pressure, fertility, immune function, and filament cytoskeleton. These categories and their associated genes could play a significant role in determining susceptibility vs. resistance to BRD.

The ACE2 gene, assigned to the regulation of systemic arterial blood pressure (S2 Table), is well-known to be involved in different respiratory diseases such as COVID-19, severe acute respiratory syndrome (SARS), and Influenza [49], having an important role in body homeostasis, not only through the blood pressure regulation but also with the electrolyte balance [50]. Nine of the 13 genes assigned to immune function were already associated with BRD, as reported in Table 2. Among the other four genes, TIGIT and CSF3 could affect the efficiency of the immune response against BRD pathogens, with an opposite contribution. In fact, CSF3 gene encodes for G-CSF growth factor that regulates different aspects of neutrophil biology. A study in mice reported that a loss of G-CSF signaling impaired neutrophil mediated immunity and caused an increased susceptibility to challenge with bacterial pathogens [51]. TIGIT seem to promote T cell dysfunction by altering their phenotype and cytokine profile during chronic viral infection [52].

Both in humans and cattle, the interplay between immunology and fertility has been reported in several studies, and it is well-known that reproductive hormones are important for the immune system function [53], as well as that the immune system affects fertility efficiency and subsequently the pregnancy outcomes [54]. For instance, the ABHD2 gene is considered a modulator of sperm fertility, and it is also involved in virus propagation and immune response [55]. Different functions have also been recognized also for the MFGE8 gene, including promoting the apoptotic cells’ phagocytosis and angiogenesis and its implication in human implantation as well as in other species [56,57]. Finally, the Keratin protein genes (KRTs, n. 15) annotated in the downstream position of QTLR_20 and added within the filament cytoskeleton category by ClueGO, are abundant in lung epithelial cells, and act modulating the cell functions as receptor signalling, protein proliferation and migration, inflammatory and immune responses [58].

The comparison with the current literature (Table 3), focused on discovering the genetic basis of BRD resistance or susceptibility, allowed us to identify a total of 61 protein-coding genes (45% of the 141 genes annotated within our QTLRs) that were already associated with this trait. Among these genes, 14 were found in at least two different studies, and ASB9, BMX, EPSTI1, and OLFM4 were reported as common candidate genes in three and four studies focused on BRD, respectively. The BMX gene also resulted as one of the differentially expressed genes identified when comparing BRD with non-BRD animals, as reported by Gohari et al. (2023) [59]. Interestingly, the largest number of genes found in this study and literature refers to Tizioto et al. (2015), who identify genes involved in immune response after an experimental challenge with BRD agents, making this approach a possible benchmark for a case control study for BRD [60]. The phenotyping strategy used here to cluster S-BRD vs. R-BRD was accurate because our results largely overlap those Tiozoto et al. (2015) found in experimental conditions [62].

The success in GWAS studies to identify QTL related to phenotype variation is in fact directly related to the capability to objectively score a trait. This is very simple for quantitative measures such as milk yield, but more difficult for complex traits related to immune response. Among these traits, BRD or other disease related traits of great impact in dairy cattle where knowledge of genomic components may open the opportunities and challenges in sustainability of farming systems [21].

Figure 3 shows a gene network built with the Genemania tool (implemented in Cytoscape) using 58 of the 61 genes already associated with BRD (the ones reported in Table 3, black full circle in Figure 3). In the same network, the gray full-circle genes (n. 20) were added by the Genemania tool representing the potential interaction partners predicted genes according to the references available in its database (background Homo sapiens). Seven of these gray genes (SH2D1A, TRIB2, MED1, TOP1, WEE1, MSH2, and NF1) resulted in differentially expressed comparing the control and various-pathogen-challenged animals involved in BRD [60], along with the EFNB2 (gray gene circled in red) is a differentially expressed in a different study performed with healthy and BRD affected animals [59], supporting our results in identifying QTL involved in immune response to BRD.

Conclusions

In this study, a total of 240 Holstein calves were scored with TUS and WISC to diagnose BRD and identify genes related to the immune response to the pathogen. These methods allowed an objective classification of resistant vs. susceptible calves. Overall, when disease phenotypes are used in mapping the genomic basis of immune response to infectious agents, classification is one of the most critical steps. The results of the present GWAS and a study developed in controlled experimental conditions [62] largely overlap, indicating that the clustering based on the attributed score well matches with the controlled experimental conditions. As the exploitation of genomes through association traits is a fundamental pillar of the correct phenotype assessment, we may relate the comprehensive list of genes identified in other research to the accuracy of the classification method used here. We can state that the TUS scoring is a valid option for classifying animals across studies, recalling that it is based on an objective classification.

With the Cytoscape software, 141 genes were used in total, and 40% of them appeared in more than one GO term with biological domains in immune response, fertility, and blood pressure regulation processes, as well as the filament cytoskeleton ones, suggesting a role in resistance or susceptibility to this disease.

The results of this study show promising outcomes, indicating that TUS diagnosis could be a valuable phenotype for GWAS efforts to understand the genetic variation contributing to BRD. classification in studies on genome variations underpinning BRD. The straightforward TUS approach combined with the WISC scores in filed conditions could become a standard method for phenotyping calves in herds. Genotyping all individuals with SNP chips could offer additional information and potential longitudinal data throughout the productive life of calves and cows, helping to further explore the relationship between innate and adaptive immune responses to BRD.

Acknowledgements

The authors gratefully acknowledge the farmers involved in this study and the support of the University of Milan's departmental fund, "Line 2 Research Support Plan."

References

[1] Overton MW. Economics of respiratory disease in dairy replacement heifers. Anim Health Res Rev 2020;21:143–8.

[2] Baptiste KE, Kyvsgaard NC. Do antimicrobial mass medications work? A systematic review and meta-analysis of randomised clinical trials investigating antimicrobial prophylaxis or metaphylaxis against naturally occurring bovine respiratory disease. Pathog Dis 2017;75:ftx083.

[3] USDA. Dairy 2014 Nutrient Management Practices on U.S. Dairy Operations, 2014. 2018.

[4] Urie NJ, Lombard JE, Shivley CB, Kopral CA, Adams AE, Earleywine TJ, et al. Preweaned heifer management on US dairy operations: Part V. Factors associated with morbidity and mortality in preweaned dairy heifer calves. J Dairy Sci 2018;101:9229–44.

[5] Panciera RJ, Confer AW. Pathogenesis and pathology of bovine pneumonia. Veterinary Clinics: Food Animal Practice 2010;26:191–214.

[6] Buczinski S, Buathier C, Bélanger AM, Michaux H, Tison N, Timsit E. Inter-rater agreement and reliability of thoracic ultrasonographic findings in feedlot calves, with or without naturally occurring bronchopneumonia. J Vet Intern Med 2018;32:1787–92.

[7] Berman J, Francoz D, Dufour S, Buczinski S. Bayesian estimation of sensitivity and specificity of systematic thoracic ultrasound exam for diagnosis of bovine respiratory disease in pre-weaned calves. Prev Vet Med 2019;162:38–45.

[8] Berman J, Masseau I, Fecteau G, Buczinski S, Francoz D. Comparison of thoracic ultrasonography and thoracic radiography to detect active infectious bronchopneumonia in hospitalized dairy calves. J Vet Intern Med 2021;35:2058–68.

[9] Rhodes V, Ryan EG, Hayes CJ, McAloon C, O'Grady L, Hoey S, et al. Diagnosis of respiratory disease in preweaned dairy calves using sequential thoracic ultrasonography and clinical respiratory scoring: Temporal transitions and association with growth rates. J Dairy Sci 2021;104:11165–75.

[10] Sáadatnia A, Mohammadi GR, Azizzadeh M, Mirshahi A, Mohieddini AA, Buczinski S. Effect of ultrasonographic lung consolidation on health and growth in dairy calves: A longitudinal study. J Dairy Sci 2023;106:8047–59.

[11] Jourquin S, Lowie T, Debruyne F, Chantillon L, Vereecke N, Boyen F, et al. Dynamics of subclinical pneumonia in male dairy calves in relation to antimicrobial therapy and production outcomes. J Dairy Sci 2023;106:676–89.

[12] Hinnant HR, Elder LA, Claus-Walker R, Mandella CM, Slanzon GS, Parrish LM, et al. Comparative diagnoses of respiratory disease in preweaned dairy calves using sequential thoracic ultrasonography and clinical respiratory scoring. Aust Vet J 2024;102:187–99.

[13] Quick AE, Ollivett TL, Kirkpatrick BW, Weigel KA. Genomic analysis of bovine respiratory disease and lung consolidation in preweaned Holstein calves using clinical scoring and lung ultrasound. J Dairy Sci 2020;103:1632–41.

[14] Elder LA, Hinnant HR, Mandella CM, Claus-Walker RA, Parrish LM, Slanzon GS, et al. Differential gene expression in peripheral leukocytes of pre-weaned Holstein heifer calves with respiratory disease. PLoS One 2023;18:e0285876.

[15] Lipkin E, Strillacci MG, Eitam H, Yishay M, Schiavini F, Soller M, et al. The use of Kosher phenotyping for mapping QTL affecting susceptibility to Bovine respiratory disease. PLoS One 2016;11. https://doi.org/10.1371/journal.pone.0153423.

[16] Henderson L, Miglior F, Sewalem A, Wormuth J, Kelton D, Robinson A, et al. Genetic parameters for measures of calf health in a population of Holstein calves in New York State. J Dairy Sci 2011;94:6181–7.

[17] Neibergs HL, Seabury CM, Wojtowicz AJ, Wang Z, Scraggs E, Kiser JN, et al. Susceptibility loci revealed for bovine respiratory disease complex in pre-weaned holstein calves. BMC Genomics 2014;15:1–19.

[18] Keele JW, Kuehn LA, McDaneld TG, Tait Jr RG, Jones SA, Smith TPL, et al. Genomewide association study of lung lesions in cattle using sample pooling. J Anim Sci 2015;93:956–64.

[19] Haagen IW, Hardie LC, Heins BJ, Dechow CD. Genetic parameters of calf morbidity and stayability for US organic Holstein calves. J Dairy Sci 2021;104:11770–8.

[20] Zhang H, Wang K, An T, Zhu L, Chang Y, Lou W, et al. Genetic parameters for dairy calf and replacement heifer wellness traits and their association with cow longevity and health indicators in Holstein cattle. J Dairy Sci 2022;105:6749–59.

[21] Lynch C, Leishman EM, Miglior F, Kelton D, Schenkel F, Baes CF. Opportunities and challenges for the genetic selection of dairy calf disease traits. Animal 2024:101141.

[22] Darvasi A, Soller M. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theoretical and Applied Genetics 1992;85:353–9.

[23] Darvasi A, Soller M. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 1994;138:1365–73.

[24] Kurz JP, Yang Z, Weiss RB, Wilson DJ, Rood KA, Liu GE, et al. A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. Immunogenetics 2019;71:35–47.

[25] Peletto S, Strillacci MG, Capucchio MT, Biasibetti E, Modesto P, Acutis PL, et al. Genetic basis of Lipomatous Myopathy in Piedmontese beef cattle. Livest Sci 2017;206. https://doi.org/10.1016/j.livsci.2017.09.027.

[26] Strillacci MG, Punturiero C, Milanesi R, Bernini F, Mason T, Bagnato A. Antibiotic treatments and somatic cell count as phenotype to map QTL for mastitis susceptibility in Holstein cattle breed. Ital J Anim Sci 2023;22:190–9.

[27] Boccardo A, Ferraro S, Sala G, Ferrulli V, Pravettoni D, Buczinski S. Bayesian evaluation of the accuracy of a thoracic auscultation scoring system in dairy calves with bronchopneumonia using a standard lung sound nomenclature. J Vet Intern Med 2023;37:1603–13.

[28] Boccardo A, Ferrulli V, Sala G, Scavone D, Paltrinieri S, Filippone Pavesi L, et al. Blood gases, acid-base, and metabolic alterations in calves with bronchopneumonia diagnosed via clinical signs and thoracic ultrasonography: A cross-sectional study. J Vet Intern Med 2024.

[29] Ollivett TL, Caswell JL, Nydam D V, Duffield T, Leslie KE, Hewson J, et al. Thoracic Ultrasonography and Bronchoalveolar Lavage Fluid Analysis in H olstein Calves with Subclinical Lung Lesions. J Vet Intern Med 2015;29:1728–34.

[30] Ollivett TL, Buczinski S. On-farm use of ultrasonography for bovine respiratory disease. Veterinary Clinics: Food Animal Practice 2016;32:19–35.

[31] Nicolazzi EL, Caprera A, Nazzicari N, Cozzi P, Strozzi F, Lawley C, et al. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. BMC Genomics 2015;16:283. https://doi.org/10.1186/s12864-015-1497-1.

[32] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. Genome Biol 2016;17:1–14.

[33] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

[34] Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009;25:1091–3.

[35] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010;38:W214–20.

[36] Wright S. Genetics of populations. vol. 10. Benton Chicago; 1948.

[37] Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet 2006;15:789–95.

[38] Pravettoni D, Buczinski S, Sala G, Ferrulli V, Bianchi F, Boccardo A. Diagnostic accuracy of focused lung ultrasonography as a rapid method for the diagnosis of respiratory disease in dairy calves. J Dairy Sci 2021;104:4929–35.

[39] Baxter-Smith K, More J, Hyde R. Use of thoracic ultrasound on Scottish dairy cattle farms to support the diagnosis and treatment of bovine respiratory disease in calves. Veterinary Record 2022;190:no-no.

[40] Rhodes V, Ryan EG, Hayes CJ, McAloon C, O'Grady L, Hoey S, et al. Diagnosis of respiratory disease in preweaned dairy calves using sequential thoracic ultrasonography and clinical respiratory scoring: Temporal transitions and association with growth rates. J Dairy Sci 2021;104:11165–75. https://doi.org/https://doi.org/10.3168/jds.2021-20207.

[41] Kamel MS, Davidson JL, Verma MS. Strategies for Bovine Respiratory Disease (BRD) Diagnosis and Prognosis: A Comprehensive Overview. Animals 2024;14:627.

[42] Mugambe J, Ahmed R, Thaller G, Schmidtmann C. Impact of inbreeding on production, fertility, and health traits in German Holstein dairy cattle utilizing various inbreeding estimators. J Dairy Sci 2024.

[43] Ablondi M, Summer A, Stocco G, Finocchiaro R, van Kaam J-T, Cassandro M, et al. The role of inbreeding depression on productive performance in the Italian Holstein breed. J Anim Sci 2023;101:skad382.

[44] Gohari K, Kazemnejad A, Mostafaei S, Saberi S, Sheidaei A. Chronic obstructive pulmonary disease: novel genes detection with penalized logistic regression. Cell Journal (Yakhteh) 2023;25:203.

[45] Schumann K, Raju SS, Lauber M, Kolb S, Shifrut E, Cortez JT, et al. Functional CRISPR dissection of gene networks controlling human regulatory T cell identity. Nat Immunol 2020;21:1456–66.

[46] Braga LG, Chud TCS, Watanabe RN, Savegnago RP, Sena TM, do Carmo AS, et al. Identification of copy number variations in the genome of Dairy Gir cattle. PLoS One 2023;18:e0284085.

[47] Parfenova OK, Kukes VG, Grishin D V. Follistatin-like proteins: structure, functions and biomedical importance. Biomedicines 2021;9:999.

[48] Kim Y-H, Lee J-R, Hahn M-J. Regulation of inflammatory gene expression in macrophages by epithelial-stromal interaction 1 (Epsti1). Biochem Biophys Res Commun 2018;496:778–83.

[49] Ahmed Z, Renart EG, Zeeshan S. Investigating underlying human immunity genes, implicated diseases and their relationship to COVID-19. Per Med 2022;19:229–50.

[50] Gintoni I, Adamopoulou M, Yapijakis C. The impact of ACE and ACE2 gene polymorphisms in pulmonary diseases including COVID-19. In Vivo (Brooklyn) 2022;36:13–29.

[51] Martin KR, Wong HL, Witko-Sarsat V, Wicks IP. G-CSF–a double edge sword in neutrophil mediated immunity. Semin Immunol, vol. 54, Elsevier; 2021, p. 101516.

[52] Schorer M, Rakebrandt N, Lambert K, Hunziker A, Pallmer K, Oxenius A, et al. TIGIT limits immune pathology during viral infections. Nat Commun 2020;11:1288.

[53] Lutton B, Callard I. Evolution of reproductive–immune interactions. Integr Comp Biol 2006;46:1060–71.

[54] Sen A, Kushnir VA, Barad DH, Gleicher N. Endocrine autoimmune diseases and female infertility. Nat Rev Endocrinol 2014;10:37–50.

[55] Baggelaar MP, den Dulk H, Florea BI, Fazio D, Bernabò N, Raspa M, et al. ABHD2 inhibitor identified by activity-based protein profiling reduces acrosome reaction. ACS Chem Biol 2019;14:2295–304.

[56] Schmitz C, Yu L, Bocca S, Anderson S, Cunha-Filho JS, Rhavi BS, et al. Role for the endometrial epithelial protein MFG-E8 and its receptor integrin αvβ3 in human implantation: results of an in vitro trophoblast attachment study using established human cell lines. Fertil Steril 2014;101:874–82.

[57] Barua S, Macedo A, Kolb DS, Wynne-Edwards KE, Klein C. Milk-fat globule epidermal growth factor 8 (MFGE8) is expressed at the embryo–and fetal–maternal interface in equine pregnancy. Reprod Fertil Dev 2018;30:585–90.

[58] Randell SH, Dang H, Broers JL V. The Use of Keratins as Lung Cell Differentiation Markers and as a Diagnostic Tool in Lung Cancer. Encyclopedia of Respiratory Medicine, Second Edition, Elsevier; 2022, p. 694–701.

[59] Jiminez J, Timsit E, Orsel K, Van der Meer F, Guan LL, Plastow G. Whole-blood transcriptome analysis of feedlot cattle with and without bovine respiratory disease. Front Genet 2021;12:627623.

[60] Tizioto PC, Kim J, Seabury CM, Schnabel RD, Gershwin LJ, Van Eenennaam AL, et al. Immunological response to single pathogen challenge with agents of the bovine respiratory disease complex: an RNA-sequence analysis of the bronchial lymph node transcriptome. PLoS One 2015;10:e0131459.

[61] Johnston D, Earley B, McCabe MS, Kim J, Taylor JF, Lemon K, et al. Elucidation of the host bronchial lymph node miRNA transcriptome response to bovine respiratory syncytial virus. Front Genet 2021;12:633125.

[62] Scott MA, Woolums AR, Swiderski CE, Finley A, Perkins AD, Nanduri B, et al. Hematological and gene co-expression network analyses of high-risk beef cattle defines immunological mechanisms and biological complexes involved in bovine respiratory disease and weight gain. PLoS One 2022;17:e0277033.

[63] Li J, Mukiibi R, Jiminez J, Wang Z, Akanno EC, Timsit E, et al. Applying multi-omics data to study the genetic background of bovine respiratory disease infection in feedlot crossbred cattle. Front Genet 2022;13:1046192.

[64] Green MM, Woolums AR, Karisch BB, Harvey KM, Capik SF, Scott MA. Influence of the At-Arrival Host Transcriptome on Bovine Respiratory Disease Incidence during Backgrounding. Vet Sci 2023;10:211.

Tables

Table 1. QTLRs identified in this study. QTLRs Start and End positions according to the ARS-UCD1.2 assembly.

QTLR_n.	Start	End	SNP	RS SNP code	CHR	BP	distance	value	Gene	SNP Position	upstream	in-between	downstream
QTLR_1	48950364	49950364	BovineHD0100014033	rs3423105562	1	49450364		5.69E-07		intergenic			ALCAM
QTLR_2	52682936	53682936	BovineHD0100015084	rs3423105354	1	53182936	3732572	4.80E-07	MYH15	intron	IFT57, HHLA2		DZIP3, CIP2A, TRAT1
QTLR_3	58199902	59199902	Hapmap47450-BTA-96948	rs41592061	1	58626471	5443535	2.02E-07	ZDHHC23	intron	SIDT1, USF3, NAA50, ATP6V1A, GRAMD1C		DRD3, TIGIT, MIR568
			BovineHD0100016729	rs109974499	1	58678558	52087	1.85E-07	CCDC191	intron			
			BTA-96949-no-rs	rs41663660	1	58773333	94775	2.02E-07	QTRT2	intron			
QTLR_4	102359117	103359117	BTA-39406-no-rs	rs3423093242	1	102859117	44085784	4.05E-07		intergenic	SI		
QTLR_5	75188352	76188352	BovineHD0300022027	rs109698273	3	75688352		3.13E-07		intergenic	LRRC7		
QTLR_6	91226361	92226361	BovineHD0500026165	rs5369491638	5	91695368		4.68E-09	PIK3C2G	intronic	CAPZA3, PLCZ1, PIK3C2G	PIK3C2G	PIK3C2G, RERGL
			UA-IFASA-4080	rs29027239	5	91757353	61985	5.16E-09	PIK3C2G	intronic			
QTLR_7	92296641	93296641	ARS-BFGL-NGS-116999	rs110267314	5	92796641	1039288	3.71E-07		intergenic			LMO3
QTLR_8	15128799	16128799	ARS-BFGL-NGS-1357	rs109782091	9	15129365		4.22E-07	FILIP1	intron	FILIP1	SENP6, MYO6, IMPG1	
			BovineHD0900004394	rs133351462	9	16128232	998867	6.46E-07		intergenic			
QTLR_9	17082837	18082837	BovineHD0900004814	rs135595961	9	17544244	1416012	4.22E-07	MEI4	intron	HTR1B, MEI4		
			BovineHD0900004841	rs136778805	9	17621429	77185	4.41E-07		intergenic			
QTLR_10	18370767	19370767	ARS-BFGL-NGS-46104	rs110957535	9	18870767	1249338	5.94E-07		intergenic	IRAK1BP1, PHIP, HMGN3		LCA5, SH3BGRL2
QTLR_11	20739564	21739564	BovineHD0900005695	rs42252864	9	20825772	1955005	1.94E-07		intergenic			FAM46A
			BTB-00384566	rs43588529	9	21653355	827583	9.50E-08		intergenic			
QTLR_12	79422330	80422330	BovineHD0900022313	rs137663324	9	79439745	57786390	4.22E-07		intergenic		NMBR, GJE1, VTA1, ADGRG6, MIR2284AA-4, HIVEP2	
			BovineHD0900022633	rs109987604	9	80355614	915869	4.60E-07	HIVEP2	intron			
			BovineHD0900022635	rs135739881	9	80357230	1616	4.60E-07	HIVEP2	intron			
			BovineHD0900022654	rs43732735	9	80404914	47684	1.49E-07	HIVEP2	intron			
QTLR_13	94091694	95091694	ARS-BFGL-NGS-103694	rs109848700	9	94591694	14186780	1.30E-07	SYNJ2	intron	ZDHHC14, SNX9, SYNJ2		SERAC1U60, GTF2H5, TULP4, TMEM181, DYNLT1, SYTL3
QTLR_14	98308661	99308661	BovineHD0900028984	rs110848300	9	98567810	3976116	3.71E-07		intergenic		QKI	
			BovineHD0900029149	rs109299906	9	99049511	481701	2.72E-09		intergenic			
QTLR_15	100708383	101708383	BovineHD0900030047	rs109315371	9	101208383	2158872	2.21E-07		intergenic	PDE10A, TBXT		PRR18, SFT2D1, MPC1, RPS6KA2
QTLR_16	10342117	11342117	BovineHD1200003098	rs110241533	12	10835926		3.41E-07		intergenic	OLFM4		PCDH8, MIR759, CNMD, SUGT1, ELF1, WBP4, KBTBD6
			BovineHD1200003101	rs109988176	12	10848307	12381	1.04E-07		intergenic			
QTLR_17	12562194	13562194	BovineHD1200003802	rs109433223	12	13062194	2213887	4.80E-07	EPSTI1	intron	TNFSF11, FAM216B		DNAJC15, ENOX1
QTLR_18	824812	1824812	BovineHD1300000152	rs137599388	13	1033957		2.87E-07	PLCB1	intron	PLCB1	MIR2285M-1	PLCB1

			BovineHD1300000159	rs136868008	13	1055763	21806	1.89E-08	PLCB1	intron			
			BovineHD1300000219	rs134875747	13	1205835	150072	5.23E-07	PLCB1	intron			
			BovineHD1300000319	rs135872126	13	1526992	321157	4.00E-08	PLCB1	intron			
			BovineHD1300000354	rs109072042	13	1615667	88675	3.33E-10	PLCB1	intron			
QTLR_19	36558263	37558263	BovineHD1700010310	rs110869101	17	36952264		6.04E-08	FSTL5	intron	FSTL5	FSTL5	FSTL5
			BovineHD1700010314	rs109399341	17	36959999	7735	5.77E-08	FSTL5	intron			
			Hapmap43738-BTA-99532	rs41567083	17	37164261	204262	1.64E-08	FSTL5	intron			
QTLR_20	40272628	41272628	BovineHD1900011769	rs137486957	19	40711083		2.28E-08	TNS4	intron	LRRC3, GSDMA, PSMD3, CSF3, MED24, THRA, NR1D1, MSL1, CASC3, RAPGEFL1, WIPF2, CDC6, RARA, GJD3, TOP2A, IGFBP4	CCR7	KRT22, KRT24, KRT25, KRT26, KRT27, KRT28, KRT10, KRT12, KRT20, KRT23, KRT39, KRT40, KRTAP3-3, KRTAP3-1, KRTAP1-1
			Hapmap54523-ss46526232	rs41255371	19	40834172	123089	1.75E-08	SMARCE1	3_prime_UTR_variant			
QTLR_21	19594619	20594619	BovineHD2100005997	rs137755071	21	20094619		5.01E-07		intergenic	AEN, ISG20		ACAN, HAPLN3, MFGE8, ABHD2
QTLR_22	59783958	60783958	BovineHD2100018046	rs110097968	21	60283958	40189339	2.08E-08	SYNE3	intron	DICER1, CLMN		GLRX5, TCL1B, TCL1A
QTLR_23	9684831	10684831	BovineHD2800003135	rs135642552	28	10184831		1.42E-07	RYR2	intron	RYR2		RYR2, ZP4
QTLR_24	16953971	17953971	BovineHD2800004904	rs135022518	28	17453971	7269140	1.62E-07		intergenic	TMEM26		CABCOCO1, ARID5B
QTLR_25	32764941	33764941	BovineHD2900010039	rs3423598291	29	33264941		7.25E-08		intergenic	JAM3, IGSF9B, SPATA19		OPCML
QTLR_26	34222160	35222160	BovineHD2900010637	rs42183295	29	34722160	1457219	3.84E-09	NTM	intron	OPCML		NTM
QTLR_27	35244229	36244229	BovineHD4100019037	rs136936924	29	35744229	1022069	1.15E-10		intergenic	NTM		
QTLR_28	127916138	128916138	BovineHD3000038779	rs109808194	X	128374647		2.31E-06		intergenic	ACE2, BMX, PIR, VEGFD, PIGA, ASB11, ASB9		MOSPD2, FANCB, GLRA2
			BovineHD3000038809	rs134798226	X	128457628	82981	5.77E-06		intergenic			

QTLR, serial number of the QTLR; Length, length of the QTLR in bp; Distance, the distance between the first in 3’ and the last in 5’ SNP of a QTLR located on the same chromosome; Upstream, In-between, Downstream, are genes mapped within the QTLR.

Table 2. Inbreeding coefficients calculated for R-BRD and R-BRD calves.

Group	**FHOM** **Mean (SD)**	**FHOM** **Min–Max values**	**FROH** **Mean (SD)**	**FROH** **Min–Max values**
R-BRD	-0.013 (0.037)	-0.083–0.094	0.155 (0.031)	0.097–0.233
S-BRD	0.008 (0.056)	-0.078–0.301	0.170 (0.045)	0.093–0.401

Table 3. List of genes annotated in QTLRs in the present study (Table 1) and references to previous studies where they were associated with BRD.

QTLR_N	Genes	Tizioto et al., 2015 (Tizioto et al., 2015)	Lipkin et al., 2017 (Lipkin et al., 2016b) and 2024*	Johnston et al., 2019 (Johnston et al., 2021)	Scott et al., 2021(Scott et al., 2022)	Li et al., 2022 (J. Li et al., 2022)	Green et al., 2023 (Green et al., 2023)
QTLR_1	ALCAM	BVDV, IBR					AUCvsDIR
QTLR_2	HHLA2						AUCvsDIR
QTLR_2	MYH15						AUCvsDIR
QTLR_2	DZIP3	BVDV					
QTLR_2	TRAT1	MYCO, PASTE, MANNHE, IBR					AUCvsDIR
QTLR_3	SIDT1	BVDV, PASTE, IBR					
QTLR_3	NAA50	BVDV					
QTLR_3	GRAMD1C						AUCvsDIR
QTLR_3	ZDHHC23	BVDV, IBR					
QTLR_3	CCDC191						AUCvsDIR
QTLR_8	FILIP1	BVDV, MANNHE					
QTLR_8	MYO6	BVDV, MANNHE, IBR					
QTLR_10	IRAK1BP1	BVDV, MANNHE					
QTLR_10	PHIP	BVDV, PASTE, IBR					
QTLR_10	HMGN3			BRSV vs CT			
QTLR_11	FAM46A						AUCvsDIR
QTLR_12	VTA1						AUCvsDIR
QTLR_12	HIVEP2	BVDV, IBR					
QTLR_13	TULP4	IBR					
QTLR_13	DYNLT1	MANNHE, IBR					
QTLR_13	SYTL3	IBR					
QTLR_14	QKI	BVDV, IBR					
QTLR_15	PDE10A	IBR					
QTLR_15	RPS6KA2		RPS6KA2				
QTLR_16	OLFM4	PASTE, MANNHE		BRSV vs CT		BRD vs NO-BRD	AUCvsDIR
QTLR_16	PCDH8						AUCvsDIR
QTLR_16	CNMD						AUCvsDIR
QTLR_16	ELF1	BVDV					
QTLR_16	KBTBD6	BVDV, IBR					
QTLR_17	FAM216B			BRSV vs CT			AUCvsDIR
QTLR_17	EPSTI1	MANNHE, IBR			IBR		AUCvsDIR
QTLR_17	DNAJC15						AUCvsDIR
QTLR_17	ENOX1						AUCvsDIR
QTLR_18	PLCB1	MANNHE					AUCvsDIR

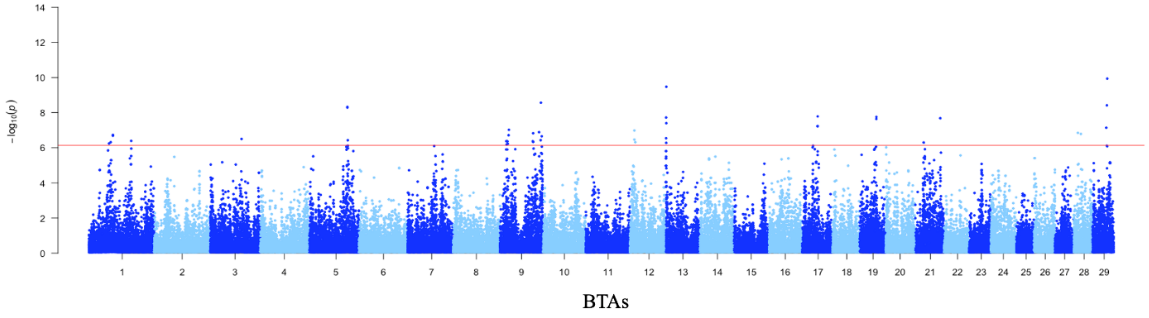
QTLR_20	GSDMA	BVDV, MANNHE, IBR					
QTLR_20	NR1D1	BVDV					
QTLR_20	MSL1						AUCvsDIR
QTLR_20	CASC3				BRSV		
QTLR_20	WIPF2	BVDV, IBR					
QTLR_20	CDC6	MANNHE		BRSV vs CT			
QTLR_20	RARA						AUCvsDIR
QTLR_20	TOP2A	BVDV, IBR			IBR, MYCO, MANNHE		
QTLR_20	IGFBP4	BVDV, PASTE, MANNHE, IBR					
QTLR_20	CCR7	BVDV, MANNHE					
QTLR_20	KRT23						AUCvsDIR
QTLR_21	ISG20	MANNHE, IBR					AUCvsDIR
QTLR_21	MFGE8	MYCO, MANNHE, IBR					
QTLR_21	ABHD2						AUCvsDIR
QTLR_22	DICER1	BVDV, IBR					
QTLR_22	CLMN						AUCvsDIR
QTLR_24	ARID5B	BVDV, IBR					
QTLR_25	JAM3	BVDV, MYCO, PASTE, MANNHE					AUCvsDIR
QTLR_28	ACE2	BVDV, PASTE, MANNHE	ACE2				
QTLR_28	BMX	BVDV, PASTE, MANNHE	BMX			BRD vs NO-BRD	
QTLR_28	PIR	BVDV	PIR				
QTLR_28	VEGFD		VEGFD				
QTLR_28	PIGA	IBR	PIGA				
QTLR_28	ASB11		ASB11				
QTLR_28	ASB9	BVDV, MYCO	ASB9	BRSV vs CT			
QTLR_28	MOSPD2	IBR					
QTLR_28	FANCB	BVDV, PASTE, IBR					

Abbreviations: IBR: bovine rhinotracheitis; BVDV: bovine viral diarrhea virus; BRSV: bovine respiratory syncytial virus; MANNH: Mannheimia haemolytica; PASTE: Pasteurella multocida; MYCO: Mycoplasma bovis; CT: control animal; AUCvsDIR: cattle which experienced a commercial auction setting (AUC) vs cattle from the cow–calf phase (DIR).

*Paper submitted and under revision: Study performed on BRD using SNPs mapped on BTX

Figures

A) Manhattan plot – Autosomes



B) Manhattan plot – Chromosome X

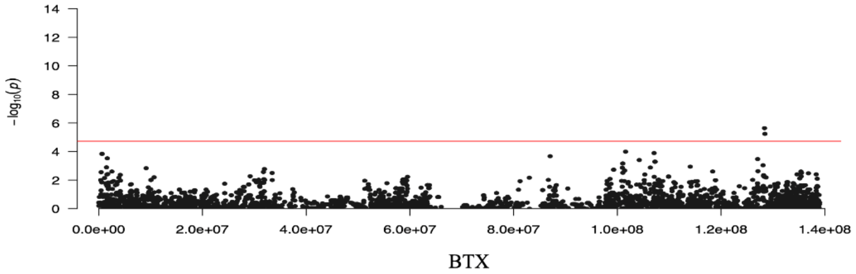


Fig. 1. Manhattan plots of the GWAS performed on BTAs (A) and on BTX (B). The Red lines represent the Bonferroni p-value threshold of significance, set at 5% genome wide.

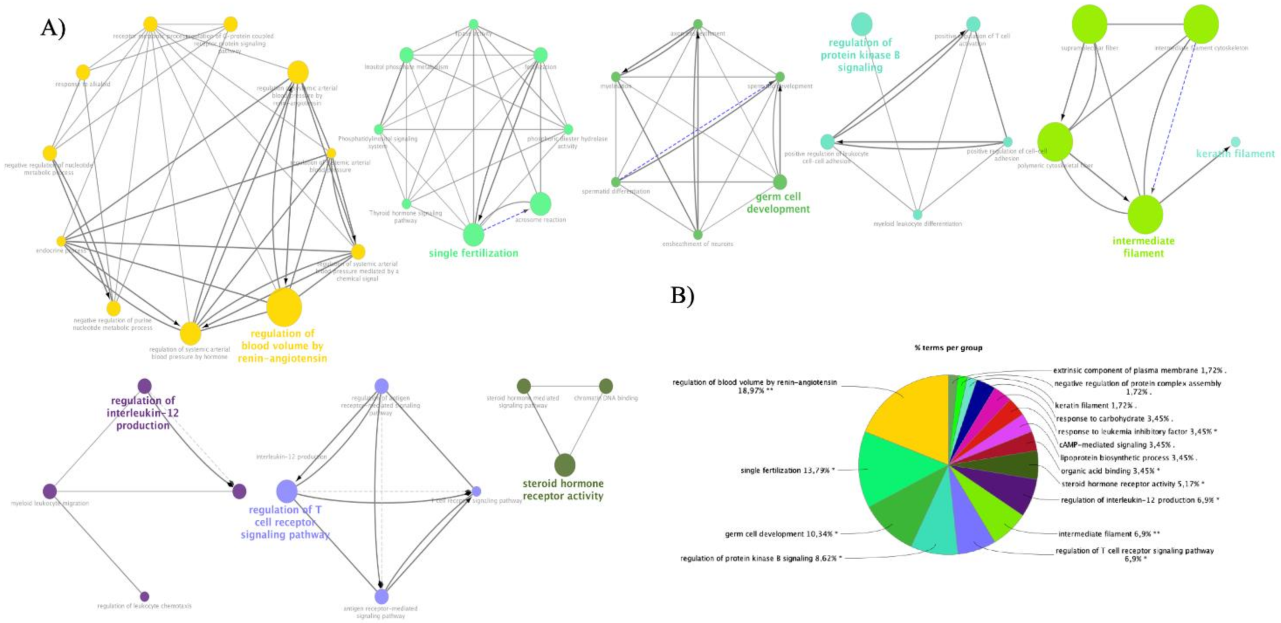


Fig. 2. Annotation Results. ClueGo Gene networks with at least three terms (A); Go-terms with higher % of associated genes (B) (Complete list of functional terms is reported in S2 Table).

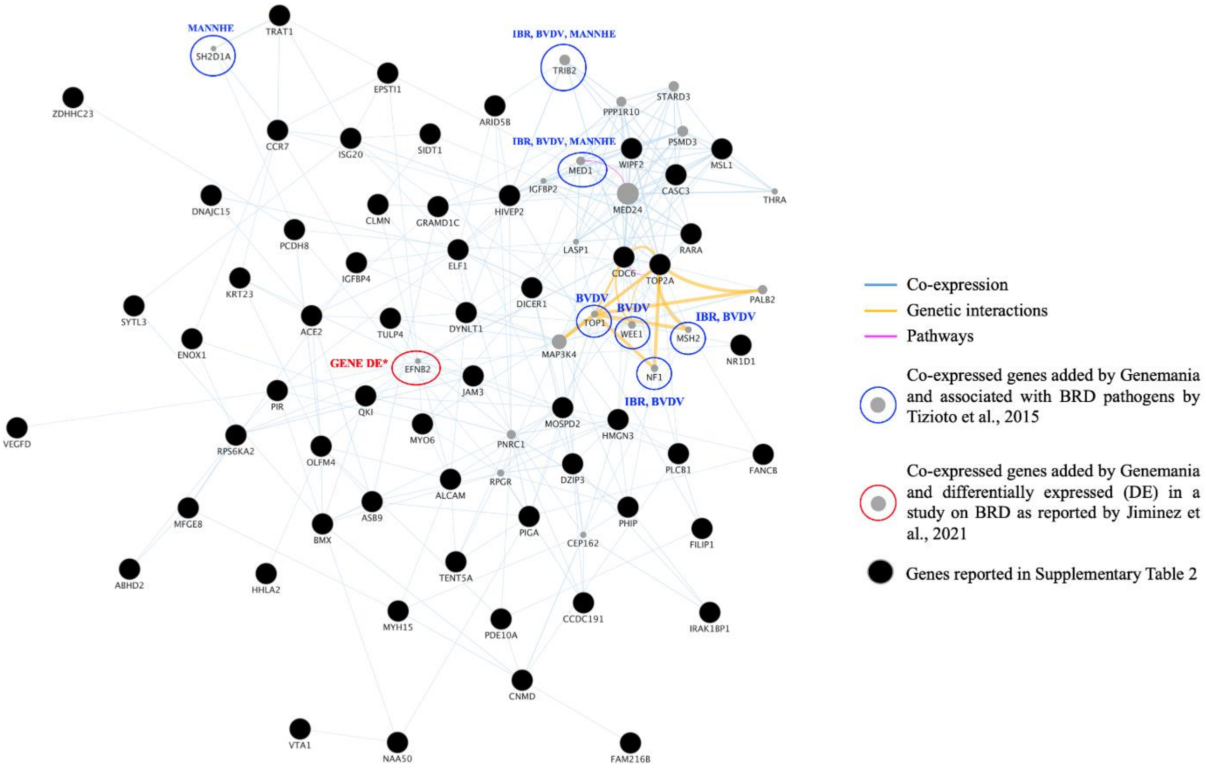


Fig. 3. Genes interacting found by Genemania tool implemented in Cytoscape. Co-expressed genes are the ones found in the entire Cytoscape database (*Homo sapiens* species as background)

Supporting information

S1 Table. List of S-BRD and R-BRD calves.

Samples	State	Age (days)	Farm	TUS	WISC for temperature	WISC for cough	WISC for nasal discharge	WISC for ocular discharge	WISC for ear position	Total WISC
1	R_BRD	62	2	1	2	0	0	0	0	2
2	R_BRD	47	7	1	1	0	2	1	0	4
3	R_BRD	51	7	1	3	0	1	0	0	4
4	R_BRD	47	7	1	2	0	1	0	0	3
5	R_BRD	32	8	1	0	0	0	1	0	1
6	R_BRD	33	8	1	1	0	0	0	0	1
7	R_BRD	35	8	1	1	0	1	1	0	3
8	R_BRD	40	8	1	0	0	1	1	0	2
9	R_BRD	41	9	1	0	0	2	2	0	4
10	R_BRD	63	9	1	2	0	0	1	0	3
11	R_BRD	63	1	1	0	0	0	0	0	0
12	R_BRD	53	4	1	1	0	0	1	0	2
13	R_BRD	71	5	1	1	0	0	1	0	2
14	R_BRD	45	5	1	2	0	0	1	0	3
15	R_BRD	37	5	1	2	0	0	1	0	3
16	R_BRD	36	5	1	1	0	0	0	0	1
17	R_BRD	65	6	1	2	0	0	2	0	4
18	R_BRD	84	10	1	3	0	0	0	1	4
19	R_BRD	92	6	0	1	0	1	0	0	2
20	R_BRD	60	6	1	2	0	1	1	0	4
21	R_BRD	47	6	1	1	0	0	1	0	2
22	R_BRD	59	6	1	1	0	0	0	0	1
23	R_BRD	54	7	1	3	0	1	0	0	4
24	R_BRD	49	7	1	2	0	1	0	0	3
25	R_BRD	49	7	1	2	0	1	0	0	3
26	R_BRD	53	7	1	2	0	0	1	0	3
27	R_BRD	49	7	1	1	0	0	0	0	1
28	R_BRD	51	7	1	2	0	0	0	0	2
29	R_BRD	33	8	1	2	0	2	0	0	4
30	R_BRD	67	8	1	2	0	0	0	0	2
31	R_BRD	72	8	1	1	0	0	0	0	1
32	R_BRD	53	9	1	2	0	1	1	0	4
33	R_BRD	70	9	1	2	0	0	1	1	4
34	R_BRD	55	9	1	2	0	0	0	0	2
35	R_BRD	49	9	1	2	0	0	0	0	2
36	R_BRD	60	9	1	2	0	0	0	0	2
37	R_BRD	64	9	1	2	0	0	1	0	3
38	R_BRD	71	9	1	1	0	1	1	0	3
39	R_BRD	70	9	1	2	0	0	0	0	2
40	R_BRD	76	9	1	2	0	0	0	0	2
41	R_BRD	73	9	1	2	0	0	1	0	3
42	R_BRD	81	9	1	0	0	0	0	0	0
43	R_BRD	81	9	1	2	0	0	1	0	3
44	R_BRD	72	9	1	0	0	0	1	0	1
45	R_BRD	85	6	1	1	0	1	0	0	2
46	R_BRD	90	10	1	1	0	0	0	0	1
47	R_BRD	73	6	1	2	0	1	0	0	3
48	S-BRD	39	1	5	3	2	0	1	3	9
49	S-BRD	60	1	5	3	2	1	0	0	6
50	S-BRD	52	1	5	2	0	0	0	0	2
51	S-BRD	76	1	5	3	3	1	1	0	8
52	S-BRD	83	1	5	3	2	0	0	0	5
53	S-BRD	53	1	5	2	2	0	1	0	5
54	S-BRD	68	1	5	1	0	0	1	0	2
55	S-BRD	57	1	5	1	0	0	0	0	1
56	S-BRD	44	1	5	1	0	0	0	0	1
57	S-BRD	41	1	5	1	0	0	0	0	1
58	S-BRD	56	1	5	1	2	0	1	0	4
59	S-BRD	25	3	5	3	1	0	0	0	4
60	S-BRD	79	4	5	2	2	0	0	0	4
61	S-BRD	75	4	5	1	2	0	0	0	3
62	S-BRD	73	4	5	3	0	0	1	0	4
63	S-BRD	83	4	5	2	2	1	1	1	7
64	S-BRD	86	4	5	3	0	0	0	0	3
65	S-BRD	12	6	5	1	1	1	0	0	3
66	S-BRD	61	6	5	2	2	0	1	0	5
67	S-BRD	36	7	5	3	3	2	0	0	8
68	S-BRD	27	7	5	1	3	1	1	0	6
69	S-BRD	48	8	5	1	0	1	1	0	3
70	S-BRD	36	8	5	2	0	0	0	0	2
71	S-BRD	69	8	5	1	0	0	0	0	1
72	S-BRD	79	9	5	2	2	1	1	0	6
73	S-BRD	62	9	5	0	2	0	1	0	3
74	S-BRD	45	6	5	0	2	0	0	0	2
75	S-BRD	83	10	5	2	2	0	0	1	5
76	S-BRD	87	10	5	3	3	0	0	0	6
77	S-BRD	69	10	5	1	2	0	2	1	6
78	S-BRD	88	10	5	2	2	0	2	0	6
79	S-BRD	94	10	5	2	2	0	1	0	5
80	S-BRD	86	10	5	3	0	0	0	0	3
81	S-BRD	71	10	5	3	0	1	0	0	4
82	S-BRD	71	10	5	2	2	0	0	0	4
83	S-BRD	79	2	5	2	3	0	2	0	7
84	S-BRD	78	2	5	2	2	0	0	0	4
85	S-BRD	78	2	5	2	2	0	1	0	5
86	S-BRD	66	2	5	3	0	1	0	0	4
87	S-BRD	66	2	5	2	0	0	1	0	3
88	S-BRD	64	2	5	1	0	0	0	0	1
89	S-BRD	65	2	5	2	2	1	1	0	6
90	S-BRD	42	2	5	2	0	1	1	0	4
91	S-BRD	28	2	5	2	0	0	0	0	2
92	S-BRD	50	2	5	1	0	0	0	0	1
93	S-BRD	41	2	5	1	2	0	1	1	5
94	S-BRD	29	2	5	1	2	0	2	0	5

S2 Table. Genes functional results provided by Cytoscape Software.

ID	Term	Ontology Source	Term PValue	Group PValue Bonferroni Corrected (< 0.05)	Associated Genes Found
GO:0002016	regulation of blood volume by renin-angiotensin	GO BP	3.23E-06	1.96E-03	[ACE2, DRD3, RPS6KA2]
GO:0003081	regulation of systemic arterial blood pressure by renin-angiotensin	GO BP	2.07E-04	1.96E-03	[ACE2, DRD3, RPS6KA2]
GO:0001990	regulation of systemic arterial blood pressure by hormone	GO BP	9.41E-04	1.96E-03	[ACE2, DRD3, RPS6KA2]
GO:0003044	regulation of systemic arterial blood pressure mediated by a chemical signal	GO BP	1.97E-03	1.96E-03	[ACE2, DRD3, RPS6KA2]
GO:0043112	receptor metabolic process	GO BP	1.92E-03	1.96E-03	[ACE2, DRD3, HTR1B, NR1D1, TRAT1]
GO:0008277	regulation of G-protein coupled receptor protein signaling pathway	GO BP	2.12E-03	1.96E-03	[DRD3, DYNLT1, HTR1B, PLCB1]
GO:0043279	response to alkaloid	GO BP	2.30E-03	1.96E-03	[DRD3, HTR1B, RYR2]
GO:1900543	negative regulation of purine nucleotide metabolic process	GO BP	2.87E-03	1.96E-03	[DNAJC15, DRD3, HTR1B]
GO:0045980	negative regulation of nucleotide metabolic process	GO BP	3.07E-03	1.96E-03	[DNAJC15, DRD3, HTR1B]
GO:0003073	regulation of systemic arterial blood pressure	GO BP	1.04E-02	1.96E-03	[ACE2, DRD3, RPS6KA2]
GO:0050886	endocrine process	GO BP	8.76E-03	1.96E-03	[ACE2, DRD3, RPS6KA2]
GO:0043401	steroid hormone mediated signaling pathway	GO BP	3.12E-03	1.18E-02	[ABHD2, LMO3, NR1D1, RARA, THRA]
GO:0007338	single fertilization	GO BP	6.02E-04	1.61E-02	[ABHD2, MFGE8, PLCB1, PLCZ1, ZP4]
GO:0007340	acrosome reaction	GO BP	1.15E-03	1.61E-02	[ABHD2, PLCB1, ZP4]
GO:0009566	fertilization	GO BP	2.43E-03	1.61E-02	[ABHD2, MFGE8, PLCB1, PLCZ1, ZP4]
GO:0032655	regulation of interleukin-12 production	GO BP	2.13E-03	2.28E-02	[CCR7, PLCB1, TIGIT]
GO:0032615	interleukin-12 production	GO BP	2.48E-03	2.28E-02	[CCR7, PLCB1, TIGIT]
GO:0097529	myeloid leukocyte migration	GO BP	2.67E-03	2.28E-02	[CCR7, JAM3, PLCB1, TNFSF11, VEGFD]
GO:0002688	regulation of leukocyte chemotaxis	GO BP	2.47E-02	2.28E-02	[CCR7, JAM3, VEGFD]
GO:0050856	regulation of T cell receptor signaling pathway	GO BP	7.54E-04	2.92E-02	[CCR7, ELF1, TRAT1]
GO:0050851	antigen receptor-mediated signaling pathway	GO BP	2.92E-02	2.92E-02	[CCR7, ELF1, TRAT1]
GO:0050854	regulation of antigen receptor-mediated signaling pathway	GO BP	2.13E-03	2.92E-02	[CCR7, ELF1, TRAT1]
GO:0050852	T cell receptor signaling pathway	GO BP	8.37E-03	2.92E-02	[CCR7, ELF1, TRAT1]
GO:0007281	germ cell development	GO BP	2.47E-03	3.15E-02	[ABHD2, CAPZA3, JAM3, QKI, RARA, RPS6KA2]
GO:0007272	ensheathment of neurons	GO BP	2.00E-02	3.15E-02	[ADGRG6, JAM3, QKI]
GO:0008366	axon ensheathment	GO BP	2.00E-02	3.15E-02	[ADGRG6, JAM3, QKI]
GO:0042552	myelination	GO BP	1.88E-02	3.15E-02	[ADGRG6, JAM3, QKI]
GO:0007286	spermatid development	GO BP	9.50E-03	3.15E-02	[ABHD2, CAPZA3, JAM3, QKI]
GO:0048515	spermatid differentiation	GO BP	1.03E-02	3.15E-02	[ABHD2, CAPZA3, JAM3, QKI]
GO:0051896	regulation of protein kinase B signaling	GO BP	3.00E-02	3.91E-02	[CSF3, DRD3, TNFSF11]
GO:0050870	positive regulation of T cell activation	GO BP	2.43E-03	3.91E-02	[CCR7, HHLA2, RARA, TNFSF11, ZP4]
GO:1903039	positive regulation of leukocyte cell-cell adhesion	GO BP	2.85E-03	3.91E-02	[CCR7, HHLA2, RARA, TNFSF11, ZP4]

GO:0002573	myeloid leukocyte differentiation	GO BP	4.29E-03	3.91E-02	[CCR7, CSF3, PIR, RARA, TNFSF11]
GO:0022409	positive regulation of cell-cell adhesion	GO BP	5.60E-03	3.91E-02	[CCR7, HHLA2, RARA, TNFSF11, ZP4]
GO:1990823	response to leukemia inhibitory factor	GO BP	4.49E-03	4.04E-02	[ARID5B, MPC1, PIGA, TNFSF11]
GO:1990830	cellular response to leukemia inhibitory factor	GO BP	4.49E-03	4.04E-02	[ARID5B, MPC1, PIGA, TNFSF11]
GO:0005882	intermediate filament	GO CC	3.06E-15	5.12E-06	[KRT10, KRT12, KRT20, KRT222, KRT23, KRT24, KRT25, KRT26, KRT27, KRT28, KRT39, KRT40, KRTAP1-1, KRTAP3-1, KRTAP3-3]
GO:0045111	intermediate filament cytoskeleton	GO CC	1.30E-13	5.12E-06	[KRT10, KRT12, KRT20, KRT222, KRT23, KRT24, KRT25, KRT26, KRT27, KRT28, KRT39, KRT40, KRTAP1-1, KRTAP3-1, KRTAP3-3]
GO:0099513	polymeric cytoskeletal fiber	GO CC	1.06E-08	5.12E-06	[DYNLT1, KRT10, KRT12, KRT20, KRT222, KRT23, KRT24, KRT25, KRT26, KRT27, KRT28, KRT39, KRT40, KRTAP1-1, KRTAP3-1, KRTAP3-3, MYO6]
GO:0099512	supramolecular fiber	GO CC	1.51E-07	5.12E-06	[DYNLT1, KRT10, KRT12, KRT20, KRT222, KRT23, KRT24, KRT25, KRT26, KRT27, KRT28, KRT39, KRT40, KRTAP1-1, KRTAP3-1, KRTAP3-3, MYO6, RYR2]
GO:0003707	steroid hormone receptor activity	GO MF	7.01E-04	1.18E-02	[ABHD2, NR1D1, RARA, THRA]
GO:0031490	chromatin DNA binding	GO MF	3.15E-03	1.18E-02	[HMGN3, RARA, SMARCE1, THRA]
GO:0016298	lipase activity	GO MF	2.62E-02	1.61E-02	[ABHD2, PLCB1, PLCZ1]
GO:0008081	phosphoric diester hydrolase activity	GO MF	1.43E-02	1.61E-02	[PDE10A, PLCB1, PLCZ1]
GO:0043177	organic acid binding	GO MF	2.13E-03	2.34E-02	[ACAN, GLRA2, HAPLN3, RARA, RYR2]
GO:0031406	carboxylic acid binding	GO MF	1.25E-02	2.34E-02	[ACAN, GLRA2, HAPLN3, RARA]
KEGG:00562	Inositol phosphate metabolism	KEGG	1.84E-03	1.61E-02	[PIK3C2G, PLCB1, PLCZ1, SYNJ2]
KEGG:04070	Phosphatidylinositol signaling system	KEGG	5.37E-03	1.61E-02	[PIK3C2G, PLCB1, PLCZ1, SYNJ2]
KEGG:04919	Thyroid hormone signaling pathway	KEGG	8.45E-03	1.61E-02	[MED24, PLCB1, PLCZ1, THRA]

ID	**Term**	**Ontology Source**	**Term PValue**	**Group PValue Corrected with Bonferroni (> 0.05)**	**Associated Genes Found**
GO:0031333	negative regulation of protein complex assembly	GO BP	2.69E-02	5.39E-02	[CAPZA3, DNAJC15, THRA]
GO:0009743	response to carbohydrate	GO BP	2.47E-02	7.42E-02	[HMGN3, NR1D1, RPS6KA2]
GO:0071322	cellular response to carbohydrate stimulus	GO BP	8.37E-03	7.42E-02	[HMGN3, NR1D1, RPS6KA2]
GO:0042158	lipoprotein biosynthetic process	GO BP	2.13E-02	8.53E-02	[PIGA, ZDHHC14, ZDHHC23]
GO:0006497	protein lipidation	GO BP	1.82E-02	8.53E-02	[PIGA, ZDHHC14, ZDHHC23]
GO:0019933	cAMP-mediated signaling	GO BP	1.48E-02	8.79E-02	[ADGRG6, DRD3, PDE10A]
GO:0019935	cyclic-nucleotide-mediated signaling	GO BP	1.76E-02	8.79E-02	[ADGRG6, DRD3, PDE10A]
GO:0045095	keratin filament	GO CC	7.62E-03	5.33E-02	[KRTAP1-1, KRTAP3-1, KRTAP3-3]
GO:0019897	extrinsic component of plasma membrane	GO CC	1.32E-02	7.90E-02	[BMX, MFGE8, SNX9, SYTL3]

Abbreviations. BP: Biological Process; MF: Molecular Function; CC: Cellular Component.

5. General discussion

The present thesis explores the genomic peculiarities of autochthonous cattle breeds and their potential to address future challenges in sustainable livestock farming. As climate change and resource scarcity become increasingly pressing issues, the ability of cattle to thrive under suboptimal conditions will be critical. Local breeds, such as the Aosta cattle, have historically adapted to harsh environments, demonstrating resilience in terms of disease resistance, metabolic efficiency, and robustness. Understanding their genetic architecture is fundamental to developing breeding strategies that not only preserve their unique biodiversity but also enhance their efficiency and economic viability.

Genetic Diversity and Its Implications for Breeding Strategies

One of the key contributions of this thesis is the characterization of genetic diversity in both local and cosmopolitan breeds. The comparison between the Aosta breeds and Holstein cattle revealed a stark contrast in their levels of inbreeding, with the Holstein breed displaying significantly higher values of F_{ROH} and longer runs of homozygosity (ROH). This result highlights the challenges faced by intensively selected breeds, where genetic bottlenecks and high selection pressure have led to a reduction in genetic variability. The findings suggest that despite their smaller population size, Aosta breeds maintain a more balanced genetic structure, likely due to careful farm breeding and to the guidelines provided by the breeding association (i.e. ANABORAVA) to its associates.

Moreover, the analysis of Mendelian disease-associated variants indicated a lower presence of detrimental mutations in the Aosta breeds compared to Holstein. This finding further supports the hypothesis that local breeds, having undergone natural selection in harsher environments, may carry fewer deleterious mutations linked to production-related diseases. Such insights lay the foundation for designing genomic selection (GS) strategies that incorporate both productivity and resilience traits.

Milk Production Efficiency and Genomic Selection Prospects

To enhance the economic sustainability of autochthonous breeds, one study in this thesis investigated the genomic regions associated with milk production traits in Aosta cattle. While significant QTL were not detected for milk, fat, and protein yield, QTL affecting fat and protein content were successfully identified. These loci overlapped with known QTL in cosmopolitan breeds, reinforcing the idea that milk composition traits are governed by conserved genetic mechanisms across breeds.

One particularly interesting finding was the presence of a significant SNP peak on BTA 14 near the DGAT1 gene, a well-known regulator of milk fat content. While the SNPs on DGAT1 were fixed in Aosta breeds, the detected signal may indicate variability in regulatory elements influencing its activity. Although we could not demonstrate that Aosta cattle have distinct QTL for milk production compared to cosmopolitan breeds, their ability to sustain milk yield even under challenging pasture conditions suggests potential physiological differences in nutrient utilization and metabolic efficiency. These findings underscore the importance of developing GS models tailored to the specific needs of local breeds rather than simply extrapolating from cosmopolitan cattle.

Structural Variability and CNVs in Dairy Breeds

This thesis also explored structural genomic variation, particularly Copy Number Variants (CNVs), as a source of genetic diversity in dairy cattle. CNVs have been increasingly recognized as functionally relevant genetic elements, and their role in adaptive traits warrants further investigation. The Holstein population was used as a reference to map CNVs in animals raised in Lombardy, and comparisons were made with Brown Swiss and Jersey breeds. The same approach will be applied to the Aosta breeds to assess their CNV landscape and determine whether structural variants contribute to their resilience and efficiency.

Genetic Basis of Disease Resistance and Resilience

A major aspect of this thesis was the identification of QTL associated with resilience traits, particularly disease resistance. The selective genotyping and DNA pooling strategy used for Holstein cattle proved to be an effective approach for detecting QTL linked to mastitis and bovine respiratory disease (BRD). Given the high costs of genotyping large populations, this method provides a valid experimental design for identifying genetic variants in smaller datasets while maintaining statistical power.

The application of this experimental design to production and efficiency traits, such as dry matter intake and methane emissions, will enhance our understanding of the genomic factors influencing resource efficiency in dairy cattle. These findings pave the way for integrating resilience traits into breeding programs, which is critical for balancing production goals with sustainability objectives.

Future Directions and Practical Implications

The results of this thesis provide a strong basis for further research and practical applications in cattle breeding:

1. *Integration of Resilience Traits in Genomic Selection* – Future GS models should incorporate resilience-related genomic markers identified in this study to enhance the robustness of selection programs, particularly in autochthonous breeds.
2. *Optimization of Breeding Strategies for Small Populations* – The observed genetic diversity in Aosta breeds suggests that selective breeding can be optimized to maintain genetic variability while improving productivity. A focus on balancing selection pressure will be essential.
3. *Expansion of Structural Variant Analyses* – Further characterization of CNVs in local breeds may uncover additional adaptive variants that contribute to their efficiency and hardiness.
4. *Longitudinal Studies on Physiological and Metabolic Efficiency* – Future research should aim to characterize how Aosta cattle sustain productivity under extensive pasture conditions, potentially identifying unique metabolic pathways that could be leveraged in breeding programs.

Conclusions

This thesis demonstrates the value of integrating genomic and phenotypic data to improve our understanding of genetic diversity, resilience, and productivity in cattle breeds. The findings highlight that while cosmopolitan breeds like Holstein have undergone intense selection for production efficiency, they also suffer from high levels of inbreeding and reduced genetic variability. In contrast, local breeds, such as Aosta cattle, maintain a broader genetic base and exhibit traits linked to resilience, making them promising candidates for sustainable breeding programs.

By leveraging genomic tools, we can develop breeding strategies that not only enhance productivity but also ensure long-term sustainability and adaptability of livestock populations in the face of global challenges. The insights gained from this research provide a stepping stone toward a more comprehensive and resilience-focused approach to cattle breeding, with direct implications for both scientific advancements and practical applications in animal husbandry.

6. Next steps

One of the objectives of this project was the identification of structural variants proprietary of the Aosta cattle breeds. To do so a total of 225 WGS have been aligned, and the variant calling was performed during my abroad period at the CRG (from January to June 2024). The analysis has been performed using nf-core pipeline Sarek, which allows quality control, alignment and variant calling using Nextflow. The run finished on the 13th of August and the final vcf files still need to be processed and analysed. For this reason, one of the next steps will be to study the structural variants of the Aosta cattle breeds and to identify the proprietary variations that characterize this population respect to others. We are also working in collaboration with the University of Padova to develop the genomic selection in these autochthonous breeds thanks to the funding of the GENOVAL project: Prin 2022 "GENOmic Breeding VALue Estimation in a Native Alpine Cattle Breed" codice CUP G53D23004040006 - Finanziato dall'Unione europea – Next Generation EU". At best of our knowledge this will be the first autochtonous cattle breed with a genomic estimation of EBV in routine. The Genomic Evaluation of reproducers has been in fact introduced and approved as a routine process in Decembre 2024.

7. Acknowledgement

My deepest gratitude goes to my supervisor, Professor Alessandro Bagnato, who made this journey possible and guided me in every step of the way. I highly regard you as a researcher and hope to one day reach your level of knowledge. Your passion for animal sciences, your perseverance, and your unwavering determination are a true inspiration. I would also like to thank Professor Maria Giuseppina Strillacci for always encouraging me to do my best.

A heartfelt thank you to Dr. Cedric Notredame for welcoming me into his wonderful research group. I learned so much during my time at CRG, and my stay there was truly inspiring. It helped me overcome a difficult period and rekindled my enthusiasm for the path I have chosen. Thanks also to all the lab members for helping me with bioinformatics and for making me feel part of the team - you are excellent colleagues, and I hope to see you again soon, wherever you may be in the world!

A huge thank you to my two external reviewers, Dr. Frank Panitz and Dr. María J. Carabaño, for accepting the task of reviewing and evaluating my PhD thesis.

My gratitude also goes to the Valdostana Cattle Breeders Association (ANABORAVA), particularly to Director Dr. Mario Vevey and Dr. Veruska Blanchet, who provided all the essential data for this research and were always available to answer my questions. Without you, this thesis would not have been possible, your professionalism and wisdom will always guide me.

Now coming to “the big stuff”

An infinite thanks to Chiara, who has been by my side throughout this entire journey - supporting me, helping me, and being my biggest supporter in the most challenging moments. Andrea and Carlotta later joined this adventure, and I thank you both for the same reasons. You are not only wonderful colleagues but also true friends, and I hope we will always stay in touch. Thank you to Raffaella for your guidance and patience. You are a great source of inspiration, and I wish you to find the strength to never give up in difficult times and to achieve everything you deserve. To all of you, I hope for a future filled with a fulfilling career that meets all your expectations - expectations that I know are high, just as you deserve, ad maiora semper.

An immeasurable thank you to my family, to those who are still with me and those who have left along the way. I know for certain that you will always be by my side, and that is why you are the foundation of who I am and who I aspire to be. It is especially thanks to you that I have reached this point - your constant support in countless ways, your encouragement to always aim higher and never settle. Thank you to Fabio, who has been by my side for the past three years, enduring my moments of discouragement and frustration, always trying to lift my spirits and make me smile, even when there was little to smile about. I love you, and I am proud to have you by my side. Thank you for

never stopping believing in me and for accepting all the sacrifices this PhD has demanded of us. I am not the best at expressing my feelings, but please know that I carry you all in my heart, because you are my strength.

Thank you also to Samu and Rache, my two wonderful nieces and nephews, who always manage to recharge me and give me energy even when I am completely exhausted. I love you, and I hope to be a more present aunt for you from now on.

To my lifelong friends, thank you for never forgetting me, despite my countless refusals to go out or spend weekends together due to my work. I love you, and you are truly precious to me. Thank you for being the ones to absorb all my negative emotions and for laughing with me over my workplace misadventures. You have the incredible ability to turn bad situations into things not worth crying over, and you have been an enormous support during these years. I wish you all the best, my dear friends.

The last special thanks go to Rebecca and Gian, who welcomed me as true friends during my six months in Barcelona and showed me a city that I fell in love with. Thank you for your company and support, you were fundamental to my experience. I hope to visit Madrid soon to see you again, and I wish you all the best in your future adventures!

It is thanks to all of you that I have made it this far. You are all part of my “big family”, each of you leaving something with me that I hope to carry forward always - some by genetics, some by emulation.

I love you all, and above all, I wish you the strength to always pursue and achieve what you desire.