UNIVERSITÀ DEGLI STUDI DI SASSARI

CORSO DI DOTTORATO DI RICERCA

Scienze Agrarie

Curriculum Scienze e Tecnologie Zootecniche

Ciclo XXXVIII

FROM GENETIC DIVERSITY TO BREEDING INDICES: STUDIES IN LOCAL AND COSMOPOLITAN CATTLE POPULATIONS

dr. Paola Erminia Milia

Direttore della Scuola	prof. Severino Zara
Referente di Indirizzo	prof. Corrado Dimauro
Docente Guida	prof. Corrado Dimauro
Tutor	dr. Alberto Cesarani

Anno Accademico 2025-2026

To my family

Acknowledgements

I would like to sincerely thank my PhD supervisor, Prof. Corrado Dimauro, for his guidance, teachings, passion, and patience throughout these three years of doctoral research.

I want to express my gratitude to Prof. Nicolò Pietro Paolo Macciotta for all the opportunities he provided me during my PhD journey.

A very special thanks goes to Dr. Alberto Cesarani for his patience, mentorship, and constant support throughout this path.

I am grateful to the Genetic Research Group and to all the members of the “Scienze Zootecniche” unit of the Department of Agriculture at the University of Sassari for their help, collaboration and, above all, for their support.

I would like to thank Dr. Clara Díaz and the entire Animal Breeding Department from INIA-CSIC in Madrid for hosting me and for the opportunity to broaden my research experience within their group.

Finally, I would like to thank Dr. Fernando Macedo, Dr. Gabriel Soares Campos, and all the colleagues at the Interbull Centre and Department of Animal Biosciences of the Swedish University of Agricultural Sciences for welcoming me and for sharing with me the enthusiasm and passion they put into every project.

Index

CHAPTER 1

General introduction

Introduction

Modern domestic cattle derive from the aurochs (*Bos primigenius*), a large wild bovid that became extinct in the 17th century (Ajmone-Marsan et al., 2010; Taboret et al., 2011). Within this ancestral species, at least two major subspecies had differentiated well before domestication: *Bos primigenius taurus*, which gave rise to the taurine type (*Bos taurus*), humpless and primarily adapted to temperate climates and today the basis of most high-producing dairy and beef cattle breeds; and *Bos primigenius indicus*, the ancestor of the indicine type (*Bos indicus* or zebu), characterized by a shoulder hump, greater resistance to heat and parasites, and a predominantly tropical distribution (Felius et al., 2014; Saravanana et al. 2020).

The domestication of *Bos taurus taurus* took place in the Near East, within the so-called Fertile Crescent—probably in the area corresponding to present-day southeastern Turkey and Syria—around 8,000–10,000 years BC, shortly after the domestication of sheep and goats (Taberlet et al., 2011; Saravanana et al. 2020). Genetic analyses indicate that domestic taurine populations descend from a very small number of founding females suggesting a strong initial bottleneck (Laurent et al., 2020). After domestication, cattle accompanied the migrations of early farmers (Hiemstra et al., 2010; Felius et al., 2014; Saravanana et al. 2020).

Taurine cattle reached Europe around 8,500 years ago, first appearing in Greece (Thessaly) and subsequently spreading along two major expansion routes: the Danubian route, through Central Europe, and the Mediterranean costal route (Hiemstra et al., 2010; Ajmone-Marsan et al., 2010). Along this westward expansion, domestic cattle arrived in Central Europe about 7,500 years ago and only roughly a millennium later reached Northern Europe (Hiemstra et al., 2010; Felius et al., 2014). This progressive expansion from the domestication center was accompanied by repeated founder effects, reflected in a gradual decline in genetic diversity with increasing distance from Anatolia/the Near East (Groeneveld et al., 2010). Over this long period, European *Bos taurus* differentiated into distinct regional groupings — Balkan, Podolian, Alpine and Nordic, Iberian breeds — shaped by the interplay of migration, environmental adaptation, and traditional husbandry practices (Taberlet et al., 2011; Felius et al., 2011). Morphological changes occurred in parallel, such as the reduction in body size from the Neolithic to the Middle Ages and the spread of short-horned types, easier to manage in stall-based systems (Hiemstra et al., 2010; Felius et al., 2011).

The earliest cattle introduced in Africa were also *Bos taurus*, which spread westwards and south-westwards, adapting to a wide variety of environments including subtropical and tropical regions (Ajmone-Marsan et al., 2010). Later, the arrival of *Bos taurus indicus* bulls from Asia led to the formation of numerous taurine-indicine hybrid populations, now particularly widespread in East Africa (Bradley et al., 1998; Ajmone-Marsan et al., 2010; Feliu et al., 2014). Traces of African cattle introgression can also be detected in some southern European populations—such as Iberian and Italian breeds—where mitochondrial haplogroups typical of Africa attest to ancient gene flow across the Mediterranean (Beja-Pereiraa et al., 2006). *Bos taurus* was finally introduced to the Americas from 1493 onwards with Iberian expeditions: the first herds gave rise to the Criollo cattle, a European-derived population that subsequently adapted to highly diverse ecological conditions across South America (Giovambattista et al., 2001; Ajmone-Marsan et al., 2010).

From the Neolithic until the 18th century, domestic cattle were not organized into “breeds” in the modern sense, but rather into relatively open local populations (Hiemstra et al., 2010). The so-called local breeds emerged through slow processes of dispersal and partial isolation, combined with natural selection and mostly unconscious human selection: farmers retained and bred animals showing higher survival, better adaptation to local production systems, and an adequate response to the multifunctional needs of rural communities, such as draught power, milk, meat, manure, and cultural values (Hiemstra et al., 2010; Taberlet et al., 2011). This led to the formation of regionally adapted populations, often characterized by high internal genetic diversity and distinctive traits such as extreme cold tolerance or resistance to tropical diseases (Saravanana et al. 2020). In Europe, the diversity pattern reflects this long history: populations closer to the domestication center (Anatolia and the Balkans) display greater variability than those in north-western Europe (Medugorac et al., 2009). In Italy, numerous local breeds—such as Cinisara, Modicana and Reggiana—represent examples of these native historical breeds, today often small in census size but recognized as important reservoirs of adaptive genetic variants potentially useful for future challenges such as climate change or emerging diseases (Mastrangelo et al., 2016).

However, this picture changed dramatically from the 18th–19th centuries onward, with the industrial revolution and the emergence of the modern concept of the cattle “breed” (Felius et al., 2014). During this period, some herdbooks were established (Felius et al., 2014) and systematic artificial selection programs were developed, including controlled crossbreeding and structured inbreeding practices. The originally continuous gene pool became fragmented into an increasing number of formally defined breeds, with reduced gene flow among them and intense selection directed towards production specialization: dairy and beef breeds (Taberlet et al., 2011). In subsequent centuries, the development of reproductive technologies such as artificial insemination and, more recently, the use of genomic data, further intensified this selection pressure (Ajmone-Marsan et al., 2010; Groeneveld et al., 2010; Taberlet et al., 2011). The global spread of a few high-producing cosmopolitan breeds often resulted in the replacement of, or extensive crossbreeding with, local populations, leading to the creation of synthetic or cosmopolitan breeds and contributing to the erosion of traditional cattle biodiversity (Taberlet et al., 2011; Felius et al., 2014). In this sense, the history of *Bos taurus* can be seen as a transition from a mosaic of locally adapted populations shaped by natural selection and co-evolution with their environments to a system dominated by a small number of highly specialized industrial breeds (Giovambattista et al., 2008; Ajmone-Marsan et al., 2010; Taberlet et al., 2011).

Local cattle breeds

Local breeds are animal populations that have historically developed within restricted geographical areas, in close co-evolution with the ecological, climatic, and socio-economic constraints of their regions of origin and with traditional livestock management systems (Hiemstra et al., 2010). Their identity does not stem from a modern selective breeding program aimed at maximizing productive performance, but rather from a long and cumulative process of local adaptation in which natural selection, together with limited and largely unconscious artificial selection, have acted simultaneously (Taberlet et al., 2011; Giovambattista et al., 2008; Saravanana et al. 2020). For most of the history of domestication, these animals were selected more for their ability to survive, to exploit scarce resources and for their versatility of use rather than for the specialization in a single economic trait (Felius et al., 2014). As a result, many local breeds still exhibit a marked multifunctionality, often being reared for dual or even triple purposes (milk, meat and sometimes draught work), and remain closely associated with territorial supply chains and niche, typical products (Hiemstra et al., 2010; Taberlet et al., 2011).

From a phenotypic and management perspective, local breeds display a typically rustic profile: they are frugal, tolerant to climatic and pathogenic stressors, efficient in the use of low-quality forages or pastures, and particularly suited to extensive or semi-extensive systems where farming is integrated with natural grazing and with vegetation and landscape management functions (Felius et al., 2014). In marginal or highly variable environments—such as mountain pastures, wooded areas or low-input systems—these populations often represent the most biologically and economically sustainable option, as they convert resources poorly suited to human nutrition into livestock products and contribute to the preservation of biodiversity and cultural landscapes (Hiemstra et al., 2010). From a genetic standpoint, they constitute important reservoirs of variability, often maintaining high levels of within-breed diversity and a relatively large effective population size, consistent with their less intensive selective history and limited geographical distribution (Giovambattista et al., 2008; Taberlet et al., 2009). This broad genetic base underpins favorable functional traits—such as health, longevity and overall robustness—that are central to the resilience of traditional livestock systems (Giovambattista et al., 2008; Hiemstra et al., 2010; Felius et al., 2014; Saravanana et al. 2020).

Cosmopolitan cattle breeds

Cosmopolitan breeds represent the exact opposite of local breeds in terms of breeding goals and genetic profile. They are the outcome of deliberate and highly methodical artificial selection aimed at maximizing one or a few specific economic traits, particularly milk or meat production (Saravanana et al., 2020). Selection intensity progressively increased with the emergence of the modern concept of "breed" over the past two centuries, and it has risen dramatically in recent decades due to the widespread adoption of new modern techniques, such as artificial insemination and embryo transfer (Taberlet et al., 2011; Saravanana et al., 2020). These technologies have facilitated the rapid international dissemination of a narrow genetic base and the extensive use of a limited number of elite sires, thereby accelerating genetic improvement while also promoting an increasing homogenization of populations (Groeneveld et al., 2010). At the same time, advances in animal nutrition, infrastructure, and transportation have enabled the development of highly uniform and controlled production environments, within which these breeds can fully express their productive potential (Groeneveld et al., 2010).

The main advantages of cosmopolitan breeds are linked to their specialization and their high efficiency in high-input production systems. They exhibit outstanding performance for target traits (such as milk yield, fat and protein contents, but also meat production, daily weight gain, intramuscular fat, calving ease, and others important traits), enabling rapid genetic progress and an economic competitiveness that makes them dominant in global markets (Hiemstra et al., 2010). However, extreme specialization and the repeated use of a small number of sires have entailed substantial biological and genetic costs. First, a marked erosion of genetic diversity is observed, associated with a low effective population size and the loss of rare alleles (Medugorac et al., 2009; Groeneveld et al., 2010). Second, the reduction in variability and the increase in autozygosity heighten inbreeding and the probability of expressing deleterious variants, leading to inbreeding depression affecting fitness-related traits, an increased mutational load and the spread of recurrent genetic defects (Hiemstra et al., 2010; Frantz et al., 2020). Finally, these populations are highly dependent on standardized environments and high input levels, showing limited adaptability in marginal or low-input contexts (Groeneveld et al., 2010).

Fitness trait: Local cattle vs Cosmopolitan cattle

The contrast between local and cosmopolitan cattle breeds reflects a classic evolutionary and zootechnical trade-off between maximizing productive performance and maintaining fitness-related traits. This divergence arises primarily from the different selective pressures to which these populations have been exposed: local breeds have been shaped predominantly by natural selection under traditional, low-input management systems; on the other hand, cosmopolitan breeds have undergone intense artificial selection for a narrow set of production traits (Giovambattista et al., 2008; Taberlet et al., 2011; Saravanana et al., 2020).

Fitness traits, i.e., biological characteristics that determine an individual's ability to survive and reproduce, play a central role in differentiating local and cosmopolitan breeds (Saravanana et al., 2020). In local breeds natural selection has primarily shaped multi-locus quantitative traits that enhance survival, robustness, and reproductive efficiency (Giovambattista et al., 2008). As a result, they tend to display high levels of fertility even under harsh or variable environmental conditions, along with extended longevity, strong disease resistance, and overall resilience (Giovambattista et al., 2008; Hiemstra et al., 2010; Gandini et al., 2017). In contrast, cosmopolitan breeds have been subjected to intensive artificial selection focused on a narrow set of production traits (Egger-Danner et al., 2015). Consequently, many cosmopolitan dairy and beef lines show declining reproductive performance, lower conception rates and reduced adaptability when removed from high-input, standardized environments (Norman et al., 2009; Hiemstra et al., 2010; Egger-Danner et al., 2015). This pattern reflects the well-known antagonistic relationship between production and fitness traits (Miglior et al., 2017). Strong selection for increased productivity has often resulted in unfavorable correlated responses in fertility, health, and longevity, due to negative genetic correlations and increased metabolic demands associated with high production (Misztal et al., 2024). These effects became more evident as selection intensity increased, highlighting the importance of balancing production and functional traits in modern breeding programs (Miglior et al., 2017; Misztal et al., 2024). To address these unfavorable correlated responses, modern breeding programs have progressively adopted multi-trait selection indices that combine production and functional traits using economic or biological weights (Egger-Danner et al., 2015). By explicitly incorporating fertility, health, and longevity alongside productivity, these indices aim to balance genetic progress across antagonistic trait groups and prevent further deterioration of fitness (Miglior et al., 2017; Misztal et al., 2024). The introduction of genomic selection has further enhanced this approach by improving the accuracy of low-heritability traits and enabling earlier selection decisions, although it also requires continuous monitoring of genetic parameters to avoid unintended correlated responses (Miglior et al., 2017; Misztal et al., 2024).

Understanding the genetic basis of adaptation, relation between production and fitness traits, together with responses to selection, requires the integration of genomic and statistical tools capable of characterizing population structure and identifying regions under selection. Advances in high-density SNP genotyping and genomic methodologies now allow the investigation of genetic diversity, differentiation, and signatures of selection across cattle populations (Fan et al., 2010). In this context, approaches such as FST-based analyses, runs of homozygosity (ROH), and multivariate methods provide complementary information to identify genomic regions associated with adaptation and economically relevant traits (Holsinger & Weir, 2009; Peripolli et al., 2017). These genomic tools, together with modern genetic evaluation and validation frameworks, form the methodological basis of the analyses described in the following sections.

Single Nucleotide Polymorphisms

The development of Single Nucleotide Polymorphism (SNP) markers has been a key driver in the advancement of livestock genomics. Among domestic species, cattle (Bos taurus) were one of the first species for which a reference genome sequence became available (Fan et al., 2010). The genome of a Hereford cows was assembled using a combination of whole-genome shotgun sequencing and bacterial artificial chromosome (BAC) sequencing, yielding an estimated genome size of approximately 2.9 Gb and identifying more than 20,000 protein-coding genes (Fan et al., 2010). The release of this draft assembly in 2009 provided the foundation for large-scale SNP discovery and for the development of high-density, cattle-specific genotyping platforms (Fan et al., 2010). The most widely used are based on Illumina BeadArray technology, including the BovineSNP50 (~54K SNPs) and the BovineHD (>500K SNPs), alongside lower-density panels intended for routine breeding applications and imputation. Although Illumina BeadArray technology has been dominant in cattle, Affymetrix GeneChip platforms have also been used in livestock genomics, reflecting two major technological approaches to SNP genotyping (Fan et al., 2010).

The widespread availability of bovine SNP arrays has profoundly influenced both quantitative genetics and practical breeding programs. A major application is genomic selection, which exploits genome-wide SNP information to predict genomic estimated breeding values (GEBV) and accelerate genetic gain for economically important traits; high-density SNP panels are now routinely integrated into national and international dairy cattle evaluation programs (Ma et al., 2019). In addition, SNP arrays have been extensively applied in genome-wide association studies (GWAS) to investigate the genetic architecture of complex traits, including milk production and composition, feed efficiency, fertility, health and robustness-related traits (Bastinaasen et al., 2011; Ma et al., 2019; Liu et al., 2020). Beyond prediction and association mapping, SNP array data have enabled a broad range of population genomic analyses, including the study of linkage disequilibrium and haplotype structure, the detection of selection signatures, and the characterization of genetic diversity, admixture patterns and phylogenetic relationships among livestock species worldwide (Fan et al., 2010; Groeneveld et al., 2010; Cesarani et al., 2019). Medium-density panels such as the BovineSNP50 have proven particularly useful for distinguishing taurine and indicine lineages, assessing within-breed variation and quantifying genetic relationships among local and international breeds (Lin et al., 2010; Falchi et al., 2023).

However, commercial SNP arrays may suffer from ascertainment bias, as markers are often discovered in a limited set of intensively selected breeds, leading to reduced informativeness and underrepresentation of rare and population-specific variants in local populations (Albrechtsen et al., 2026; Warburton et a., 2026). Whole-genome sequencing (WGS), which determines nearly the complete DNA sequence of an individual, mitigates this limitation by capturing the full spectrum of genetic variation, including low-frequency and private alleles, and is therefore increasingly applied to local breeds to better characterize their genetic diversity and identify unique adaptive variants (Matukumalli et al., 2009; Warburton et a., 2026).

Univariate and multivariate methods for population genomic analyses

Population-based summary statistics and multivariate structure analyses provide complementary tools to describe genetic differentiation and demographic history from SNP data.

Wright fixation index

A first class of indices derives from Wright's fixation coefficient, which measures deviations from Hardy–Weinberg expectations (Brown, 1970). At the within-population level, the fixation index F quantifies the deficit ($F > 0$) or excess ($F < 0$) of heterozygotes relative to random mating (Wright, 1951). At the between-population level, F_{ST} measures the proportion of total genetic variation that is due to allele frequency differences among populations (Wright, 1951; Weir & Cockerham, 1984). By definition, F_{ST} ranges from 0 to 1, with values close to 0 indicating little genetic differentiation (i.e., similar allele frequencies) among the considered populations, whereas values approaching 1 indicating strong differentiation and substantial genetic separation (Weir & Cockerham, 1984). From a practical standpoint,

F_{ST} can be estimated from genotype or allele-frequency data using several alternative estimators. The most widely adopted in population genomics is the Weir and Cockerham (1984) unbiased estimator (θ), implemented in tools such as VCFtools (Danecek, 2011) and PLINK (Purcell et al. 2007) and applicable either per-SNP or in sliding windows across the genome. Other commonly used formulations include Hudson's estimator, built on average pairwise sequence/genotype differences (Hudson et al., 1992); Nei's estimator similarly based on partitioning genetic diversity into within- and total components, quantifying differentiation as the fraction of total heterozygosity attributable to among-population differences (Nei et al., 1977; Nei et al., 1984). Software such as FSTest provides a unified framework to compute and compare all these approaches under the same input format (Vahedi et al., 2024).

Estimates of F_{ST} are therefore widely used to summarize genetic differentiation among breeds and to reconstruct population relationships through pairwise-distance matrices underlying clustering, multidimensional scaling (MDS)/principal coordinates analysis (PCoA) or phylogenetic representations (Gurgul et al., 2020; Kitada et al., 2021). In genome-wide applications, locus-specific or window-based F_{ST} profiles are compared against the genomic background to identify outlier regions exhibiting unusually high differentiation (Holsinger & Weir, 2009). Such peaks are commonly interpreted as candidate signatures of diversifying selection, including local adaptation or strong artificial selection, and are often refined by smoothing procedures (e.g., LOWESS/LOESS) and thresholding strategies to enhance the signal-to-noise ratio (Pintus et al., 2014; Sorbolini et al., 2015; Cesarani et al., 2019). In livestock genomics, this approach has been extensively applied to detect breed-specific selective sweeps associated with economically relevant traits (e.g., production type, growth, fertility, or environmental resilience), and to support the identification of genomic regions contributing to divergence between dairy, beef, or dual purpose populations (Gurgul et al., 2020). Beyond selection mapping, overall F_{ST} levels provide insight into demographic processes—such as drift, isolation, migration, and admixture—and are routinely integrated with complementary statistics to disentangle selective from neutral sources of differentiation (Holsinger & Weir, 2009; Peter, 2016). Importantly, although early heuristic cut-offs (e.g., $F_{ST} \approx 0.05 - 0.15$ as "moderate" differentiation) are still reported, broad empirical surveys across breeds highlight substantial variability in F_{ST} distributions, underscoring that inference should be grounded in the trait, species, and study specific genomic context rather than in universal thresholds (Hall, 2022).

Runs of Homozygosity

Runs of homozygosity (ROH) are defined as contiguous stretches of homozygous genotypes in diploid organisms (Gibson et al., 2006). These genomic segments were first described in human populations and are thought to arise from the transmission of identical haplotypes from common ancestors (Broman and Weber, 1999). Consequently, ROH are usually interpreted as identical by descent (IBD) regions (Biscarini et al., 2014). There is no universal consensus on the criteria used to define ROH, and several parameters are typically specified in order to distinguish true IBD segments from homozygosity arising by chance (Ferenčaković et al., 2013a; Hillestad et al., 2018; Peripolli et al., 2017; Falchi et al., 2024a). A central criterion is the minimum physical length of a segment, commonly expressed in megabases (Mb). In cattle and human studies, thresholds of around 1 Mb are often applied (Gaspa et al., 2014; Marras et al., 2015), whereas analyses based on low-density SNP panels (e.g. 50K) tend to adopt longer minimum lengths (e.g., 4 Mb) to avoid short, unreliable segments (Purfield et al., 2012; Mastrangelo et al., 2016). Another key parameter is the minimum number of consecutive homozygous SNPs (e.g., 15–50) required for a segment to be declared as a true ROH (Gaspa et al., 2014; Ferenčaković et al., 2013b; Falchi et al., 2024b), which is usually chosen as a function of SNP density and sample size to control false positives (Purfield et al., 2012). A maximum gap between adjacent SNPs (e.g., 1 Mb) is also often specified; homozygous SNPs separated by larger distances are not considered part of the same ROH (Ferenčaković et al., 2013b; Gaspa et al., 2014; Biscarini et al., 2014; Marras et al., 2015). Finally, thresholds are set for the number of allowed heterozygous genotypes and missing calls within a ROH. In stricter definitions, especially when using low-density panels, heterozygous genotypes are not allowed (Gaspa et al., 2014; Marras et al., 2015; Hillestad et al., 2018), whereas with high-density data a small number of heterozygous and missing SNPs can be tolerated in order to account for genotyping errors, particularly in long segments (McQuillan et al., 2008; Ferenčaković et al., 2013a). Overall, the choice of parameters strongly depends on SNP density: low-density chips are less reliable for short ROH (<4–5 Mb), but are adequate for detecting long ROH (>5 Mb) that reflect more recent inbreeding (Purfield et al., 2012; Ferenčaković et al., 2013a).

Several software tools are available for ROH detection. Among the most widely used are PLINK (Purcell et al., 2007) and the R package "*DetectRuns*" (Biscarini et al., 2018). PLINK implements a sliding-window algorithm that scans the genome for windows with a high proportion of homozygous SNPs (Howrigan et al., 2011; Ferenčaković et al., 2013a). Depending on marker density and parameter settings, this approach may generate spurious runs or miss very long segments. LD pruning has been proposed in some studies to reduce the effect of high-LD regions; however, its use remains debated because removing correlated SNPs may also reduce marker density and affect ROH detection (Howrigan et al., 2011). "*DetectRuns*" package provides both a strictly consecutive approach (ConsecutiveRUNS) and a sliding-window approach (SlidingRUNS), offering flexibility in how ROH are defined (Biscarini et al., 2018). Other commonly used programs include SNP1101, which has been reported to produce inbreeding estimates closer to the "true" simulated inbreeding than PLINK when using a minimum window size of 20–50 SNPs (Forutan et al., 2018), and BCFtools, which uses hidden Markov models (HMMs) to exploit population sequencing information, recombination rates and allele frequencies (Forutan et al., 2018). Additional tools such as SNPandVariation Suite (SVS), GERMLINE, cgaTOH, and BEAGLE have also been applied in cattle and human studies, often in combination with specific criteria for minimum SNP count and tolerance for heterozygous or missing genotypes (Ferenčaković et al., 2013a, 2013b; Howrigan et al., 2011; Peripolli et al., 2017). HMM-based tools like BEAGLE explicitly model LD and haplotype probabilities, but tend to provide more conservative ROH calls than PLINK (Howrigan et al., 2011).

From an applied perspective, ROH are highly informative for population genetics, animal breeding and human genetics, because they capture detailed information on genomic structure and demographic history (Peripolli et al., 2017). A fundamental application is the quantification of autozygosity and the estimation of genomic inbreeding coefficients (F_{ROH}) (Peripolli et al., 2017). F_{ROH} is often regarded as a more accurate measure of individual inbreeding than pedigree-based coefficients

(F_{PED}), because it reflects the realized proportion of the genome that is IBD and incorporates stochastic variation introduced by recombination, which is not captured by pedigree-based approaches (McQuillan et al., 2008; Kim et al., 2013; Ferenčaković et al., 2013b; Forutan et al., 2018). ROH are particularly powerful in detecting inbreeding arising from distant ancestors, whose contribution is frequently absent or underestimated in pedigree records (Ferenčaković et al., 2013b; Peripolli et al., 2017).

The length distribution of ROH provides insights into past demographic events (Peripolli et al., 2017). Long ROH (e.g. >10 Mb) are indicative of recent inbreeding, as there has been limited opportunity for recombination to fragment the IBD haplotype, whereas shorter ROH (<2–4 Mb) reflect more ancient common ancestry, such as founder effects or historical bottlenecks occurring tens or hundreds of generations ago (Gibson et al., 2006; Ferenčaković et al., 2013b). By comparing the frequency and length of ROH across the genome, it is possible to infer processes such as variation in effective population size, geographic isolation, and demographic bottlenecks (Purfield et al., 2012; Kirin et al., 2010, Peripolli et al., 2017). ROH are also widely used to identify genomic regions that have been subject to strong natural or artificial selection (Peripolli et al., 2017). Genomic segments where ROH occur at high frequency in a population, often termed “ROH islands” or hotspots, are interpreted as signatures of directional selection (Nothnagel et al., 2010; Gorssen et al., 2021). Under artificial selection, for example for milk yield or growth traits, favorable alleles and their surrounding haplotypes increase in frequency and may become fixed in homozygous state; through hitchhiking effect, neighboring markers are also dragged to homozygosity, forming extended ROH (Cesarani et al., 2021; Falchi et al., 2023). Similar processes can underline adaptation to environmental conditions or resistance to infectious, metabolic or reproductive diseases (Peripolli et al., 2017; Biscarini et al., 2014; Falchi et al., 2023). In livestock and small local populations, ROH analysis is therefore central to genetic management and conservation, as it enables characterization of genomic variability, monitoring of inbreeding trends, and the design of mating strategies that reduce the transmission of long ROH to offspring, thereby preserving diversity and fitness (Mastrangelo et al., 2016). Finally, ROH analyses have become an important tool in quantitative genetics as a complement or alternative to single marker genome-wide association studies (GWAS) (Biscarini et al., 2014). The use of ROH for association purposes is motivated by the hypothesis that part of the genetic architecture of complex traits and disease susceptibility involves recessive or partially recessive variants inherited identical by descent, and by the ability of ROH to capture regional effects and hitchhiking phenomena (Biscarini et al., 2014). Building on this rationale, Cesarani et al. (2021) formalized a genome-wide ROH analysis (GWRA) to systematically test the association between ROH occurrence and quantitative production traits. Testing for associations between the presence of ROH in specific genomic regions and phenotypes (e.g., case–control status or quantitative trait values) provides a natural framework to detect recessive mutations. Unlike single-SNP GWAS, ROH-based approaches consider contiguous segments and explicitly capture hitchhiking effect, by testing the combined effect of loci that have become fixed together (Biscarini et al., 2014; Cesarani et al., 2021). In dairy cattle, for example, ROH have been used to identify genomic regions associated with susceptibility to infectious, metabolic, reproductive and locomotor diseases. Many ROH islands overlap with known QTL for milk production traits, highlighting the adverse correlation between selection for high production and animal health (Biscarini et al., 2014). Similar GWRA studies in Simmental bulls have demonstrated strong relationships between autozygosity and traits such as milk yield and fat and protein content, and have identified multiple genomic regions previously associated with these traits (Cesarani et al., 2021).

Stepwise discriminant analysis and Canonical discriminant analysis

Multivariate discriminant techniques play a central role in animal sciences when the objective is to classify individuals, discriminate predefined groups, or identify biological patterns from complex and high-dimensional datasets (Sorbolini et al., 2016; Acciaro et al., 2020; Congiu et al., 2024a). Among these techniques, Stepwise Discriminant Analysis (SDA) and Canonical Discriminant Analysis (CDA) have been extensively applied in genomics, precision livestock farming, food traceability, and behavioral studies, often in a sequential and complementary manner (Dimauro et al., 2013; Manca et al., 2020a; Congiu et al., 2024a).

The methodological foundation of SDA derives from the concept of stepwise regression, a statistical approach developed to identify an optimal subset of explanatory variables capable of describing the variability of a response variable using the minimum number of predictors (Rao et al., 1992). The primary rationale behind stepwise procedures is the principle of parsimony, whereby unnecessary or weakly informative variables are excluded to reduce noise, mitigate multicollinearity, and improve model interpretability and predictive efficiency (Miller et al., 1984). Stepwise procedures are particularly relevant in contexts characterized by a large number of correlated predictors, where computational efficiency and experimental costs are also critical considerations (Miller et al., 1984; Cellesi et al., 2016). From the conceptual idea of Fisher (1922) to the development of an algorithm by Efroymson (1960), variable selection is performed through iterative algorithms, including forward selection, backward elimination, or a combination of both, based on predefined statistical criteria for variable inclusion or removal (Rao et al., 1992). In forward selection, variables are entered sequentially by choosing, at each step, the predictor that yields the largest improvement in model fit, commonly quantified as the greatest increase in explained variance (e.g., partial R^2) relative to the current model. Conversely, backward elimination starts from the full model and removes, step by step, the variable associated with the smallest contribution to explained variance (i.e., the smallest decrease in partial R^2). Finally, the combined stepwise procedure alternates inclusion and removal by applying separate entry and stay thresholds to retain only variables that continue to provide a meaningful incremental contribution (Efroymson, 1960; Miller, 1984). This stepwise logic has been extended to discriminant analysis (DA), a supervised approach for classifying observations into predefined groups, through SDA, which is specifically designed to select the subset of variables that maximizes discrimination among those groups (Dimauro et al., 2013). In the context of SDA, the selection process aims to identify variables that contribute most effectively to group separation, often based on criteria such as partial squared correlation, Wilks' lambda, or related F-statistics (Miller et al., 1984; Dimauro et al., 2015). A key advantage of SDA lies in its ability to address high-dimensional datasets, where the number of variables largely exceeds the number of observations, a common scenario in animal genomics, metabolomics, and sensor-based data (Manca et al., 2020a; Acciaro et al., 2020). By selecting a reduced set of linearly independent and highly informative variables, SDA ensures that subsequent multivariate analyses can be performed on full-rank data matrices (Manca et al., 2020a). Also in SDA, the three computational methods used in multiple regression analysis can be exploited. However, in genomic contest, where the number of SNP-variables far exceeds that of the genotyped animals, the forward algorithm is recommended.

CDA represents the natural analytical continuation following SDA. The discriminant analysis was originally introduced by Fisher in 1936 as a method to classify observations into predefined groups based on quantitative traits (Cellesi et al., 2016). CDA, in particular, is a supervised multivariate technique that combines classification and dimensionality reduction, and is conceptually related to canonical correlation analysis and principal component analysis (Manca et al., 2020a). Unlike Principal Component Analysis (PCA), which maximizes total variance of data, without considering group structure, CDA explicitly aims to maximize the variance between predefined groups while minimizing within-group variability (Cellesi et al., 2016). CDA derives a set of canonical functions, defined as linear combinations of the original variables, that summarize the multivariate differences among groups. The coefficients associated with each canonical function quantify the relative contribution of each variable to group separation, allowing for both statistical discrimination and biological interpretation (Cellesi et al., 2016; Acciaro et al., 2020). The statistical significance of group separation is typically evaluated using the Mahalanobis distance and the associated Hotelling's T^2 test, ensuring a rigorous assessment of discriminant performance (Mardia, 1975; De Maesschalck et al., 2000) .

Within animal sciences, SDA is frequently implemented as a preliminary filtering step to identify a reduced set of highly informative predictors, which are then used as input for CDA to enhance model stability, interpretability, and statistical reliability. This sequential SDA–CDA framework has been successfully applied across a range of biological and zootechnical contexts. In animal genomics, SDA has been used to derive limited panels of SNPs capable of discriminating among cattle or sheep breeds, and CDA has subsequently been applied to maximize group separation and support breed classification and food traceability, including the assignment of meat or dairy products to breed, geographical origin, or production system (Dimauro et al., 2012; Dimauro et al., 2015; Acciaro et al., 2020). Similar SDA-guided variable selection strategies have been adopted in metabolomic and lipidomic studies, where subsets of fatty acids with high discriminant power were first identified through SDA and then analyzed using CDA to characterize multivariate patterns associated with conditions such as milk fat depression, thereby highlighting candidate biomarker profiles (Conte et al., 2018). Beyond these applications, SDA and CDA have been combined within multivariate genome-wide association study pipelines — often together with classical discriminant analysis — to identify SNP sets associated with quantitative traits (e.g., body weight, growth rate, carcass traits, and meat quality), particularly under limited sample size conditions (Manca et al., 2020a). In precision livestock farming, SDA has been used to select accelerometer-derived features, while CDA has been employed to discriminate and classify behavioral states such as grazing, walking, and resting, demonstrating the utility of the SDA–CDA sequence for both high-dimensional "omics" data and sensor-based phenotyping (Congiu et al., 2024a). Finally, in addition to genomic and behavioral applications, CDA has also been applied to lactation curve data to support the early identification of cows with potentially persistent lactations (Manca et al., 2020b). In this approach, parameters estimated from mathematical models fitted to early daily milk-yield records are used as multivariate predictors, and CDA is then used to test group separation and classify lactations to inform herd management decisions (Manca et al., 2020b). In several applications, multivariate discriminant techniques have been further integrated with univariate approaches, such as Wright's fixation index (F_{ST}) or Analysis of variance (ANOVA), to increase robustness and confidence in the detected signals. In genomic studies, the comparison between multivariate SDA–CDA results and univariate F_{ST}-based scans has allowed the identification of shared SNPs, thereby strengthening evidence for true signatures of divergent selection between populations (Congiu et al., 2024b). Similarly, in behavioral, metabolomic, and lipidomic studies, univariate analyses are often used as an initial screening step, while multivariate methods account for correlation structures among variables and reveal discriminant patterns that may not emerge from single-variable testing alone (Conte et al., 2020; Congiu et al., 2024a).

Genetic evaluation and validation in quantitative genetics

Genetic evaluation models in quantitative genetics enable the estimation of breeding values, whereas validation approaches quantify the consistency and predictive ability of the resulting genetic indices.

Genetic evaluation models

Livestock selection has been practiced for millennia, as evidenced by early descriptions of cattle breeding choices in classical sources such as Virgil's "Georgica" (Cellesi, 2014). The theoretical basis of modern breeding, however, emerged much later with the formalization of variation and selection in domestic animals, emphasizing the cumulative effect of sustained human choice over generations (Darwin, 1869; Cellesi, 2014). Early quantitative thinking was further strengthened by

Galton's "law of regression", which introduced a statistical framework for predicting offspring performance from parental values and anticipated later developments in selection indices (Galton, 1886; Hazel, 1943). The first rigorous application of statistics to genetic evaluation in dairy cattle is commonly attributed to Lush (1931), who promoted progeny testing as a more accurate basis for ranking sires than individual performance or pedigree alone, thereby linking the average records of daughters to the sire's underlying genetic merit. Shortly thereafter, the selection index theory was formalized, defining aggregate genotype (H) as a weighted combination of economically important traits and providing a principled method to optimize genetic gain under multi-trait breeding goals (Hazel, 1943). A major methodological breakthrough was achieved by Henderson (1975) with the development of Best Linear Unbiased Prediction (BLUP) for mixed models, which became the international standard for routine genetic evaluation. Crucially, Henderson's efficient computation of the inverse numerator relationship matrix (A^{-1}) enabled practical implementation of the animal model by exploiting the full set of pedigree relationships, thereby improving the accuracy and consistency of estimated breeding values across large populations (Henderson, 1975).

On the basis of these developments, modern genetic evaluation in livestock breeding is commonly performed using either single-trait models (STM) or multiple-trait models (MTM), both typically implemented within the animal model framework. In STM genetic parameters and breeding values are estimated for one trait at a time, without explicitly accounting for genetic or environmental correlations with other traits (Gengler et al., 1999). While this approach is computationally efficient and relatively straightforward to implement, it does not exploit the information contained in correlated traits and may therefore lead to suboptimal accuracy, particularly in the presence of selection on correlated characteristics (Sorensen & Johansson, 1992; Negussie et al., 2006). In contrast, MTM jointly analyze two or more traits, explicitly modelling non-zero genetic and environmental covariances among them. By incorporating covariance structures for additive genetic, permanent environmental and, when relevant, herd-by-sire interaction effects across traits, MTM are able to borrow information from correlated phenotypes and provide more reliable estimates of breeding values (Schaeffer, 1984; Thompson & Meyer, 1986). In this perspective, the main advantage of MTM lies in its ability to increase the accuracy of genetic evaluation, especially for traits with low heritability or limited phenotypic information (Srivastava et al., 2019). Moreover, when multiple traits are under selection, MTM better captures the underlying genetic trends, including correlated responses (Negussie et al., 2006; Srivastava et al., 2019). In such cases, information from genetically correlated and more heritable traits can substantially improve the reliability of estimated breeding values, an effect that is particularly pronounced when genetic correlations differ markedly from residual correlations (Negussie et al., 2006; VanRaden et al., 2014; Srivastava et al., 2019). MTM are also well suited to situations with missing data, as they can accommodate unbalanced recording schemes by exploiting the available information from correlated traits (VanRaden et al., 2014; Srivastava et al., 2019). Furthermore, MTM play a crucial role in mitigating selection bias when selection decisions have been based on traits not explicitly included in a single-trait analysis, a common scenario in modern breeding programs where production, fertility and functional traits are often genetically antagonistic (Negussie et al., 2006; Srivastava et al., 2019). Despite these advantages, the increased computational complexity of MTM represents a practical limitation, as the number of parameters to be estimated grows rapidly with the number of traits (Mistzal, 1993). Consequently, STM may still be adequate when traits are weakly correlated or when computational resources are constrained (VanRaden et al., 2014). Nevertheless, advances in numerical algorithms and model parameterization have progressively reduced these limitations, making MTM increasingly feasible for routine genetic evaluations (Mistzal, 1993; Gengler et al., 1999; VanRaden et al., 2014).

Genetic evaluation under both STM and MTM frameworks is typically implemented using mixed-model methodology based on BLUP and its extensions, including pedigree- and genomic-based formulations (Calus et al., 2009; Srivastava et al., 2019). In practice, large-scale evaluations are commonly performed using dedicated software such as the BLUPF90 family of programs, which provides a flexible environment for fitting single- and multiple-trait animal models and for estimating (co)variance components via REML (Mistzal et al., 2022). In addition, MiX99 has been widely adopted in national and international routine evaluations, particularly for computationally demanding multiple-trait analyses, where efficient solvers such as preconditioned conjugate gradient algorithms are required to handle large datasets (Lidauer et al., 2022). The availability of dense SNP panels enabled the development of genomic selection, in which genome-wide marker information is used to predict breeding values (Meuwissen et al., 2001; Misztal et al., 2024). Early genomic prediction approaches relied on SNP-based regression models such as SNP-BLUP or RR-BLUP, which assume a large number of markers with small and homogeneous effects (VanRaden et al., 2008). In contrast, Bayesian methods (e.g. BayesA, BayesB, and related extensions) allow heterogeneous marker variances and perform variable selection, thereby accommodating situations where a subset of loci may have larger effects (Meuwissen et al., 2001; Fernando et al., 2014; Misztal et al., 2020). Subsequently, genomic BLUP (GBLUP) was proposed as a computationally efficient alternative, replacing the pedigree-based relationship matrix (A) with a genomic relationship matrix (G) derived from SNP data, thereby improving the accuracy of genetic evaluations, particularly for young animals without phenotypes (VanRaden, 2008; Misztal et al., 2020). More recently, the single-step GBLUP (ssGBLUP) approach has become the industry standard, as it integrates pedigree, phenotypic and genomic information simultaneously within a unified mixed-model framework (Misztal et al., 2020). This method avoids biases associated with separate evaluations and allows the inclusion of both genotyped and non-genotyped animals, making it particularly suitable for routine large-scale genetic evaluations (Misztal et al., 2020).

Validation approaches: from Interbull dairy procedures to beef cattle applications

The validation of genetic evaluation models represents a critical step in animal breeding, as statistical models used to estimate breeding values are, by definition, simplified representations of complex biological processes and are therefore only approximate (Kennedy et al., 1988). Model validation is essential to ensure that estimated breeding values are accurate, unbiased, and reliable for supporting selection decisions (Palucci & Dürr, 2014). In particular, the effectiveness of a breeding program depends on the accuracy of genetic predictions, which directly influences the expected response to selection and the ranking of animals (Legarra & Reverter, 2018; Macedo et al., 2020a). Systematic bias or inappropriate dispersion of estimated breeding values may lead to distorted genetic trends and suboptimal selection choices, such as the incorrect identification of superior young animals or an overestimation of genetic progress (Legarra & Reverter, 2018; Macedo et al., 2020a; Macedo et al., 2020b). Beyond supporting optimal selection decisions at the national level, model validation is also fundamental to guarantee consistency and comparability across breeding programs (Boichard et al., 1995). In international contexts, biased or poorly validated genetic evaluations can severely compromise the exchange of genetic material and the interpretation of international comparisons, especially when evaluations are converted across countries (Bonaiti et al., 1993). For this reason, the International Bull Evaluation Service (Interbull) was established to harmonize genetic evaluations and to ensure the reliability of estimated breeding values exchanged internationally (Palucci & Dürr, 2014). Interbull has adopted a standardized validation framework, based on three complementary methods (Method I, Method II and Method III) proposed by Boichard et al. (1995), to detect bias, inflation or deflation of breeding values, and inconsistencies arising from changes in national evaluation systems. These validation tests are mandatory whenever a genetic evaluation model is modified, and they provide a formal framework to assess the stability and credibility of genetic indices prior to their international use (Palucci & Dürr, 2014). Within this framework, the three methods target different but complementary aspects of potential bias and instability in genetic evaluations. Method I compares genetic trends obtained using first-lactation records only with those derived from evaluations based on all available lactations in routine national genetic evaluations. Because first-lactation data are less sensitive to potential misspecification of age or parity adjustment factors, substantial discrepancies between the two trends may indicate the presence of bias when later lactations are included. Method II investigates within-sire temporal patterns in daughter-based performance measures, such as daughter deviations (DD) or daughter yield deviations (DYD), which are expected to be free from systematic non-genetic year effects once the environmental components included in the model have been accounted for. Under an unbiased evaluation system, the estimated year effect should fluctuate around zero; consistent increasing or decreasing patterns therefore suggest, respectively, overestimation or underestimation of the national genetic trend. Method III evaluates the stability of official sire breeding values across successive routine evaluation runs—typically separated by several years—by assessing whether observed changes exceed what would be expected from random variation due to the addition of new progeny information. Owing to its general applicability and limited data requirements, Method III has been widely adopted across traits and evaluation systems (Boichard et al., 1995; Bonaiti et al., 1993; Palucci & Dürr, 2014). These Interbull validation procedures are, however, primarily implemented within dairy cattle evaluations, whereas equivalent standardized frameworks are not yet routinely established for beef cattle populations. The direct transfer of these procedures to beef cattle is not straightforward, mainly because they were designed around the data structure, recording schemes and trait definition typical of dairy evaluations (Boichard et al., 1995; Bonaiti et al., 199). In contrast, beef cattle evaluations often rely on more heterogeneous recording systems, smaller population sizes, and traits that are frequently measured once in the animal’s lifetime or expressed on categorical scales (Eriksson et al., 2004; VanRaden et al., 2014). These differences complicate the direct and standardized application of validation frameworks originally designed for dairy cattle and highlight the need for adapted or alternative approaches when assessing the reliability of genetic evaluations in beef populations.

More recently, Legarra and Reverter (2018) proposed the Linear Regression method (LRM) as a general framework for validating and comparing genetic evaluation models, with particular relevance for populations where traditional validation approaches are difficult to implement. The LRM is based on the comparison of estimated breeding values obtained for the same group of animals at two different time points, using an earlier evaluation based on partial information and a subsequent evaluation incorporating additional data (Legarra & Reverter, 2018; Macedo et al., 2020b). By analyzing how breeding values change as new information becomes available, the method allows the assessment of systematic bias, dispersion, and relative gain in accuracy without requiring highly accurate reference breeding values or pre-corrected phenotypes (Legarra & Reverter, 2018). This feature makes the LRM especially suitable for beef cattle evaluations, which are often characterized by smaller progeny groups, heterogeneous recording schemes, low-heritability traits, and limited availability of conventional reference measures such as daughter yield deviations (DYD), deregressed estimated breeding values (dEBV), or adjusted phenotypes used as independent predictands for validation (Macedo et al., 2020b). Because LRM relies on internal consistency between successive evaluations rather than on external validation datasets, it can be applied to a wide range of traits, including reproductive and fitness-related traits, as well as indirectly observed effects commonly encountered in beef breeding programs, such as maternal genetic effects, permanent maternal environmental effects, interactions between direct and maternal effects, and herd-management related effects (Legarra & Reverter, 2017; Legarra & Reverter, 2018).

In this context, the overall aim of this thesis was to investigate genetic diversity, population structure, and genomic signatures of differentiation in local cattle breeds, and to explore methodological approaches for genetic evaluation and validation in cosmopolitan populations. Specifically, the thesis aims to (i) characterize genetic diversity and relationships among local cattle populations, (ii) identify shared genomic backgrounds and candidate regions associated with population differentiation, (iii) develop tools for the genetic evaluation of functional traits, and (iv) assess approaches for the validation of genetic indices in beef cattle.

Outline of the Thesis

This thesis is organized into a general introduction, four experimental chapters, and a final chapter providing overall conclusions.

Chapter 1 introduces local and cosmopolitan cattle breeds and reviews the main statistical and genetic approaches used to investigate biodiversity and genetic relatedness among local populations, as well as economically relevant traits and validation methods for genetic indices in cosmopolitan breeds.

Chapters 2 and 3 focus on local cattle breeds. Chapter 2 investigates genetic diversity among Spanish cattle populations and subpopulations, whereas Chapter 3 examines the shared genomic background among three local Sardinian cattle populations.

Chapters 4 and 5 address quantitative genetics in cosmopolitan cattle populations. Chapter 4 proposes a fertility index for Italian Jersey cows, while Chapter 5 evaluates the application of a novel approach for validating genetic indices in beef cattle.

Finally, the concluding chapter summarizes the main findings from the four experimental studies and discusses their broader implications.

References

Acciaro, M., Decandia, M., Sitzia, M., Manca, C., Giovanetti, V., Cabiddu, A., ... & Dimauro, C. (2020). Discriminant analysis as a tool to identify bovine and ovine meat produced from pasture or stall-fed animals. Italian Journal of Animal Science, 19(1), 1065-1070. doi:10.1080/1828051X.2020.1816507

Ajmone-Marsan P., Garcia J. F., and Lenstra J. A. 2010. On the origin of cattle: how aurochs became cattle and colonized the world. Evolutionary Anthropology:Issues, News, and Reviews 19(4):148–157. doi:10.1002/evan.20267.

Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Molecular biology and evolution*, *27*(11), 2534-2547. doi: 10.1093/molbev/msq148

Bastinaasen, J., Wijga, S., Calus, M., Veerkamp, R., Bovenhuis, H., Berry, D., ... & Wall, E. (2011). GWAS for robustness traits. *Interbull Bulletin*, (44).

Beja-Pereira A., Caramelli D., Lalueza-Fox C., Vernesi C., Ferrand N., Casoli A., Goyache F., Royo L. J., Conti S., Lari M., Martini A., Ouragh L., Magid A., Atash A., Zsolnai A., Boscato P., Triantaphylidis C., Ploumi K., Sineo L., Mallegni F., Taberlet P., Erhardt G., Sampietro L., Bertranpetit J., Barbujani G., Luikart G., and Bertorelle G. 2006. The origin of European cattle: evidence from modern and ancient DNA. Proceedings of the National Academy of Sciences 103(21):8113–8118. doi:10.1073/pnas.0509210103.

Biscarini, F., Biffani, S., Nicolazzi, E. L., Morandi, N., & Stella, A. (2014). Applying runs of homozygosity to the detection of associations between genotype and phenotype in farm animals. *Reproduction*, *5*(6656617), 6976839.

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). *detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. R package version 0.9. 6*.

Boichard, D., Bonaiti, B., Barbat, A., & Mattalia, S. (1995). Three methods to validate the estimation of genetic trend for dairy cattle. *Journal of dairy science*, *78*(2), 431-437. doi: https://doi.org/10.3168/jds.S0022-0302(95)76652-8

Bonaiti, B., Boichard, D., Barbat, A., & Mattalia, S. (1993). Problems arising with genetic trend estimation in dairy cattle. *Interbull Bulletin*, (8).

Broman, K. W., & Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *The American Journal of Human Genetics*, *65*(6), 1493-1500.

Brown, A. H. D. (1970). The estimation of Wright's fixation index from genotypic frequencies. *Genetica*, *41*(1), 399-406. doi: 10.1007/BF00958921

Calus, M. P. (2010). Genomic breeding value prediction: methods and procedures. *animal*, *4*(2), 157-164. doi: 10.1017/S1751731109991352

Cellesi, M. (2014). Statistical tools for genome-wide studies.

Cellesi, M., Dimauro, C., Gaspa, G., & Macciotta, N. P. P. (2016). *Analisi statistica multivariata*. Teaching material for the course of Multivariate Statistical Analysis, Department of Agriculture, University of Sassari, Italy.

Cesarani A, Sechi T, Gaspa G, Usai MG, Sorbolini S, Macciotta NPP, Carta A. 2019. Investigation of genetic diversity and selection signatures between Sarda and Sardinian Ancestral black, two related sheep breeds with evident morphological differences. Small Ruminant Res. 177:68– 75. doi: 10.1016/j.smallrumres.2019.06.014.

Cesarani, A., Gaspa, G., Pauciullo, A., Degano, L., Vicario, D., & Macciotta, N. P. (2021). Genome-wide analysis of homozygosity regions in European Simmental bulls. *Journal of Animal Breeding and Genetics*, *138*(1), 69-79. doi: 10.1111/jbg.12502.

Congiu, M., Cesarani, A., Macciotta, N. P. P., Rassu, S. P. G., Giovanetti, V., & Dimauro, C. (2024a). Using tri-axial accelerometers data to predict behavior activity of grazing donkeys. Computers and Electronics in Agriculture, 227, 109582. doi: 10.1016/j.compag.2024.109582

Congiu, M., Cesarani, A., Falchi, L., Macciotta, N. P. P., & Dimauro, C. (2024b). Combined use of univariate and multivariate approaches to detect selection signatures associated with milk or meat production in cattle. Genes, 15(12), 1516. doi: 10.3390/genes15121516

Conte, G., Dimauro, C., Serra, A., Macciotta, N. P. P., & Mele, M. (2018). A canonical discriminant analysis to study the association between milk fatty acids of ruminal origin and milk fat depression in dairy cows. *Journal of Dairy Science*, *101*(7), 6497-6510. doi: 10.3168/jds.2017-13941

Darwin CR (1869) On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life. London: John Murray.

De Maesschalck R, Jouan-Rimbaud D, Massart DL. 2000. The mahalanobis distance. Chemom Intell Lab Syst. 50(1):1–18. doi: 10.1016/S0169-7439(99)00047-7.

Dimauro, C., Cellesi, M., Steri, R., Gaspa, G., Sorbolini, S., Stella, A., & Macciotta, N. P. P. (2013). Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes. *Animal genetics*, *44*(4), 377-382. doi: 10.1111/age.12021

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

Efroymson, M.A. (1960): Multiple Regression Analysis, Mathematical Methods for Digital Computers, ed. A. Ralston and H.S. Wilf, John Wiley & Sons Inc., New York.

Egger-Danner, C., Cole, J. B., Pryce, J. E., Gengler, N., Heringstad, B., Bradley, A., & Stock, K. F. (2015). Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal*, *9*(2), 191-207. doi: 10.1017/S1751731114002614

Eriksson, S., Nasholm, A., Johansson, K., & Philipsson, J. (2004). Genetic relationships between calving and carcass traits for Charolais and Hereford cattle in Sweden. *Journal of Animal Science*, *82*(8), 2269-2276. doi: 10.2527/2004.8282269x

Falchi, L. (2024a). Various approaches for the genomic analysis of the genetic structure of different cattle populations.

Falchi, L., Cesarani, A., Criscione, A., Hidalgo, J., Garcia, A., Mastrangelo, S., & Macciotta, N. P. P. (2024b). Effect of genotyping density on the detection of runs of homozygosity and heterozygosity in cattle. *Journal of Animal Science*, *102*. doi: 10.1093/jas/skae147

Falchi, L., Cesarani, A., Mastrangelo, S., Senczuk, G., Portolano, B., Pilla, F., & Macciotta, N. P. (2023). Analysis of runs of homozygosity of cattle living in different climate zones. Journal of animal science, 101. doi: 10.1093/jas/skad061

Fan, B., Du, Z. Q., Gorbach, D. M., & Rothschild, M. F. (2010). Development and application of high-density SNP arrays in genomic studies of domestic animals. *Asian-Australasian journal of animal sciences*, *23*(7), 833-847. doi:10.5713/ajas.2010.r.03

Felius M., Beerling M. L., Buchanan D. S., Theunissen B., Koolmees P. A., and Lenstra J. A. 2014. On the history of cattle genetic resources. Diversity 6(4):705–750. doi:10.3390/d6040705.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I., & Sölkner, J. (2013b). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics*, *130*(4), 286-293. doi: 10.1111/jbg.12012

Ferenčaković, M., Sölkner, J., & Curik, I. (2013a). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution*, *45*(1), 42. doi: 10.1186/1297-9686-45-42

Fernando, R. L., Dekkers, J. C., & Garrick, D. J. (2014). A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genetics Selection Evolution*, *46*(1), 50. doi: 10.1186/1297-9686-46-50

Forutan, M., Ansari Mahyari, S., Baes, C. *et al.* Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics* 19, 98 (2018). doi:10.1186/s12864-018-4453-z.

Frantz LA, Bradley DG, Larson G, Orlando L. 2020. Animal domestication in the era of ancient genomics. Nat Rev Genet. 21(8):449–460. doi: 10.1038/s41576-020-0225-0.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. The Journal of the Anthropological Institute of Great Britain and Ireland, 15, 246-263.

Gandini, G., Turri, F., Rizzi, R., Crotta, M., Minozzi, G., and Pizzi, F. (2017). Economic evaluation of genetic improvement in local breeds: The case of the Verzaschese goat. Ital. J. Anim. Sci. 16:199–207. doi: 10.1080/1828051X.2017.1279034

Gaspa G., Marras G., Sorbolini S., Ajmone Marsan P., Williams J. L., Valentini A., Dimauro C., and Macciotta N. P. P. 2014. Genome-wide Homozygosity in Italian Holstein cattle using HD SNP panel. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada. p.1-3. doi: 10.13140/2.1.1142.7200.

Gengler, N., Wiggans, G. R., & Wright, J. R. (1999). Animal model genetic evaluation of type traits for five dairy cattle breeds. *Journal of Dairy Science*, *82*(6), 1350-e1.

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

Giovambattista G., Ripoli M. V., Peral-Garcia P., and Bouzat J. L. 2001. Indigenousdomestic breeds as reservoirs of genetic diversity: the Argentinean Creole cattle. Animal Genetics 32:240–247. doi:10.1046/j.1365-2052.2001.00774.x

Gorssen, W., Meyermans, R., Janssens, S., & Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genetics Selection Evolution*, *53*(1), 2. doi: 10.1186/s12711-020-00599-7

Groeneveld L. F., Lenstra J. A., Eding H., Toro M. A., Scherf B., Pilling D., Negrini R., Finlay E. K., Jianlin H., Groeneveld E., Weigend S. and The GLOBALDIV Consortium. 2010. Genetic diversity in farm animals – a review. Animal Genetics 41:6–31. doi:10.1111/j.1365-2052.2010.02038.x

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., ... & Globaldiv Consortium. (2010). Genetic diversity in farm animals–a review. *Animal genetics*, *41*, 6-31. doi: 10.1111/j.1365-2052.2010.02038.x

Gurgul, A., Jasielczuk, I., Szmatoła, T., Sosin-Bzducha, E., Majewska, A., & Litwińczuk, Z. (2020). Divergent selection signatures of phenotypic and production traits among conserved and commercial cattle breeds. *Livestock Science*, *239*, 104174. doi: 10.1016/j.livsci.2020.104174

Hall SJ. 2022. Genetic differentiation among livestock breeds—values for Fst. Animals. 12(9):1115. doi: 10.3390/ ani1209

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of dairy science*, *92*(2), 433-443. doi: 10.3168/jds.2008-1646

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, *28*(6), 476-490.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 423-447. doi: jstor.org/stable/2529430

Hiemstra SJ, de Haas Y, Meaki-Tanila A, Gandini G, eds. 2010. Local cattle breeds in Europe: development of policies and strategies for self-sustaining breeds. Wageningen, Netherlands: Wageningen Academic Publishers.

Hillestad, B., Woolliams, J. A., Boison, S. A., Grove, H., Meuwissen, T., Våge, D. I., & Klemetsdal, G. (2017). Detection of runs of homozygosity in Norwegian Red: Density, criteria and genotyping quality control. *Acta Agriculturae Scandinavica, Section A—Animal Science*, *67*(3-4), 107-116. doi: 10.1080/09064702.2018.1501088

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F ST. *Nature Reviews Genetics*, *10*(9), 639-650. doi: 10.1038/nrg2611

Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC genomics*, *12*(1), 460. doi: 10.1186/1471-2164-12-460

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*(2), 583-589. doi: 10.1093/genetics/132.2.583

Kennedy, B. W., Schaeffer, L. R., & Sorensen, D. A. (1988). Genetic properties of animal models. *Journal of Dairy science*, *71*, 17-26. doi:10.1016/S0022-0302(88)79975-0

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., ... & Sonstegard, T. S. (2013). Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PloS one*, *8*(11), e80813. doi: 10.1371/journal.pone.0080813

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M., & Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PloS one*, *5*(11), e13996.

Legarra, A., & Reverter, A. (2017). Can we frame and understand cross-validation results in animal breeding?.

Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution*, *50*(1), 53. doi: 10.1186/s12711-018-0426-6

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M., & Strandén, I. (2022). *MiX99: A software package for solving large mixed model equations*. Natural Resources Institute Finland (Luke), Jokioinen, Finland. Release I/2022. FI-31600.

Lin, B. Z., Sasazaki, S., & Mannen, H. (2010). Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers. *Animal science journal*, *81*(3), 281-289. doi: 10.1111/j.1740-0929.2010.00744.x

Liu, L., Zhou, J., Chen, C. J., Zhang, J., Wen, W., Tian, J., ... & Gu, Y. (2020). GWAS-based identification of new loci for milk yield, fat, and protein in Holstein cattle. *Animals*, *10*(11), 2048.doi: 10.3390/ani10112048

Lush, J. L. (1931). The number of daughters necessary to prove a sire. *Journal of Dairy Science*, *14*(3), 209-220.

Ma, L., Cole, J. B., Da, Y. A. N. G., & VanRaden, P. M. (2019). Symposium review: Genetics, genome-wide association study, and genetic improvement of dairy fertility traits. *Journal of dairy science*, *102*(4), 3735-3743. doi: 10.3168/jds.2018-15269

Macedo, F. L., Christensen, O. F., Astruc, J. M., Aguilar, I., Masuda, Y., & Legarra, A. (2020a). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genetics Selection Evolution*, *52*(1), 47. doi: 10.1186/s12711-020-00567-1

Macedo, F. L., Reverter, A., & Legarra, A. (2020b). Behavior of the linear regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *Journal of Dairy Science*, *103*(1), 529-544. Doi: 10.3168/jds.2019-16603

Manca, E., Cesarani, A., Gaspa, G., Sorbolini, S., Macciotta, N. P., & Dimauro, C. (2020a). Use of the multivariate discriminant analysis for genome-wide association studies in cattle. *Animals*, *10*(8), 1300. doi: 10.3390/ani10081300

Manca, E., Cesarani, A., Macciotta, N. P., Atzori, A. S., Pulina, G., & Dimauro, C. (2020b). Use of discriminant statistical procedures for an early detection of persistent lactations in dairy cows. Computers and Electronics in Agriculture, 176, 105657. doi: 10.1016/j.compag.2020.105657

Mardia KV, Kent JT, Bibby JM. 2000. Multivariate analysis. London: Academic Press.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone‐Marsan, P., Valentini, A., ... & Macciotta, N. P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal genetics*, *46*(2), 110-121. doi: 10.1111/age.12259

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M. T., & Portolano, B. (2016). Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal*, *10*(5), 746-754. doi:10.1017/S1751731115002943

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., ... & Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PloS one*, *4*(4), e5350. doi: 10.1371/journal.pone.0005350

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., ... & Wilson, J. F. (2008). Runs of homozygosity in European populations. The American Journal of Human Genetics, 83(3), 359-372.

Medugorac I., Medugorac A., Russ I., Veit-Kensch C. E., Taberlet P., Luntz B., Mix H. M., and Förster M. 2009. Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. Molecular Ecology 18:3394–3410. doi:10.1111/j.1365-294X.2009.04286.x

Meuwissen, T. H. E., I. van den Berg, and M. Goddard. 2021. On the use of wholegenome sequence data for across-breed genomic prediction and fine-scale mapping of QTL. Genet. Sel. Evol. 53:19. doi: 10.1186/s12711-021-00607-4

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. doi: 10.3168/jds.2017-12968.

Miller, A. J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *147*(3), 389-410. doi: 10.2307/2981576

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, LegarraA, Vitezica Z. 2022. BLUPF90 family of programs. Availablefrom: http://nce.ads.uga.edu/software/.

Misztal, I., Lawlor, T. J., & Short, T. H. (1993). Implementation of single-and multiple-trait animal models for genetic evaluation of Holstein type traits. *Journal of dairy science*, *76*(5), 1421-1432. doi:10.3168/jds.S0022-0302(93)77473-1

Misztal, I., & Lourenco, D. (2024). Potential negative effects of genomic selection. *Journal of animal science*, *102*, skae155. doi: 10.1093/jas/skae155.

Negussie, E., Koivula, M., & Mäntysaari, E. A. (2006). Genetic parameters and single versus multi-trait evaluation of udder health traits. *Acta Agriculturae Scand Section A*, *56*(2), 73-82. doi: 10.1080/09064700600979693

Nei, M., & Chakravarti, A. (1977). Drift variances of FST and GST statistics obtained from a finite number of isolated populations. *Theoretical population biology*, *11*(3), 307-325. doi:10.1016/0040-5809(77)90014-4

Norman, H. D., Wright, J. R., Hubbard, S. M., Miller, R. H., & Hutchison, J. L. (2009). Reproductive status of Holstein and Jersey cows in the United States. Journal of dairy science, 92(7), 3517-3528. doi: 10.3168/jds.2008-1768

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

Palucci, V., & Dürr, J. (2014). Trend validation procedures applied by Interbull–historical overview. *population*, *1*(2), 5.

Palucci, V., Haugaard, K., & Macedo, F. (2023). Interbull new services: Current and Future. *Interbull bulletin*, (59), 158-163.

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

Peter, B. M. (2016). Admixture, population structure, and F-statistics. *Genetics*, *202*(4), 1485-1501. doi: 10.1534/genetics.115.183913

Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools, *Bioinformatics*, Volume 27, Issue 15, August 2011, Pages 2156–2158. doi:10.1093/bioinformatics/btr330

Pintus E, Sorbolini S, Albera A, Gaspa G, Dimauro C, Steri R, Marras G, Macciotta NPP.,. 2014. Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piemontese and Italian Brown cattle breeds. Anim Genet. 45(1):1–11. doi: 10.1111/age.12076.

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559–575. doi: 10.1086/519795

Purfield D. C., Berry D. P., McParland S., and Bradley D. G. 2012. Runs of homozygosity and population history in cattle. BMC Genetics 13:1–11. doi:10.1186/1471-2156-13-70

Rao, C. R. (1992). RA Fisher: The founder of modern statistics. *Statistical science*, 34-48.

Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. 2020. Selection signatures in livestock genome: a review of concepts, approaches and applications. Livestock Sci. 241:104257. doi: 10.1016/j.livsci.2020.104257

Schaeffer, L. R. (1984). Sire and cow evaluation under multiple trait models. *Journal of Dairy Science*, *67*(7), 1567-1580. doi: 10.3168/jds.S0022-0302(84)81479-4

Shuichi Kitada, Reiichiro Nakamichi, Hirohisa Kishino, Understanding population structure in an evolutionary context: population-specific F_{ST} and pairwise F_{ST}, *G3 Genes|Genomes|Genetics*, Volume 11, Issue 11, November 2021. doi:10.1093/g3journal/jkab316

Sorbolini, S., Gaspa, G., Steri, R., Dimauro, C., Cellesi, M., Stella, A., ... & Macciotta, N. P. P. (2016). Use of canonical discriminant analysis to study signatures of selection in cattle. Genetics Selection Evolution, 48(1), 58. doi: 10.1186/s12711-016-0236-7

Sorbolini, S., Marras, G., Gaspa, G., Dimauro, C., Cellesi, M., Valentini, A., & Macciotta, N. P. (2015). Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories. Genetics Selection Evolution, 47(1), 52. doi: 10.1186/s12711-015-0128-2

Sorensen, D. A., & Johansson, K. (1992). Estimation of direct and correlated responses to selection using univariate animal models. *Journal of animal science*, *70*(7), 2038-2044. doi: 10.2527/1992.7072038x

Srivastava, S., Lopez, B. I., Heras-Saldana, S. D. L., Park, J. E., Shin, D. H., Chai, H. H., ... & Lim, D. (2019). Estimation of genetic parameters by single-trait and multi-trait models for carcass traits in Hanwoo cattle. *Animals*, *9*(12), 1061. doi: 10.3390/ani9121061

Taberlet P., Coissac E., Pansu J., and Pompanon F. 2011. Conservation genetics of cattle, sheep, and goats. Comptes rendus biologies 334(3):247–254. doi: 10.1016/j.crvi.2010.12.007

Takahata, N., & Nei, M. (1984). FST and GST statistics in the finite island model. Genetics, 107(3), 501. doi: 10.1093/genetics/107.3.501

Thompson, R., & Meyer, K. (1986). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science*, *15*(4), 299-313. doi: 10.1016/0301-6226(86)90071-0

Vahedi, S.M., Salek Ardestani, S. FSTest: an efficient tool for cross-population fixation index estimation on variant call format files. *J Genet* 103, 4 (2024). doi:10.1007/s12041-023-01459-1

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414-4423. doi:10.3168/jds.2007-0980

VanRaden, P. M., Tooker, M. E., Wright, J. R., Sun, C., & Hutchison, J. L. (2014). Comparison of single trait to multi-trait national evaluations for yield, health, and fertility. *Journal of dairy science*, *97*(12), 7952-7962. doi: 10.3168/jds.2014-8489

Warburton, C. L., & Hayes, B. J. (2026). Ascertainment Bias in Cattle SNP Arrays and Implications for Multibreed Genomic Predictions. *Animal Genetics*, *57*(2), e70081. doi: 10.1002/age.70081

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370. doi:10.2307/2408641

CHAPTER 2

Evaluation of genetic differences between two autochthonous Spanish cattle breeds

This is the published version in **Italian Journal of Animal Science** of:

Milia, PE., Cesarani, A., Diaz, C., Meneses, C., Parejo, J. C., Falchi, L., Macciotta, NPP., Dimauro, C. (2025). Evaluation of genetic differences between two autochthonous Spanish cattle breeds. *Italian Journal of Animal Science*,*24*(1), 361–375. https://doi.org/10.1080/1828051X.2024.2449454

This version is used by permission of the **Italian Journal of Animal Science**.

Abstract

Genetic diversity was assessed in five subpopulations (Negra-MORB and Cárdena-MORC for Morucha breed; Jarda-JAR, Bociblanca-BOC, and Avileña-Negra Ibérica breed-ANI for Avileña Negra Ibérica main breed) belonging to two local Spanish cattle breeds (Avileña Negra Ibérica - AVI and Morucha-MOR) by using univariate and multivariate statistical approaches. A total of 4,137 animals (2,785 ANI, 333 BOC, 127 JAR, 263 MORB and 629 MORC) were genotyped with the Affymetrix Beadchip Microarray HD. After quality control, 30,391 common SNPs were retained. In the univariate approach, the Wright fixation index (FST) was computed between and within breeds. In the multivariate approach, the stepwise discriminant analysis and the canonical discriminant analysis (CDA) were applied. The highest weighted FST value (0.05) was computed between BOC and MORB. Instead, the lowest values were found between ANI and JAR and between ANI and BOC. In the multivariate approach, CDA significantly discriminated the five breeds by using 1,144 markers. Of the four Canonical functions (CAN1, CAN2, CAN3 and CAN4) extracted by CDA, CAN1 fully separated AVI and MOR whereas CAN2 disentangled the two sub- populations of MOR. Considering CAN3 and CAN4, the CDA separated AVI subpopulations. Among markers selected by both FST and CDA, the top 10 discriminant SNPs were mapped to genomic regions harboring genes involved in coat colour, behaviour, and morphological traits. This study confirmed that the two breeds, different at phenotypic level, are different also at genetic level. Moreover, the results showed that MOR subpopulations are more different from each other than the AVI ones.

Keywords

Genetic diversity; multivariate analysis; autochthonous cattle; population genetics

Highlights

- Univariate and multivariate techniques highlighted significant SNP that could potentially be used to certify and commercially enhance the products of local breeds.
- Spanish local breeds can be distinguished not only by their region of origin but also by their phenotypic and genetic characteristics.

Introduction

Modern livestock biodiversity exists due to evolutionary processes originating from natural selection and animal domestication (Hall 2004). Domestication started around 8000 years B.C. when human hunter- gatherer societies were replaced by pastoral and agricultural societies. After its domestication, cattle became one of the most important domestic animal species and among the earliest forms of capital, appreciated for providing milk, draught power, and meat (Rifkin 1992; Felius et al. 2011). Throughout the years, events such as genetic drift, migration, cross- breeding, and artificial selection were responsible for the differences between and within the current cattle populations (Cesarani et al. 2019; Frantz et al. 2020; Saravanan et al. 2020). Other factors, such as geo- graphical barriers, local management, and cultural practices, contributed to change the animals' genetic structure and hence the breeds characteristics (Cesarani et al. 2019; Saravanan et al. 2020).

Most cattle breeds farmed in Europe are native or local populations (Hiemstra et al. 2010). These breeds are valued as more frugal, healthy, and hardy than cosmopolitan and highly selected cattle breeds, due to their environmental adaptation and suitability for extensive management, natural grazing, and vegetation management (Felius et al. 2014). Besides representing cultural, historical, socio-economic, and environmental value, autochthonous local cattle populations carry a large amount of genetic variation (Hiemstra et al. 2010). Over the years, however, the preference for highly productive breeds has caused the abandonment of less productive local breeds (Felius et al. 2014). The decline of local breeds leads to an inevitable loss of genetic diversity, which is fundamental to sustain the wellbeing of millions of farm- ing and pastoral communities, understanding the process of evolution, and developing appropriate conservation strategies (Edea et al. 2013; Osada 2015). The loss of genetic diversity leads to negative effects within a population. According to several studies, populations with poor genetic variability show a reduced ability to adapt to environmental changes (Hoffmann and Parsons 1997), higher probabilities of extinction (Mills and Smouse 1994; Lacy 1997; Frankham et al. 2002), and a greater susceptibility to inbreeding depression (Osada 2015). For all these reasons, the interest in studying the genetic variability of cattle breeds has reached, in the last years, very high levels (Notter 1999). In Spain, autochthonous cattle breeds are considered an important reservoir of genetic variability and a cultural asset (González-Rodríguez et al. 2017). The differentiation between and within these taurine breeds started when groups of individuals were confined to specific environments and became reproductively isolated from each other (González-Rodríguez et al. 2017). Historically, Iberian cattle were raised for meat and milk production, and work. Nowadays, they are appreciated only for their beef production (Felius et al. 2014; González-Rodríguez et al. 2017).

a. Avileña Negra Ibérica main breed (ANI) (MAPA, 2023)

b. Bociblanca (BOC) (MAPA, 2023)

c. Jarda (JAR)

d. Morucha Negra (MORB) (MAPA, 2023)

e. Morucha Cárdena (MORC) (MAPA, 2023)

Figure 1. Pictures (retrieved from the Ministerio de Agricultura, Pesca y Alimentación, MAPA) of the five cattle populations analyzed in the study.

Avileña Negra-Ibérica and Morucha are two local Spanish breeds raised under an extensive production system closely linked to the Mediterranean forest area known as the 'dehesa'. During their development, and even today, both populations share nearby geographic areas in the Castilla-León region. Nowadays, Morucha (MOR) is mainly farmed in the province of Salamanca, where it is locally adapted to the pasture area and appreciated for its high-quality meat (IGP Carne de Morucha de Salamanca). Within MOR, two sub-populations can be identified according to the coat colour: the black (MORB) and the 'cárdeno' (MORC) (Figure 1). Historically, MORB was the most prevalent subpopulation, but in the 1980s, MORC became the predominant one. This probably happened due to the better adaptability of MORC to 'less harsh' environments where it was possible to raise larger herds over which breeders had more control. Over the years, the number of MORB individuals has considerably declined, to the point where they are now protected by a genetic conservation program (Fernández et al. 2019; MAPA 2023).

On the contrary, the Avileña Negra-Ibérica (AVI) is more numerous and geographically widespread, inhabiting more diverse areas ranging from mountainous regions to flat "dehesas". AVI is valued for its longevity and rusticity, as well as high fertility in difficult environments and high-quality of meat (IGP Carne de Ávila). Three sub-populations of this breed differ based on coat colour: the main breed Avileña Negra-Ibérica (ANI), characterised by a uniform black coat; the Bociblanca (BOC), which presents a black to dark brown coat with some discoloration of the dorsolumbar line and the groin area, and the Jarda (JAR), which differs from ANI and BOC in its coat colour, which is the so-called 'Spot Pinzaguer' phenotype, recently studied in the Pinzaguer (Küttel et al. 2019) and Gloucester (Artesi et al. 2020) cattle breeds (Figure 1). The aim of this study was to investigate the genetic diversity and the genetic architecture underlying phenotypic differences between subpopulations of AVI and MOR cattle breeds using genome-wide SNP data and various statistical approaches.

Materials and methods

Animals and genomic data

Genomic data from 4,137 animals belonging to AVI and MOR were used in this study (Table 1). Supplementary Table 1 summarizes the distinguishing features of the analyzed subpopulations. Animals were genotyped with the Affymetrix Beadchip Microarray MD 70K (Affymetrix Inc., Santa Clara CA). Quality control on genotypes was carried out by breed using PLINK 1.9 (Purcell et al. 2007) with the following criteria: minor allele frequency greater than 2% (plink –cow –maf 0.02), and animal and SNP call rate greater than 95% (plink –cow –mind 0.95 –geno 0.95). Moreover, SNPs not in Hardy-Weinberg Equilibrium ($P < 1e\text{-}6$) (plink –cow –hwe 1e−6), mapped on sexual chromosome or unmapped according to the ARSUCD1.2 release were also discarded. The datasets of the two breeds (3,245 for AVI and 892 for MOR) were then merged, by retaining the 30,391 SNPs in common.

Table 1. Number of animals by breed and subpopulation.

Breed	Subpopulation	Individuals
Avileña Negra Ibérica main breed (AVI)	Avileña Negra Ibérica (ANI)	2,785
	Bociblanca (BOC)	333
	Jarda (JAR)	127
Morucha (MOR)	Morucha Cárdena (MORC)	629
	Morucha Negra (MORB)	263
Total		4,137

Wright fixation index

Using PLINK 1.9 (Purcell et al. 2007), the pairwise Wright fixation index (F_{ST}) was computed between the two breeds and within the five subpopulations by using the equation proposed by Weir and Cockerham (1984). The raw estimated F_{ST} values were analyzed with LOWESS, a local weighted regression technique (Pintus et al. 2014) that fit a smooth curve through points. LOWESS is a reliable non-parametric technique that avoids making any significant assumptions about the distribution of the data (Jacoby 2000). The regression was fitted using a window of 20 SNPs and markers with smoothed F_{ST} value larger than 3 standard deviations from the chromosomal mean were selected and considered as significant. These parameters were adopted to have results more consistent with literature (Pintus et al. 2014, Cesarani et al. 2019) and that can be more easily discussed.

Canonical discriminant analysis

The multivariate approach proposed by Manca et al. (2020) was exploited to highlight genetic differences between the two populations, MOR and AVI, and the relative subpopulations. Genomic data was arranged in a multivariate manner: the rows were the involved animals, the columns were the SNPs plus one column indicating the population (AVI and MOR) and one column indicating the related subpopulation (ANI, BOC, JAR, MORB and MORC) to which the animals belonged. Two multivariate statistical techniques were applied to data: the canonical discriminant analysis (CDA) and the stepwise discriminant analysis (SDA). The CDA is a multivariate statistical technique aimed to a) highlight differences between groups of individuals and b) improve understanding of the relationships among the involved variables. If p indicates the number of groups, the CDA derives p−1 equations, named canonical functions (CAN), that are linear combinations of the original variables (Xn). The structure of a CAN is:

$$CAN = C_1 X_1 + C_2 X_2 + \ldots\ldots + C_2 X_2$$

where C_i are the canonical coefficients (CC) that provide information about the contribution of each variable Xi in composing the CAN. To better decipher the role of a X_i in the separation of groups, Rencher and Scott (1990) standardized the CC values to take into account the multiple correlations among the original variables (X_n). The absolute value of the standardized CC can be used to rank the corresponding variables thus highlighting their contribution to the CAN.

Differences between groups, in the space of the CANs, were tested by using the Hotelling's T-square test (De Maesschalck et al. 2000). CANs were then exploited to predict the corresponding group to each animal belongs. In practice, the CANs are applied to each individual thus producing a discriminant score. An animal is assigned to one specific group if its discriminant score is lower than the cut-off value obtained by calculating the weighted mean distance between the centroids of the two groups (Mardia et al. 2000). The CDA, however, can be developed only if, in the data matrix, the number of columns does not exceed the number of rows. In the present research, having only 4,137 animals (the rows) and 30,391 SNPs (the variables/columns), a reduction of the space of variables was needed. For this reason, the SDA, a statistical technique specifically conceived to select the subset of variables that better separate groups, was applied to the data. In particular, in the STEPDISC procedure of SAS software, release 9.4 (SAS Inst. Inc., Cary, NC, USA), the 'forward' method with the partial R2 indicator for adding variables was chosen. The algorithm used to discriminate groups can be summarized as follows. The SDA, with the 'forward' option for variable selection, was applied, separately, to the 29 autosomes by tuning the R2 threshold value until obtaining a total number of SNP near equal to the number of involved animals. After, the selected markers were merged in a single dataset, and a new run of the SDA was developed to obtain a number of most discriminant markers very similar to that of the markers selected in the univariate approach (F_{ST}). Finally, by using SNPs selected in the last run of the SDA, the CDA, using the CANDISC procedure of SAS, was developed to test if the considered populations (and subpopulations) were significantly separated, and animals correctly assigned to the groups. SNPs in common between CDA and F_{ST} were ranked according to their standardized absolute CC value, and, among them, the top ten most discriminating markers were submitted to genes mapping. The entire procedure was first used to determine whether CDA was able to discriminate the five subpopulations when they are studied together. Then, the procedure was applied to investigate the genetic differences between (i) AVI and MOR, (ii) MORB and MORC, (iii) ANI and BOC, (iv) ANI and JAR and (v) BOC and JAR. Finally, in each comparison, the minimum number of markers able to significantly separate the two groups was obtained. These SNPs were selected by ranking the absolute value of the standardized CCs and deleting SNPs with the lower CCs' value until the error in assigning animals to the true group was lower than 0.5%.

Gene discovery

Gene discovery was carried out only for genes close (250 kb downstream and upstream; Manca et al. 2020) to the ten most discriminant SNPs, selected both in the univariate (FST) and the multivariate (CDA) approaches, in the following comparisons: AVI versus MOR, ANI versus JAR, ANI versus BOC, BOC versus JAR and MORB versus MORC. Genes were identified using online databases (NCBI, genome browser), and the Bos Taurus genome assembly ARS-UCD1.2 was used as reference.

Results

Genetic differentiation between populations

F_{ST} value between AVI and MOR breeds was 0.026. The F_{ST} values for all the pairwise comparisons are reported in Table 2. The highest weighted F_{ST} value (0.051) was observed between BOC and MORB, subpopulations that belong to the two different investigated breeds. The lowest values were found between ANI vs JAR (0.010) and ANI vs BOC (0.028). Table 3 reports, for each comparison, the significant SNPs identified by FST analysis. The largest number of markers (562) was found between ANI and JAR, whereas the lowest (299) between AVI and MOR.

Table 2. Results of the weighted Wright Fixation index (F_{ST}) analysis between Avileña and Morucha subpopulations.

	ANI	BOC	JAR	MORB
BOC	0.028			
JAR	0.010	0.033		
MORB	0.039	0.051	0.045	
MORC	0.032	0.044	0.038	0.035

ANI = Avileña Negra Ibérica; BOC = Bociblanca; JAR = Jarda; MORB = Morucha Negra; MORC = Morucha Cárdena.

Table 3. Number of markers able to discriminate between the groups (number in parenthesis represent the minimum number of SNP) selected by the CDA and number of significant markers from the F_{ST} analysis.

Type of comparison	Comparison	CDA	FST	Common
Between breed	AVI vs MOR	306 (42)	299	19 (6)
Within AVI	ANI vs BOC	380 (93)	410	21 (7)
	ANI vs JAR	531 (33)	562	68 (12)
	JAR vs BOC	369 (19)	399	24 (0)
Within MOR	MORB vs MORC	324 (49)	354	26 (5)

ANI = Avileña Negra Ibérica; BOC = Bociblanca; JAR = Jarda; MORB = Morucha Negra; MORC = Morucha Cárdena.

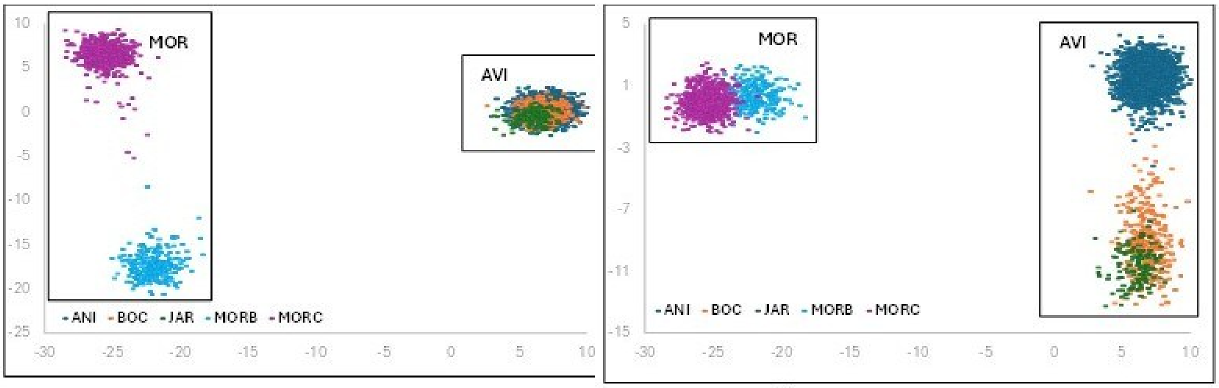
The multivariate approach

In the multivariate approach, the CDA significantly discriminated (Hotelling's T-test, P-value) the five breeds by using only 1,144 markers, previously selected by the SDA. Since five subpopulations were considered, CDA extracted four CANs (Table 4). CAN1 explained most of the total variation (around 80%) whereas CAN2 explained about 12%. The remaining variation was shared by CAN3 (6%) and CAN4 (3%). Distances among the five subpopulations in the space of the four CANs are listed in Table 4. The mean distances computed between subpopulations belonging to different breeds were larger than the distances computed between subpopulations within breed: for example, the average distance among the three subpopulations of AVI (i.e., ANI, BOC, and JAR) was of 255 markers, whereas the average distance between subpopulations of Avileña and Morucha was 1,221.

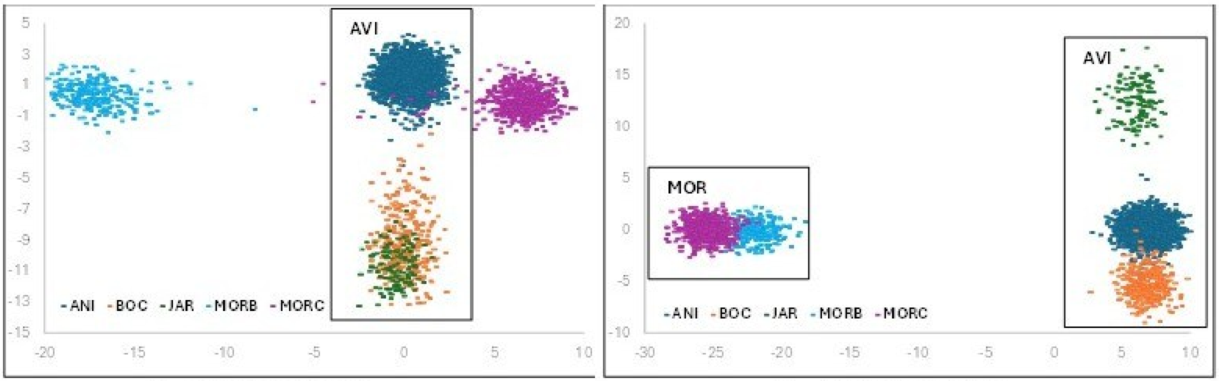
Table 4. Distances and explained variance among the five subpopulations in the space of the four CANs.

		Distances				Eigenvalue	Explained variance	
Breed	Subpopulation[1]	ANI	BOC	JAR	MORB		Proportion	Cumulative
AVI	BOC	145				λ_1	0.79	0.79
	JAR	303	318			λ_2	0.12	0.91
MOR	MORB	1,144	1,243	1,342		λ_3	0.06	0.97
	MORC	1,092	1,196	1,306	578	λ_4	0.03	1.00

ANI = Avileña Negra Ibérica; BOC = Bociblanca; JAR = Jarda; MORB = Morucha Negra; MORC = Morucha Cárdena.



a. CAN1 vs CAN2 **b. CAN1 vs CAN3**



c. CAN2 vs CAN3 **d. CAN1 vs CAN4**

Figure 2. Graph of the four canonical functions (CAN1, CAN2, CAN3 and CAN4) obtained in a genome-wide canonical discriminant analysis using restricted number (1,144) of most discriminant SNP.

Figure 2 shows the five subpopulations in the space of CANs. In Figure 2a, the two populations (AVI and MOR), are perfectly separated by CAN1, whereas CAN2 is able to disentangle the two subpopulations of MOR (MORB with negative values and MORC with positive values), but not the three subpopulations of AVI. CAN3 was able to discriminate the three subpopulations of AVI: ANI with positive values, BOC and JAR with negative values (Figure 2b–c). Finally, Figure 2d shows how the CAN4 separated JAR (positive values) from ANI and BOC (negative values).

The SDA and the CDA were applied to highlight pairwise differences both between the two main populations and among their subpopulations (Table 3). AVI and MOR were perfectly separated by using 306 SNPs with 19 of them in common with the markers selected in the FST approach. The minimum number of markers able to separate the two breeds was 42 with 6 of them in common with FST. Similar results were obtained when the various sub-populations were compared with each other (Table 3). In particular, regarding the AVI sub-populations, the most distant groups were JAR and BOC (Table 4) and only 19 markers where enough to separate them (Table 3). On the other hand, the nearest breeds were ANI and BOC: in this case the distance was 145 and the minimum number of SNPs to separate them was 93. Finally, ANI and JAR had an intermediate distance and CDA used 33 markers to separate them. The two subpopulations of MOR (i.e. MORB and MORC), were well separated using 49 markers.

Gene discovery

All genes mapped close to the 10 most discriminant SNPs (selected by both CDA and F_{ST}) for each pairwise comparison are listed in Table 5. For AVI and MOR, 34 genes, located on seven different chromosomes (i.e., BTA 1, 3, 6, 10, 11, 12, and 27) were found. The highest number of genes (9) was found on BTA12, whereas the lowest was found on BTA6 where only one gene, GABRG1, was detected. Thirty-three genes were found in the comparison between the two MOR subpopulations (MORB and MORC). These genes were located on BTA3, BTA6, BTA10, BTA14, BTA18, BTA22, BTA25, BTA28, and BTA29. On BTA22, only one gene, the RYBP, was found. Moreover, two genomic regions with six genes in total located on BTA29 were found. In ANI vs BOC, 37 genes were identified on chromosomes 1, 3, 4, 6, 12, 13, 15, 23, and 28. Nine genes were found on BTA28, whereas only one gene, OR5M10, was found on BTA23. Twenty-seven genes, located on BTA2, BTA3, BTA5, BTA6, and BTA17 were found in the comparison involving ANI and JAR subpopulations. The highest number of genes was found on BTA17, with 13 genes in total. Only one gene, GALNT13, was found in BTA2. In the last comparison, BOC versus JAR, 29 genes in total were found. The highest number of genes (7) was found on BTA18 whereas the lowest on BTA6 and BTA19 where genes ADGRL3 and KIF2B were found, respectively. Other 6 genes were found on BTA19, and the remaining genes were located on BTA4, BTA8, BTA9, BTA14, and BTA15.

Table 5. Genomic location and potentially associated genes found in common between FST and CDA approaches.

Comparison	BTA	SNP	Region (Mb)	Gene acronym	Gene name
AVI vs MOR	1	Affx-113745935	82.1-82.6	VPS8	VPS8 Subunit Of CORVET Complex
				MAGEF1	MAGE Family Member F1
	3	Affx-113731918	116.5-117.0	COPS8	COP9 Signalosome Subunit 8
				COL6A3	Collagen Type VI Alpha 3 Chain
				MLPH	Melanophilin
				RAB17	RAB17, Member RAS Oncogene Family
				PRLH	Prolactin Releasing Hormone
	6	Affx-106529231	64.08-64.5	GABRG1	Gamma-Aminobutyric Acid Type A Receptor Subunit Gamma1
	10	Affx-113724979	2.44-2.94	YTHDC2	YTH N6-Methyladenosine RNA Binding Protein C2
				KCNN2	Potassium Calcium-Activated Channel Subfamily N Member 2
	11	Affx-106511665	59.28-59.78	LRRTM4	Leucine Rich Repeat Transmembrane Neuronal 4
		Affx-113715544	21.11-21.6	DHX57	DExH-Box Helicase 57
				MORN2	MORN Repeat Containing 2

			ARHGEF33	Rho Guanine Nucleotide Exchange Factor 33
			SOS1	SOS Ras/Rac Guanine Nucleotide Exchange Factor 1
			CDKL4	Cyclin Dependent Kinase Like 4
			MAP4K3	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 3
12	Affx-106516149	86.72-87.22	TFDP1	Transcription Factor Dp-1
			ATP4B	ATPase H+/K+ Transporting Subunit Beta
			GRK1	G Protein-Coupled Receptor Kinase 1
			TMEM255B	Transmembrane Protein 255B
			GAS6	Growth Arrest Specific 6
			RASA3	RAS P21 Protein Activator 3
			CDC16	Cell Division Cycle 16
			UPF3A	UPF3A Regulator of Nonsense Mediated MRNA Decay
			CHAMP1	Chromosome Alignment Maintaining Phosphoprotein 1
14	Affx-113745752	61.26-61.76	BAALC	BAALC Binder Of MAP3K1 And KLF4
			ATP6V1C1	ATPase H+ Transporting V1 Subunit C1
			AZIN1	Antizyme Inhibitor 1

	14	Affx-113742795	53.68-53.93	-	-
	27	Affx-113723433	14.66-15.16	ENPP6	Ectonucleotide Pyrophosphatase/Phosphodiesterase 6
				IRF2	Interferon Regulatory Factor 2
				CASP3	Caspase 3
				PRIMPOL	Primase And DNA Directed Polymerase
				CENPU	Centromere Protein U
MORB vs MORC	3	Affx-106508034	70.17-70.67	ERICH3	Glutamate Rich 3
				TNNI3K	TNNI3 Interacting Kinase
				FPGT	Fucose-1-Phosphate Guanylyltransferase
				LRRIQ3	Leucine Rich Repeats And IQ Motif Containing 3
	6	Affx-46217064	6.47-6.97	SEC24D	SEC24 Homolog D, COPII Coat Complex Component
				METTL14	Methyltransferase 14, N6-Adenosine-Methyltransferase Subunit
				PRSS12	Serine Protease 12
				NDST3	N-Deacetylase And N-Sulfotransferase 3
	10	Affx-106515394	12.83-13.33	MEGF11	Multiple EGF Like Domains 11

			DIS3L	DIS3 Like Exosome 3'-5' Exoribonuclease
			TIPIN	TIMELESS Interacting Protein
			MAP2K1	Mitogen-Activated Protein Kinase Kinase 1
			SNAPC5	Small Nuclear RNA Activating Complex Polypeptide 5
14	Affx-106499747	60.07-60.57	LRP12	LDL Receptor Related Protein 12
			DPYS	Dihydropyrimidinase
			DCSTAMP	Dendrocyte Expressed Seven Transmembrane Protein
			RIMS2	Regulating Synaptic Membrane Exocytosis 2
18	Affx-257167599	44.80-45.30	GPI	Glucose-6-Phosphate Isomerase
			PDCD2L	Programmed Cell Death 2 Like
			UBA2	Ubiquitin Like Modifier Activating Enzyme 2
			WTIP	WT1 Interacting Protein
22	Affx-106522427	29.23-29.73	RYBP	RING1 And YY1 Binding Protein
25	Affx-106528006	28.98-28.48	CALN1	Calneuron 1
			GALNT17	Polypeptide N-Acetylgalactosaminyltransferase 17
28	Affx-113736666	8.30 – 8.80	GNG4	G Protein Subunit Gamma 4

				LYST	Lysosomal Trafficking Regulator
				NID1	Nidogen 1
	29	Affx-257087119	7.12-7.62	GRM5	Glutamate Metabotropic Receptor 5
				CTSC	Cathepsin C
				RAB38	RAB38, Member RAS Oncogene Family
	29	Affx-113731799	21.68-22.18	CCDC179	Coiled-Coil Domain Containing 179
				SVIP	Small VCP Interacting Protein
				GAS2	Growth Arrest Specific 2
ANI vs BOC	1	Affx-113744755	52.57-53.07	CD47	CD47 Molecule
				IFT57	Intraflagellar Transport 57
				HHLA2	HHLA2 Member Of B7 Family
	2	Affx-41965081	7.25-7.75	COL5A2	Collagen Type V Alpha 2 Chain
				COL3A1	Collagen Type III Alpha 1 Chain
				DIRC1	Disrupted In Renal Carcinoma 1
	3	Affx-113714896	95.91-96.41	ELAVL4	ELAV Like RNA Binding Protein 4
				AGBL4	AGBL Carboxypeptidase 4

4	Affx-113719554	114.5-115.01	GALNTL5	Polypeptide N-Acetylgalactosaminyltransferase Like 5
			GALNT11	Polypeptide N-Acetylgalactosaminyltransferase 11
			KMT2C	Lysine Methyltransferase 2C
			CCT8L2	Chaperonin Containing TCP1 Subunit 8 Like 2
6	Affx-106529527	58.53-59.03	KLB	Klotho Beta
			RPL9	Ribosomal Protein L9
			LIAS	Lipoic Acid Synthetase
			UGDH	UDP-Glucose 6-Dehydrogenase
			SMIM14	Small Integral Membrane Protein 14
			UBE2K	Ubiquitin Conjugating Enzyme E2 K
			PDS5A	PDS5 Cohesin Associated Factor A
12	Affx-39474242	5.28-5.78	-	-
13	Affx-113739114	21.08-21.58	MALRD1	MAM And LDL Receptor Class A Domain Containing 1
			PLXDC2	Plexin Domain Containing 2
15	Affx-106518852	64.68-65.18	CAPRIN1	Cell Cycle Associated Protein 1
			NAT10	N-Acetyltransferase 10

				ABTB2	Ankyrin Repeat And BTB Domain Containing 2
				CAT	Catalase
				ELF5	E74 Like ETS Transcription Factor 5
				EHF	ETS Homologous Factor
	23	Affx-106524428	21.64-22.14	OR5M10	Olfactory Receptor Family 5 Subfamily M Member 10
	28	Affx-44470577	24.7-25.20	PBLD	Phenazine Biosynthesis Like Protein Domain Containing
				HNRNPH3	Heterogeneous Nuclear Ribonucleoprotein H3
				RUFY2	RUN And FYVE Domain Containing 2
				DNA2	DNA Replication Helicase/Nuclease 2
				SLC25A16	Solute Carrier Family 25 Member 16
				TET1	Tet Methylcytosine Dioxygenase 1
				CCAR1	Cell Division Cycle And Apoptosis Regulator 1
				STOX1	Storkhead Box 1
				DDX50	DExD-Box Helicase 50
ANI vs JAR	2	Affx-41965081	7.25-7.75	COL5A2	Collagen Type V Alpha 2 Chain
				COL3A1	Collagen Type III Alpha 1 Chain

			DIRC1	Disrupted In Renal Carcinoma 1
2	Affx-106525249	41.57-42.07	GALNT13	Polypeptide N-Acetylgalactosaminyltransferase 13
3	Affx-113714896	95.91-96.41	ELAVL4	ELAV Like RNA Binding Protein 4
			AGBL4	AGBL Carboxypeptidase 4
3	Affx-106514806	0.36-0.86	TIPRL	TOR Signaling Pathway Regulator
			GPR161	G Protein-Coupled Receptor 161
			DCAF6	DDB1 And CUL4 Associated Factor 6
			MPC2	Mitochondrial Pyruvate Carrier 2
			ADCY10	Adenylate Cyclase 10
5	Affx-106555873	89.41-89.91	PDE3A	Phosphodiesterase 3A
	Affx-115859690			
6	Affx-113718412	70.18-70.72	KIT	KIT Proto-Oncogene, Receptor Tyrosine Kinase
			KDR	Kinase Insert Domain Receptor
	Affx-113717031			
11	Affx-113745563	56.90-57.40	-	-
17	Affx-113748841	69.0-69.50	MTMR3	Myotubularin Related Protein 3
			HORMAD2	HORMA Domain Containing 2

				LIF	LIF Interleukin 6 Family Cytokine
				OSM	Oncostatin M
				CASTOR1	Cytosolic Arginine Sensor For MTORC1 Subunit 1
				TBC1D10A	TBC1 Domain Family Member 10A
				SF3A1	Splicing Factor 3a Subunit 1
				CCDC157	Coiled-Coil Domain Containing 157
				RNF215	Ring Finger Protein 215
				SEC14L2	SEC14 Like Lipid Binding 2
				MTFP1	Mitochondrial Fission Process 1
				SEC14L3	SEC14 Like Lipid Binding 3
				SEC14L4	SEC14 Like Lipid Binding 4
BOC vs JAR	4	Affx-113737260	67.75-68.25	CREB5	CAMP Responsive Element Binding Protein 5
				JAZF1	JAZF Zinc Finger 1
	6	Affx-113726065	76.38-76.88	ADGRL3	Adhesion G Protein-Coupled Receptor L3
	8	Affx-47478317	102.85-103.33	ZNF618	Zinc Finger Protein 618

			AMBP	Alpha-1-Microglobulin/Bikunin Precursor
			KIF12	Kinesin Family Member 12
9	Affx-113719507	15.37-15.87	MYO6	Myosin VI
			IMPG1	Interphotoreceptor Matrix Proteoglycan 1
	Affx-40413795			
14	Affx-113723997	66.36-66.91	LAPTM4B	Lysosomal Protein Transmembrane 4 Beta
			MTDH	Metadherin
			TSPYL5	TSPY Like 5
15	Affx-106541489	64.03-64.53	KIAA1549L	KIAA1549 Like
			CD59	CD59 Molecule (CD59 Blood Group)
			FBXO3	F-Box Protein 3
			LMO2	LIM Domain Only 2
18	Affx-106534272	14.73-15.23	DEF8	Differentially Expressed In FDCP 8 Homolog
			DBNDD1	Dysbindin Domain Containing 1
			GAS8	Growth Arrest Specific 8
			SHCBP1	SHC Binding And Spindle Associated 1
			VPS35	VPS35 Retromer Complex Component

			ORC6	Origin Recognition Complex Subunit 6
			MYLK3	Myosin Light Chain Kinase 3
19	Affx-257168465	3.52-4.02	KIF2B	Kinesin Family Member 2B
19	Affx-113715201	8.69-9.19	CUEDC1	CUE Domain Containing 1
			VEZF1	Vascular Endothelial Zinc Finger 1
			SRSF1	Serine And Arginine Rich Splicing Factor 1
			DYNLL2	Dynein Light Chain LC8-Type 2
			OR4D2	Olfactory Receptor Family 4 Subfamily D Member 2
			MKS1	MKS Transition Zone Complex Subunit 1

Discussion

Genotyped animals of two Spanish cattle breeds, AVI and MOR, and their subpopulations, ANI, BOC, JAR for AVI, and MORB and MORC for MOR, were studied to highlight genetic differences among them. According to previous studies, Iberian cattle breeds can be mainly distinguished by their geographical area rather than their morphotypes (Mateus et al. 2004; Ginja et al. 2010; Martín-Burriel et al. 2011). Due to the close geographical origin and production system, AVI and MOR have always been grouped together (Cañon et al. 2011; Cañas-Álvarez et al. 2016; González-Rodríguez et al. 2016; 2017). Cañas-Álvarez et al. (2015) hypothesized that AVI fairly contributed to MOR because of the transhumance adopted by its breeders to take advantage of the seasonal lag between regions. Consequently, this phenomenon, combined with geographic proximity, led to admixture between these two breeds and among their subpopulations. Results of the present research fully confirm that AVI and MOR can be considered as two different breeds, and the respective subpopulations are also genetically different. As autochthonous populations, AVI and MOR were recently involved in a breeding program by the National Breeders Associations. This program is intended to improve both amount and quality of the meat and enhance the economic performance, maintaining the hardiness of these breeds (MAPA 2023).

Wright fixation index is widely used in population and evolutionary genetics studies due to its important indication on evolutionary process that influenced genetic variation within and among populations. Moreover, F_{ST} is directly related to the variance in allele frequencies among populations individuals giving an important information about the degree of resemblance among them (Holsinger and Weir 2009). Thanks to F_{ST} information, it was possible to better understand the genetic relationships among the cattle populations analyzed in this study. F_{ST} value estimated between AVI and MOR (0.026) was slightly lower compared to the ones obtained in other studies conducted on these breeds. Cañon et al. (2001) reported a F_{ST} value of 0.045, whereas Cañas-Álvarez et al. (2015) computed an average F_{ST} value of 0.029. F_{ST} values among other comparisons highlighted the genetic proximity among AVI subpopulations, whereas the MOR ones appeared to be more distant. In general, the F_{ST} values obtained in this study are quite low. Originally, F_{ST} scores ranging from 0.05 to 0.15 were considered to show moderate differentiation among populations (Wright 1965). However, Hall (2022) suggested that the value of 0.10 for between-breed F_{ST} cannot be considered always as significant, because of its significant variability. Considering this, the results obtained in literature by Cañas-Álvarez et al. (2015) and the result obtained in the present study may suggest a real genetic distinctiveness between the investigated cattle populations. The significant SNPs highlighted through the LOWESS procedure are reported in Table 3. The highest number of significant SNPs was found among ANI and JAR subpopulations, whereas the lowest was found between the two main breeds, AVI and MOR. Results obtained in the univariate approach were confirmed by the multivariate one. The CDA developed involving the five subpopulations is synthesized in Figure 2. In particular, in Figure 2a, the AVI and MOR populations are perfectly separated by the CAN1, which accounts for the 79% of the total variation. However, the CAN2, which explains 12% of the variability, clearly disentangles the two subpopulations of MOR (MORB and MORC), but the three subpopulations of AVI appear completely overlapping. This result confirms what was obtained in the univariate approach. Only considering the CAN3 and CAN4 (which, together, add up to only 9%) CDA was able to separate the three subpopulations.

This result indicates that, from the genomic point of view, MOR subpopulations are more different than the AVI subpopulations. As mentioned above, the investigated subpopulations have been traditionally differentiated according to their coat colour. MORB and MORC have a black and a 'cárdeno' coat, respectively, whereas ANI has a black coat, BOC has a black/ brown coat and JAR has a 'spot Pinzaguer' coat (Supplementary Table 1 and Figure 1). Table 3 lists the number of significant markers selected by FST for any comparison. In the multivariate approach, we voluntarily selected an approximately equal number of markers for any comparison. The two main breeds, AVI and MOR, were significantly separated with 306 markers selected by the SDA, but only 19 SNPs were found in common between the two approaches. The 19 common SNPs were considered highly discriminant and the top 10 were submitted to gene discovery. For CDA, 42 markers were enough to significantly separate the two populations. In this case, only 6 SNPs were in common between FST and CDA. These differences between markers selected by the two statistical techniques can be explained considering that CDA is a multivariate technique and, thus, consider the multiple correlations among markers. As a consequence, some markers that in the univariate approach are not significant can be significant in the multivariate one. However, from a mathematical point of view, the fewer markers are needed to separate two groups, the greater is the distance between them. Therefore, since 42 is a truly small number of markers, the two main populations appear quite different, thus confirming the result obtained in the univariate approach. Inside the single main populations, MORB and MORC were significantly separated by 49 SNPs with 5 of them in common to the two approaches (Table 3). For AVI, in the space of CANs, the distance of ANI from BOC was 145, and the minimum number of markers required to significantly separate the two subpopulations was 93. On the other hand, the distance of JAR from ANI and BOC was 303 and 318, respectively. JAR was significantly discriminated from ANI and BOC by using only 33 and 19 markers, respectively. In consequence, ANI and BOC are more similar than JAR and ANI and JAR and BOC. These results contrast with those obtained based on the FST index, where the nearest subpopulations were ANI and JAR.

Gene discovery was carried out analyzing genome regions in the surroundings of the top 10 most discriminant markers found in common between F_{ST} and CDA approaches when the two cattle breeds and the related subpopulations were compared. It is known that main differences among the analyzed subpopulations can be found in morphological traits (Fernández et al. 2009). Several genes related to these traits and, also, to meat quality, were found in this study. Starting with the first comparison involving the two main breeds, AVI and MOR, two genes related with coat colour, MLPH and RAB17, were found on BTA3. Mutations within the MLPH gene are known to cause colour dilution due to a disruption of peripheral transport of melanosomes in numerous species (Philipp et al. 2005; Cieslak et al. 2011). This was confirmed by a study conducted on Belgian Blue cattle where this gene was found to be related with the 'cool grey' coat colour (Li et al. 2016). It is possible to speculate that mutations of this gene are responsible for the typical "cárdeno" coat of MORC. Moreover, Nicoloso et al. (2009) found this gene to be among the ones with the highest discriminant power among cattle characterized by different coat colour. The MLPH gene was also reported to be related with coat colour in Korean brindle cattle (Lee et al. 2015; Taye et al. 2017). The RAB17 gene was found to be related with hair coat type and colour properties in African and Chinese indigenous cattle (Lee et al. 2015; Li et al. 2020). Three other genes mapped in this same genomic region: two genes, COPS8 and COL6A3, related to meat quality (Ariwaka et al. 2024; Dinh et al. 2021), and one gene, PRLH, related to heat tolerance (Taye et al. 2017; Zeng et al. 2018). In BTA11, the LRRTM4 gene was reported to be associated with behaviour together with other traits of interest such as milk temperament and production, fertility, and rusticity (Chen et al. 2020; Verardo et al. 2021). In BTA11, we found several genes involved in the body weight determinism, such as DHX57, MORN2, ARHGEF33, and MAP4K3 (Edea et al. 2020; Zepeda-Batista et al. 2021), and one gene, the SOS1, associated to residual feed intake (Lam et al. 2021).

Also in the comparison between MORB and MORC subpopulations, one gene related to coat colour, MAP2K1, was found (Gutiérrez-Gil et al. 2015; Mei et al. 2018). The presence of significant SNP close to a gene involved in the coat colour determinism was expected since these two subpopulations have different colors. Moreover, we found two genes related with grazing behaviour (i.e. ERICH3 and GRM5) located on BTA3 and BTA29. In particular, the ERICH3 gene is involved in the regulation of the serotonin metabolic pathway in humans (Gupta et al. 2016), and it has been recently associated with walking minutes during grazing in cattle (Stegemiller et al. 2021). The GRM5 gene has been studied in recent studies on free range management beef cattle by Moreno García et al. (2022, 2024); this gene was found to be associated with grazing personality behaviours, such as the use of steep and rugged rangelands. Since these two subpopulations are mainly raised at pasture in marginal areas, we can assume that they could have different behaviors while grazing. Several other genes associated with body weight, feed efficiency, and meat quality traits were found in the comparison between MORB and MORC (Supplementary Table 2).

In ANI and BOC comparison, several genes related to body conformation were highlighted. The IFT57 gene, located on BTA1, was found to be related with feet and leg conformation in Nellore cattle (Vargas et al. 2018). One gene related to muscularity, COL5A2, mapped on BTA2 (Doyle et al. 2020; 2021). In this genomic region, one meat quality related gene, COL3A1 (Jeong et al. 2017), was also found. Other meat quality related genes, located on BTA6, were LIAS, SMIM14, UBE2K, and PDS5A (Ben-Jemaa et al. 2021; Muniz et al. 2022; Sood et al. 2023). Three genes located on BTA4, GALNTL5, KMT2C, and CCT8L2, were found to be related with chest width conformation in cattle (Lee et al. 2024). One gene located on BTA15, CAPRIN1, was associated to important biological processes and metabolic pathways related to behavioral traits, social interactions, and aggressiveness in Angus cattle (Araujo et al. 2021). This gene was also found to be related with meat quality, together with the gene NAT10 (Sood et al. 2023).

In ANI and JAR comparison, two genes related to coat colour, KIT and KDR, were found on BTA6. The KIT gene was found to be responsible for the Pinzgauer cattle spotting pattern, which is indeed the typical colour pattern of the JAR subpopulation (Küttel et al. 2019). This coat colour arises due to the presence of structural variants in the nearby of coat colour related genes (such as KIT, in this case) that result in modifications in gene functionality (Baxter et al. 2004), such as the absence of the main coat colour in some regions of the body. The KDR gene was found to be related with coat colour in two studies investigating the domestication-related phenotypes in beef cattle (Qanbari et al. 2014; Moravčíková et al. 2019). We found some genes associated with meat production traits, such as feed efficiency, muscle development, growth, and also for this comparison (Supplementary Table 2).

In the last comparison, BOC versus JAR, only one coat colour potentially related gene was found. The DBNDD1 gene has not been directly associated with coat colour in literature. However, Bertolini et al. (2022) found this gene while studying selection signatures in close autochthonous cattle breeds mainly distinguished by their coat colors. Furthermore, Kim et al. (2018) found this gene in a similar study conducted on three Korean cattle breeds. It could be of interest to develop further analysis to verify whether this gene could be involved in coat colour determinism. As in previous subpopulation comparison, several genes related to beef production and carcass quality, residual feed intake, feed conversion ratio and conception rate were found (Supplementary Table 2). All the cattle analyzed in the present study are mainly raised for meat production (González-Rodríguez et al. 2017): the presence of several genes associated with this trait could be due to differences in beef performances among the various subpopulations.

The differences in morphological, grazing behaviour, and production traits reported among these populations could be due to recent events. In fact, despite the similar geographical origin, the genetic conservation programs (implemented for BOC and MORB) and the breeding programs (implemented for AVI and MOR) pursued in the last decade could have reduced the genetic flow and contributed to the isolation of the genetic heritage of these subpopulations, enhancing the differences among them (Fernández et al. 2009; MAPA 2023). Contextually, factors such as the diversity of grazing areas, herd management practices, and local environmental conditions may have strengthened the genetic differences among the subpopulations.

Overall, this study gives important insights on the understanding of the genetic dynamics and adaptive mechanisms that shaped the genome of these cattle populations. The identification of genetic traits associated with morphological and behavioral characteristics are essential for understanding the evolutionary mechanisms and adaptations that have allowed these breeds to thrive in local environments with particular ecological characteristics. The identification and valorization of these genetic traits highlight the importance of preserving these breeds, which not only represent a unique genetic heritage (that might mitigate the risks associated with the loss of genetic diversity in the context of global environmental changes) but also play a crucial role in sustainable and resilient agricultural systems.

Conclusion

The analysis of genetic differences between the AVI and MOR populations, developed by using both univariate and multivariate approaches, confirmed that, despite the common origin and use, they appear genetically different. This occurs also for the relative subpopulations. Several genomic regions were highlighted harboring genes involved in the determinism of morphological traits, such as body weight and coat colour, and behaviour traits. Some genes related to feed intake, milk and meat traits were also detected. These results are compatible with the hypothesis that these breeds, traditionally classified through the colour coat, can also be distinguished by a genomic point of view. The differences observed among these subpopulations could reflect the impact of genetic conservation and selection programs, as well as environmental and management factors, in shaping their genomic characteristics. This study provides key information on genetic traits related to adaptation and productivity, pointing out the importance of preserving these breeds to ensure biodiversity, agricultural sustainability and resilience to environmental change.

Acknowledgments

The authors would like to acknowledge the Real Asociación Española de Avileña Negra Ibérica (RAEANI) and the Asociación Nacional de Criadores de Ganado Vacuno de Raza Morucha for providing the genotypes.

References

Abo-Ismail MK, Lansink N, Akanno E, Karisa BK, Crowley JJ, Moore SS, Plastow GS. 2018. Development and validation of a small SNP panel for feed efficiency in beef cattle. Journal of Animal Science 96(2):375-397.

Akanno EC, Chen L, Abo-Ismail MK, Crowley JJ, Wang Z, Li C, Plastow GS. 2018. Genome-wide association scan for heterotic quantitative trait loci in multi-breed and crossbred beef cattle. Genetics Selection Evolution 50:1-13.

Araujo AC, Carneiro PL, Alvarenga AB, Oliveira HR, Miller SP, Retallick K, Brito LF. 2021. Haplotype-based single-step GWAS for yearling temperament in American Angus cattle. Genes 13(1):17.

Arikawa LM, Mota LFM, Schmidt PI, Frezarim GB, Fonseca LFS, Magalhães AFB, de Albuquerque LG. 2024. Genome-wide scans identify biological and metabolic pathways regulating carcass and meat quality traits in beef cattle. Scientific Reports 14:1 -14.

Barrett JC, Fry B, Maller JD, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263-265.

Baxter LL, Hou L, Loftus SK, Pavan WJ. 2004. Spotlight on spotted mice: a review of white spotting mouse mutants and associated human pigmentation disorders. Pigment Cell Research 17(3):215-224.

Ben-Jemaa S, Senczuk G, Ciani E, Ciampolini R, Catillo G, Boussaha M, Mastrangelo S. 2021. Genome-wide analysis reveals selection signatures involved in meat traits and local adaptation in semi-feral Maremmana cattle. Frontiers in Genetics 12:675569.

Bertolini F, Moscatelli G, Schiavo G, Bovo S, Ribani A, Ballan M, et al. 2022. Signatures of selection are present in the genome of two close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for their coat colours. *Journal of Animal Breeding and Genetics* 139(3):307-319.

Biscarini F, Cozzi P, Gaspa G, Marras G. 2019. detectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes. R package version 0.9.6, https://CRAN.R-project.org/package=detectRUNS.

Brunes LC, Baldi F, Lopes FB, Lôbo RB, Espigolan R, Costa MF, Magnabosco CU. 2021. Weighted single-step genome-wide association study and pathway analyses for feed efficiency traits in Nellore cattle. Journal of Animal Breeding and Genetics 138(1):23-44.

Bruscadin JJ, de Souza MM, de Oliveira KS, Rocha MIP, Afonso J, Cardoso TF, de Almeida Regitano LC. 2021. Muscle allele-specific expression QTLs may affect meat quality traits in Bos indicus. Scientific Reports 11(1):7321.

Cañas-Álvarez JJ, González-Rodríguez A, Munilla S, Varona L, Díaz C, Baro JA, Altarriba J, Molina A, Piedrafita J. 2015. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip. J Anim Sci. 93(11):5164–5174. doi: 10.2527/jas.2015-9271.

Cañas-Álvarez JJ, Mouresan EF, Varona L, Díaz C, Molina A, Baro JA, Altarriba J, Carabaño MJ, Casellas J, Piedrafita J. 2016. Linkage disequilibrium, persistence of phase, and effective population size in Spanish local beef cattle breeds assessed through a high-density single nucleotide polymorphism chip. J Anim Sci. 94(7):2779–2788. doi: 10. 2527/jas.2016-0425

Cañon J, Alexandrino P, Bessa I, Carleos C, Carretero Y, Dunner S, Ferran N, Garcia D, Jordana J, Laloë D, et al. 2001. Genetic diversity measures of local European beef cattle breeds for conservation purposes. Genet Sel Evol. 33(3):311–332. doi: 10.1186/1297-9686-33-3-311.

Cañon J, García D, Delgado JV, Dunner S, Telo da Gama L, Landi V, Martín-Burriel I, Martínez A, Penedo C, Rodellar C, et al. 2011. Relative breed contributions to neutral genetic diversity of a comprehensive representation of Iberian native cattle. Animal. 5(9):1323–1334. doi: 10.1017/ S1751731111000267.

Cesarani A, Sechi T, Gaspa G, Usai MG, Sorbolini S, Macciotta NPP, Carta A. 2019. Investigation of genetic diversity and selection signatures between Sarda and Sardinian Ancestral black, two related sheep breeds with evident morphological differences. Small Ruminant Res. 177:68– 75. doi: 10.1016/j.smallrumres.2019.06.014.

Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, Hu S, Huang S, Zhang H, Zheng Z, et al. 2018. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun. 9(1):2337. doi: 10.1038/s41467-018-04737-0.

Chen SY, Oliveira HR, Schenkel FS, Pedrosa VB, Melka MG, Brito LF. 2020. Using imputed whole-genome sequence variants to uncover candidate mutations and genes affecting milking speed and temperament in Holstein cattle. J Dairy Sci. 103(11):10383–10398. doi: 10.3168/jds.2020- 18897.

Cieslak M, Reissmann M, Hofreiter M, Ludwig A. 2011. Colours of domestication. Biol Rev Camb Philos Soc. 86(4): 885–899. doi: 10.1111/j.1469-185X.2011.00177.x.

De Maesschalck R, Jouan-Rimbaud D, Massart DL. 2000. The mahalanobis distance. Chemom Intell Lab Syst. 50(1):1–18. doi: 10.1016/S0169-7439(99)00047-7.

Dinh PTN, Chung Y, Lee DJ, Kang DH, Park BH, Chung KY, Lee SH. 2021. Association study of functional candidate genes for meat tenderness in Hanwoo cattle. J Anim Breed Genomics. 5(4):1–10.

Do DN, Bissonnette N, Lacasse P, Miglior F, Sargolzaei M, Zhao X, Ibeagha-Awemu EM. 2017. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. J Dairy Sci. 100(3): 1955–1970. doi: 10.3168/jds.2016-11910.

Doyle JL, Berry DP, Veerkamp RF, Carthy TR, Evans RD, Walsh SW, Purfield DC. 2020. Genomic regions associated with muscularity in beef cattle differ in five contrasting cattle breeds. Genet Sel Evol. 52(1):2. doi: 10.1186/s12711-020-0523-1.

Doyle JL, Purfield DC, Moore T, Carthy TR, Walsh SW, Veerkamp RF, Evans RD, Berry DP. 2021. Identification of genomic regions that exhibit sexual dimorphism for size and muscularity in cattle. J Anim Sci. 99(5):1–19. doi: 10. 1093/jas/skab070.

Edea Z, Dadi H, Kim S-W, Dessie T, Lee T, Kim H, Kim J-J, Kim K-S. 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. Front Genet. 4:35. doi: 10.3389/fgene.2013.00035.

Edea Z, Jung KS, Shin SS, Yoo SW, Choi JW, Kim KS. 2020. Signatures of positive selection underlying beef production traits in Korean cattle breeds. J Anim Sci Technol. 62(3):293–305. doi: 10.5187/jast.2020.62.3.293.

Felius M, Beerling ML, Buchanan DS, Theunissen B, Koolmees PA, Lenstra JA. 2014. On the history of cattle genetic resources. Diversity. 6(4):705–750. doi: 10.3390/d6040705.

Felius M, Koolmees PA, Theunissen B, Lenstra JA, European Cattle Genetic Diversity Consortium. 2011. On the breeds of cattle—historic and current classifications. Diversity 3(4):660–692. doi: 10.3390/d3040660.

Fernández M, Gómez M, Delgado JV, Adán S, Jiménez M. 2009. Guía de campo de las razas autoctónas españolas. Madrid: Ministerio de Medio Ambiente y Medio Rural y Marino. for Statistical Computing, Vienna, Austria. https:// www.R-project.org/.

Frankham R, Briscoe DA, Ballou JD. 2002. Introduction to conservation genetics. Cambridge (UK): Cambridge University Press. Frantz LA, Bradley DG, Larson G, Orlando L. 2020. Animal domestication in the era of ancient genomics. Nat Rev Genet. 21(8):449–460. doi: 10.1038/s41576-020-0225-0.

Ginja C, Telo Da Gama L, Penedo MCT. 2010. Analysis of STR markers reveals high genetic structure in Portuguese native cattle. J Hered. 101(2):201–210. doi: 10.1093/jhered/ esp104.

González-Rodríguez A, Munilla S, Mouresan EF, Cañas-Álvarez JJ, Díaz C, Piedrafita J, Altarriba J, Baro JÁ, Molina A, Varona L. 2016. On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. Genet Sel Evol. 48(1):1–12. doi: 10.1186/s12711-016-0258-1.

González-Rodríguez A, Munilla S, Mouresan EF, Cañas-Álvarez JJ, Baro JA, Molina A, Díaz C, Altarriba J, Piedrafita J, Varona L. 2017. Genomic differentiation between Asturiana de los Valles, Avileña-Negra Ibérica, Bruna dels Pirineus, Morucha, Pirenaica, Retinta and Rubia Gallega cattle breeds. Animal. 11(10):1667–1679. doi: 10.1017/ S1751731117000398 .

Gupta M, Neavin D, Liu D, Biernacka J, Hall-Flavin D, Bobo WV, Frye MA, Skime M, Jenkins GD, Batzler A, et al. 2016. TSPAN5, ERICH3 and selective serotonin reuptake inhibitors in major depressive disorder: pharmacometabolomics-informed pharmacogenomics. Mol Psychiatry. 21(12): 1717–1725. doi: 10.1038/mp.2016.6.

Gurgul A, Jasielczuk I, Szmatoła T, Sosin-Bzducha E, Majewska A, Litwińczuk Z. 2020. Divergent selection signatures of phenotypic and production traits among conserved and commercial cattle breeds. Livestock Science. 239:104174. doi: 10.1016/j.livsci.2020.104174.

Gutiérrez-Gil B, Arranz JJ, Wiener P. 2015. An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. Front Genet. 6:167. doi: 10.3389/fgene. 2015.0016.

Hall SJ. 2004. Livestock biodiversity. Oxford: Blackwell.

Hall SJ. 2022. Genetic differentiation among livestock breeds—values for Fst. Animals. 12(9):1115. doi: 10.3390/ ani12091115.

Hiemstra SJ, de Haas Y, Mäki-Tanila A, Gandini G, eds. 2010. Local cattle breeds in Europe: development of policies and strategies for self-sustaining breeds. Wageningen, Netherlands: Wageningen Academic Publishers. Hoffmann AA, Parsons PA. 1997. Extreme environmental change and evolution.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet. 10(9):639–650. doi: 10.1038/nrg2611.

Jacoby WG. 2000. LOESS: a nonparametric, graphical tool for depicting relationships between variables. Electoral Studies. 19(4):577–613. doi: 10.1016/S0261-3794(99)00028-1.

Jeong JY, Ibrahim M, Kim MJ, So K, Jeong YD, Park S, Kim M, Lee H-J. 2017. Comparisons of extracellular matrix-related gene expression levels in different adipose tissues from Korean cattle. Livestock Science. 198:138–146. doi: 10. 1016/j.livsci.2017.02.020.

Kambal S, Tijjani A, Ibrahim SA, Ahmed MKA, Mwacharo JM, Hanotte O. 2023. Candidate signatures of positive selection for environmental adaptation in indigenous African cattle: a review. Anim Genet. 54(6):689–708. doi: 10.1111/ age.133.

Kim S, Cheong HS, Shin HD, Lee SS, Roh HJ, Jeon DY, Cho CY. 2018. Genetic diversity and divergence among Korean cattle breeds assessed using a BovineHD single-nucleotide polymorphism chip. Asian-Australas J Anim Sci. 31(11): 1691–1699. doi: 10.5713/ajas.17.0419.

Küttel L, Letko A, Häfliger IM, Signer-Hasler H, Joller S, Hirsbrunner G, Mészáros G, Sölkner J, Flury C, Leeb T, et al. 2019. A complex structural variant at the KIT locus in cattle with the Pinzgauer spotting pattern. Anim Genet. 50(5):423–429. doi: 10.1111/age.12821.

Lacy RC. 1997. Importance of genetic variation to the viability of mammalian populations. J. Mammal. 78(2):320–335. doi: 10.2307/1382885.

Lam S, Miglior F, Fonseca PAS, Gómez-Redondo I, Zeidan J, Suárez-Vega A, Schenkel F, Guan LL, Waters S, Stothard P, et al. 2021. Identification of functional candidate variants ITALIAN JOURNAL OF ANIMAL SCIENCE 373 and genes for feed efficiency in Holstein and Jersey cattle breeds using RNA-sequencing. J Dairy Sci. 104(2):1928– 1950. doi: 10.3168/jds.2020-18241.

Lee HL, Park JH, Kim JG. 2015. Expression of coat color associated genes in Korean brindle cattle by microarray analysis. J Embryo Transfer. 30(2):99–107.

Lee J, Mun H, Koo Y, Park S, Kim J, Yu S, Shin J, Lee J, Son J, Park C, et al. 2024. Enhancing genomic prediction accuracy for body conformation traits in Korean Holstein cattle. Animals. 14(7):1052. doi: 10.3390/ani14071052.

Li R, Li C, Chen H, Li R, Chong Q, Xiao H, Chen S. 2020. Genome-wide scan of selection signatures in Dehong humped cattle for heat tolerance and disease resistance. Anim Genet. 51(2):292–299. doi: 10.1111/age.12896.

Li W, Sartelet A, Tamma N, Coppieters W, Georges M, Charlier C. 2016. Reverse genetic screen for loss-of-function mutations uncovers a frameshifting deletion in the melanophilin gene accountable for a distinctive coat color in Belgian Blue cattle. Anim Genet. 47(1):110–113. doi: 10. 1111/age.12383.

Liao H, Zhang XH, Qi YX, Wang YQ, Pang YZ, Zhang ZB, Liu P. 2018. The relationships of collagen and ADAMTS2 expression levels with meat quality traits in cattle. Indian J Anim Res. 52(1):167–172.

Makina SO, Taylor JF, van Marle-Köster E, Muchadeyi FC, Makgahlela ML, MacNeil MD, Maiwashe A. 2015. Extent of linkage disequilibrium and effective population size in four South African Sanga cattle breeds. Front Genet. 6: 337. doi: 10.3389/fgene.2015.00337.

Manca E, Cesarani A, Gaspa G, Sorbolini S, Macciotta NPP, Dimauro C. 2020. Use of the multivariate discriminant analysis for genome-wide association studies in cattle. Animals. 10(8). doi: 10.3390/ani10081300

Mardia KV, Kent JT, Bibby JM. 2000. Multivariate analysis. London: Academic Press.

Martín-Burriel I, Rodellar C, Cañon J, Cortés O, Dunner S, Landi V, Martínez-Martínez A, Gama LT, Ginja C, Penedo MCT, et al. 2011. Genetic diversity, structure, and breed relationships in Iberian cattle. J Anim Sci. 89(4):893–906. doi: 10.2527/jas.2010-3338.

Mateus JC, Penedo MCT, Alves VC, Ramos M, Rangel-Figueiredo T. 2004. Genetic diversity and differentiation in Portuguese cattle breeds using microsatellites. Anim Genet. 35(2):106–113. doi: 10.1111/j.1365-2052.2004.01089.x.

Mei C, Wang H, Liao Q, Wang L, Cheng G, Wang H, Zhao C, Zhao S, Song J, Guang X, et al. 2018. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. Mol Biol Evol. 35(3):688–699. doi: 10. 1093/molbev/msx322.

Mills LS, Smouse PE. 1994. Demographic consequences of inbreeding in remnant populations. Am Natural. 144(3): 412–431. doi: 10.1086/285684. Ministerio de Agricultura, Pesca y Alimentación (MAPA). 2023. [accessed 2023 Jun 15] https://www.mapa.gob.es/.

Moravčíková N, Kasarda R, Vostrý L, Krupová Z, Krupa E, Lehocká K, Olšanská B, Trakovická A, Nádaský R, Židek R, et al. 2019. Analysis of selection signatures in the beef cattle genome. Czech J Anim Sci. 64(12):491–503. doi: 10. 17221/226/2019-CJAS.

Moreno García CA, Zhou H, Altimira D, Dynes R, Gregorini P, Jayathunga S, Maxwell TMR, Hickford J. 2022. The glutamate metabotropic receptor 5 (GRM5) gene is associated with beef cattle home range and movement tortuosity. J Anim Sci Biotechnol. 13(1):111. doi: 10.1186/s40104-022- 00755-7.

Muller HJ, Jacobs-Muller JM. 1925. The standard errors of chromosome distances and coincidence. Genetics. 10(6): 509–524. doi: 10.1093/genetics/10.6.509.

Muniz MMM, Simielli Fonseca LF, Scalez DCB, Vega AS, Silva DBDS, Ferro JA, Chardulo AL, Baldi F, Cánovas A, de Albuquerque LG. 2022. Characterization of novel lncRNA muscle expression profiles associated with meat quality in beef cattle. Evol Appl. 15(4):706–718. doi: 10.1111/eva.13365.

Nicoloso L, Milanesi E, Passerotti S, Malavolta M, Gilmozzi I, Crepaldi P. 2009. Assessing SNPs in coat colour genes for cattle breed traceability. Ital J Anim Sci. 8(Suppl 2):120– 122. doi: 10.4081/ijas.2009.s2.120.

Notter DR. 1999. The importance of genetic diversity in livestock populations of the future. J Anim Sci. 77(1):61–69. doi: 10.2527/1999.77161x.

Osada N. 2015. Genetic diversity in humans and non-human primates and its evolutionary consequences. Genes Genet Syst. 90(3):133–145. doi: 10.1266/ggs.90.133.

Pacheco HA, Rossoni A, Cecchinato A, Peñagaricano ~ F. 2023. Identification of runs of homozygosity associated with male fertility in Italian Brown Swiss cattle. Front Genet. 14: 1227310. doi: 10.3389/fgene.2023.1227310.

Pedrosa VB, Schenkel FS, Chen SY, Oliveira HR, Casey TM, Melka MG, Brito LF. 2021. Genomewide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. Genes (Basel). 12(11):1830. doi: 10.3390/ genes12111830.

Pertoldi C, Purfield DC, Berg P, Jensen TH, Bach OS, Vingborg R, Kristensen TN. 2014. Genetic characterization of a herd of the endangered Danish Jutland cattle. J Anim Sci. 92(6):2372–2376. doi: 10.2527/jas.2013-7206.

Philipp U, Hamann H, Mecklenburg L, Nishino S, Mignot E, Günzel-Apel A-R, Schmutz SM, Leeb T. 2005. Polymorphisms within the canine MLPH gene are associated with dilute coat color in dogs. BMC Genet. 6(1):34. doi: 10.1186/1471-2156-6-34.

Pimentel ECG, Bauersachs S, Tietze M, Simianer H, Tetens J, Thaller G, Reinhardt F, Wolf E, Konig € S. 2011. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Anim Genet. 42(3):251–262. doi: 10.1111/j.1365-2052.2010.02148.x.

Pintus E, Sorbolini S, Albera A, Gaspa G, Dimauro C, Steri R, Marras G, Macciotta NPP.,. 2014. Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piemontese and Italian Brown cattle breeds. Anim Genet. 45(1):1–11. doi: 10.1111/age.12076.

Porto-Neto LR, Kijas JW, Reverter A. 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genet Sel Evol. 46(1):1–5. doi: 10.1186/ 1297-9686-46-22.

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

Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, Nielsen R, Simianer H. 2014. Classic selective sweeps 374 P. E. MILIA ET AL. revealed by massive sequencing in cattle. PLoS Genet. 10(2):e1004148. doi: 10.1371/journal.pgen.1004148.

Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H. 2010. The pattern of linkage disequilibrium in German Holstein cattle. Anim Genet. 41(4): 346–356. doi: 10.1111/j.1365-2052.2009.02011.x.

R Core Team. 2024. R: a language and environment for statistical computing. R Foundation.

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

Ramey HR, Decker JE, McKay SD, Rolf MM, Schnabel RD, Taylor JF. 2013. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics. 14(1):382. doi: 10.1186/1471-2164-14-382. 816:146165. doi: 10.1016/j.gene.2021.146165.

Rencher AC, Scott DT. 1990. Assessing the contribution of individual variables following rejection of a multivariate hypothesis. Commun. Stat. Simul. Comput. 19(2):535–553. doi: 10.1080/03610919008812874.

Rifkin J. 1992. Biosphere politics: a new consciousness for a new century.

Salatta BM, Muniz MMM, Fonseca LFS, Mota LFM, de Souza Teixeira C, Frezarim GB, de Albuquerque LG. 2024. Skeletal muscle lncRNA profile associated with fatty acids in beef cattle. [in press].

Salleh MS, Mazzoni G, Höglund JK, Olijhoek DW, Lund P, Løvendahl P, Kadarmideen HN. 2017. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high-and lowresidual feed intake in Nordic dairy cattle. BMC Genomics. 18(1):258. doi: 10.1186/s12864-017-3622-9.

Salleh MS, Mazzoni G, Nielsen MO, Løvendahl P, Kadarmideen H. 2018. Identification of expression QTLs targeting candidate genes for residual feed intake in dairy cattle using systems genomics. J Genet Genome Res. 5(1): 1–14. doi: 10.23937/2378-3648/1410035.

Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. 2020. Selection signatures in livestock genome: a review of concepts, approaches and applications. Livestock Sci. 241:104257. doi: 10.1016/j.livsci.2020.104257.

Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, Bhattarai EK, Molaei M, Freetly HC, Hansen SL, et al. 2017. Genome-wide association studies for feed efficiency and growth traits in US beef cattle. BMC Genomics. 18(1):386. doi: 10.1186/s12864-017-3754-y.

Serão NV, González-Peña D, Beever JE, Faulkner DB, Southey BR, Rodriguez-Zas SL. 2013. Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. BMC Genet. 14(1):94. doi: 10.1186/1471-2156- 14-94.

Singh A, Kumar A, Mehrotra A, A K, Pandey AK, Mishra BP, Dutt T. 2021. Estimation of linkage disequilibrium levels and allele frequency distribution in crossbred Vrindavani cattle using 50K SNP data. PLoS One. 16(11):e0259572. doi: 10.1371/journal.pone.0259572.

Sood V, Rodas-González A, Valente TS, Virtuoso MCS, Li C, Lam S, López-Campos Ó, Segura J, Basarab J, Juárez M. 2023. Genome-wide association study for primal cut lean traits in Canadian beef cattle. Meat Sci. 204:109274. doi: 10.1016/j.meatsci.2023.109274.

Sorbolini S, Marras G, Gaspa G, Dimauro C, Cellesi M, Valentini A, Macciotta NP. 2015. Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories. Genet Sel Evol. 47(1):1–13. doi: 10.1186/s12711- 015-0128-2.

Stegemiller MR, Ellison MJ, Hall JB, Sprinkle JE, Murdoch BM. 2021. Identifying genetic variants affecting cattle grazing behavior experiencing mild heat load. Transl Anim Sci. 5(Supplement_S1):S61–S66. doi: 10.1093/tas/txab151.

Taye M, Lee W, Caetano-Anolles K, Dessie T, Hanotte O, Mwai OA, Kemp S, Cho S, Oh SJ, Lee H-K, et al. 2017. Whole genome detection of signature of positive selection in African cattle reveals selection for thermotolerance. Anim Sci J. 88(12):1889–1901. doi: 10.1111/asj.12851.

Vanvanhossou SFU, Scheper C, Dossa LH, Yin T, Brügemann K, König S. 2020. A multi-breed GWAS for morphometric traits in four Beninese indigenous cattle breeds reveals loci associated with conformation, carcass and adaptive traits. BMC Genomics. 21(1):783. doi: 10.1186/s12864-020- 07170-0.

Vargas G, Neves HH, Camargo GMF, Cardoso V, Munari DP, Carvalheiro R. 2018. Genome-wide association study and functional analysis of feet and leg conformation traits in Nellore cattle. J Anim Sci. 96(5):1617–1627. doi: 10.1093/ jas/sky079.

Verardo LL, E Silva FF, Machado MA, do Carmo Panetto JC, de Lima Reis Faza DR, Otto PI, de Almeida Regitano LC, da Silva LOC, do Egito AA, do Socorro Maués Albuquerque M, et al. 2021. Genome-wide analyses reveal the genetic architecture and candidate genes of indicine, taurine, synthetic crossbreds, and locally adapted cattle in Brazil. Front Genet. 12:702822. doi: 10.3389/fgene.2021.702822.

Vieira J, Verardo L, Magalhães A, Bonafé C, Silva M, Machado M, de Morais LR. 2022. Candidate genes for tick resistance in cattle: a systematic review combining post-GWAS analyses with sequencing data. J Appl Anim Res. 50: 460–470.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38(6):1358– 1370. doi: 10.1111/j.1558-5646.1984.tb05657.x.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 19(3):395–420. doi: 10.2307/2406450.

Zeng L, Chen N, Ning Q, Yao Y, Chen H, Dang R, Zhang H, Lei C. 2018. PRLH and SOD1 gene variations associated with heat tolerance in Chinese cattle. Anim Genet. 49(5): 447–451. doi: 10.1111/age.12702.

Zepeda-Batista JL, Nuñez-Domínguez R, Ramírez-Valverde R, Ruíz-Flores A, Jahuey-Martínez FJ, Parra-Bracamonte GM. 2021. Genome-wide association analysis for liveweight traits in Braunvieh cattle.

Zhang H, Liu A, Wang Y, Luo H, Yan X, Guo X, Li X, Liu L, Su G. 2021. Genetic parameters and genome-wide association studies of eight longevity traits representing either full or partial lifespan in Chinese Holsteins. Front Genet. 12:634986. doi: 10.3389/fgene.2021.634986.

Zhao F, McParland S, Kearney F, Du L, Berry DP. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet Sel Evol. 47(1):1–12. doi: 10.1186/s12711-015-0127-3.

Appendix Chapter 2

Evaluation of genetic differences between two autochthonous Spanish cattle breeds

Supplementary Table 1. Distinguishing characteristics of the five populations (Ministerio de Agricultura, Pesca y Alimentación, MAPA).

Supplementary Table 2. Genes and associated traits found in each comparison.

Supplementary Table 1. Distinguishing characteristics of the five populations (Ministerio de Agricultura, Pesca y Alimentación, MAPA).

		AVI			MOR	
	Distinctive features	ANI	BOC	JAR	MORB	MORC
Morphological traits	Coat color	Black	Black to dark brown coat with some discoloration of the dorsolumbar line and the groin area	Brown with the absence of color of the main layer in certain body regions (“Spot Pinzaguer”)	Black (white in the undercoat)	Brown coat represented by a mixture of black and white hairs in different proportions
	Horn	Black or white with black tips	-	-	White on the stock and shovel, and black at their tips	Dirty white on the stock and shovel and black on the tip or piton
	Height withers males (cm)	148	-	-	-	143
	Height withers females (cm)	140	-	-	-	137
	Weight males (kg)	900	1,000	-	-	900

Production data	Weight females (kg)	500	550	-	-	500
	Production	Meat	Meat		Meat	Meat
	Average daily gain (gr/day)	1	950	-	1,240	1
Others	Breeding system	Extensive	Extensive	-	Extensive	Extensive
	Special features	High meat quality, mobility and longevity	High rusticity	-	More rustic and smaller than Cárdena with more developed udders	High fecundity

Supplementary Table 2. Genes and associated traits found in each comparison.

Comparison	Gene	Associated trait	References
AVI vs MOR	VPS8, MAGEF1	Reproduction	Dubon et al., (2021)
	MAGEF1	Growth and carcass	Kim et al., (2012)
	COPS8, COL6A3	Meat quality	Dinh et al., (2021); Arikawa et al., (2024)
	MLPH, RAB17	Coat color	Nicoloso et al., (2009); Lee et al., (2015); Li et al., (2016); Taye et al., (2017); Li et al., (2020)
	PRLH	Heat tolerance	Taye et al., (2017); Zeng et al., (2018)
	GABRG1	Milk production	Pedrosa et al., (2021)
	YTHDC2, KCNN2	Feed efficiency	Olivieri et al., (2016)
	LRRTM4	Milk production, Behavior, Rusticity, and Fertility	Chen et al., (2020); Verardo et al., (2021)

DHX57, MORN2, ARHGEF33, MAP4K3	Body weight, Weaning weight	Edea et al., (2020); Zepeda-Batista et al., (2021)
SOS1	Residual feed intake, Meat and carcass quality	Lam et al., (2021); Naserkheil et al., (2021)
CDKL4	Growth	Sheet et al., (2024)
TFDP1, ATP4B, GRK1	Tick resistance	Pickering et al., (2017)
GAS6	Somatic cell count	Jeretina et al., (2020)
RASA3	Immune system, Heat stress	Kiser et al., (2017); Del Corvo et al., (2021)
CDC16	Meat and Carcass trait (Pig)	Dall'Olio, (2020)
CHAMP1	Growth and carcass quality (Sheep)	Cardona Tobar et al., (2020)

BAALC	Immune system	Neibergs et al., (2014)
AZIN1	Lactation persistency	Neibergs et al., (2014); Do et al., (2017)
ATP6V1C1	Meat quality, Lactation persistency	Jeong et al., (2013); Do et al., (2017); Li et al., (2020)
ENPP6	Milking speed, Trypanotolerance	Chen et al., (2020); Goyache et al., (2021)
IRF2	Immune system	Liu et al., (2012);
CASP3	Reproduction	Kamemori et al., (2011)
PRIMPOL	Carcass quality	Roberts, (2018)
CENPU	Fertility	Hozé et al., (2020)

MORB vs MORC	ERICH3	Grazing behavior	Stegemiller et al., (2021)
	TNNI3K	Udder depth	Kramer et al., (2014)
	TNNI3K, FPGT	Lameness	Sánchez-Molano et al., (2019)
	LRRIQ3	Body weight, Feed efficiency, Feed intake	Rajawat et al., (2022); Abo-Ismail et al., (2018); Salleh et al., (2022)
	PRSS12	Meat quality	Beak et al., (2022)
	NDST3	Clinical ketosis	Lunn et al., (2022); Soares et al., (2021)
	MEGF11	Feed efficiency	Vanvanhossou et al., (2020)
	DIS3L	Milk production	Saravanan et al., (2021); Rajawat et al., (2022)

DIS3L, TIPIN	Meat quality	Roberts, (2018)
MAP2K1	Coat color	Gutiérrez-Gil et al., (2015); Mei et al., (2018)
SNAPC5	Milk composition	Cecchinato et al., (2019)
LRP12	Feed efficiency	Serão, (2013)
LRP12, DPYS	Metabolic weight	Seabury et al., (2017)
DCSTAMP	Immune system	Wathes et al., (2021)
RIMS2	Carcass and meat quality	Wang et al., (2018); Liu et al., (2019)
GPI	Endocrinal and classical fertility	Tarekegn et al., (2021)
PDCD2L, UBA2, WTIP	Reproduction	Ahmad et al., (2023)

RYBP	Body conformation	Haque et al., (2023)
CALN1	Longevity, Memory and Learning	Zhang et al., (2021); Sorbolini et al., (2015)
GALNT17	Meat quality	Arikawa et al., (2024)
GNG4	Residual feed intake	Onteru et al., (2013)
LYST	Hereditary disease	Kunieda et al., (2005)
NID1	Growth, Trypanotolerance	Mao et al., (2016); Tijjani et al., (2019)
GRM5	Home range and Beef tortuosity, Grazing behavior	Moreno García et al., (2022); Moreno García et al., (2024);
CTSC	Immune system	Han et al., (2020)
RAB38	Coat color	Hollmann et al., (2017)

	CCDC179	Somatic cell count	Count et al., (2019)
	GAS2	Fertility	Lefebvre et al., (2021)
ANI vs BOC	CD47	Respiratory disease	Newman et al., (2011)
	IFT57	Feet and leg conformation	Vargas et al., (2018)
	COL5A2	Linear type	Doyle et al., (2020); Doyle et al., (2021)
	COL3A1	Meat quality, Growth	Jeon et al., (2017); Chen et al., (2021)
	ELAVL4	Fertility	Jiang et al., (2019)
	AGBL4	Growth Body conformation	Marín-Garzón et al., (2021) Lee et al., (2024)

GALNTL5, KMT2C, CCT8L2		
KLB	Lactation persistency	Do et al., (2017)
RPL9	Fertility	Mohammadi et al., (2020); Pacheco et al., (2023)
LIAS	Fertility, Meat	Ben-Jemaa et al., (2021); Pacheco et al., (2023)
SMIM14, PDS5A	Meat	Ben-Jemaa et al., (2021)
UBE2K	Meat quality and Carcass	Srikanth et al., (2020); Muniz et al., (2022); Sood et al., (2023)
UGDH	Milk production	Xu et al., (2012)
MALRD1	Meat and Carcass, Body conformation	Vostry et al., (2023); Haque et al., (2023)

PLXDC2, ABTB2	Carcass	Strillacci et al., (2014); Karisa et al., (2013)
NAT10	Meat quality	Sood et al., (2023)
CAPRIN1	Respiratory disease, Behavior, Meat quality	Neupane et al., (2018); Araujo et al., (2021); Sood et al., (2023)
CAT	Fertility	Abdollahi-Arpanahi et al., (2019)
ELF5	Milk protein synthesis	Sigl et al., (2014); Xia et al., (2018)
EHF	Milk production	Pedrosa et al., (2021)
OR5M10	Growth	Higgins et al., (2018) Higgins et al., (2019)
PBLD, HNRNPH3	Respiratory disease	Neupane et al., (2018); Sun et al., (2020)
SLC25A16	Behavior	Macleod et al., (2019)
TET1		Zhang et al., (2023)

		Skeletal muscle growth and development	
	CCAR1	Mammary gland development	Wetzel-Gastal et al., (2018)
	RUFY2, STOX1, DDX50	Whorls (Horse)	Lima et al., (2021)
ANI vs JAR	COL5A2	Linear type	Doyle et al., (2020)
	COL3A1	Meat quality	Liao et al., (2018); Jeong et al., (2017)
	GALNT13	Endoparasite infection, Reproduction	Wolf et al., (2021); Borowska et al., (2018)
	ELAVL4	Reproduction, Calf mortality	Jiang et al., (2019); Garzón, (2019)
	AGBL4	Growth and Carcass, Fertility	Hering et al., (2014); Akanno et al., (2018);

		Raza et al., (2020)
GPR161, DCAF6, MPC2	Fertility	Alves et al., (2022)
ADCY10	Fertility, Milk quality	Alves et al., (2022); Laodim et al., (2024)
PDE3A	Respiratory disease, Fertility	Neupane et al., (2018); Liu et al., (2017)
KIT, KDR	coat color	Tazzoli et al., (2007); Fontanesi et al., (2010); Brenig et al., (2013); Quanbari et al., (2014); Küttel et al., (2019); Moravčíková et al., (2019)
MTMR3	Milk production, Carcass, Feed efficiency	Pimentel et al., (2011); Bekele et al., (2023); Brunes et al., (2021)
HORMAD2	Calving performance and conformation	Abo-Ismail et al., (2017)

	LIF	Pregnancy, Heat Stress	Oshima et al., (2003); Otto et al., (2019)
	OSM	Heat stress, Feed efficiency	Otto et al., (2019); Worku et al., (2023); Brunes et al., (2021)
	SF3A1, SEC14L2, MTFP1	Feed efficiency	Brunes et al., (2021); Salleh et al., (2018)
	CASTOR1	Heat stress	Yue et al., (2015)
BOC vs JAR	CREB5	Fertility, Dry matter intake	Tarekegn et al., (2021); Tarekegn et al., (2022)
	JAZF1	Body size, Sole ulcer	Zhao et al., (2015); Sölzer et al., (2022)
	ADGRL3	Milk production, Meat quality	Jiang et al., (2019); Reis et al., (2023); Yodklaew et al., (2017)
		Meat quality(Pig)	Cinar et al., (2012)

AMBP		
KIF12	Meat quality (Sheep)	Ekawati et al., (2022)
MYO6	Tick resistance, Growth, Clinical mastitis, weaning weight, lon- gissimus muscle area, residual feed intake, milk fat yield	Vieira et al., (2022); Mustafa et al., (2018); Makina et al., (2015)
IMPG1	Milk fat yield (Buffalo)	Du et al., (2019)
LAPTM4B	Fertility	Yathish et al., (2017); Hodge et al., (2023)
MTDH	Skeletal and muscle disordes, Genetic disorders	Newman et al., (2011)
TSPYL5	Meat quality, Fertility	Lee et al., (2011); Hodge et al., (2023)

KIAA1549L	Meat quality, Fertility	Soo et al., (2023); Abril-Parreño et al., (2023)
FBXO3, LMO2	Meat quality	Sood et al., (2023)
CD59	Immune response	Cao et al., (2023)
DEF8, GAS8, ORC6	Beef production and carcass quality, Residual feed intake, Feed conversion ratio, Conception rate	Edea et al., (2020)
SHCBP1	Meat quality, Beef production and carcass quality, Residual feed intake, Feed conversion ratio, Conception rate	Braz et al., (2019)
ORC6	Carcass	Srikanth et al., (2020)
DBNDD1	Growth	Gurgul et al., (2020)
	Milk quality, Milk production	

VPS35		Hylén et al., (2018); Saravanan et al., (2021)
MYLK3	Muscle development , Reproduction	Liu et al., (2022); King et al., (2022)
KIF2B	Adaptation	Chen et al., (2018)
CUEDC1	Meat quality, Milk production	Pedrosa et al., (2021); Briscadin et al., (2021); Salatta et al., (2024)
VEZF1, SRSF1	Meat quality , Lactation persistency	Briscadin et al., (2021); Salatta et al., (2024)
DYNLL2	Meat quality	Salatta et al., (2024)
MKS1	Lactation persistency	Do et al., (2017)

CHAPTER 3

Tracing shared genomic regions among local cattle breeds from Sardinia

This is the published version in **Frontiers in Animal Science** of:

Milia, PE., Cesarani, A., Falchi, L., Macciotta, NPP., Dimauro, C. (2025). Tracing shared genomic regions among local cattle breeds from Sardinia. *Frontiers in Animal Science*, *6*. https://doi.org/10.3389/fanim.2025.1685894

This version is used by permission of the **Frontiers in Animal Science**.

Abstract

Three local cattle breeds are farmed in the island of Sardinia (Italy): Sarda (SAR), Sardo-Bruna (SB), and Sardo-Modicana (SM). Historically, SAR was the Sardinian autochthonous breed, known for its resilience despite low productivity. To improve its performance, SAR cows were crossbred with Brown Swiss (BR) and Modicana (MOD) bulls, originating SB and SM. The aim of this research was to investigate how much SAR genomic background remains in SB and SM. A sample of 239 animals (64 SAR, 57 SM, 20 SB, 50 BR and 48 MOD) were genotyped using the 50K SNP Illumina BeadChip. Initially, univariate approaches were used to investigate the genomic relationship of SB and SM with SAR, MOD, and BR. Admixture analysis (K = 2) revealed a greater genetic similarity of SB with SAR than with BR, whereas SM showed a greater genetic proximity with MOD than with SAR. This was also reflected in the Weighted FST values: SAR and SB showed the lowest FST (0.015), confirming their closer relationship, whereas SB and BR showed the highest (0.058), consistent with their greater divergence. For SM, the FST with MOD (0.027) was lower than with SAR (0.030), supporting its closer affinity with MOD. SNP in ROH (SNPROH) analyses identified 21 genomic regions shared among Sardinian cattle breeds, where 3 genes (COMMD1, B3GNT2, and FAM161A) linked to environmental adaptation were found. Further analyses combining univariate and multivariate methods identified SNP with low discriminant power between SAR and the derived populations (SM and SB). Some of these regions overlapped with shared runs of homozygosity (ROHREP), and contained genes potentially associated with adaptation (CTNNA2, ITPR2, NTN1, and USP43). This study demonstrates that part of the SAR genome is still present in SB and SM, particularly in regions related to adaptive traits.

Keywords

Autochthonous cattle breeds, adaptive genomic signatures, runs of homozygosity (ROH), Wright’s fixation index (FST), stepwise discriminant analyses (SDA), canonical discriminant analyses (CDA).

Introduction

The highest number of endangered autochthonous cattle breeds is found in Europe and in the Caucasus, due to the widespread replacement of local populations with a limited number of highly specialized cosmopolitan breeds (Curone, 2019). Despite the dominance of these high-yielding and economically efficient breeds, some indigenous populations have survived, primarily due to the strong connection between their productive and reproductive traits and the environment in which they have evolved. Compared to highly selected cosmopolitan breeds, local populations are characterized by a greater frugality, fertility, disease resistance, longevity and resilience (Gandini et al., 2017), that make them particularly suitable for extensive management, natural grazing, and vegetation control (Felius et al., 2014). For this reason, they often represent the best option for economically exploiting marginal areas such as mountain pastures, woodlands, and foothills (De Haan et al., 1997; Ajmone-Marsan et al., 2001). These breeds have not undergone intensive genetic selection to enhance a single specific trait; as a result, many of them retain a dual-purpose (milk and meat) or even triple-purpose aptitude (milk, meat and work).

Sardinia is the second main island of Italy. Three local cattle breeds are farmed in this region: the Sarda (SAR), the Sardo-Bruna (SB), and the Sardo-Modicana (SM). Until the first half of the last century, the predominant breed was the Sarda (Figure 1A). This breed is present in Sardinia since the Neolithic, after the introduction of animals of Bos Macroceros species from Iberia, North Africa and Syria (Della Maria, 1936; Brandano et al., 1983). The breed is currently characterized by a marked phenotypic heterogeneity, especially in coat color and size (Bigi and Zanon, 2008). The Sarda is farmed in marginal and impervious areas, permitting the exploitation of agronomically infertile soils, mostly for meat production. The herd book was established in 1987.

In order to improve its milk and meat production, since the second half of the XIX century Sarda cows started to be crossed with Brown Swiss (BR; Figure 1D) bulls imported from Switzerland, originating the dual aptitude Sardo-Bruna breed (Figure 1B) (Brandano et al., 1983). This breed, valued for its productivity and rusticity, was also exported to other Italian regions as well as Corsica and North Africa. It is medium size cattle characterized by a grey coat of various shades. Sardo Bruna cattle are mostly farmed in hilly and mountainous areas, often together with sheep in mixed systems (Bigi and Zanon, 2008). The family herd book was established in 1933, while the technical population standards were approved in 1996 (Regione Sardegna, 2007).The last breed, the Sardo-Modicana (Figure 1C), originated with the aim of improving the working aptitude of the Sarda. This breed originated from the crossbreeding between Modicana (MOD; Figure 1E) bulls imported from Sicily and native Sarda cows. Before the development of agricultural mechanization, this breed was very appreciated for its working aptitude. The herd book, established in 1927, became effective in 1935. Instead, the breed registry was instituted in 1987 to safeguard the genetic variability of the breed, and it was later renewed in 1995 (Regione Sardegna, 2007).

Due to their easy calving and maternal aptitude, Sardinian local breeds are now mostly used for meat production by crossbreeding with specialized beef breeds, destined for fattening centers. Thus, the production of purebred animals is limited. According to FAO's Domestic Animal Diversity Information System (DAD-IS, last update on 7 October 2024), the numbers of breeding males and females for Sarda are 1192 and 16597, raised across 898 herds with an average herd size of 24 animals. The Sardo-

Bruna, a total of 1002 and 20741 individuals are registered as breeding males and females, respectively. The breed is farmed in 1,352 herds, with an average value of 21 animals per herd. The Sardo-Modicana, which is classified as “at risk”, has only 48 breeding males and breeding 1,493 females, across 118 herds averaging 19 animals. These figures highlight the vulnerability of purebred populations and the potential risk of genetic erosion. The three Sardinian cattle breeds are not under breeding programs, whereas in situ conservation programs are currently active (FAO, 2024).

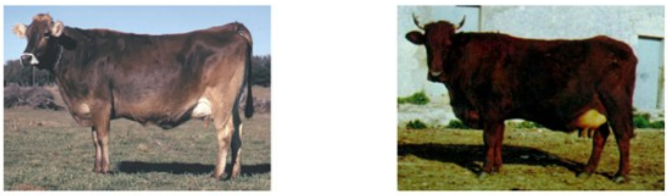
Recent advances in genomic technologies, such as high-density SNP arrays and next-generation sequencing, provided exceptional insights into the evolutionary dynamics and genetic architecture of cattle populations, enabling detailed investigations of genetic diversity and historical evolution (Acciaro, 2013; Taye, 2018). The main approaches currently applied include the detection of selection signatures through allele frequency analysis (e.g., FST), linkage disequilibrium and haplotype length (XP-EHH), composite methods (XP-CLR), and allele spectrum analysis (Tajima’s D) (Taye, 2018). These tools allow the identification of genes associated with production and adaptation. Analyses of population structure and admixture carried out by the use of PCA or STRUCTURE and ADMIXTURE software, enable the identification of homogeneous genetic groups and ancestral contributions (e.g., Bos taurus vs. Bos indicus) (Edea et al., 2014; Taye et al., 2017; Cesarani et al., 2018). Runs of homozygosity (ROH)—continuous homozygous segments—are useful for estimating inbreeding (FROH), reconstructing demographic history, and detecting regions under selection (Fabbri et al., 2022). Linkage disequilibrium (LD), i.e., the non-random association of alleles at nearby loci, provides insights into the evolutionary history of populations: extended LD may signal recent selection events or demographic bottlenecks, whereas low LD indicates high genetic variability and absence of recent inbreeding (Gibson et al., 2006; Taye, 2018; Fabbri et al., 2022).

Application of these methodologies to Sardinian local breeds has revealed a detailed picture of their genetic variability. With respect to genetic structure and diversity, Sardinian breeds exhibit high intra-breed heterogeneity (Cesarani et al., 2018). Multidimensional scaling and admixture analyses confirm the genetic proximity of SM to MOD and SB to SAR (Cesarani et al., 2018; Mastrangelo et al., 2018). Inbreeding estimates based on ROH analyses (Fabbri et al., 2022) revealed lower values for Sardinian breeds when compared with other Italian local breeds (e.g., Calvana, Pisana and Pontremolese). Furthermore, the distribution of ROH suggests that inbreeding events in these breeds were less recent and less intense (Fabbri et al., 2022). Both ROH based approach and FST highlight genes likely subjected to strong selective pressure: the EIF6 gene on chromosome 13 (involved in glycolysis and lipid synthesis) and the DHRS7 gene on chromosome 10 (linked to steroid metabolism and intramuscular fat deposition) (Cesarani et al., 2018). Contemporary effective population size (cNe), which represents the idealized size of a population where genetic variability is maintained constantly across generations, varied among breeds: SB showed the largest cNe, indicating greater genetic variability and, consequently, higher evolutionary potential than SAR and SM (Mastrangelo et al., 2018). Studies on mitochondrial origins and African introgression revealed that most Sardinian mitochondrial DNA haplotypes belong to haplogroup T3 (typical of Europe), although T1 haplotypes, characteristic of African cattle, were also detected, including one unique haplotype specific to Sardinian breeds (Petretto et al., 2022). This evidence points to historical

African introgression, either directly via Mediterranean routes or indirectly through the Iberian Peninsula.

Although previous studies have provided insights into the population structure, genetic diversity, and adaptation signatures of Sardinian cattle breeds, no investigation has yet addressed to what extent the SAR is still retained in the derived SB and SM populations. The scientific question of the paper was to estimate how much of SAR genomic background remains in the genetic heritage of SB and SM. For this reason, genotypes of SB and SM were compared not only with SAR, but also with those of the purebred populations from which they originated (BR and MOD).

a. Sarda (SAR) **b.** Sardo – Bruna (SB) **c.** Sardo – Modicana (SM)



d. Brown Swiss (BR) (Encyclopedia Britannica, 2025) **e.** Modicana (MOD) (Agraria.org, 2025)

Figure 1. Pictures of the cattle breeds involved in the study. **a.** Sarda cattle; **b.** Sardo–Bruna cattle; **c.** Sardo–Modicana cattle; **d.** Brown Swiss cattle (source: *Brown Swiss*, Encyclopædia Britannica, https://www.britannica.com/animal/Brown-Swiss, accessed 31 Jan 2025); **e.** Modicana cattle (source: Agraria.org, *Modicana*, https://www.agraria.org/razzebovineminori/modicana.htm, accessed 31 Jan 2025).

Material and methods

Animals and genomic data

Five different breeds (SAR, BR, MOD, SB, and SB) were considered in this study. The Supplementary Table 1 summarizes the distinguishing features of the analyzed subpopulations (Bigi et al., 2008; Biodiversità Sardegna, 2025; Braunvieh Switzerland, 2025; DAD-IS, 2025). All animals were genotyped with the 50K SNP Illumina BeadChip (Illumina, San Diego, CA, USA). The initial dataset contained 242 animals (50 BR, 50 MOD, 64 SAR, 20 SB, and 58 SM) and 35,364 common SNP. Quality control on genotypes was carried out breed by breed using PLINK 1.9 (Purcell et al., 2007) according to the following criteria: minor allele frequency greater than 2%, animal and SNP call rate greater than 95%. Moreover, SNP not in Hardy-Weinberg Equilibrium ($P<1e-6$), mapped on sexual chromosomes or unmapped according to the ARS1.2URC release, were also discarded. After quality control, 239 animals (50 BR, 48 MOD, 64 SAR, 20 SB, and 57 SM) and 35,057 SNP were retained for the subsequent analyses. Two different statistical approaches, univariate and multivariate, were employed to identify markers and corresponding genomic regions containing genes specifically associated with SAR.

Population genomic structure and admixture

Admixture analysis was conducted using Admixture software (v1.3), following the method described by Alexander et al. (2015) in the related software manual. To facilitate the interpretation of genetic structure among the analyzed populations, genetic admixture was estimated assuming $K = 2$ ancestral populations and calculated by the maximum likelihood method. For this reason, three Admixture analyses were conducted considering: i) SAR, SB, and SM; ii) MOD, SAR, and SM, iii) BR, SAR, and SB. By setting K = 2 the animals of three breeds were assigned to just two ancestral groups and, thus, we could better evaluate whether the two "composite" breeds (SB and SM) are closer to SAR or to the other two purebreds. In parallel, the genomic relationship matrix (GRM) was built with GCTA v1.92.1 (Yang et al., 2011) and used to estimate pairwise genomic relationships among individuals. Average relationships between breeds were calculated to assess genetic similarity. A principal component analysis (PCA) was then performed on the GRM in R v4.2.2 (R Core Team, 2022), and the first components were visualized using the package "ggplot2" (Wickham, 2016) to further explore the population structure.

Linkage disequilibrium and effective population size

Linkage disequilibrium (LD) and effective population size (Ne) were estimated for the five cattle breeds considered in this study (SAR, SB, SM, MOD, and BR). LD values were calculated with HAPLOVIEW (Barrett et al., 2005), considering markers within a physical distance of 1000 kb, following the procedure described by McKay et al. (2007). Estimates of Ne were derived from LD values according to the equation of Sved (1971), modified to account for finite population size. In addition, for the three Sardinian local breeds (SAR, SB, and SM), Ne was also computed using demographic data retrieved by the FAO database (DAD-IS, 2024). In this case, effective population size was estimated according to the equation proposed by Falconer and Mackay (1996):

$$Ne = \frac{4 * N_m * N_f}{(N_m + N_f)}$$

where N_m and N_f are the number of adult males and females, respectively.

Wright fixation index

Using PLINK 1.9, the pairwise Wright fixation index (FST) was computed for the following pairwise comparisons: SAR vs SB, SAR vs SM, BR vs SB and MOD vs SM using the equation proposed by Weir and Cockerham (1984). To identify genomic regions potentially conserved or shared among Sardinian cattle breeds, FST values from the SAR vs SB and SAR vs SM pairwise comparisons were retained for further analyses. Given the strong asymmetry of the FST distribution, with values highly concentrated near zero, we applied a modified threshold to select SNP with lower FST values (FST_{LOW}) than expected under neutrality. Specifically, we defined a lower-bound threshold as:

$$FST_{LOW} = average(FST) - 0.5 * IQR$$

where average (FST) is the average weighted FST value among the considered comparison and IQR is its interquartile range. This choice departs from the conventional criterion (1.5 × IQR) which would typically be used to identify outliers. However, the standard threshold proved too stringent for our dataset, resulting in the near-complete exclusion of SNP. The adjusted threshold preserved the statistical rationale of selecting variants based on deviation from the central tendency, while being tailored to the specific distribution observed in our data. This approach allowed us to retain SNP with unusually low genetic differentiation between SAR and the other breeds, which may reflect regions under balancing selection or shared ancestral polymorphisms.

Runs of Homozygosity

Consecutive ROH were computed using the "*detectruns*" R package (Biscarini et al., 2018). The following parameters were adopted to define a ROH: at least 15 homozygote SNP in a row covering a minimum of 1 million base pairs; no heterozygote or missing SNP allowed. Number of ROH per animal (nROH), average ROH length (meanMb), number of ROH per class of length (1-2 Mb, 2-4 Mb, 4-8 Mb, 8-16 Mb, >16 Mb, respectively) and the genomic coefficients of inbreeding based on ROH (F_{ROH}) were computed. A ROH was defined as unique when it originated and ended at the same chromosomal positions. These unique ROH could occur in multiple individuals: when the same region was detected in more than one sample, it was classified as a repeated ROH (ROH_{REP}) (Cesarani et al., 2018; Macciotta et al., 2021; Falchi et al., 2023). ROH_{REP} were investigated to identify genomic regions shared among individuals from Sardinian breeds. A genomic region was considered as shared (shared ROH_{REP}) only if it was present in at least one individual per breed (SAR, SB, and SM). To ensure the specificity of these shared segments to Sardinian cattle, only shared ROH_{REP} that were not detected in any BR and MOD individual were retained. While ROH_{REP} required identical start and end positions across individuals, we also adopted a SNP-level method based on the frequency of ROH occurrence. This strategy corresponds to the concept of ROH islands (Peripolli et al., 2018), allowing us to capture broader genomic regions consistently shared among Sardinian breeds, beyond strictly defined ROH boundaries. For each SNP, the proportion of individuals in which the SNP was located within a ROH was calculated and defined as SNP_{ROH}. To identify SNP_{ROH} shared among the Sardinian breeds, markers with a SNP_{ROH} value larger than the 97th percentile in all three breeds (i.e., SAR, SB, and SM) were selected. This stringent cutoff was chosen to capture only the strongest signals of shared genomic regions across Sardinian breeds, thereby avoiding the inclusion of weakly represented markers. Higher thresholds (98th and 99th percentiles) were also tested but did not yield common markers, suggesting the 97th percentile as the maximum threshold at which shared SNP_{ROH} could be identified.

The multivariate approach

The multivariate analysis, according to the algorithm proposed by Manca et al. (2020), was used to explore genetic similarities in the pairwise comparisons involving the three Sardinian cattle breeds (i.e., SAR vs. SB and SAR vs. SM). The genomic data were structured in a multivariate format, where the rows represented the animals and the columns contained the SNP, along with an additional column indicating the breed. Two multivariate statistical techniques were applied to the data: the Stepwise Discriminant Analysis (SDA) and the Canonical Discriminant Analysis (CDA). The SDA is a statistical technique specifically conceived to identify the subset of variables that best differentiate groups. However, in this paper, SDA was used with an opposite purpose: SNP selected in different runs of SDA were subtracted from the dataset until the remaining markers had no discriminating power. With this aim, the STEPDISC procedure of SAS software (version 9.4, SAS Inst. Inc., Cary, NC, USA) was used with the ‘forward’ method and variables were selected basing on the partial R^2 criterion.

CDA is a multivariate statistical approach aimed at highlighting differences between groups of individuals and improving the understanding of relationships among variables. If p represents the number of groups, the CDA generates $p-1$ equations, known as canonical functions (CAN), which are linear combinations of the original variables (Xn). A CAN is structured as follows:

$$CAN = C_1 X_1 + C_2 X_2 + \ldots.. + C_2 X_2$$

where Ci are the canonical coefficients (CC), indicating the contribution of each variable Xi in composing the CAN. To better interpret the role of Xi in group separation, Rencher and Scott (1990) standardized the CC values to account for correlations among the original variables. The absolute values of the standardized CCs can be used to rank variables based on their contribution to the CAN. Group differences within the CAN space were assessed using the Hotelling's T-square test (De Maesschalck et al. 2000). The CANs were also employed to predict the group membership of each animal. In practice, the CANs were applied to individual animals, generating a discriminant score. An animal was assigned to a specific group if its score was below the cutoff value, which was determined by calculating the weighted mean distance between the centroids of the two groups (Mardia et al., 2000).

The overall multivariate approach followed these five steps: i) the SDA was applied to SNP located in genomic regions previously individuated by FST_{LOW}. ii) Different runs of the SDA were subsequently applied to that data. At each run, the partial R^2 threshold was set to a very low value (0.0000001) to select the maximum number of discriminant markers that, in any case, cannot be greater than the number of involved animals. iii) After each SDA run, the CDA was applied to the remaining SNP to test if those markers were able to significantly discriminate groups. iv) The procedure stopped when the residual SNP were not able to significantly separate groups. Genomic regions surrounding these SNP could harbor genes that can be ascribed to SAR. Finally, the markers identified through this analysis were compared with those derived from the univariate approach (shared ROH_{REP} and SNP_{ROH}) with the aim of identifying genomic regions commonly detected by both methods.

Gene and QTL discovery

Gene and QTL discovery was carried out on SNP_{ROH} identified in the comparison involving SAR, SB, and SM, as well as in genomic regions shared across both univariate and multivariate approaches. Genes located within 250 kb upstream and downstream of these variants were considered to account for local linkage disequilibrium (Manca et al., 2020). Annotated genes and QTL were retrieved from the UCSC Genome Browser Gateway (http://genome.ucsc.edu/) and the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov) databases, using the *Bos taurus* genome assembly ARS-UCD1.2 as reference.

Results

Population genomic structure and admixture

Figure 2a reports the Admixture analysis involving the three Sardinian autochthonous breeds (SAR, SB, and SM). SAR and SB shared a more similar ancestral composition (green bars in the Figure), with proportions of 0.80 and 0.89 for K=1, and 0.20 and 0.11 for K=2, respectively. In contrast, SM exhibited a clearly distinct ancestral component (blue bars in the Figure), with proportions of 0.09 for K=1 and 0.91 for K=2, suggesting a divergent genetic background. Figure 2b shows the comparison among MOD, SAR, and SM. SM displayed greater ancestral similarity with MOD (0.94 and 0.65 for K=1, green color) than with SAR, which showed proportions of 0.10 for K=1 and 0.90 for K=2 (blue color). Finally, Figure 2c presents the analysis including BR, SAR, and SB. SB showed a closer and nearly equal genetic affinity (blue bars) with SAR (0.10 and 0.12 for K=1, 0.90 and 0.88 for K=2, respectively), while BR showed a more distinct profile (0.93 for K=1 green color).

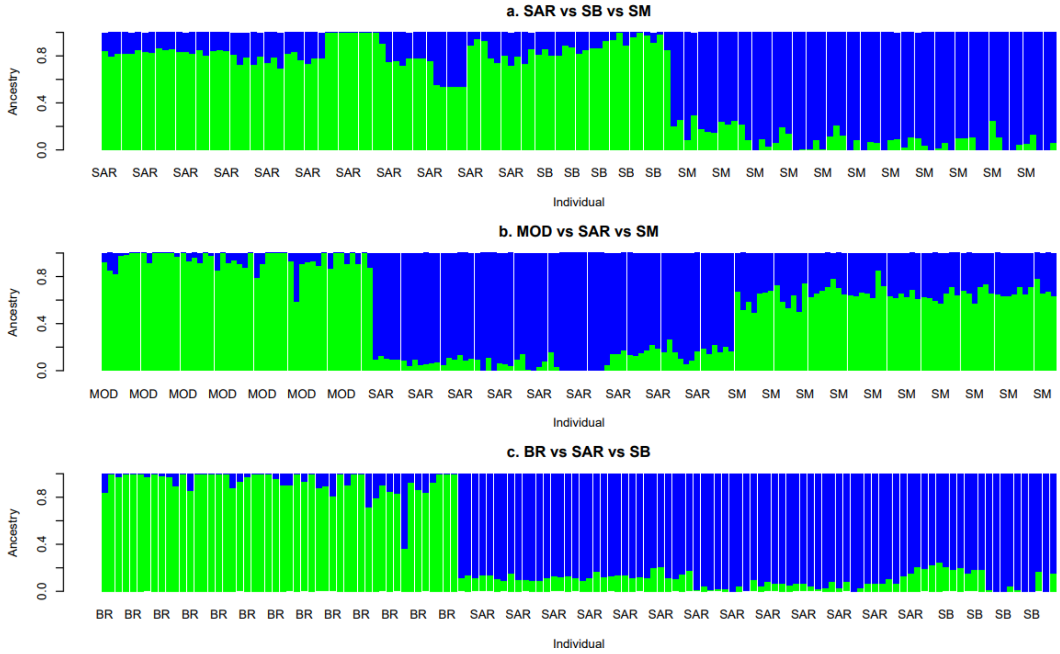


Figure 2. Admixture plots of the admixture analyses results at K = 2 in the five analyzed cattle breeds: Bruna (BR), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

A clear separation among the breeds was observed along the first principal component (PC1), which explained about 4.2% of the total variance (Figure 3).

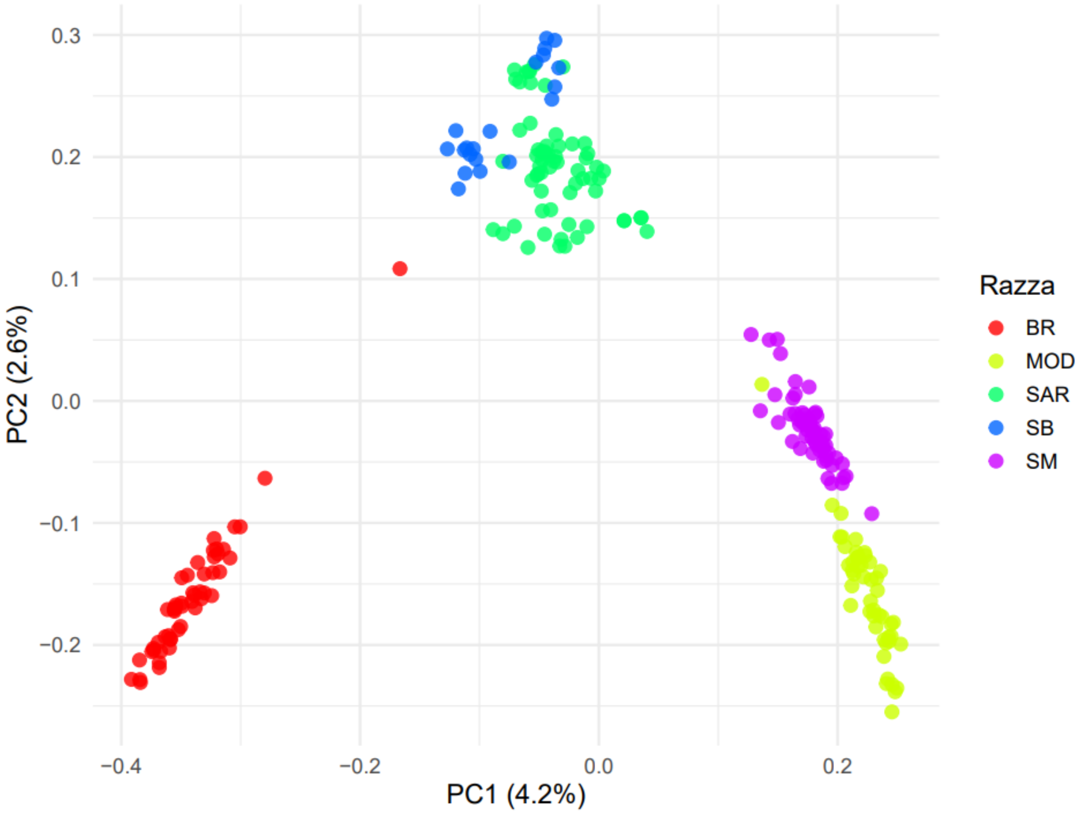


Figure 3. Plot of the Principal component analysis (PCA) based on the genomic relationship matrix of the five investigated breeds: Bruna (BR), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

Along this axis, individuals from BR cluster (red points) were plotted at the left extreme, whereas MOD and SM were positioned on the right (yellow and fuchsia points, respectively). SAR and SB breeds (blue and green points, respectively) occupied intermediate positions. The second component (PC2), explaining 2.6% of the total variance, further corroborated the differentiation among the investigated breeds, with SB close to SAR and SM close to MOD, confirming their closer genetic similarity. Estimates of average genomic relationships among breeds revealed generally low values, confirming clear genetic differentiation across groups. The highest mean relationship was observed between SAR and SB (0.026), whereas the lowest was detected between SM and SB (mean = -0.027). A positive value was also found for the SM vs MOD comparison (mean = 0.013), indicating a slightly higher degree of genomic similarity compared with other breed pairs. The remaining comparison showed values close to zero, suggesting limited genomic relatedness among the analyzed populations (Table 1).

Table 1. Genomic relationship results obtained from the genomic kinship matrix and the weighted Wright's fixation index (FST) analysis for the pairwise comparisons considered in the five analyzed cattle breeds: Bruna (BR), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

Comparison	Genomic relationship	FST
BR vs SB	-0.011	0.058
MOD vs SM	0.013	0.027
SAR vs SB	0.026	0.015
SAR vs SM	-0.021	0.030

Linkage disequilibrium and effective population size

The decay of linkage disequilibrium (LD) showed clear differences among the five breeds analyzed (Figure 4).

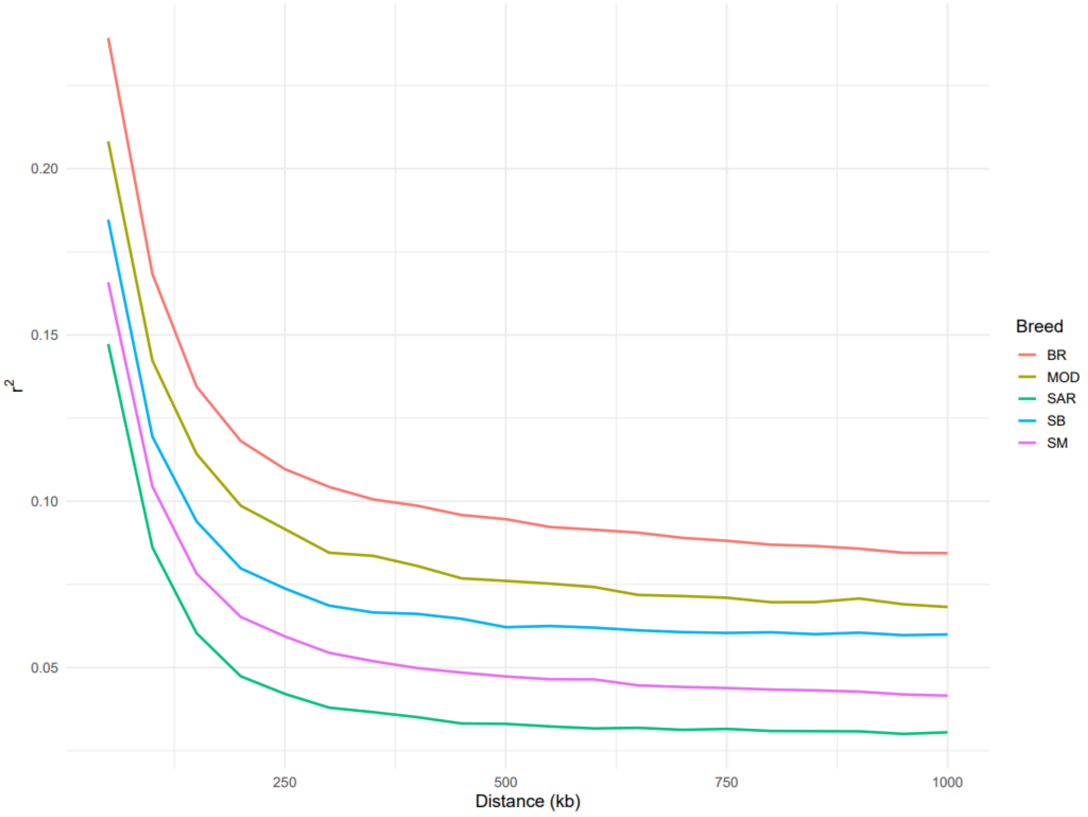


Figure 4. Average linkage disequilibrium (LD; r^2) between markers within an interval of 1000 kb in the five analyzed cattle breeds Bruna (BR), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

As expected, BR displayed the highest levels of LD across all distance classes. MOD showed intermediate levels of LD, whereas the three Sardinian local breeds (SAR, SB, and SM) exhibited lower overall LD values, with SAR showing the fastest decay of LD, followed by SB and SM.

Estimates of effective population size (Ne) revealed consistent patterns with the LD decay profiles (Figure 5).

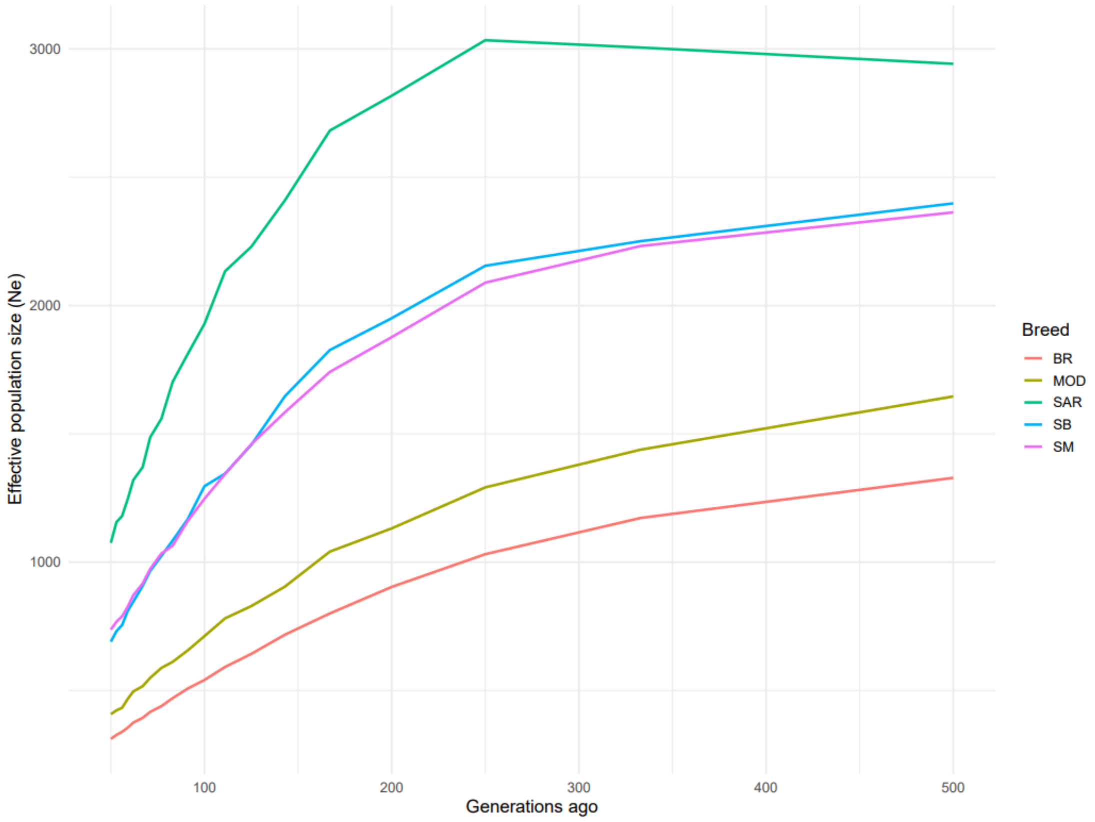


Figure 5. Effective population size based on LD according to the generation ago considered in the five analyzed cattle breeds Bruna (BR), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

The SAR breed exhibited the highest Ne across generations, followed by SB and SM, while MOD and BR showed the lowest values. For the three Sardinian local breeds, Ne estimated using the equation of Falconer and MacKay (1996), showed that SAR had the largest effective population size (Ne = 4449), closely followed by the SB (Ne = 3823), whereas the SM had a substantially lower Ne (Ne = 186).

Wright fixation Index

The FST values for all the pairwise comparisons are reported in Table 1. The highest weighted FST value (0.058) was observed between BR vs SB and between SAR vs

SM (0.030). The lowest value was instead found between SAR vs SB (0.015). Applying the FST_{LOW} threshold to the initial dataset of 35,057 SNP yielded 16,048 SNP in the SAR vs SB comparison and 15,598 in the SAR vs SM comparison. These SNP represent genomic variants with significantly reduced genetic differentiation and were thus considered as candidate markers of conserved genomic regions between these Sardinian breeds.

Runs of Homozygosity

Basic statistics of ROH analyses are reported in Table 2.

Table 2. Basic statistics and results of ROH statistical analyses in the five analyzed cattle breeds: Bruna (BR), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

	Animals		ROH				ROH class^3					
	N	%	N	%	nROH^1	meanMb^2	<2Mb	2 - 4 Mb	4 - 8 Mb	8 - 16 Mb	> 16 Mb	F_{ROH}
SAR	64	27	3,241	21	55.26±29.22^{BC}	3.9±5.97^{A}	60	17	11	8	4	0.080±0.085
SB	20	8	1,107	7	56.44±26.03^{BC}	3.36±4.96^{B}	58	22	11	5	3	0.075±0.064
SM	57	24	3,343	22	61.43±19.12^{C}	3.29±4.68^{B}	58	23	12	6	2	0.078±0.053
BR	50	21	4,109	27	81.23±12.41^{A}	4.07±4.9^{A}	47	25	15	9	4	0.135±0.001
MOD	48	20	3,477	23	72.47±17.49^{AB}	3.46±4.57^{B}	55	22	14	6	2	0.101±0.044

A total of 15,277 ROH were identified across 239 (64 SAR, 20 SB, 57 SM, 50, BR and 50 MOD) animals, all of which had at least five ROH. In terms of distribution among the breeds, BR, MOD, and SM exhibited larger number of ROH accounting for 72% of the total identified ROH. In contrast, SB displayed the lowest number, with only 1,107 ROH, representing just 7% of the total. Regarding the length of the ROH, SM showed the lowest average ROH length (33.29±4.68 Mb). SAR and SB exhibited a significantly similar average number of ROH per animal (55.26±29.22 and 56.44±26, respectively), whereas BR showed the highest number of ROH (81.23 ± 12.41^{A}) followed by MOD (72.47 ± 17.49^{AB}) and SM (61.43 ± 19.12^{C}). Regarding the ROH length classes, all breeds exhibited a higher proportion of ROH segments shorter than 2 Mb and a very similar distribution in the other classes (Table 2). The highest ROH-based inbreeding coefficient (F_{ROH}) was computed for BR (0.135±0.001), whereas the lowest and similar values were observed for the three Sardinian breeds (0.080±0.085 for SAR, 0.075 ± 0.064 for SB and 0.078±0.053 for SM).

The results of the shared ROH_{REP} analysis are presented in Supplementary Table 2. Four genomic regions—located on BTA4, BTA14, BTA28, and BTA29—were identified as shared by at least one animal of the three Sardinian breeds, encompassing a total of 95 SNP. In the pairwise comparisons involving SAR vs SB or SAR vs SM, 32 and 40 genomic regions were detected in at least one animal per breed, respectively. The analysis based on SNP_{ROH} values allowed the identification of SNP frequently included within ROH in individuals of SAR, SB, and SM. A Total of eleven markers, 6 on BTA1 and 5 on BTA11, had a SNP_{ROH} larger than 97th of its distribution in the three breeds.

The multivariate approach

The SDA and CDA were applied to assess pairwise similarities between SAR vs SB and SAR vs SM. Of the 16,048 and 15,598 SNP previously identified in the univariate approach, the SDA retained 74 and 79 SNP, in the two comparisons, respectively (Supplementary Table 3). These SNP were considered as those that most strongly unified the breeds involved in the two comparisons. This is because the CDA was not able to significantly separate breeds (Hotelling's T-square test > 0.05) when these SNP were used, with an error rate in assigning animals to the correct breed of around 20%. This result ultimately confirms that only common SNP remained among all three breeds. Notably, among the 74 SNP identified between SAR vs SB, one SNP on BTA4, two on BTA9, and one on BTA11 overlapped with the shared ROH$_{REP}$ detected in the same comparison (Table 3). Similarly, of the 79 SNP found between SAR and SM, one SNP on BTA5 and one on BTA19 were located within the corresponding shared ROH$_{REP}$ comparison (Table 3).

Table 3. Genomic regions and potentially associated genes found in common between CDA and shared ROH approaches in SAR vs SB and SAR vs SM comparisons (SAR, Sarda; SB, Sardo-Bruna; SM, Sardo-Modicana).

Comparison	BTA	SNP	Region (Mb)	Gene Acronym	Gene name	Associated trait	Reference
SAR vs SB	4	BTA-90533-no-rs	1.74-2.24	-	-	-	-
	9	ARS-BFGL-NGS-32882	102.28-102.39	AFDN	Afadin, Adherens Junction Formation Factor	-	-
	9	ARS-BFGL-NGS-32882	102.41-102.44	KIF25	Kinesin Family Member 25	-	-
	9	ARS-BFGL-NGS-32882 ARS-BFGL-NGS-115046	102.45-102.46	FRMD1	FERM Domain Containing 1	-	-
	9	ARS-BFGL-NGS-32882 ARS-BFGL-NGS-115046	102.55-102.56	DACT2	Dishevelled Binding Antagonist Of Beta Catenin 2	Blood β-Hydroxybutyrate	Wang et al., (2024)

	9	ARS-BFGL-NGS-32882 ARS-BFGL-NGS-115046	102.65-102.80	SMOC2	SPARC Related Modular Calcium Binding 2	Fertility	Höglund et al., (2015)
	11	Hapmap42798-BTA-109203	54.80-55.99	CTNNA2	Catenin Alpha 2	Milk and fat yield Trypanotolerance Tick resistance Adaptation Fertility	Laodim et al., (2023) Kim et al., (2017) Otto et al., (2018) Flori et al., (2019) Fonseca et al., (2020)
SAR vs SM	5	Hapmap51303-BTA-74377	82.98-83.57	ITPR2	Inositol 1,4,5-Trisphosphate Receptor Type 2	Milk production Milk and fat yield Tail length Adaptation	Gangwar et al., (2025) Laodim et al., (2023) Wang et al., (2025) Terefe, (2018)
	19	ARS-BFGL-NGS-79203	28.35-28.43	PIK3R5	Phosphoinositide-3-Kinase Regulatory Subunit 5	Adaptation	Wu et al., (2024)
	19	ARS-BFGL-NGS-79203	28.45-28.65	NTN1	Netrin 1	Adaptation	Del Corco et al., (2021)
	19	ARS-BFGL-NGS-79203	28.66-28.87	STX8	Syntaxin 8	-	-
	19	ARS-BFGL-NGS-79203	28.87-28.89	CFAP52	Cilia And Flagella Associated Protein 52	Fertility	Liang et al., (2024)

	19	ARS-BFGL-NGS-79203	28.90-28.95	USP43	Ubiquitin Specific Peptidase 43	Adaptation (sheep)	Mohamadipoor Saadatabadi et al., (2021)

Gene Discovery

Among the SNP_{ROH} found in the comparison involving SAR, SB and SM, a total of 21 genes (of which 15 located on BTA1 and 6 located on BTA11, Table 4) and 38 QTL (14 on BTA1 and 24 on BTA11, Supplementary Table 5) were identified. In the genomic regions highlighted both by univariate and multivariate approach, 6 genes (5 on BTA 9 and 1 on BTA11, Table 3) and 37 QTL (Supplementary Table 4) were found between SAR vs SB whereas 6 genes (1 on BTA5 and 5 on BTA19, Table 3) and 40 QTL (Supplementary Table 4) were found between SAR vs SM.

Table 4. Genomic regions, percentage of SNP_{ROH} in each breed (% SNP_{ROH}) and potentially associated genes found in SNP_{ROH} among the three Sardinian cattle breeds (SAR, Sarda; SB, Sardo-Bruna; SM, Sardo-Modicana).

			% SNP_{ROH}			Gene			
CHR	SNP	Region (Mb)	SAR	SB	SM	Acronym	Name	Associated trait	References
1	ARS-BFGL-NGS-116572	82.97-82.98	27	25	19	AP2M1	Adaptor Related Protein Complex 2 Subunit Mu 1	Milk and fat yield	Laodim et al., (2023)
1	ARS-BFGL-NGS-116572	82.98-83.00	27	25	19	DVL3	Dishevelled Segment Polarity Protein 3	Milk and fat yield	Laodim et al., (2023)
1	ARS-BFGL-NGS-116572	83.00-83.01	27	25	19	EIF2B5	Eukaryotic Translation Initiation Factor 2B Subunit Epsilon	Liver functional genomics	Laporta et al., (2014)
								Age at first calving	Dubon et al., (2021)
1	ARS-BFGL-NGS-116572	83.06-83.07	27	25	19	HTR3C	5-Hydroxytryptamine Receptor 3C	Age at first calving	Dubon et al., (2021)
								Milk production	Korkuc et al., (2023)

1	ARS-BFGL-NGS-116572	83.12-83.21	27	25	19	ABCC5	ATP Binding Cassette Subfamily C Member 5	-	-
1	ARS-BFGL-NGS-116572	83.25-83.30	27	25	19	PARL	Presenilin Associated Rhomboid Like	-	-
1	ARS-BFGL-NGS-116572	83.30-83.31	27	25	19	MAP6D1	MAP6 Domain Containing 1	Milk production	Korkuc et al., (2023)
1	ARS-BFGL-NGS-116572	83.31-83.42	27	25	19	YEATS2	YEATS Domain Containing 2	Milk production	Korkuc et al., (2023)
1	ARS-BFGL-NGS-116572	83.45-83.48	27	25	19	KLHL24	Kelch Like Family Member 24	Reproduction	Pal et al., (2023)
1	Hapmap41527-BTA-38460	83.55-83.60	27	25	19	KLHL6	Kelch Like Family Member 6	Reproduction	Yang et al., (2022)
1	Hapmap40527-BTA-38457	83.65-83.91	30	25	19	MCF2L2	MCF.2 Cell Line Derived Transforming Sequence-Like 2	-	-
1	Hapmap52416-rs29016842	83.75-83.83	33	25	19	B3GNT5	UDP-GlcNAc:BetaGal Beta-1,3-N-Acetylglucosaminyltransferase 5	Immunity (pig) Milk quality	Pal et al., (2019) Poulsen et al., (2019)
1	BTB-00036444	83.92-83.95	31	25	19	LAMP3	Lysosomal Associated Membrane Protein 3	-	-

1	BTB-00036444	83.96-84.02	31	25	19	MCCC1	Methylcrotonyl-CoA Carboxylase Subunit 1	Meat quality	Song et al., (2024)
1	BTB-01702174	84.04-84.07	31	25	19	DCUN1D1	Defective In Cullin Neddylation 1 Domain Containing 1	Milk yield and milk quality	Nayak et al., (2023)
								Sire conception rate	Li et al., (2012)
11	Hapmap59053-rs29015825	59.86-60.06	28	25	19	USP34	Ubiquitin Specific Peptidase 34	Body conformation	Ogunbawo et al., (2025)
11	Hapmap32032-BTA-128772	60.14-60.18	28	25	19	XPO1	Exportin 1	Body conformation	Ogunbawo et al., (2025)
11	BTA-99888-no-rs	60.43-60.44	28	25	19	FAM161A	FAM161 Centrosomal Protein A	Body conformation	Ogunbawo et al., (2025)
								Adaptation	Zhong et al., (2024)
11	BTA-99888-no-rs	60.47-60.48	28	25	19	CCT4	Chaperonin Containing TCP1 Subunit 4	Body conformation	Ogunbawo et al., (2025)
11	Hapmap47238-BTA-30090	60.50-60.67	28	25	21	COMMD1	Copper Metabolism Domain Containing 1	Body conformation	Ogunbawo et al., (2025)

								Finishing precocity	Surati et al., (2024)
								Adaptation (buffalo)	Zhong et al., (2024)
								Immunity	
11	ARS-BFGL-BAC-11723	60.72-60.76	26	30	19	B3GNT2	UDP-GlcNAc:BetaGal Beta-1,3-N-Acetylglucosaminyltransferase 2	Body conformation	Ogunbawo et al., (2025)
								Adaptation (buffalo)	Surati et al., (2024)

Discussion

Local breeds represent an important component of the zootechnical and cultural heritage of the regions where they originated. Despite their remarkable resilience and deep-rooted adaptation to local environments, shaped by specific ecological conditions and traditional practices (Felius et al., 2014), many native breeds face the threat of extinction. This is largely due to the widespread introduction of high-yield cosmopolitan breeds, which have progressively displaced local populations (Curone, 2019). One of the major dangers is the genetic erosion, driven by the crossbreeding with more productive non-native cattle (Bigi & Zanon, 2008). Preserving and promoting autochthonous breeds goes beyond safeguarding biodiversity and it also means protecting a distinctive cultural legacy (Nonić et al., 2021). While cosmopolitan breeds offer higher economic returns, they are often ill-suited to the particular demands of marginal areas. In this context, Sardinian autochthonous breeds represent an exemplary case of livestock populations genetically adapted to marginal environments, whose survival is, however, increasingly threatened by the widespread diffusion of cosmopolitan breeds. It is therefore of paramount importance to support the conservation and valorization of these autochthonous cattle, ensuring both their genetic continuity and their role within the cultural and economic identity of Sardinia. Admixture analysis is a population genetics method used to estimate the proportion of ancestry of individuals or populations from a presumed number of ancestral populations (Kumar et al., 2003). In this study, admixture analysis was employed to assess genetic similarities between Sarda cattle, two composite (SB and SM), and the other two purebreds (BR and MOD) used in the crossbreeding schemes. In the analysis, the number of ancestral population (i.e., the K parameter) was set to 2, consistently with the idea that genetically similar populations tend to show comparable proportions of ancestry components inferred at K=2 (Kumar et al., 2003). Admixture analyses results showed that SB is more strongly related with SAR than with BR, whereas SM resulted to be genetically closer to MOD rather than SAR (Figure 2a, 2b and 2c). Similar results were reported by Mastrangelo et al. (2018), where admixture analyses provided valuable insights into the genetic composition of Sardinian cattle breeds: at lower K values, these breeds tended to share a common ancestral component with autochthonous Sicilian cattle breeds, indicating a degree of shared ancestry or historical gene flow. On the other hand, the Sardinian breed, at K=24, exhibited a less distinct cluster compared to other breeds, suggesting a more complex genetic background potentially shaped by past admixture events. As a result, SAR shares genetic components with both SB and SM, but in different proportions. However, other studies showed how SAR tends to be closer to SB rather than SM when analyzed with BR and MOD (e.g., Cesarani et al., 2018). Determining the population structure of Sardinian cattle breeds provided useful insights into their genetic relationships The PCA based on the genomic relationship matrix separated BR from the other populations along PC1, reflecting its well-known divergence from local breeds (Cesarani et al., 2018; Mastrangelo et al., 2018). In agreement with previous reports, MOD and SM clustered together, supporting the historical preservation of MOD genetic component in SM (Cesarani et al., 2018; Mastrangelo et al., 2018). Interestingly, SAR and SB tend to overlap in the PCA space, contrary to earlier studies that reported a clearer separation between SB and SAR (Cesarani et al., 2018; Mastrangelo et al., 2018). This result, probably influenced by different sampling procedures than earlier studies, suggests that SB retains a substantial genetic background shared with SAR, possibly due to local breeding practices or more recent genetic exchanges. Overall, our findings support a complex genetic landscape in which both distinct differentiation (BR vs. other breeds) and shared ancestry (SAR and SB, SM and MOD) coexist, consistent with the historical management of these local cattle populations.

The results of LD and effective population size analyses highlighted differences between local and cosmopolitan populations. Sardinian breeds showed lower LD consistent with a larger effective population size, mainly due to the use of natural matings (i.e., no artificial insemination) and the limited exchange of bulls among farms. In this context, Sardinian breeds, even when compared to other local and non-local Italian breeds, tend to show a higher Ne (Mastrangelo et al., 2018).

The Wright fixation index is a key tool in population and evolutionary genetics research, as it provides valuable insights into the evolutionary processes that have shaped genetic variation both within and between populations. Additionally, FST is closely linked to the variance in allele frequencies across populations, offering important information regarding the degree of genetic similarity among individuals (Holsinger and Weir, 2009). FST weighted values estimates among the considered populations confirmed what was already observed with Admixture: SB seems to be more genetically closer to SAR rather than to BR. The FST value between SAR and SB (0.015) closely aligns with the findings of Mastrangelo et al. (2018), who reported a value of 0.016. However, in the comparisons between SAR vs SM and SM vs MOD, the FST values obtained in this study (0.030 for SAR vs SM and 0.027 for SM vs MOD, respectively) were slightly lower than those obtained by Mastrangelo et al.

(2018) (0.040 for SAR vs SM and 0.025 for SM vs MOD, respectively). These discrepancies could be due to differences in sample size or allele calling methodologies, that may influence FST estimates and account for the slight discrepancies observed between the two studies (Holsinger & Weir, 2009).

ROH are uninterrupted stretches of homozygous genotypes that originate from common ancestors, indicating regions identical by descent (Gibson et al., 2006). The length of these segments can offer insight into the timing of inbreeding events (Gibson et al., 2006.), whereas their pattern can also reflect selective pressures, both natural and artificial, as selection can lead to the fixation of beneficial alleles, thereby increasing homozygosity at specific loci (Szmatola et al., 2019; Falchi et al., 2023). In the present study, all breeds exhibited a higher concentration of ROH in the short and medium length classes, indicating that most inbreeding events likely occurred in the more distant past (Gibson et al., 2006); comparable results were observed in the same breeds by Cesarani et al. (2018). The inbreeding coefficients (FROH) were relatively similar among the three Sardinian breeds – SAR (0.080 ± 0.085), SB (0.075 ± 0.064), and SM (0.078 ± 0.053) – suggesting comparable levels of autozygosity and inbreeding history (Mastrangelo et al., 2018). In contrast, BR and MOD showed higher FROH value (0.135 ± 0.001 for BR and 0.101 ± 0.044 for MOD), reflecting a lower genetic diversity. Similar ROH-based inbreeding coefficients were reported by Mastrangelo et al., (2018) for MOD cattle (0.105 ± 0.067), whereas slightly lower values were reported by Signer-Hasler et al. (2017) in Brown Swiss cattle (0.115 ± 0.037). The current results are broadly consistent with previous studies, in which the Sarda breed was reported to have a low mean FROH (0.060 ± 0.063), reflecting a larger within-breed genetic variability (Mastrangelo et al., 2018). Likewise, moderate inbreeding levels were already reported in literature for SB and SM (Mastrangelo et al., 2018).

The analysis of ROH across breeds or individuals provides valuable insights into population genetics and animal breeding, helping to identify genomic regions under selection or linked to important traits (Peripolli et al., 2017; Szmatoła et al., 2019; Falchi et al., 2022). The results obtained in this study highlighted shared genomic regions (ROH_{REP}) and genetic markers (SNP_{ROH}) shared among Sardinian breeds, representing candidate loci potentially underlying common adaptive traits or breed-specific selection. Notably, both shared ROH_{REP} and SNP_{ROH} analyses revealed homozygous signals exclusive to the Sardinian populations, i.e., absent in BR and MOD, supporting the concept of a “Sardinity” genetic signature. This approach diverges from the majority of previous studies conducted on these local breeds, which typically focus on identifying genetic differences rather than similarities (e.g., Cesarani et al., 2018; Mastrangelo et al., 2018; Fabbri et al., 2022). Instead, here we emphasize the common genomic features that unify the Sardinian cattle, providing novel insights into the shared evolutionary histories and potential adaptive convergence. Complementing this univariate approach, the multivariate analysis using SDA and CDA in the SNP previously identified with FST_{low} in the pairwise comparisons involving SAR vs SB and SAR vs SM, led to the identification of SNP subsets representing shared genomic components. CDA, when SNP selected at the end of the SDA runs were used, was not able to significantly separate breeds (Hotelling’s T-square test > 0.05) with 20% of misclassification rate. This result suggests that these markers largely capture common genetic features across the breeds, possibly reflecting shared ancestry, gene flow, or convergent selection. Remarkably, both methodologies independently highlighted genomic regions (Table 3) where markers were consistently retained and mapped within shared ROH segments between SAR vs SB and SAR vs SM. Their consistent detection across distinct analytical frameworks reinforces its biological relevance and points to the need for further investigation. Overall, this convergence of univariate and multivariate evidence highlights the power of integrating complementary methods to identify robust genomic signatures with implications for breed conservation and genetic improvement strategies.

In this study, population stratification and genetic diversity and similarity in this study were investigated using different approaches based on SNP analysis. A more comprehensive approach, based on haplotypes, might deliver clearer insights on past crossbreeding events and proportion of SAR genome still conserved in SB and SM. However, most of the haplotypes-based analyses require information on ancestral populations and reference haplotypes (e.g., phased genotypes) that were not available for the current study.

Gene and QTL discovery

Among the approaches employed to detect genomic regions shared across Sardinian breeds, the SNP-based method ($\mathrm{SNP_{ROH}}$) deserves particular attention as it allows the identification of broader and more biologically meaningful regions of homozygosity (Peripolli et al., 2018). As already pointed out, the three breeds under investigation in this study (SAR, SB, and SM) are characterized by a high level of resilience and adaptation to harsh environments. Using this approach, in two genomic region highlighted among SAR, SB, and SM breeds, two genes, *COMMD1* (BTA11: 60.50–60.67 Mb) and *B3GNT2* (BTA11: 60.72–60.76 Mb) (Table 3), were previously identified in Chilika buffalo as part of the genetic signature associated with adaptation to saline environments (Surati et al., 2024). Chilika is an indigenous Indian buffalo breed adapted to brackish water conditions around the Chilika Lake, characterized by high salinity levels and saline vegetation. The *COMMD1* and *B3GNT2* genes were functionally linked to biological pathways such as MAPK signaling, renin secretion, and endocytosis, all potentially contributing to physiological resilience in high-salinity habitats (Surati et al., 2024). Notably, *COMMD1* has also been identified as a candidate gene involved in oxidative stress response and anti-inflammatory regulation in cattle (Zhong et al., 2024). Specifically, *COMMD1* plays a key role in modulating the NF-κB signaling pathway, which is central to immune response and inflammation control. Reduced levels of COMMD1 protein have been associated with prolonged activation of NF-κB, leading to increased expression of pro-inflammatory cytokines (Muller et al., 2007). By fine-tuning NF-κB activity, *COMMD1* may thus help maintain immune homeostasis under environmental stressors (Taye et al., 2017). *FAM161A* gene (BTA11: 60.43-60.44) has been reported as a candidate gene for environmental adaptation in cattle (Zhong et al., 2024). This gene plays a critical role in retinal structure and function, with its normal expression being essential for maintaining the integrity and performance of retinal cells (Häfliger et al., 2021). In tropically adapted breeds, such as the Hainan yellow cattle, the functional integrity of *FAM161A* is thought to contribute to environmental fitness by supporting effective vision — a key trait for navigating complex, wild habitats. Some QTL found in the regions highlighted by SNP_{ROH} among the three Sardinian cattle breeds were associated with traits of reproductive and adaptive relevance (Supplementary Table 5). In particular, QTL linked to fertility and reproduction were detected on BTA1 (non-return rate and birth index) and BTA11 (inseminations per conception). For health and resistance, QTL associated with tick resistance and bovine tuberculosis susceptibility were identified on BTA11.

For completeness, we also investigated the genomic regions identified by both univariate and multivariate approaches. These regions, as well as those identified by the SNP_{ROH} approach, revealed genes potentially associated with adaptation, fertility, and immunity (Table 4). In the SAR vs SB comparison, the gene *CTNNA2* (BTA11: 54.80–55.99 Mb) was found. In Mediterranean cattle breeds, this gene was previously reported within genomic regions under positive selection related to climate variables, including the Temperature-Humidity Index (Flori et al., 2019). These findings suggest a potential role for *CTNNA2* in thermoregulation and heat stress response, two key traits for adaptation to arid and semi-arid environments. Beyond its involvement in climate resilience, *CTNNA2* has also been associated with trypanotolerance (Kim et al., 2017), tick resistance (Otto et al., 2018), fertility (Fonseca et al., 2020), and milk and fat yield in tropical dairy cattle (Laodim et al., 2023). The convergence of multiple lines of evidence supports the hypothesis that *CTNNA2* plays a pleiotropic role in conferring resilience to environmental stressors while contributing to productive and reproductive performance in poor environments. In the SAR vs SM comparison, three genes—*NTN1* (BTA1: 28.45-28.65 bp), *PIK3R5* (BTA19: 28.35-28.43 BP), and *USP43* (BTA19: 28.90-28.95 bp)— all of which previously associated with adaptation to environmental stressors, were identified. The *NTN1* (Netrin 1) gene was reported by Del Corvo et al. (2021) as differentially methylated in response to heat stress in Nellore and Angus cattle. Specifically, it was hypomethylated in Nellore individuals during periods of thermal challenge, suggesting a potential role in enhancing gene expression under stress conditions. The *PIK3R5* gene was identified in Hainan cattle by Hui et al. (2024) within a region under positive selection. A missense mutation in this gene was associated with altered mRNA and protein structures, potentially modulating the PI3K/Akt/mTOR pathway. This signaling cascade plays a central role in cellular metabolism and stress response, and thus, its regulation may contribute to improved heat tolerance in cattle from tropical regions. Lastly, the *USP43* gene was found under selection in Iranian sheep breeds adapted to high temperatures (Mohamadipoor Saadatabadi et al., 2021). One fertility related gene, *CFAP52* (BTA19: 28.90-28.95) was also found (Liang et al., 2024). *ITPR2* gene (BTA5: 82.98–83.97 Mb was reported to play a role in calcium ion signaling pathways, which are involved in several adaptive and physiological processes (Taye, 2018; Wang et al., 2025). Notably, *ITPR2* has been identified as a gene under positive selection in African cattle, potentially contributing to thermal tolerance through the regulation of eccrine sweat gland function (Taye, 2018). The protein encoded by *ITPR2* modulates intracellular calcium release, a process critical for sweat secretion. Loss-of-function mutations in this gene have been associated with anhidrosis, indicating its essential role in thermoregulation (Klar et al., 2014; Cui and Schlessinger, 2015). In Brown Swiss cattle, *ITPR2* has been implicated in skeletal traits such as tail length, with enriched functional annotations in calcium ion binding (GO:0005509), suggesting involvement in bone development and remodeling (Wang et al., 2025). Furthermore, it was found within the cAMP-mediated signaling pathway (GO:0071320), which is associated with anti-inflammatory responses (Tavares et al., 2020). In addition to gene-based signals, several QTL identified within the overlapping ROH$_{\text{REP}}$-CDA regions were associated with traits of economic and adaptive importance

(Supplementary Table 4). On BTA1, QTL linked to growth and feed efficiency were found between 83.49–83.96 Mb (average daily gain, residual feed intake, methane production), while QTL for calving ease and somatic cell score clustered at 84.06 Mb. On BTA11, some QTL between 59.73–60.86 Mb were related to fertility (inseminations per conception) and health (tick resistance, susceptibility to bovine tuberculosis).

The identification of genes and QTL associated with adaptation, fertility and immunity, reinforces the long-standing hypothesis that Sardinian cattle have undergone natural and human-driven selection for resilience in marginal environments. In particular, the Sarda breed's ability to thrive on poor pastures and under low-input systems may have favored the retention of genetic variants linked to adaptive features, disease resistance and reproductive efficiency. These results are consistent with previous observations on the adaptability of local breeds across the Mediterranean, where extensive farming systems and pasture-based diets have shaped genetic profiles associated with improved meat and milk quality (e.g., fatty acid composition, kappa-casein content), metabolic efficiency, and reduced susceptibility to stress (Acciaro, 2013, Gatellier et al., 2005; Curone et al., 2016; Mastrangelo et al., 2018; Stanton et al., 2021; Alothman et al., 2019; Magan et al., 2021). Moreover, the co-occurrence of production-related QTL (growth, feed efficiency, milk yield) and adaptive signals in the same genomic regions suggests that local breeds such as the Sarda may combine modest productivity with functional robustness—a valuable trait combination in the context of sustainable and low-input agriculture.

Conclusion

This study confirms the retention of a substantial genomic background from the Sarda (SAR) breed in the derived Sardo-Bruna (SB) and Sardo-Modicana (SM) breeds. Admixture patterns and shared ROH regions suggest a clear genetic continuity, further supported by the identification of selective signatures associated with traits of environmental adaptation, immunity, and productive performance. The combined use of univariate and multivariate methods proved to be essential in capturing both locus-specific differentiation and broader genomic patterns, enhancing the resolution and reliability of the results. These findings highlight the adaptive potential embedded within local cattle breeds and their relevance for sustainable livestock systems, particularly in marginal environments. Indigenous breeds like SAR, SB, and SM are not only a reservoir of unique genetic diversity, but also represent a cultural and ecological asset, deeply connected to the Sardinian territory. The ongoing genetic erosion and risk of extinction of these breeds pose serious threats, not only to biodiversity but also to the resilience of local agropastoral systems and to the environmental balance. Therefore, conservation and valorization strategies for local breeds should be prioritized within broader frameworks of sustainable agriculture and rural development.

Acknowledgements

This study was supported by the project "*Conoscenza e gestione sostenibile dei sistemi agricoli e forestali con il miglioramento sostenibile delle produzioni primarie: il caso dell'allevamento bovino in Sardegna*" (University of Sassari, progetti di ricerca interdisciplinare – DM 737/2021, risorse 2021–2022; Finanziato dall'Unione Europea, Next-GenerationEU) and by the project "*Analisi genomica con piattaforme SNP ad alta densità e sequenziamento del DNA genomico e mitocondriale del bovino di razza Sarda ai fini dello studio della variabilità genetica e storia evolutiva*" of the Fondazione Banco di Sardegna (2016) . This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 7 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

Acciaro, M. (2013). Role of pasture in the growth, intake and meat quality of Sardinian cattle breeds PhD thesis, University of Sassari, Sassari, Italy.

Ajmone-Marsan P, Negrini R, Crepaldi P, Milanesi E, Gorni C, Valentini A et al. (2001). Assessing genetic diversity in Italian goat populations using AFLP® markers. *Animal Genetics* 32(5): 281–288. doi: 10.1046/j.1365-2052.2001.00789.x

Alexander, D. H., Shringarpure, S. S., Novembre, J., and Lange, K. (2015). Admixture 1.3 software manual. Software manual, UCLA Human Genetics Software Distribution, Los Angeles, CA, United States.

Alothman, M., Hogan, S. A., Hennessy, D., Dillon, P., Kilcawley, K. N., O'Donovan, M., and O'Callaghan, T. F. (2019). The "grass-fed" milk story: Understanding the impact of pasture feeding on the composition and quality of bovine milk. Foods 8:350. doi: 10.3390/foods8080350

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-5. doi:10.1093/bioinformatics/bth457.

Bigi, D. and Zanon, A. (2008). Atlante delle razze autoctone: Bovini, Equini, Ovicaprini, Suini allevati in Italia. Edagricole, Bologna, Italy.

Biodiversità Sardegna (2025). Vacca Sardo-Bruna. Available online at: https://www.biodiversitasardegna.it/laore/it/agrobiodiversita/repertorio-regionale/risorsa/Vacca-sardo-bruna/ (Accessed January 31, 2025).

Biscarini, F., Cozzi, P., Gaspa, G., and Marras, G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version

0.9.6. Available online at: https://CRAN.R-project.org/package=detectRUNS (accessed July 31, 2025).

Brandano, P., Asara, P., Pulina, G., Bolla, P., and Crimella, C. (1983). The Sardinian cattle. 1. Morphological and biological characters. Ann. Fac. Agric. Univ. Sassari 30:161–177.

Braunvieh Switzerland (2025). Braunvieh Switzerland homepage. Available online at: https://homepage.braunvieh.ch/en/ (Accessed January 31, 2025).

Cesarani A, Sorbolini S, Criscione A, Bordonaro S, Pulina G, Battacone G, et al. Genome-wide variability and selection signatures in Italian island cattle breeds. Anim Genet. 2018;49(5):371-83.

Cui, C. Y. and Schlessinger, D. (2015). Eccrine sweat gland development and sweat secretion. Exp. Dermatol. 24:644–650. doi: 10.1111/exd.12761

Curone, G. (2019). Physiological approach to the study of autochthonous cattle breeds of Northern Italy. PhD thesis, Department of Veterinary Medicine (DIMEVET), University of Milan, Milan, Italy.

De Haan C, Steinfeld H and Blackburn H (1997). *Livestock and the environment: Finding a balance.* Rome, Italy: Food and Agriculture Organization of the United Nations, United States Agency for International Development, and The World Bank, 115 pp.

De Maesschalck R, Jouan-Rimbaud D and Massart DL (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems* 50(1): 1–18. doi: 10.1016/S0169-7439(99)00047-7

Del Corvo, M., Lazzari, B., Capra, E., Zavarez, L., Milanesi, M., Utsunomiya, Y. T., and Ajmone-Marsan, P. (2021). Methylome patterns of cattle adaptation to heat stress. Front. Genet. 12:633132. doi: 10.3389/fgene.2021.633132

Della Maria, G. (1936). Ancient stories of Sardinian bovine breed. Riv. Zootec. 13:47–57.

Dubon MAC, Pedrosa VB, Feitosa FLB, Costa RB, de Camargo GMF, Silva MR and Pinto LFB (2021). Identification of novel candidate genes for age at first calving in Nellore cows using a SNP chip specifically developed for *Bos taurus indicus* cattle. *Theriogenology* 173: 156–162. doi: 10.1016/j.theriogenology.2021.08.011

Edea Z, Dadi H, Kim SW, Park JH, Shin GH, Dessie T and Kim KS (2014). Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in Ethiopia. Journal of Animal Breeding and Genetics 131(5): 358–366. doi: 10.1111/jbg.12083

Fabbri MC, Dadousis C, Tiezzi F, Maltecca C, Lozada-Soto E, et al. (2022). Genetic diversity and population history of eight Italian beef cattle breeds using measures of autozygosity. *PLOS ONE* 17(10): e0248087. doi: 10.1371/journal.pone.0248087

Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Essex: Prentice Hall; 1996.

FAO (2024). DAD-IS: Domestic Animal Diversity Information System. Data on Italian goat breeds. Available at: https://fao.org/dad-is (Accessed in October 2024).

Felius, M., Beerling, M. L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., and Lenstra, J. A. (2014). On the history of cattle genetic resources. Diversity 6, 705–750. doi: 10.3390/d6040705.

Flori L, Moazami-Goudarzi K, Alary V, Araba A, Boujenane I, Boushaba N, et al. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology* 28(5): 1009–1029. doi: 10.1111/mec.15004

Fonseca, P. A., Suárez-Vega, A., and Cánovas, A. (2020). Weighted gene correlation network meta-analysis reveals functional candidate genes associated with high- and sub-fertile reproductive performance in beef cattle. Genes 11:543. doi: 10.3390/genes11050543

Gandini, G., Turri, F., Rizzi, R., Crotta, M., Minozzi, G., and Pizzi, F. (2017). Economic evaluation of genetic improvement in local breeds: The case of the Verzaschese goat. Ital. J. Anim. Sci. 16:199–207. doi: 10.1080/1828051X.2017.1279034

Gangwar, M., Kumar, S., Ahmad, S. F., Singh, A., Agarwal, S., A. P. L., and Kumar, A. (2025). Exploring genetic variants affecting milk production traits through genome-wide association study in Vrindavani crossbred cattle of India. Trop. Anim. Health Prod. 57:104. doi: 10.1007/s11250-025-04348-0

Gatellier, P., Mercier, Y., Juin, H., and Renerre, M. (2005). Effect of finishing mode (pasture-or mixed-diet) on lipid composition, colour stability and lipid oxidation in meat from Charolais cattle. Meat Sci. 69:175–186. doi: 10.1016/j.meatsci.2004.06.022

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

Häfliger, I. M., Marchionatti, E., Stengård, M., Wolf-Hofstetter, S., Paris, J. M., Jacinto, J. G., and Drögemüller, C. (2021). CNGB3 missense variant causes recessive achromatopsia in Original Braunvieh cattle. Int. J. Mol. Sci. 22:12440. doi: 10.3390/ijms222212440

Höglund JK, Guldbrandtsen B, Lund MS and Sahana G (2015). Identification of genomic regions associated with female fertility in Danish Jersey using whole genome sequence data. *BMC Genetics* 16: 60. doi: 10.1186/s12863-015-0210-3

Holsinger, K. E., and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat. Rev. Genet. 10, 639–650. doi: 10.1038/nrg2611

Hui W, Li S, Jiang J, Wang H, Wu G, Li L, et al. (2024). Genetic diversity of Hainan cattle population based on whole genome resequencing and screening of candidate genes related to heat adaptation. *Journal of Southern Agriculture* 55(12): 3707–3717.

Kim SJ, Ka S, Ha JW, Kim J, Yoo D, Kim K, et al. (2017). Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama. *BMC Genomics* 18(1): 371. doi: 10.1186/s12864-017-3742-2

Klar, J., Hisatsune, C., Baig, S. M., Tariq, M., Johansson, A., Rasool, M., Malik, N. A., et al. (2014). Abolished InsP3R2 function inhibits sweat secretion in both humans and mice. J. Clin. Invest. 124, 4773–4780. doi: 10.1172/JCI70720.

Korkuć, P., Neumann, G. B., Hesse, D., Arends, D., Reißmann, M., Rahmatalla, S., et al. (2023). Whole-genome sequencing data reveal new loci affecting milk production in German black pied cattle (DSN). Genes 14, 581. doi: 10.3390/genes14030581.

Kumar, P., Freeman, A. R., Loftus, R. T., Gaillard, C., Fuller, D. Q., and Bradley, D. G. (2003). Admixture analysis of South Asian cattle. Heredity 91, 43–50. doi: 10.1038/sj.hdy.6800277.

Laodim T, Koonawootrittriron S, Elzo MA, Suwanasopee T, Jattawa D and Sarakul M (2023). Genetic factors influencing milk and fat yields in tropically adapted dairy cattle: insights from quantitative trait loci analysis and gene associations. *Animal Bioscience* 37(4): 576. doi: 10.5713/ab.23.0246

Laporta J, Rosa GJ, Naya H and Carriquiry M (2014). Liver functional genomics in beef cows on grazing systems: novel genes and pathways revealed. Physiological Genomics 46(4): 138–147. doi: 10.1152/physiolgenomics.00155.2013

Li, G., Peñagaricano, F., Weigel, K. A., Zhang, Y., Rosa, G., and Khatib, H. (2012). Comparative genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. J. Dairy Sci. 95, 6122–6129. doi: 10.3168/jds.2012-5591.

Liang, H., Chen, W., Liu, X., Han, Y., Khan, A., Wang, C., and Khan, M. Z. (2024). Genetic polymorphisms in genes associated with mammalian semen quality traits: A review. Agriculture 14, 2137. doi: 10.3390/agriculture14122137.

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

Magan, J. B., O’Callaghan, T. F., Kelly, A. L., and McCarthy, N. A. (2021). Compositional and functional properties of milk and dairy products derived from cows fed pasture or concentrate-based diets. Compr. Rev. Food Sci. Food Saf. 20, 2769–2800. doi: 10.1111/1541-4337.12751.

Manca, E., Cesarani, A., Gaspa, G., Sorbolini, S., Macciotta, N. P., and Dimauro, C. (2020). Use of the multivariate discriminant analysis for genome-wide association studies in cattle. Animals 10, 1300. doi: 10.3390/ani10081300.

Mardia KV, Kent JT and Bibby JM (2000). *Multivariate analysis.* London, United Kingdom: Academic Press.

Mastrangelo, S., Ciani, E., Ajmone Marsan, P., Bagnato, A., Battaglini, L., Bozzi, R., et al. (2018). Conservation status and historical relatedness of Italian cattle breeds. Genet. Sel. Evol. 50, 1–16. doi: 10.1186/s12711-018-0406-x.

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, et al. (2007). Whole genome linkage disequilibrium maps in cattle. BMC Genetics 8(1): 74. doi: 10.1186/1471-2156-8-74

Mohamadipoor Saadatabadi, L., Mohammadabadi, M., Amiri Ghanatsaman, Z., Babenko, O., Stavetska, R., Kalashnik, O., et al. (2021). Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds. BMC Vet. Res. 17, 1–9. doi: 10.1186/s12917-021-03077-4.

Muller, P., van Bakel, H., van de Sluis, B., Holstege, F., Wijmenga, C., and Klomp, L. W. (2007). Gene expression profiling of liver cells after copper overload in vivo and in vitro reveals new copper-regulated genes. J. Biol. Inorg. Chem. 12, 495–507. doi: 10.1007/s00775-006-0201-y.

Nayak, S. S., Panigrahi, M., Rajawat, D., Ghildiyal, K., Sharma, A., Parida, S., et al. (2023). Comprehensive selection signature analyses in dairy cattle exploiting purebred and crossbred genomic data. Mamm. Genome 34, 615–631. doi: 10.1007/s00335-023-10021-4.

Nonić M and Šijačić-Nikolić M (2021). Genetic diversity: sources, threats, and conservation. In: *Life on Land*, 421–435. doi: 10.1007/978-3-319-95981-8_53

Ogunbawo, A. R., Hidalgo, J., Mulim, H. A., Carrara, E. R., Ventura, H. T., Souza, N. O., et al. (2025). Applying the algorithm for Proven and young in GWAS reveals high polygenicity for key traits in Nellore cattle. Front. Genet. 16, 1549284. doi: 10.3389/fgene.2025.1549284.

Otto, P. I., Guimarães, S. E. F., Varardo, L. L., et al. (2018). Genome-wide association studies for tick resistance in *Bos taurus* × *Bos indicus* crossbred cattle: A deeper look into this intricate mechanism. J. Dairy Sci. 101, 11020–11032. doi: 10.3168/jds.2017-14223.

Pal, A., and Chakravarty, A. K. (2019). Disease resistance for different livestock species. In *Genetics and Breeding for Disease Resistance of Livestock*, p. 271. doi: 10.1016/B978-0-12-816406-8.00019-X.

Pal, L., Cesarani, A., Mastrangelo, S., Senczuk, G., Portolano, B., Pilla, F., and Macciotta, N. P. P. (2023). Analysis of runs of homozygosity of cattle living in different climate zones. J. Anim. Sci. 101:skad061. doi: 10.1093/jas/skad061

Pal, P. (2023). Genome wide detection of selection signatures and association of SNP with stature and milk production traits in Karan Fries cattle [Doctoral dissertation]. National Dairy Research Institute.

Peripolli E, Stafuzza NB, Munari DP, Lima ALF, Irgang R, Machado MA, et al. (2018). Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genomics* 19(1): 34. doi: 10.1186/s12864-017-4413-z

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

Petretto E, Dettori ML, Pazzola M, Manca F, Amills M, Vacca GM. Mitochondrial DNA diversity of the Sardinian local cattle stock.. Sci Rep. 2022;12(1):2486. doi: 10.1038/s41598-022-06420-3.

Poulsen NA, Robinson RC, Barile D, Larsen LB and Buitenhuis B (2019). A genome-wide association study reveals specific transferases as candidate loci for bovine milk oligosaccharides synthesis. *BMC Genomics* 20(1): 404. doi: 10.1186/s12864-019-5786-y

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81(3): 559–575. doi: 10.1086/519795

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (accessed July 31, 2025)

Regione Sardegna (2007). Programma di sviluppo rurale 2007-2013 REG. (CE) N. 1698/2005. Available at: https://www.regione.sardegna.it/documenti/1_26_20090930155030.pdf (accessed May 15, 2024)

Rencher AC and Scott DT (1990). Assessing the contribution of individual variables following rejection of a multivariate hypothesis. *Communications in Statistics – Simulation and Computation* 19(2): 535–553. doi: 10.1080/03610919008812874

Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, et al. (2017). Population structure and genomic inbreeding in nine Swiss dairy cattle populations. *Genetics Selection Evolution* 49(1): 83. doi: 10.1186/s12711-017-0358-6

Song, X., Yao, Z., Zhang, Z., Lyu, S., Chen, N., Qi, X., et al. (2024). Whole-genome sequencing reveals genomic diversity and selection signatures in Xia'nan cattle. BMC Genomics 25, 559. doi: 10.1186/s12864-024-10463-3.

Stanton, C. A. T. H. E. R. I. N. E., Mills, S., Ryan, A., Di Gioia, D., and Ross, R. P. (2021). Influence of pasture feeding on milk and meat products in terms of human health and product quality. Ir. J. Agric. Food Res. 59, 292–302. Available at: https://www.jstor.org/stable/27115418.

Surati, U., Niranjan, S. K., Pundir, R. K., Koul, Y., Vohra, V., Gandham, R. K., and Kumar, A. (2024). Genome-wide comparative analyses highlight selection signatures underlying saline adaptation in Chilika buffalo. Physiol. Genomics 56, 609–620. doi: 10.1152/physiolgenomics.00028.2024.

Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor Popul Biol. 1971;2(2):125-41. doi:10.1016/0040-5809(71)90011-6.

Szmatoła, T., Gurgul, A., Jasielczuk, I., Ząbek, T., Ropka-Molik, K., Litwińczuk, Z., and Bugno-Poniewierska, M. (2019). A comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. Animals 9, 1024. doi: 10.3390/ani9121024.

Tavares, L. P., Negreiros-Lima, G. L., Lima, K. M., Silva, P. M. R. E., Pinho, V., Teixeira, M. M., et al. (2020). Blame the signaling: role of cAMP for the resolution of inflammation. Pharmacol. Res. 159, 105030. doi: 10.1016/j.phrs.2020.105030.

Taye MT (2018). *Identification of adaptive signatures in the cattle genome.* Doctoral dissertation, Seoul National University, South Korea. Available at: https://s-space.snu.ac.kr/handle/10371/140789

Taye, M., Lee, W., Jeon, S., Yoon, J., Dessie, T., Hanotte, O., Mwai, O. A., Kemp, S., Cho, S., Oh, S. J., et al. (2017). Exploring evidence of positive selection signatures in cattle breeds selected for different traits. Mamm. Genome 28, 528–541. doi: 10.1007/s00335-017-9715-6.

Wang Y, Wang Z, Liu W, Xie S, Ren X, Yan L, et al. (2024). Genetic background of blood β-hydroxybutyrate acid concentrations in early-lactating Holstein dairy cows based on genome-wide association analyses. *Genes* 15(4): 412. doi: 10.3390/genes15040412

Wang, J., Shen, N., Zhao, K., Liao, J., Jiang, G., Xiao, J., et al. (2025). Revealing study and breeding implications for production traits and tail characteristics in

Simmental cattle by GWAS. Front. Genet. 16, 1491816. doi: 10.3389/fgene.2025.1491816.

Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38(6): 1358–1370. doi: 10.2307/2408641

Wickham H (2016). Programming with ggplot2. In: *ggplot2: Elegant graphics for data analysis.* Cham, Switzerland: Springer International Publishing, 241–253.

Wu, H., Li, S. Y., Jiang, J. M., Wang, H., Wu, G. S., Li, L. B., Chen, Q. L., Du, L., Wang, F. Y., and Chen, S. (2024). Genetic diversity of Hainan cattle population based on whole genome resequencing and screening of candidate genes related to heat adaptation. J. Southern Agric. 55, 3707–3717. doi: 10.3969/j.issn.2095-1191.2024.12.020

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan 7;88(1):76-82. doi: 10.1016/j.ajhg.2010.11.011. Epub 2010 Dec 17. PMID: 21167468; PMCID: PMC3014363.

Yang, Y., Chang, C., Baiyin, B., Liu, Z., Guo, L., Zhou, L., et al. (2022). Blood transcriptome analysis of beef cow with different parity revealed candidate genes and gene networks regulating the postpartum diseases. Genes 13, 1671. doi: 10.3390/genes13091671.

Zhong, Z., Wang, Z., Xie, X., Pan, D., Su, Z., Fan, J., et al. (2024). Insights into adaption and growth evolution: genome-wide copy number variation analysis in Chinese Hainan Yellow cattle using whole-genome re-sequencing data. Int. J. Mol. Sci. 25, 11919. doi: 10.3390/ijms252211919.

Appendix Chapter 3

Tracing shared genomic regions among local cattle breeds from Sardinia

Supplementary Table 1. Distinguishing characteristics of the cattle breeds included in the study (Bigi et al, 2008; Biodiversità Sardegna, 2025; Braunvieh Switzerland, 2025).

Supplementary Table 2. Shared ROH$_{\text{REP}}$ among the pairwise comparison SAR vs SB and SAR vs SM.

Supplementary Table 3. Genomic regions identified through multivariate analyses for SAR vs SB and SAR vs SM comparisons. For each region, the table reports the chromosome (BTA), SNP ID, and physical position (in base pairs).

Supplementary Table 4. QTL and associated traits identified in genomic regions highlighted both by CDA and shared ROH$_{\text{REP}}$ analyses in the pairwise comparisons SAR vs SB and SAR vs SM.

Supplementary Table 5. QTL and associated traits found in the SNP$_{\text{ROH}}$ region found among SAR, SB and SM.

Supplementary Table 1. Distinguishing characteristics of the cattle breeds included in the study (Bigi et al, 2008; Biodiversità Sardegna, 2025; Braunvieh Switzerland, 2025).

	Distinctive features	BR	MOD	SAR	SB	SM
Morphological traits	Size	Medium	Large	Small	Medium	Medium - Large
	Coat color	Brown with white ear and muzzle	Dark red	Various shades of fawn, sometimes with the presence of spots and streaks	Grayish coat of various shades	Fawn-colored coat, with shades ranging from light red to dark wine-colored
	Height withers males (cm)	-	150	119-121	140	160
	Height withers females (cm)	-	140	-	120-123	145
	Weight males (kg)	1000 - 1100	≥ 800	200-250	800	800-900

	Weight females (kg)	700	480-550	150-200	500	450-550
Production data	Production	Milk	Milk, Meat	Meat	Meat	Meat
Others	Breeding system	-	Extensive	Extensive	Semi-intensive, Semi-extensive, Extensive	Extensive

Supplementary Table 2. Shared ROH$_{REP}$ among the pairwise comparison SAR vs SB and SAR vs SM.

Comparison	BTA	Region (Mb)	Animals	SAR	SB	SM
SAR vs SB vs SM	4	69.61-70.61	4	2	1	1
	14	49.53-50.96	3	1	1	1
	28	1.03-3.52	3	1	1	1
	29	23.53-25.53	3	1	1	1
SAR vs SB	1	30.35-31.42	2	1	1	
	1	83.23-85.28	2	1	1	
	4	0.37-4.93	2	1	1	
	4	49.32-50.38	2	1	1	
	4	69.61-70.61	2	1	1	
	7	102.09-103.14	2	1	1	
	7	18.06-19.29	3	1	1	1
	7	59.37-62.76	4	2	2	
	7	62.89-65.21	2	1	1	
	9	45.14-46.84	2	1	1	
	9	63.49-64.59	2	1	1	
	9	99.92-104.23	2	1	1	
	11	45.27-46.69	4	3	1	
	11	55.82-57.58	2	1	1	
	12	73.06-74.34	2	1	1	
	13	53.73-55.66	2	1	1	
	14	49.53-50.96	2	1	1	
	16	36.00-38.14	3	1	1	1
	16	44.13-46.16	3	2	1	
	18	59.90-61.69	2	1	1	

	18	61.54-65.79	3	1	1	1
	20	41.39-42.81	2	1	1	
	21	19.58-20.81	2	1	1	
	21	31.39-33.60	2	1	1	
	24	11.45-12.47	4	2	1	1
	24	38.22-39.45	2	1	1	
	27	44.10-45.45	2	1	1	
	28	1.03-3.52	2	1	1	
	28	3.64-6.09	2	1	1	
	29	18.50-23.39	2	1	1	
	29	23.53-25.53	3	2	1	
	29	39.33-41.49	2	1	1	
SAR vs SM	1	107.93-109.44	2	1		1
	1	30.35-31.48	2	1		1
	1	44.38-45.56	2	1		1
	1	64.19-65.33	2	1		1
	1	66.61-67.62	2	1		1
	2	0.27-1.56	2	1		1
	2	73.13-74.59	2	1		1
	4	69.61-70.61	4	2	1	1
	4	69.73-71.03	4	2		2
	5	82.45-83.51	2	1		1
	6	26.82-28.84	2	1		1
	7	15.11-16.18	2	1		1
	7	52.89-53.98	2	1		1
	8	40.14-41.21	2	1		1
	9	4.47-5.56	2	1		1
	10	44.77-45.97	2	1		1

11	56.90-57.99	5	1		4
11	59.86-61.04	2	1		1
12	60.19-61.25	3	1		2
14	21.69-22.83	3	2		1
14	49.53-50.96	3	1	1	1
14	50.35-51.53	3	1		2
14	50.78-51.79	3	2		1
14	65.67-66.81	3	2		1
14	78.36-81.73	2	1		1
15	4.93-6.05	2	1		1
16	3.13-4.14	2	1		1
16	39.14-40.41	4	1		3
16	45.66-46.71	3	2		1
18	17.32-18.37	2	1		1
19	27.59-28.92	3	1		2
21	40.72-41.79	3	2		1
23	51.03-52.28	2	1		1
26	1.07-3.83	2	1		1
26	34.16-36.18	2	1		1
27	33.02-34.22	2	1		1
27	44.20-45.45	2	1		1
28	1.03-3.52	3	1	1	1
28	31.45-36.02	2	1		1
29	23.53-25.53	3	1	1	1

Supplementary Table 3. Genomic regions identified through multivariate analyses for SAR vs SB and SAR vs SM comparisons. For each region, the table reports the chromosome (BTA), SNP ID, and physical position (in base pairs).

Comparison	BTA	SNP	Position (bp)
SAR vs SB	1	Hapmap50634-BTA-26738	36.99
	1	ARS-BFGL-NGS-19097	43.9
	1	BTB-01159966	5.72
	1	BTB-00032648	66.52
	1	ARS-BFGL-NGS-110564	145.75
	2	BTB-01290636	17.4
	2	BTB-00083524	20.99
	2	ARS-BFGL-NGS-36416	30.8
	2	BTB-00088621	34.01
	2	Hapmap42250-BTA-55589	36.75
	2	Hapmap53891-rs29021231	45.45
	2	ARS-BFGL-NGS-82679	66.07
	2	ARS-BFGL-NGS-38248	126.46
	3	ARS-BFGL-NGS-38423	12.67
	3	BTA-90312-no-rs	30.35
	3	ARS-BFGL-NGS-113746	31.34
	3	Hapmap43441-BTA-103289	58.07
	3	UA-IFASA-9204	59.62
	3	BTB-01830087	60.11
	4	BTA-90533-no-rs	1.99
	4	ARS-BFGL-NGS-70391	9.11

4	Hapmap51036-BTA-71487	18.56
4	UA-IFASA-5600	20.64
4	Hapmap27013-BTA-158242	23.6
4	BTB-00203399	94.13
5	ARS-BFGL-NGS-49659	39.49
5	BTA-15560-no-rs	91.68
6	ARS-BFGL-NGS-116880	14.45
6	Hapmap23419-BTC-059652	80.59
6	Hapmap27308-BTC-043338	80.76
6	ARS-BFGL-NGS-1110	103.85
7	BTB-00319289	71.67
8	BTB-01881993	21.55
8	Hapmap41646-BTA-81057	39.68
8	ARS-BFGL-NGS-34028	99.2
9	Hapmap49089-BTA-23167	32.37
9	BTB-00404235	89.2
9	ARS-BFGL-NGS-32882	102.5
9	ARS-BFGL-NGS-115046	102.71
10	ARS-BFGL-NGS-5722	16.16
10	Hapmap41917-BTA-61237	21.96
10	ARS-BFGL-NGS-23266	26.12
10	Hapmap53311-rs29018232	38.78
10	BTB-01101698	91.86
11	ARS-BFGL-NGS-13472	4.73
11	ARS-BFGL-NGS-119597	48.05
11	Hapmap38739-BTA-25571	50.35
11	Hapmap42798-BTA-109203	55.89

	11	ARS-BFGL-NGS-31946	66.58
	11	ARS-BFGL-NGS-82102	105.08
	12	BTA-31543-no-rs	11.72
	12	ARS-BFGL-NGS-72624	35.2
	12	BTA-87771-no-rs	78.07
	13	ARS-BFGL-NGS-113966	25.76
	13	ARS-BFGL-NGS-14452	36.42
	13	ARS-BFGL-NGS-80055	48.23
	13	BTA-34052-no-rs	80.69
	14	ARS-BFGL-BAC-26939	78.49
	14	ARS-BFGL-BAC-20217	8.76
	14	Hapmap23304-BTC-069412	11.3
	14	BTB-00554463	16.97
	14	Hapmap23524-BTC-065402	25.19
	15	Hapmap39348-BTA-63139	33.11
	15	ARS-BFGL-BAC-27770	39.98
	15	ARS-BFGL-NGS-1186	51.73
	16	ARS-BFGL-NGS-21857	60.7
	16	ARS-BFGL-NGS-15423	72.26
	17	ARS-BFGL-NGS-22143	46.92
	17	ARS-BFGL-NGS-39993	60.78
	18	BTB-00691673	3.33
	18	Hapmap51229-BTA-42956	33.15
	19	BTA-28147-no-rs	4.51
	19	ARS-BFGL-NGS-100358	32.75
	20	Hapmap36217-SCAFFOLD290026_21689	8.03
SAR vs SM	1	ARS-BFGL-NGS-49634	48.61

1	BTB-01165120	50.06
1	BTB-01991572	4.99
1	BTB-01086841	94.24
2	BTA-47963-no-rs	66.61
2	ARS-BFGL-NGS-112281	67.35
2	ARS-BFGL-NGS-118505	126.37
2	ARS-BFGL-NGS-74920	0.9
3	ARS-BFGL-NGS-740	10.4
3	BTA-68578-no-rs	84.52
3	BTA-68864-no-rs	85.96
3	BTA-108673-no-rs	90.82
3	BTB-00156691	110.4
4	ARS-BFGL-NGS-89632	14.08
4	BTB-01007411	36.97
4	ARS-BFGL-NGS-78389	48.8
4	BTA-70724-no-rs	57.38
4	ARS-BFGL-NGS-105821	57.71
4	ARS-BFGL-NGS-60615	75.17
4	ARS-BFGL-NGS-7597	100.96
5	ARS-BFGL-NGS-18989	41.96
5	ARS-BFGL-NGS-67203	54.39
5	BTA-74527-no-rs	88.18
5	ARS-BFGL-NGS-365	90.54
5	ARS-BFGL-NGS-99256	104.33
5	ARS-BFGL-NGS-8495	115.92
6	BTB-00243845	15.7
6	BTB-00252870	40.07

6	ARS-BFGL-NGS-112970	53.31
6	BTB-01899984	107.48
7	ARS-BFGL-NGS-8346	3.16
7	BTB-00291698	5.85
7	ARS-BFGL-NGS-6108	22.49
7	Hapmap58358-rs29011315	36.72
7	Hapmap47734-BTA-78811	37.46
7	ARS-BFGL-NGS-108501	57.42
7	Hapmap42072-BTA-113330	74.49
8	BTA-121823-no-rs	7.79
8	UA-IFASA-5756	29.69
8	ARS-BFGL-NGS-114764	44.09
8	ARS-BFGL-NGS-25805	44.13
8	BTA-38275-no-rs	53.72
9	ARS-BFGL-NGS-115580	36.09
9	ARS-BFGL-NGS-3314	96.46
9	ARS-BFGL-NGS-6479	102.29
10	ARS-BFGL-NGS-112670	3.54
10	ARS-BFGL-NGS-16712	10.23
10	BTA-59405-no-rs	19.78
10	ARS-BFGL-NGS-118433	26.02
10	BTA-30638-no-rs	53.39
10	ARS-BFGL-NGS-97889	68.29
10	ARS-BFGL-BAC-16717	89.27
10	BTA-82461-no-rs	95.84
11	ARS-BFGL-NGS-108649	1.47
11	BTB-00753528	19.35

11	ARS-BFGL-NGS-112640	29.42
11	ARS-BFGL-NGS-110808	39
11	ARS-BFGL-NGS-52860	65.71
11	Hapmap48671-BTA-17038	67.41
11	ARS-BFGL-NGS-65485	74.37
13	Hapmap15258-rs29013198	72.1
13	ARS-BFGL-NGS-118453	73.46
14	BTB-01252321	20.99
15	ARS-BFGL-NGS-114766	75.63
15	BTB-00622270	82.64
15	ARS-BFGL-NGS-119665	31.21
15	ARS-BFGL-NGS-115263	37.43
15	ARS-BFGL-NGS-39267	44.88
15	ARS-BFGL-NGS-22122	45.34
16	ARS-BFGL-NGS-69979	69.11
17	ARS-BFGL-NGS-110680	4.42
17	ARS-USMARC-583	29.92
17	ARS-BFGL-NGS-62271	68.64
17	ARS-BFGL-NGS-44494	72.61
18	ARS-BFGL-NGS-73872	47.75
18	ARS-BFGL-NGS-98851	61.27
19	UA-IFASA-6519	17.8
19	ARS-BFGL-NGS-101671	27.76
19	ARS-BFGL-NGS-88746	39.07

Supplementary Table 4. QTL and associated traits identified in genomic regions highlighted both by CDA and shared ROHREP analyses in the pairwise comparisons SAR vs SB and SAR vs SM.

Comparison	CHR	Region (Mb)	QTL	Associated trait
SAR vs SB	4	1,810,205-1,810,209	Health	Bovine tuberculosis susceptibility
	4	1,871,614-1,871,618	Health	Bovine tuberculosis susceptibility
	4	1,926,015-1,926,019	Health	Bovine tuberculosis susceptibility
	4	1,928,631-1,928,635	Health	Bovine tuberculosis susceptibility
	4	1,938,310-1,938,314	Health	Bovine tuberculosis susceptibility
	4	1,982,261-1,982,265	Health	Bovine tuberculosis susceptibility
	4	2,043,547-2,043,551	Exterior	Heel horn erosion
	4	2,086,601-2,086,605	Health	Bovine tuberculosis susceptibility
	4	2,116,927-2,116,931	Health	Bovine tuberculosis susceptibility
	9	102283186-102283190	Reproduction	Inseminations per conception
	9	102,286,975-1022,86,979	Reproduction	Non-return rate
	9	102,289,573-102,289,577	Milk	Milk caproic acid content
	9	102,338,302-102,338,306	Meat and Carcass	Carcass weight
	9	102,363,248-102,363,252	Meat and Carcass	Carcass weight
	9	102,451,270-102,451,274	Milk	Milk tetracosanoic acid content
	9	102,499,278-102,499,282	Production	Length of productive life

	9	102,594,523-102,594,527	Milk	Milk cis-7,10,13,16,19-Docosapentaenoic acid content
	9	102,683,648-102,683,652	Health	Insulin level
	9	102,943,187-102,943,191	Meat and Carcass	Tenderness score
	9	102,943,211-102,943,215	Health	Somatic cell count
	11	55,639,867-55,639,871	Milk	Milk protein percentage
	11	55,667,712-55,667,716	Milk	Milk protein percentage
	11	55,760,091-55,760,095	Milk	Milk fat yield
	11	55,882,191-55,882,195	Milk	Milk protein percentage
	11	55,888,309-55,888,313	Reproduction	Calving to conception interval
	11	55,895,259-55,895,263	Milk	Milk protein percentage
	11	55,925,390-55,925,394	Milk	Milk protein percentage
	11	55,929,094-55,929,098	Milk	Milk protein percentage
	11	55,937,617-55,937,621	Milk	Milk protein percentage
	11	55,947,075-55,947,079	Milk	Milk protein percentage
	11	55,947,309-55,947,313	Milk	Milk protein percentage
	11	55,953,090-55,953,094	Milk	Milk protein percentage
	11	55,959,149-55,959,153	Milk	Milk fat percentage
	11	55,980,898-55,980,902	Milk	Milk fat percentage
	11	55984,604-55,984608	Health	Somatic cell score
	11	55,996,332-55,996,336	Milk	Milk kappa-casein percentage
	11	56,001,469-56,001,473	Milk	Milk protein percentage
SAR vs SM	5	83,132,517-83,132,521	Milk	Milk fat yield
	5	83,132,517-83,132,521	Reproduction	Pregnancy rate

5	83,283,304-83,283,308	Milk	Milk alpha-S2-casein percentage
5	83,283,304-83,283,308	Reproduction	Pregnancy rate
5	83,286,196-83,286,200	Milk	Milk alpha-S2-casein percentage
5	83,286,796-83,286,800	Exterior	Dairy capacity composite index
5	83,286,796-83,286,800	Reproduction	Retained placenta
5	83,328,652-83,328,656	Milk	Milk fat percentage
5	83,342,894-83,342,898	Milk	Milk fat yield
5	83,355,126-83,355,130	Exterior	Udder texture
5	83,374,185-83,374,189	Milk	Milk fat percentage
5	83,414,111-83,414,115	Exterior	Udder structure
5	83,414,111-83,414,115	Milk	Milk fat yield
5	83,421,912-83,421,916	Milk	Milk kappa-casein percentage
5	83,492,453-83,492,457	Milk	Milk fat percentage
5	83,512,415-83,512,419	Meat and Carcass	Connective tissue amount
5	83,513,072-83,513,076	Meat and Carcass	Connective tissue amount
5	83,515,951-83,515,955	Meat and Carcass	Connective tissue amount
5	83,516,444-83,516,448	Meat and Carcass	Connective tissue amount
5	83,517,607-83,517,611	Meat and Carcass	Connective tissue amount
5	83,518,453-83,518,457	Meat and Carcass	Connective tissue amount
5	83,519,646-83,519,650	Meat and Carcass	Connective tissue amount
5	83,539,520-83,539,524	Meat and Carcass	Connective tissue amount
5	83,540,581-83,540,585	Meat and Carcass	Connective tissue amount
5	83,544,273-83,544,277	Meat and Carcass	Connective tissue amount
5	83,544,995-83,544,999	Meat and Carcass	Connective tissue amount
5	83,545,094-83,545,098	Meat and Carcass	Connective tissue amount
5	835,47,552-83,547,556	Meat and Carcass	Connective tissue amount
5	83,547,652-83,547,656	Meat and Carcass	Connective tissue amount

5	83,547,790-83,547,794	Meat and Carcass	Connective tissue amount
5	83,547,866-83,547,870	Meat and Carcass	Connective tissue amount
5	83,547,917-83,547,921	Meat and Carcass	Connective tissue amount
5	83,548,428-83,548,432	Meat and Carcass	Connective tissue amount
5	83,549,005-83,549,009	Meat and Carcass	Connective tissue amount
19	28,575,556-28,575,560	Exterior	Dairy capacity composite index
19	28,598,939-28,598,943	Exterior	Dairy capacity composite index
19	28,599,856-28,599,860	Milk	Milk protein percentage
19	28,705,038-28,705,042	Production	Residual feed intake
19	28,767,495-28,767,499	Health	Somatic cell count
19	28,896,081-28,896,085	Milk	Dry period length

Supplementary Table 5. QTL and associated traits found in the SNP_{ROH} region found among SAR, SB and SM.

CHR	Region (Mb)	QTL	Associated trait
1	83,498,164-83,498,168	Production	Average daily gain
1	83,810,341-83,810,345	Reproduction	Non-return rate
1	83,871,153-83,871,157	Production	Methane production
1	83,966,653-83,966,657	Production	Residual feed intake
1	84,036,034-84,036,038	Meat and Carcass	Trans-6/9-C18:1 fatty acid content
1	84,0648,67-84,064,871	Exterior	Dairy form
1	840,64,867-84,064,871	Health	Somatic cell score
1	84,064,867-84,064,871	Milk	Milk fat percentage
1	84,064,867-84,064,871	Production	Net merit
1	84,064,867-84,064,871	Reproduction	Calving ease
1	137,599,041-137,599,045	Milk	Milk iron content
1	137,645,848-137,645,852	Milk	Milk iron content
1	137,668,190-137,668,194	Exterior	Stature
1	137,996,413-137,996,417	Reproduction	Birth index
11	59,739,302-59,739,306	Milk	Milking speed
11	59,794,188-59,794,192	Milk	Milking speed
11	59,801,324-59,801,328	Milk	Milk fat yield
11	59,836,923-59,836,927	Milk	Milk fat yield
11	59,838,898-59,838,902	Milk	Milk fat yield
11	59,850,226-59,850,230	Milk	Milk fat percentage
11	59,987,392-59,987,396	Health	Bovine tuberculosis susceptibility
11	60,377,762-60,377,766	Milk	Milk kappa-casein percentage
11	60,403,983-60,403,987	Meat and Carcass	Connective tissue amount
11	60,404,364-60,404,368	Milk	Milk tridecylic acid content

11	60,426,283-60,426,287	Reproduction	Inseminations per conception
11	60,468,950-60,468,954	Production	Body weight
11	60,494,767-60,494,771	Production	Body weight
11	60,517,674-60,517,678	Production	Body weight
11	60,567,345-60,567,349	Production	Body weight
11	60,708,828-60,708,832	Health	Tick resistance
11	60,731,273-60,731,277	Health	Tick resistance
11	60,775,320-60,775,324	Production	Body weight
11	60,807,903-60,807,907	Exterior	Dairy form
11	60,807,903-60,807,907	Milk	Milk fat yield
11	60,807,903-60,807,907	Production	Body depth
11	60,832,430-60,832,434	Milk	Milk protein percentage
11	60,841,983-60,841,987	Health	Tick resistance
11	60,864,550-60,864,554	Production	Body weight gain

CHAPTER 4

Assessing fertility: Development of a selection index for fertility in Italian Jersey cattle

This is the published version in **Journal of Dairy Science** of:

Milia, PE., Cassandro, M., Cesarani, A., Fabris, A., Finocchiaro, R., Marusi, M., Macciotta, NPP., Dimauro, C. (2025). Assessing fertility: Development of a selection index for fertility in Italian Jersey cattle. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2025-27293

This version is used by permission of the **Journal of Dairy Science**.

Abstract

Selection for increased milk yield in dairy cattle has led to a decline in fertility due to unfavorable genetic correlation. However, fertility remains crucial in dairy farming for maintaining a regular productive cycle, ensuring continuous milk supply, and reducing costs related to nonproductive cows. For this reason, the aim of this study was to estimate genetic parameters for fertility traits and to assess their genetic correlation with production traits in the Italian Jersey cattle breed. Days open (DO) as a fertility trait, and kg/305-d milk (MY), protein (PY), and fat yields (FY) were considered as production traits, respectively. After data editing, the final dataset included 33,053 DO records from 16,667 cows, restricted to values between 25 and 400 d. The mean phenotypic value for DO was 121 ± 71 d. Variance components and h^2 estimates for DO were obtained using a single-trait animal model, whereas genetic correlations were estimated with a multiple-trait animal model. The models accounted for herd, permanent environment, and animal as random effects, whereas parity and month-year were considered as fixed effects. A pedigree-based relationship matrix was used in the model. Estimated h^2 for DO was low (0.04 ± 0.01), and unfavorable genetic correlations were observed with MY (0.16 ± 0.08), PY (0.12 ± 0.09), and FY (0.23 ± 0.09). Genetic trends revealed a consistent increase in production traits, whereas the trend of DO showed a greater variability over time. Despite the low h^2, these results highlight the potential to improve fertility through selection and support the inclusion of reproductive traits in breeding programs aimed at achieving a more balanced and sustainable genetic progress in Italian Jersey cattle.

Keywords

Genetic selection dairy cattle breeding, reproductive efficiency, Jersey cattle.

Introduction

Fertility is a complex, multifactorial trait in dairy cattle, influenced by genetics, physiology, management, and environment (Cassandro, 2014; Kgari et al., 2021). Its economic importance lies in the ability of cows to return to estrus, conceive efficiently, and maintain pregnancy, directly affecting farm productivity and profitability. Poor fertility increases veterinary costs, the number of inseminations, and culling rates due to reproductive failure, especially in early lactations (Shanks et al., 1981; Howard et al., 2017). Reproductive disorders are among the main reasons for involuntary culling, with replacement heifers costing substantially more than the salvage value of culled cows (Roche et al., 2020).

Several reproductive indicators are used to evaluate fertility. Calving interval reflects both the speed of conception and the ability to carry a pregnancy to term; in American Holsteins, mean calving interval in 2006 reached 422 d compared with 410 in Jerseys and 397 in Italian Simmentals (Norman et al., 2009; Cesarani et al., 2020). Days open, defined as the interval from calving to conception, averaged 117 d in Spanish Holsteins (González-Recio and Alenda, 2005). Conception rate, defined as the percentage of inseminations that resulted in a confirmed calving, has declined steadily, with first-service conception rate dropping from ~65% in 1951 to ~40% in 1996 in American herds (Lucy, 2001; Norman et al., 2009). Similarly, the average number of inseminations per conception increased from 2.1 in 1996 to 2.5 in 2006 in American Holsteins (Norman et al., 2009). In Ethiopian dairy herds, Temesgen et al. (2022) reported a median of 154 d open and a conception rate of 44.7%. These data illustrate the scale of fertility decline associated with intensive milk selection (Lucy, 2001; Norman et al., 2009). The main reason is the antagonistic genetic relationship between milk components and fertility, with unfavorable correlations consistently reported (e.g., +0.25 to +0.69 with calving interval, +0.5 to +0.7 with days open, and −0.10 to −0.35

with conception rate; Cesarani et al., 2020; Tiezzi et al., 2011; VanRaden et al., 2014). This antagonism is largely mediated by the negative energy balance in early lactation, which alters endocrine function, delays ovarian activity, and compromises oocyte quality (Lucy, 2001). Moreover, in some specific environments, fertility can be strongly affected by season of insemination, with markedly reduced conception during the hot summer months, and improved performance when insemination occurred ≤85 d postpartum, when artificial insemination was used, and in herds with higher milk production (Temesgen et al., 2022). These findings underline that, beyond genetic selection, environmental and management factors can profoundly influence reproductive outcomes. Despite their low h^2, fertility traits show substantial phenotypic variation and genetic potential for improvement (Miglior, 1999; Kgari et al., 2021). Several countries (e.g., Denmark, Sweden, and Finland) have included fertility traits in national breeding evaluations since the 1990s, and phenotypic fertility indicators have stabilized or improved where selection has been applied (Crowe, 2008). Accurate phenotyping, however, remains essential (González-Recio and Alenda, 2005).

In Italy, fertility traits are still not formally included in the national genetic evaluation of Jerseys cattle, a breed with a limited population size that is often raised alongside Holsteins to increase milk fat yield (ANAFIBJ, 2025). Although selection indices for traits such as SCC and longevity have been developed, fertility traits have not yet been addressed (Fabris et al., 2025). The aim of this study was to estimate genetic parameters for fertility and production traits and to evaluate the genetic trend for days open in the Italian Jersey population.

Materials and methods

Data and Edits

To investigate genetic variation in reproductive performance of Italian Jersey cattle, different fertility indicators were initially tested. First-service conception rate and the interval between first and last insemination were considered; however, insemination and pregnancy records were often incomplete and unreliable because most breeders tended to report only the successful service while neglecting previous unsuccessful attempts. Consequently, these measures were excluded from the analysis. Days open (**DO**), defined as the number of days from calving to the insemination that resulted in a subsequent conception and calving (Louca and Legates, 1968; Silva et al., 1992), was therefore adopted as a fertility trait.

For this study, records of the Italian Jersey cattle population were provided by Associazione Nazionale Allevatori Frisona, Bruna e Jersey (ANAFIBJ). Along with DO as fertility indicator, production traits considered were 305-d milk (MY), fat (FY), and protein (PY) yields. Cows with at least 1 calving event between 2000 and 2023 were selected. Cows that changed herds during their career, had an age at first calving greater than 36 mo, and showed a gestation length shorter than 240 d or longer than 300 d were excluded. Additionally, sires with fewer than 3 daughters and herds with fewer than 3 cows were discarded. For DO editing, cows with a calving date but no insemination record were removed. When the interval between 2 consecutive inseminations exceeded 150 d, the event was interpreted as an abortion followed by a new insemination. Insemination records of cows culled before the expected calving date were discarded. Information on the use of hormonal treatments was not available, and records of sexed semen were not considered in the analyses. The edited dataset consisted of 29,405 DO records from 14,750 cows, with values ranging between 25 and 400 d and an average of 117 ± 66 d (Table 1). For the editing of kg/305-d production traits (MY, FY, and PY), the same inclusion period (2000–2023) and general editing criteria were applied. In addition, cows with MY below 1,000 kg and PY and FY below 100 kg were excluded. Given the limited number of records available from the fourth parity onward and the similar biological performance observed, all parities beyond the fourth were grouped together with the fourth parity into a single class (4+) in both datasets.

The number of observations, animals, and descriptive statistics after these edits are shown in Table 1.

Table 1. Number of observations, number of cows, mean, and standard deviation (SD) for days open (DO), milk yield (MY), protein yield (PY) and fat yield (FY) in the considered data set

Trait	Observation (no.)	Animals (no.)	Mean	SD
Days Open, days	29,405	14,750	117	66
Milk Yield, kg	77,071	30,425	6,307	1,555
Protein Yield, kg	77,071	30,425	249	60
Fat Yield, kg	77,071	30,425	318	77

Mean values of DO and production traits by parity are presented in Table 2. Days open remained relatively stable across the first 3 parities, ranging from 115 to 116 d, with a noticeable increase observed in cows with fourth or later parity (123 d). Average MY, PY, and FY increased from the first to the third parity, with a slight decline observed from the fourth parity onward.

Table 2. Phenotypes (mean ± standard deviations) considered in the study according to the parity of cows. Lactations from 4th are aggregated in a single class.

Parity	Observation (no.)	Days open, days	Milk yield, kg	Protein yield, kg	Fat yield, kg
1	9,442	116±66	5,926±1,356	232±52	295±66
2	7,867	115±65	6,487±1,583	259±61	329±79
3	5,065	116±65	6,659±1,626	264±62	338±81
4+	3,450	123±70	6,405±1,634	254±63	327±82

Variance Components Estimation

Variance components for DO and production traits were estimated with a single-trait animal model:

$$y = X\beta + Z_h h + Z_a a + Z_{pe} pe + e \quad [1]$$

where y is the phenotypic record for the trait under analyses; X is the incidence matrix for fixed effects and β includes parity (4 levels) and month-year (276 levels for DO; 288 for production traits); $Z_h$herd (Z_hh) is the incidence matrix for the random effect of the herd (613 levels for DO and 1,353 for production traits) distributed as $\mathrm{N}(0, \mathrm{I}\sigma_{\mathrm{h}}^2)$ where I is an identity matrix and σ_h^2 is the variance associated with the herd effect; $Z_a$is the matrix for the random additive genetic effect (27,857 and 46,264 animals in the relationship matrix for DO and production traits, respectively) distributed as $\mathrm{N}(0, \mathrm{A}\sigma_{\mathrm{a}}^2)$ where A is the pedigree-based relationship matrix and σ_a^2 is the additive genetic variance; Z_{pe} is the matrix for the random effect of permanent environment (27,857 levels for DO and 46,264 for production traits) distributed as $\mathrm{N}(0, \mathrm{I}\sigma_{\mathrm{pe}}^2)$ where I is an identity matrix and σ_{pe}^2 is the permanent environmental variance; and e is the random residual. The herd effect was set as random because of its high variability in herd size (Fabris et al., 2025). Under this parametrization, the expectations and (co)variance structures of the random effects were defined as:

$$\mathrm{E}[y] = X\beta,$$

$$Var\begin{bmatrix} h \\ a \\ pe \\ e \end{bmatrix} = \begin{bmatrix} \sigma_h^2 I_H & 0 & 0 & 0 \\ 0 & \sigma_a^2 A & 0 & 0 \\ 0 & 0 & \sigma_{pe}^2\, I_P & 0 \\ 0 & 0 & 0 & \sigma_e^2 I_n \end{bmatrix},$$

Heritability (h^2) was estimated using the following formula:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_{pe}^2+\sigma_h^2+\sigma_e^2} \quad [2]$$

Given the particular structure of our data, we also estimated the intra-herd heritability (h^2_{IH}) following the methodology proposed by Tiezzi et al. (2011):

$$h^2_{\mathrm{IH}} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2}, \qquad [3]$$

In this formula, the variance component of the random herd effect is excluded from the calculation. This methodological choice was motivated by the fact that the herd effect accounted for a substantial proportion of the total phenotypic variance for the traits under investigation. As such, intra-herd heritability provides a more realistic estimate of the genetic progress that can be achieved within each specific production system.

Repeatability (r^2) was also estimated as the proportion of the total phenotypic variance explained by additive genetic and permanent environmental effects, according to the following formula:

$$r^2 = \frac{\sigma_a^2+\sigma_{pe}^2}{\sigma_a^2 + \sigma_{pe}^2 + \sigma_h^2 + \sigma_e^2}, \qquad [4]$$

Similarly, Intra-herd repeatability (r^2_{IH}) was estimated as the proportion of the total phenotypic variance explained by additive genetic and permanent environmental effects. As for intra-herd heritability (h^2_{IH}), the variance component of the random herd effect was excluded from the calculation.

Genetic correlations between fertility and production traits were estimated using a multiple-trait animal model with the same structure as equation [1], considering DO and kg/305-d yields of MY, PY, and FY. All the available records for these 4 traits were included in the analysis (27,967 records from 14,325 cows).

Variance components and (h^2 and h^2_{IH}) for single-trait animal models were estimated using a REML algorithm implemented in the BLUPF90 family programs (Misztal et al., 2022) with a pedigree relationship matrix (**A**). In the multiple-trait animal model, variance components were estimated using a Bayesian approach via the Gibbs sampling algorithm (100,000 iterations, a burn-in of 50,000, and a thinning rate of 5), implemented in the gibbsf90+ program (Misztal et al., 2014). For all the analyses, pedigree information was traced back 4 generations, resulting in a full pedigree of 63,005 animals.

In addition, the coefficient of genetic variation (CVg) for DO was calculated as the ratio between the additive genetic SD and the phenotypic mean of the trait, expressed as a percentage (Houle, 1992; Berry et al., 2017).

Official Genetic Evaluation

Mimicking an official evaluation for DO, an updated dataset that included 33,053 DO records from 16,667 cows was considered. The same model described in Equation [1] was adopted with the following effects: parity (4 levels), month-year (285 levels), herds (652 levels), animal and permanent environment (30,360 levels). Breeding values (EBV) for DO were estimated using the *MiX99* software package (Lidauer et al., 2022). As for all functional traits published by ANAFIBJ, EBV for DO were standardized to a mean of 100 and a standard deviation of 5. Higher EBV values indicate shorter Days Open; therefore, animals with EBV greater than 100 are genetically superior for this fertility trait. Based on their EBV values, bulls were grouped into three categories: low-index (EBV < 95), medium-index ($95 \leq \text{EBV} \leq 105$), and high-index (EBV > 105). To evaluate the phenotypic impact associated with different genetic merit levels, the average DO of daughters was calculated within each bull category. Furthermore, three types of comparison were performed: a) the difference between each category's average DO and the overall population mean (μ_{DO}); b) the difference between each category's average DO and the genetic base mean (μ_{gb}) - defined as the average genetic merit (EBV) of cattle born from 2017 to 2019, which represents a fixed point for evaluating genetic progress across generations; c) the direct pairwise differences among the three categories themselves (e.g., high vs low, high vs medium, medium vs low). These comparisons allowed for a quantification of the phenotypic expression associated with different genetic merit levels and offered practical interpretation tools for breeders aiming to improve reproductive efficiency through sire selection.

The ANAFIBJ developed and implemented the genetic evaluation for DO in the Italian Jersey population. The new trait was submitted to Interbull, the international center for genetic evaluation of dairy cattle, which harmonizes national evaluations and provides international genetic correlations. This procedure ensures international validation and comparability, allowing the Italian Jersey population to be directly connected with global breeding programs.

Results

Table 3 shows the variance components, both heritability and intra-herd heritability and both repeatability and intra-herd repeatability estimated using the single-trait animal model.

Table 3. Estimates herd variance, permanent environment variance, additive genetic variance, residual variance, h^2, IH h^2, repeatability and IH heritability of the considered traits obtained with the single trait models.

	Days Open	Milk yield	Protein yield	Fat yield
Herd variance	311±37	935,050±46,286	1,668±79	3,035±142
Additive genetic variance	177±32	438,300±17,244	532±23	833±39
Permanent environment variance	190±39	249,650±1,1802	351±16	562±28
Residual variance	3,703±41	700,850±4,541	1,078±70	2,144±14
Heritability	0.04±0.01	0.19±0.01	0.15±0.01	0.13±0.01
Intra-herd heritability	0.04±0.01	0.32±0.01	0.27±0.01	0.23±0.01
Repeatability	0.08±0.01	0.29±0.01	0.24±0.01	0.21±0.01
Intra-herd repeatability	0.09±0.01	0.49±0.01	0.45±0.01	0.39±0.01

The variance component estimates indicated that the herd effect had a significant influence on the production traits, with herd variance being larger than the additive genetic variance. Additionally, the highest variability for DO was absorbed by the residual variance. Production traits exhibited low-to-moderate h^2 values (0.19 ± 0.01 for MY; 0.15 ± 0.01 for PY; 0.13 ± 0.01 for FY), whereas the h^2 for DO was low (0.04 ± 0.01). The IH h^2 estimates showed distinct patterns across trait types (Table 3). For DO, h^2_{IH} remained essentially unchanged at 0.04 ± 0.01. Conversely, production traits demonstrated a marked increase in IH h^2 values (0.32 ± 0.01 for MY; 0.27 ± 0.01 for PY; 0.23 ± 0.01 for FY). Similarly, repeatability estimates followed the same pattern observed for h^2. For DO, repeatability was slightly higher than h^2 (0.08 ± 0.01), while IH repeatability increased to 0.09 ± 0.01. In contrast, production traits showed substantially higher values, with repeatability estimates of 0.29 ± 0.01 for MY, 0.24 ± 0.01 for PY, and 0.21 ± 0.01 for FY, and markedly greater IH repeatabilities of 0.49 ± 0.01, 0.45 ± 0.01, and 0.39 ± 0.01, respectively (Table 3).

The genetic correlation between DO and MY (Table 4) was low and positive. A similar positive correlation was found between DO and PY, whereas the largest was found between DO and FY. Genetic correlations among the 3 production traits were large and positive.

Table 4. Genetic correlations (below the diagonal), residual correlations (in parentheses), and phenotypic correlations (above the diagonal) among the traits considered in the multiple-trait model[1]

Item	Days Open	Milk Yield	Protein Yield	Fat Yield
Days Open		0.13***	0.13***	0.16***
Milk Yield	0.16±0.08 (0.184)		0.95*	0.80*
Protein Yield	0.12±0.09 (0.20)	0.86±0.01* (0.96)		0.86*
Fat Yield	0.23±0.09* (0.17)	0.58±0.03* (0.87)	0.78±0.02* (0.87)	

[1] Phenotypic correlations were tested for significance with a Pearson test (*P < 0.05, ***P < 0.001). Genetic correlations were considered significant when the 95% posterior credible interval (HPD95%) from the Gibbs sampling did not include zero; significant estimates are also indicated with asterisks for consistency. The HPD95% intervals for genetic correlations were as follows: MY – PY [0.84;0.88], MY – FY [0.52;0.63], FY – PY [0.74;0.81], DO – MY [0.00;0.31], DO – PY [-0.04;0.29], DO – FY [0.06;0.41].

The genetic trends over the period from 1990 to 2020 are presented in Figure 1 and were obtained from both bulls and cows included in the evaluation. Only animals with pedigree and performance records meeting the quality control thresholds described in Materials and Methods were retained.

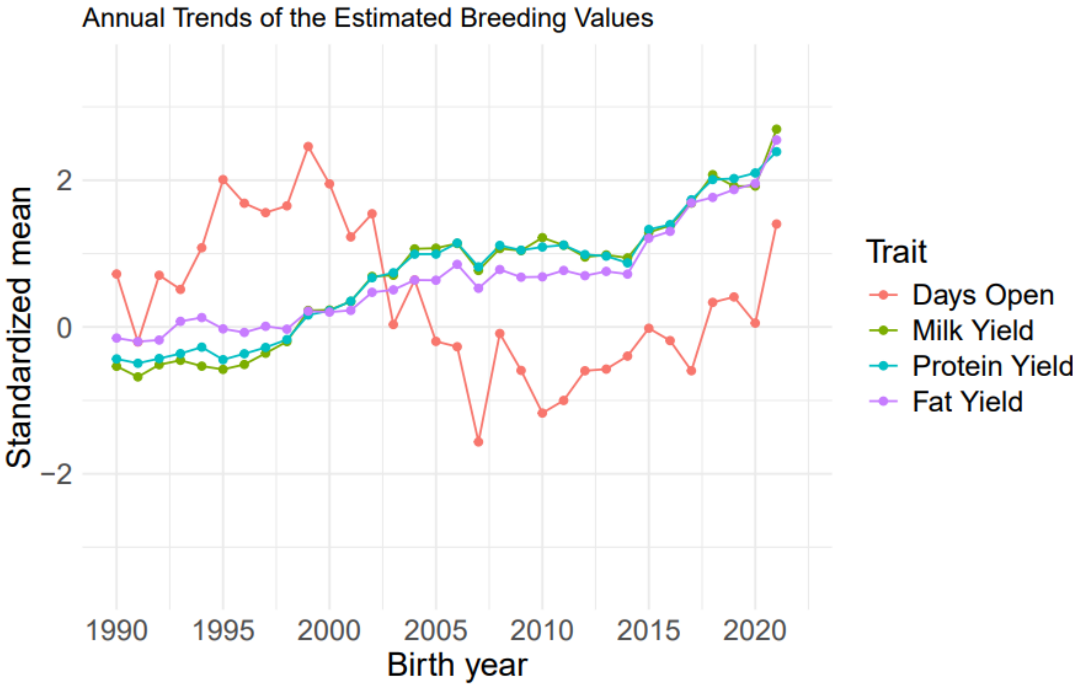


Figure 2. Standardization means of Annual EBV trends per birth year.

The EBV for production traits showed a consistently increasing trend throughout the period examined. Conversely, the evolution of EBV trend for DO presented a less definite pattern. An initial increase was observed from 1990 to 1998, followed by a declining phase from 1998 to 2006, and then a new increasing trend that continued until 2020.

Official Genetic Evaluation

In the reduced dataset used for mimicking an official evaluation for DO, the average reliabilities of EBV were 21.91 ± 6.21 and 52.87 ± 13.64 for cows and bulls, respectively. The average phenotypic DO of daughters varied according to the genetic merit of their sires for DO (Table 5). Specifically, daughters of high-index bulls exhibited the shortest average DO, followed by those of medium- and low-index bulls (Table 5), respectively. The difference in average DO between high- and medium-index categories was 14 d, and the difference between high- and low-index categories was 21 d. A difference of 7 d was observed between the medium- and low-index groups. Comparisons with the overall population (μ_{DO}) and the μ_{gb}, are presented in Table 5. The μ_{DO}, and the μ_{gb}, defined using cows born between 2017 and 2019, were both 121 d. The minimal difference between these 2 reference points suggests that the average DO in the population has remained relatively stable in the years considered.

Table 5. Comparisons among sires classified into 3 categories (high, medium, and low) according to their EBV for DO[1]

Group	No. Bulls	Daughters mean, days	Index - μ_{DO}, days	Index- μ_{gb}, days
High-index	77	110^{A}	-11	-11
Medium-index	134	124^{B}	2	2
Low-index	53	131^{C}	10	10

$^{A-C}$Superscript letters indicate significant differences ($P < 0.05$) among sire categories.

[1] Shown are the mean phenotypic DO of the daughters and the deviations of each sire category from the μDO and μgb.

Discussion

In the present study, the genetic variability of the calving-to-conception interval and its relationships with production traits in Italian Jersey cattle were investigated. Results showed a low h^2 for DO (around 0.04), and low-to-moderate positive genetic correlations with production traits, confirming the existence of a biological antagonism between fertility and production performance. The analysis of the genetic trend for DO revealed a nonlinear pattern over time, with periods of improvement followed by phases of decline. This irregularity may reflect indirect selection for production traits, the relatively small size of the Jersey population in Italy, and the fact that many herds are multibreed with predominantly Holstein cows, where less emphasis is placed on genetic improvement of Jerseys, potentially leading to greater fluctuations in their genetic trends. The phenotypic impact indexes revealed substantial differences in reproductive performance of daughters according to sire genetic merit, with differences up to 20.89 d between extreme categories.

The DO means observed in this study agree with previous reports for Jersey cattle. For instance, Silva et al. (1992) reported a mean DO of 116 d in American Jerseys, whereas Zambrano and Echeverri (2014) found a slightly higher mean of 125.62 d in Colombian Jerseys. Even higher values were reported in American Jerseys by Campos et al. (1994), who estimated a mean of 127.15 d. In 2019, according to the Council on Dairy Cattle Breeding (CDCB),the average DO for Jersey in the United States was 115 d (CDCB, 2025).

As far as production traits are concerned, phenotypic means obtained in this study are lower than those reported in other countries for Jersey cattle. The latest data from CDCB (2025) for American Jersey cattle showed substantially higher values: 8,827, 441, and 330 kg for MY, FY, and PY, respectively. Similarly, Nordic data from Nordic Cattle Genetic Evaluation (NAV, 2025) report averages of 7,713 kg for MY, 459 kg FY, and 333 kg for

PY. Differences observed between the phenotypic means reported in this study and those published in the literature or in official statistics may be attributed to several reasons. One is the data source and population coverage: the present study refers to particular subsets of herds that are not fully representative of the national average. Specifically, the herds included both farms where Jersey is the predominant breed and many multibreed farms with only a limited number of Jersey cows, which nevertheless contribute to data collection. This composition of the dataset may partly explain the lower mean yields observed in our study compared with official national statistics. Other reasons could be attributed to different management practices (nutrition, reproductive strategies, and herd health programs) across countries (Lucy, 2001). Moreover, genetic composition and intensity of selection programs may differ. The markedly higher production levels observed in American Jersey cattle may be attributed to the long-term and intensive selection for milk yield applied in the American population (Huson et al., 2020; Guinan et al., 2023). Although American genetics have been introduced into the Italian Jersey population, the overall selection intensity for milk production traits has historically been lower, which may explain the phenotypic gap.

Standard and IH h^2 estimated for DO were relatively low. Lower values of h^2 (approximately 0.03) have been reported in American Jerseys, German Holsteins, and Northern Thai dairy cattle (Campos et al., 1994; König et al., 2005; Interbull, 2025).Values ranging from 0.05 to 0.09 were observed in Colombian Holstein and Jersey populations, as well as in Indian Karan Fries cattle (Zambrano and Echeverri, 2014; Worku et al., 2021). Higher estimates (0.06–0.07) have been reported for Canadian, Danish, Dutch, Holstein, and Jersey cattle (Interbull, 2025). Repeatability estimates obtained in the present study were in line with values reported in the literature for dairy cattle fertility traits (Zambrano and Echeverri, 2014; Worku et al., 2021). Differences among studies could be ascribed to the model used; for example, in some studies the herd is treated as a fixed effect whereas in others, including the present research, it is considered random. Days open exhibited a pronounced pattern in the variance decomposition. The additive genetic variance was relatively low, whereas the residual variance was extremely high. This aligns with literature findings indicating that fertility is largely influenced by nongenetic factors such as nutrition, herd health, and reproductive management practices (Garnsworthy, 2007; Cassandro, 2014; Shortle, 2014**)**. Further differences can be the result of the widespread use of hormonal protocols for ovulation synchronization, and fixed-time artificial insemination may mask underlying genetic variation in fertility traits (Oliviera et al., 2021). Such treatments enhance reproductive performance even in less fertile cows, thereby reducing the observable phenotypic variability. As a result, key indicators become biased, leading to reduced accuracy of genetic evaluations, reranking of sires, and potential distortion of long-term genetic trends. In the absence of reliable records on the use of hormonal treatments at the herd or animal level, fertility data may thus introduce systematic bias into genetic evaluation programs (Oliviera et al., 2021). For the Italian Jersey population, the extent of the use of hormonal protocols remains unknown and therefore cannot be quantified. This pattern suggests that the scope for genetic improvement of fertility in Italian Jersey cattle population is limited, while management interventions remain crucial for achieving substantial progress in reproductive performance.

Standard and IH h^2 for production traits estimated in the present study were low to moderate. Roman et al. (2000) reported h^2 estimates of 0.26, 0.31, and 0.17 for MY, FY, and PY in Jersey cattle. Higher h^2 values were also reported for Danish Jersey cattle, whereas more comparable estimates were observed in American Jersey cattle (Interbull, 2025). The variance component estimates showed that a substantial portion of the phenotypic variance was attributable to the herd effect. This is reflected in the higher h^2_{IH} (0.32 for MY, 0.23 for FY, and 0.27 for PY), underscoring the significant influence of herd-specific environmental and management factors. These results emphasize the importance of appropriately modeling herd effects to accurately assess the genetic potential of animals across different production environments.

Moderate and unfavorable genetic correlations were observed between DO and production traits, in line with previous findings. Toghiani (2012) reported unfavorable correlations between DO and MY (0.35), PY (0.62), and FY (0.39), respectively. Comparable results for the correlation between DO and FY were also reported by Hermas et al. (1987). These genetic correlations can be attributed to the low h^2 of reproductive traits and their strong dependence on environmental and management conditions, such as feeding strategies, herd health status, reproductive management, and climatic stressors (González-Recio and Alenda, 2005; Kgari et al., 2021). This makes simultaneous genetic improvement of fertility and production traits particularly challenging, especially when selection historically favors high-yielding animals (Lucy, 2001). This antagonistic relationship may stem from competition for metabolic resources, particularly during peak lactation, when cows are expected to resume reproductive activity (Royal et al., 2000). Negative energy balance, mobilization of body fat reserves, and increased incidence of metabolic and hormonal disorders can delay the return to ovarian cyclicity. More recently, the inclusion of fertility traits in multitrait selection indexes has aimed to reverse the negative fertility trend observed in high-producing dairy cows (Crowe, 2008; Norman et al., 2009; Kgari et al., 2021; Hu et al., 2021). In fact, over the past 2 decades, breeding goals in many countries have shifted from an exclusive focus on milk production to more balanced indexes that also include fertility, health, and longevity (Miglior et al., 2017). The inclusion of functional productive life and fertility traits in national selection indexes (e.g., TPI in the US, LPI in Canada, and Nordic indexes) has allowed continued genetic progress for milk yield while reducing the negative trends previously observed in fertility and longevity (Miglior et al., 2017; VanRaden et al., 2004).

These balanced indexes demonstrated that it is possible to achieve ~90% of the genetic gain for production while maintaining or improving functional traits at the same time. In the present study, most correlations of DO with production traits were weak, and their 95% CI included zero, indicating that they cannot be truly considered different from zero. An exception was the correlation between FY and DO, which was positive and significant (Table 4). Overall, these results suggest that strong positive relationships are largely confined to production traits themselves, with only limited evidence of an antagonistic relationship with fertility in this Jersey dataset. This pattern supports the interpretation that selection for milk components has a limited correlated response on reproductive performance under the studied conditions. Taken together, our findings indicate that in the Italian Jersey population the fertility-production antagonism is less pronounced than reported in other dairy cattle populations.

Genetic and phenotypic trends for production and fertility traits were markedly different (Figure 1, Figure 2 and 2b).

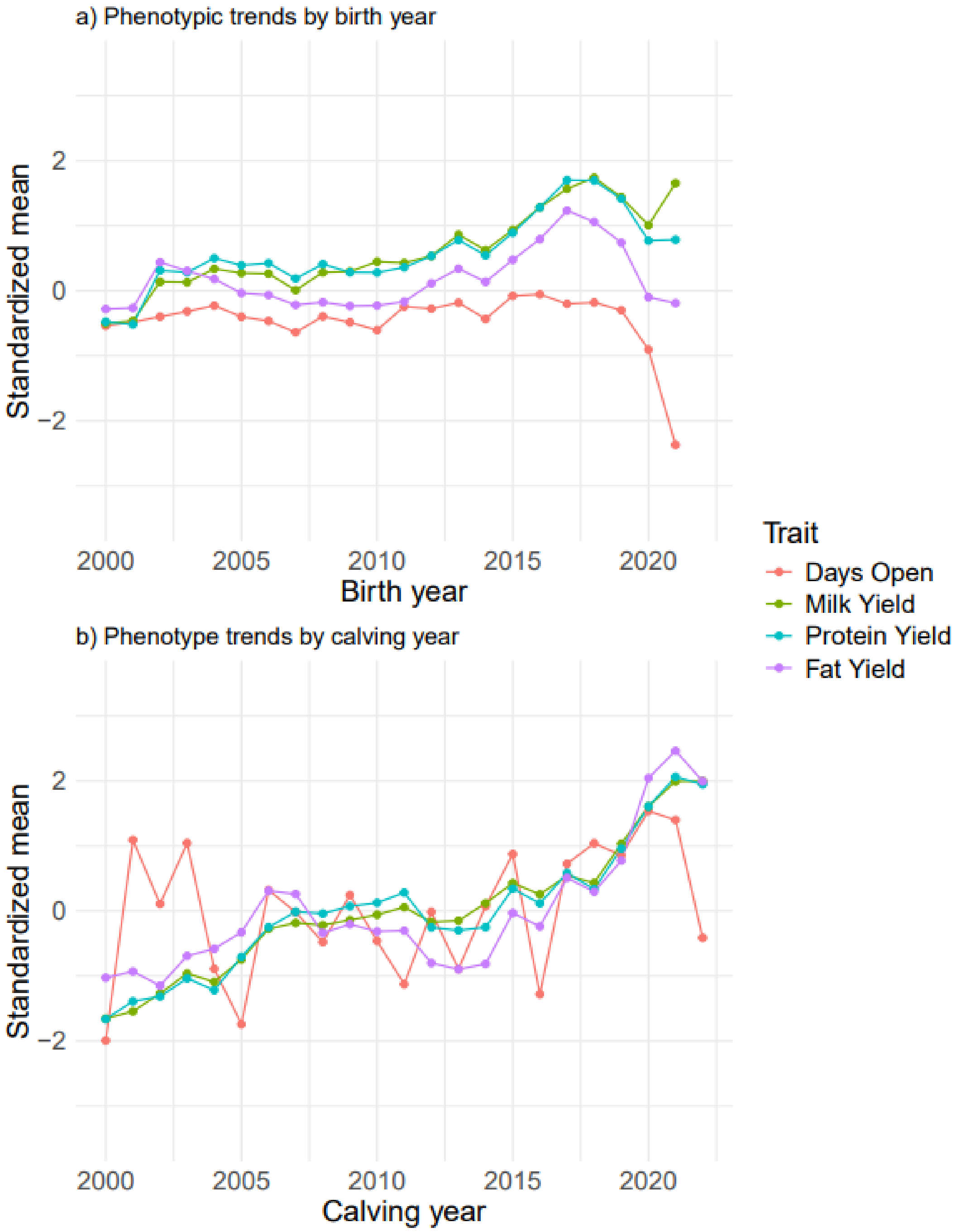


Figure 1. Annual means of phenotypic trends of the considered traits per birth year (a) and calving year (b).

The EBV for production traits (MY, FY, and PY; Figure 1) showed a steady increase over time, with a more pronounced rise between 2016 and 2020, reflecting the effectiveness of selection programs aimed at enhancing production performance. In contrast, the EBV trend for DO was more variable, with a peak in 1998, a trough in 2006, and a moderate increase from 2016 onward, indicating the absence of consistent direct selection for fertility. Phenotypic trends provided complementary insights (Figure 2). When expressed by birth year (Figure 2a), phenotypic DO remained largely stable until 2015, followed by a decline, whereas when expressed by calving year (Figure 2b), DO showed pronounced fluctuations, highlighting its strong dependence on environmental and management factors. For production traits, both genetic and phenotypic trends were positive, with phenotypic gains accelerating after 2015. The variability in DO trends likely reflects the relatively small size of the Italian Jersey population, its frequent management within multibreed herds, and the strong influence of nongenetic factors on fertility performance.

Official genetic evaluations in Italian Jersey cattle population are currently carried out for milk production traits, udder conformation, and feet and legs, whereas no fertility traits are included in the national selection index (ANAFIBJ, 2025). The present Italian Jersey Selection Index (IQJ) assigns 84% of the weight to production traits and 16% to conformation traits. Specifically, the included traits and their relative weights are milk (kg), 19%; protein (kg), 65%; udder support, 2%; fore udder attachment, 2%; and udder depth, 12%. The IQJ is calculated as follows:

$$\text{IQJ} = 1.33 \times (-0.03 \times \text{Milk (kg)} + 2.94 \times \text{Protein (kg)} + 1.55 \times \text{Fore Udder Attachment} + 1.51 \times \text{Udder Support} + 13.06 \times \text{Udder Depth})$$

This structure likely explains the steady increase observed in EBV for production traits over recent years, as well as the concurrent rise—reflecting worsening performance—in DO. By contrast, in countries such as the United States, fertility traits have been formally included in national selection indexes since the early 2000s (Norman et al., 2009). This strategic shift has contributed to a gradual reversal of previously unfavorable genetic trends for reproductive traits, resulting in moderate improvements at both the genetic and phenotypic levels (Norman et al., 2009; Guinan et al., 2023). The aim of this study was precisely to investigate fertility trends in this context, to provide the groundwork for the future inclusion of fertility traits in the national genetic evaluation system.

The observed differences in average DO across sire categories based on their EBV for DO suggest a clear association between sire genetic merit for fertility and reproductive performance. Daughters of bulls with favorable EBV for DO exhibited significantly shorter DO compared with those of medium- and low-EBV sires, with a gap of more than 20 d between the 2 extreme categories. The observed differences among sire categories, together with the corresponding phenotypic means of their daughters, suggest that indirect selection may be taking place, possibly through traits genetically correlated with fertility, such as longevity or udder health. Supporting this hypothesis, Liu et al. (2008) reported favorable genetic correlations between fertility EBV and traits like udder health and longevity in Holstein populations from Germany, Austria, and Luxembourg.

Moreover, the minimal difference between the overall population mean and the μ_{gb} for DO suggests that phenotypic progress for this trait has been stagnant in recent years. This reflects the lack of direct genetic pressure on reproductive traits, underscoring the need to revise current breeding goals. Incorporating fertility traits (such as DO) into the national selection index could enhance the effectiveness of genetic improvement programs and mitigate the negative effects historically associated with intense selection for production (Liu et al., 2008).

Fertility traits in dairy cattle generally show low h^2 estimates, reflecting the strong influence of environmental and management factors. However, low h^2 does not necessarily imply limited potential for genetic improvement. As highlighted by Berry et al. (2017), the CVg provides a more informative measure of the potential for genetic change, because it expresses additive genetic variability relative to the mean of the trait. This metric can reveal substantial evolutionary potential even in traits with low h^2. In fact, fertility traits in dairy cattle often exhibit CVg values up to 7%, comparable with those of highly heritable traits such as MY or BW (Houle, 1992; Berry et al., 2017). In our study, the CVg for DO was estimated at 11.4%. Such a value is relatively high and indicates that, despite the low h^2 of fertility traits, there is substantial exploitable genetic variability for this trait.

Looking ahead, the National Association (ANAFIBJ) will apply an internal algorithm to evaluate genetic progress over a 10-year horizon, including DO as a fertility trait, as well as SCC and longevity. The aim is to slightly shift selection toward improved functional traits, without compromising selection for MY, milk composition, and overall performance.

Conclusion

This study updates genetic parameter estimates for fertility and production traits in the Italian Jersey cattle population. The results confirm the low h^2 of reproductive traits and suggest that fertility is primarily influenced by environmental factors. The low h^2 of DO and its unfavorable genetic correlations with MY, PY, and FY reflect the typical antagonism between fertility and production, highlighting the difficulty of improving reproductive performance through selection focused solely on productive traits. The genetic trend for DO is declining, indicating reduced fertility in recent years, despite significant improvements in MY and components. Nonetheless, the clear differences in average DO across sire EBV classes indicate that there is still genetic variability for fertility traits, and therefore real potential for genetic improvement through targeted selection. Integrating fertility traits like DO into breeding programs is crucial for balanced genetic progress, improved reproductive efficiency, and long-term herd sustainability.

Acknowledgements

This study was supported by ‘Latteco2 project, sottomisura 10.2 of the PSRN-Biodiversity 2020–2023‘ (MIPAAF. D.M. no. 465907 del 24/09/2021, project unique code J12C21004080005) and by “Mining big-data to fit animal to climate change, improve welfare and mitigate the environmental impact of livestock productions” (BIGFit project, CUP J53D23010110006, Finanziamento dell'Unione Europea – NextGenerationEU). The authors gratefully acknowledge the Italian Holstein, Brown Swiss, and Jersey Association (ANAFIBJ, Cremona, Italy) for providing data and the dairy herd owners for generating and recording the data and for permitting its use in this study. No human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use

Committee or Institutional Review Board. The authors have not stated any conflicts of interest.

References

ANAFIBJ. 2025. Associazione nazionale allevatori Frisona, Bruna e Jersey. Accessed Jul. 22, 2025. https://www.anafi.it

Campos, M. S., C. J. Wilcox, C. M. Becerril, and A. Diz. 1994. Genetic parameters for yield and reproductive traits of Holstein and Jersey cattle in Florida. J. Dairy Sci. 77:867–873. https://doi.org/10.3168/jds.S0022-0302(94)77021-1

Cassandro, M. 2014. Genetic aspects of fertility traits in dairy cattle—review. Acta Agrar. Kaposváriens. 18:11–23.

Cesarani, A., G. Gaspa, Y. Masuda, L. Degano, D. Vicario, D. A. Lourenco, and N. P. Macciotta. 2020. Variance components using genomic information for 2 functional traits in Italian Simmental cattle: Calving interval and lactation persistency. J. Dairy Sci. 103:5227–5233. https://doi.org/10.3168/jds.2019-17421

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. 18:237–268. https://www.jstor.org/stable/2097132

Council on Dairy Cattle Breeding (CDCB). 2025. Accessed Jul. 22, 2025. https://uscdcb.com

Crowe, M. A. 2008. Resumption of ovarian cyclicity in post-partum beef and dairy cows. Reprod. Domest. Anim. 43:20–28. https://doi.org/10.1111/j.1439-0531.2008.01210.x

De Kruif, A. 1978. Factors influencing the fertility of a cattle population. Reproduction 54:507–518. https://doi.org/10.1530/jrf.0.0540507

Fabris, A., D. T. M. Stella Maestri, R. Finocchiaro, M. Marusi, and M. Cassandro. 2025. Genetic evaluation of longevity in Italian Jersey. Ital. J. Anim. Sci. 24:68–77. https://doi.org/10.1080/1828051X.2024.2440124

Fathoni, A., W. Boonkum, V. Chankitisakul, S. Buaban, and M. Duangjinda. 2025. Pleiotropic genes affecting milk production, fertility, and health in Thai-Holstein crossbred dairy cattle: A GWAS approach. Animals 15:1320. https://doi.org/10.3390/ani15091320

Garnsworthy, P. 2007. Body condition score in dairy cows: Targets for production and fertility. Pages 61–86 in Recent Advances in Animal Nutrition.

Garnsworthy, P. C., K. D. Sinclair, and R. Webb. 2008. Integration of physiological mechanisms that influence fertility in dairy cows. Animal 2:1144–1152. https://doi.org/10.1017/S1751731108002358

Goni, S., C. J. C. Muller, B. Dube, and K. Dzama. 2015. Milk production of Jersey and Fleckvieh × Jersey cows in a pasture-based feeding system. Trop. Anim. Health Prod. 47:139–144. https://doi.org/10.1007/s11250-014-0698-y

González-Recio, O., and R. Alenda. 2005. Genetic parameters for female fertility traits and a fertility index in Spanish dairy cattle. J. Dairy Sci. 88:3282–3289. https://doi.org/10.3168/jds.S0022-0302(05)73011-3

Guinan, F. L., G. R. Wiggans, H. D. Norman, J. W. Dürr, J. B. Cole, C. P. Van Tassell, D. Lourenco, et al. 2023. Changes in genetic trends in US dairy cattle since the implementation of genomic selection. J. Dairy Sci. 106:1110–1129. https://doi.org/10.3168/jds.2022-22205

Hansen, L. B., A. E. Freeman, and P. J. Berger. 1983. Variances, repeatabilities, and age adjustments of yield and fertility in dairy cattle. J. Dairy Sci. 66:281–292. https://doi.org/10.3168/jds.S0022-0302(83)81788-3

Hermas, S. A., C. W. Young, and J. W. Rust. 1987. Genetic relationships and additive genetic variation of productive and reproductive traits in Guernsey dairy cattle. J. Dairy Sci. 70:1252–1257. https://doi.org/10.3168/jds.S0022-0302(87)80138-8

Howard, J. T., J. E. Pryce, C. Baes, and C. Maltecca. 2017. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100:6009–6024. https://doi.org/10.3168/jds.2017-12787

Hu H, Mu T, Ma Y, Wang X, Ma Y. (2021). Analysis of longevity traits in Holstein cattle: a review. *Front. Genet.* **12**:695543. https://doi.org/10.3389/fgene.2021.695543

Huson, H. J., T. S. Sonstegard, J. Godfrey, D. Hambrook, C. Wolfe, G. Wiggans, C. P. Van Tassell, et al. 2020. A genetic investigation of Island Jersey cattle, the foundation of the Jersey breed: Comparing population structure and selection to Guernsey, Holstein, and United States Jersey cattle. Front. Genet. 11:366. https://doi.org/10.3389/fgene.2020.00366

Interbull Centre. N.d. International Bull Evaluation Service. Accessed Apr. 2025. https://interbull.org

Junior, G. O., Schaeffer, L. R., Schenkel, F., Tiezzi, F., & Baes, C. F. (2021). Potential effects of hormonal synchronized breeding on genetic evaluations of fertility traits in dairy cattle: A simulation study. *Journal of Dairy Science, 104*(4), 4404–4412. https://doi.org/10.3168/jds.2020-18944

Kgari, R. D., C. J. C. Muller, K. Dzama, and M. L. Makgahlela. 2020. Evaluation of female fertility in dairy cattle enterprises—A review. S. Afr. J. Anim. Sci. 50. http://dx.doi.org/10.4314/sajas.v50i6.8

Lidauer, M., K. Matilainen, E. Mäntysaari, T. Pitkänen, M. Taskinen, and I. Strandén. 2022. MiX99: A software package for solving large mixed model equations. Natural Resources Institute Finland (Luke), Jokioinen, Finland. Release I/2022. FI-31600. https://www.luke.fi/sites/default/files/2022-06/refguide_solver.pdf

Liu, Z., J. Jaitner, F. Reinhardt, E. Pasman, S. Rensing, and R. Reents. 2008. Genetic evaluation of fertility traits of dairy cattle using a multiple-trait animal model. J. Dairy Sci. 91:4333–4343. https://doi.org/10.3168/jds.2008-1029

Louca, A., and J. E. Legates. 1968. Production losses in dairy cattle due to days open. J. Dairy Sci. 51:573–583. https://doi.org/10.3168/jds.S0022-0302(68)87031-6

Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: Where will it end? J. Dairy Sci. 84:1277–1293.https://doi.org/10.3168/jds.S0022-0302(01)70158-0

Makgahlela, L. 2008. Calving interval now included in the national genetic evaluation. Natl. Milk Rec. Improv. Scheme Newsl. 13:20.

Miglior, F. 1999. Selection for fertility traits. Proc. 24th EHFC Meet. 110:114.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., and C. F. Baes. 2017. A 100-year review: Identification and genetic selection of economically important traits in dairy cattle. *J. Dairy Sci.* 100:10251–10271.https://doi.org/10.3168/jds.2017-12968

Misztal, I., S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2022. BLUPF90 family of programs. Accessed Jul. 2025. http://nce.ads.uga.edu/software/

Muller, C. J. C., J. P. Potgieter, S. W. P. Cloete, and K. Dzama. 2014. Non-genetic factors affecting fertility traits in South African Holstein cows. S. Afr. J. Anim. Sci. 44:54–63. http://dx.doi.org/10.4314/sajas.v44i1.8

NAV. 2025. Nordic Cattle Genetic Evaluation. Accessed Jul. 2025. https://nordicebv.info

Norman, H. D., J. R. Wright, S. M. Hubbard, R. H. Miller, and J. L. Hutchison. 2009. Reproductive status of Holstein and Jersey cows in the United States. J. Dairy Sci. 92:3517–3528. https://doi.org/10.3168/jds.2008-1768

Philipsson, J. 1981. Genetic aspects of female fertility in dairy cattle. Livest. Prod. Sci. 8:307–319. https://doi.org/10.1016/0301-6226(81)90049-X

Pritchard, T., Coffey, M., M., Mrode, R., & Wall, E. (2013). Genetic parameters for production, health, fertility and longevity traits in dairy cows. *Animal*, 7(1), 34–46. https://doi.org/10.1017/S1751731112001401

Pryce, J. E., and R. F. Veerkamp. 2001. The incorporation of fertility in genetic improvement programs. BSAP Occas. Publ. 26:237–249. https://doi.org/10.1017/S0263967X00033711

Roche JR, Burke CR, Crookenden MA, Heiser A, Loor JL, Meier S, Mitchell MD, Phyn CVC, Turner SA. (2018). Fertility and the transition dairy cow. *Reprod. Fertil. Dev.* 30(1):85–100. https://doi.org/10.1071/RD17412

Roche, J. F., D. Mackey, and M. D. Diskin. 2000. Reproductive man-agement of postpartum cows. Anim. Reprod. Sci. 60–61:703–712. https://doi.org/10.1016/S0378-4320(00)00107-X

Roman, R. M., C. J. Wilcox, and F. G. Martin. 2000. Estimates of repeatability and heritability of productive and reproductive traits in a herd of Jersey cattle. Genet. Mol. Biol. 23:113–119. https://doi.org/10.1590/S1415-47572000000100021

Royal, M. D., A. O. Darwash, A. P. F. Flint, R. Webb, J. A. Woolliams, and G. E. Lamming. 2000. Declining fertility in dairy cattle: Changes in traditional and endocrine parameters of fertility. Anim. Sci. 70:487–501. https://doi.org/10.1017/S1357729800051845

Shanks, R. D., A. E. Freeman, and F. N. Dickinson. 1981. Postpartum distribution of costs and disorders of health. J. Dairy Sci. 64:683–688. https://doi.org/10.3168/jds.S0022-0302(81)82630-6

Shortle, W. (2014). Fertility in dairy cattle: a review. http://hdl.handle.net/10832/1253

Silva, H. M., C. J. Wilcox, W. W. Thatcher, R. B. Becker, and D. Morse. 1992. Factors affecting days open, gestation length, and calving interval in Florida dairy cattle. J. Dairy Sci. 75:288–293. https://doi.org/10.3168/jds.S0022-0302(92)77764-9

Sven, K., N. Chatthong, and H.-J. Lengerken. 2005. Estimation of variance components for production and fertility traits in Northern Thai dairy cattle to define optimal breeding strategies. Arch. Tierz. 48:233–246. https://doi.org/10.5194/aab-48-233-2005

Temesgen MY, Assen AA, Gizaw TT, Minalu BA, Mersha AY. (2022). Factors affecting calving to conception interval (days open) in dairy cows located at Dessie and Kombolcha towns, Ethiopia. *PLoS One* **17**(2):e0264029. https://doi.org/10.1371/journal.pone.0264029

Tiezzi, F., C. Maltecca, M. Penasa, A. Cecchinato, Y. M. Chang, and G. Bittante. 2011. Genetic analysis of fertility in the Italian Brown Swiss population using different models and trait definitions. J. Dairy Sci. 94:6162–6172. https://doi.org/10.3168/jds.2011-4661

Toghiani, S. 2012. Genetic relationships between production traits and reproductive performance in Holstein dairy cows. Arch. Anim. Breed. 55:458–468. https://doi.org/10.5194/aab-55-458-2012

VanRaden, P. M., Tooker, M. E., Wright, J. R., Sun, C., and J. L. Hutchison. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J. Dairy Sci.* 97:7952–7962. https://doi.org/10.3168/jds.2014-8489

Worku, K., Y. Kechero, and G. P. Janssens. 2021. Measuring seasonal and agro-ecological effects on nutritional status in tropical ranging dairy cows. J. Dairy Sci. 104:4341–4349. https://doi.org/10.3168/jds.2020-18995

Zambrano, J. C., J. C. Rincón, and J. J. Echeverri. 2014. Genetic and environmental variance and covariance parameters for some reproductive traits of Holstein and Jersey cattle in Antioquia, Colombia. Rev. Bras. Zootec. 43:132–139.

CHAPTER 5

Application of the Legarra & Reverter Method to Evaluate Carcass, Reproductive, and Weaning Traits in Three beef Cattle Breeds

Abstract

The reliability and stability of breeding value predictions are central in genetic improvement programs. The Linear Regression method (LRM) provides a framework to assess prediction quality through accuracy, bias, and dispersion metrics. The aim of this study was to evaluate the quality of pedigree-based estimated breeding values (EBV) across multiple beef cattle breeds, traits, and national evaluation contexts. EBV for carcass, calving, and growth traits were validated using the LRM by comparing predictions from partial and complete datasets of young bulls. Across breeds and countries, accuracy ratio showed wide variability, ranging from 0.15 to 0.90. Traits with higher heritability consistently showed greater predictive ability, with growth traits reaching accuracies above 0.85 in several national evaluation contexts. In contrast, carcass-related traits showed lower and more heterogeneous values (as low as 0.15). Bias values varied substantially across scenarios (−3.59 to +5.83), indicating systematic over- and under-prediction depending on breed, country, and genetic effect. In growth trait, maternal effects were less stable than direct effects, as they showed opposite bias patterns across populations and larger scaling deviations from unity. Dispersion coefficients ranged from strong under-dispersion (<0.70) to over-dispersion (>1.10), confirming substantial differences in scaling behavior across evaluation systems. Weighted linear models highlighted that country was the most influential factor across all LRM metrics ($P < 0.0001$), while heritability and its interaction with trait significantly affected accuracy ratio, bias, and dispersion ($P < 0.0001$). The overall explanatory power of the models was highest for accuracy ratio ($R^2 = 64.1\%$), and lower for bias (35.6%) and dispersion (26.7%), indicating that systematic factors explain predictive ability better than scaling properties. These results highlight that EBV predictive performance is strongly context-dependent and influenced by the genetic architecture of the trait and the structure of national evaluation systems.

Keywords: Beef cattle, Linear Regression, weighted model, national evaluation

Introduction

Bias in genetic evaluations has long been a concern in animal breeding, and recent studies have renewed attention to its impact, reporting overestimation of Estimated Breeding Values (EBV) (Macedo et al., 2020a). Because genetic models are practical simplifications of reality, some deviation between estimated and true breeding values is inevitable (Kennedy et al., 1988); however, this deviation should remain limited, and its detection is central to evaluate the realized predictive ability of EBV (Palucci & Dürr, 2014). Cross-validation has become the standard approach to compare genetic evaluation methods, typically relying on correlations between predicted values and subsequent records or complete evaluations. However, the analytical interpretation of these validation measures, particularly in terms of bias and accuracy, is not always clearly formalized. Moreover, appropriate validation tools remain limited for complex models, such as those including maternal effects or threshold traits, and for breeding schemes with small progeny groups (Legarra and Reverter, 2018). For Interbull dairy evaluations, several standardized procedures are available (Interbull Methods I–III, Mendelian Sampling Test, GEBV Test) (Boichard et al., 1995; Tyrisevä et al., 2011; Sullivan et al., 2009) and, in principle, they could be extended to beef cattle. In practice, however, important differences in beef breeding schemes, such as small progeny groups, more heterogeneous recording schemes, low-heritability traits, and the inclusion of maternal effects, limit direct transferability (Eriksson et al., 2004; VanRaden et al., 2014). A recent approach proposed by Legarra and Reverted, 2018, is the Linear Regression method (LRM). This method addresses the limitations of current methods used for the validation of genetic or genomic predictions by comparing EBV or Genomic Breeding Values (GEBV) obtained from two types of datasets: a partial dataset ($data_P$), containing only older or limited information and a whole dataset ($data_W$), containing all available information (old and new).

The LRM evaluates the relationship between breeding values obtained from $data_p$ and $data_w$ datasets in order to assess the quality of genetic evaluations providing estimates of bias, dispersion (slope), and accuracy (Legarra and Reverter, 2018; Macedo et al., 2019). In this approach, breeding values from the whole dataset (EBV_w), based on more complete information, are used as the reference variable, whereas breeding values from the partial dataset (EBV_p) are used as the predictor variable. Unlike the common assumption that the regression between EBV_w and EBV_p equals one, the LRM demonstrates that this regression tends to be systematically lower than one when validation sets are small or contain closely related individuals due to family structure effects (Legarra and Reverter, 2018). For sufficiently large and independent validation sets, LRM statistics yield unbiased estimates providing a reliable framework to evaluate genetic evaluation performance (Macedo et al., 2019). In this context, the definition of the validation cohort is a crucial step because it directly affects LRM statistics and their interpretation (Legarra and Reverter, 2018). In practice, it consists of identifying the focal animals on which the comparison between evaluations is conducted. Depending on the objective, the cohort may include the whole population or a target subset—most commonly young selection candidates or sires whose information increases over time—so that model performance is assessed in the group most relevant for selection decisions (Legarra and Reverter, 2018). Cohort size and structure are key: small and highly related cohorts can distort regression outcomes, yielding slopes below one even under a correct model and thus mimicking EBV inflation (Legarra and Reverter, 2018). Large cohorts are therefore preferable whenever possible. Moreover, family structure influences the expected gain from genomics, with larger families enhancing within-family discrimination. Finally, cohort composition may introduce bias itself: if animals are pre-selected (e.g., elite young bulls chosen on parent average), their realized accuracy is expected to be lower than in an unselected population, and this must be considered when interpreting validation results (Legarra and Reverter, 2018; Macedo et al., 2020b). To avoid arbitrary cohort definition, objective criteria should be considered when constructing validation groups. Animals should belong to the same selection stage—typically young candidates—and experience a clear increase in information between partial and whole evaluations (Legarra and Reverter, 2018; Macedo et al., 2020b; Bonifazi et al., 2021). This increase can be verified using measures such as reliability, effective record contribution, progeny size, or number of phenotypic records (Misztal et al., 2020; Bonifazi et al., 2021). In addition, cohorts should be sufficiently large and genetically heterogeneous to avoid family-driven artifacts in LR statistics (Legarra and Reverter, 2018). Under these conditions, validation results reflect true changes in information rather than cohort-specific effects.

Building on the insights provided in the literature, where the LRM has been proposed as a robust and semi-parametric tool for assessing bias, slope, and accuracy in genetic evaluations, the objective of this study was to test its applicability under real-world conditions. Previous simulation studies and empirical examples have demonstrated that the LRM can overcome several limitations of traditional validation approaches, such as the reliance on pre-corrected phenotypes, the inclusion of indirect traits or the sensitivity to miss-specified heritabilities (Macedo et al., 2020b). However, further validation across different populations and traits is required to establish its practical utility.

The present work applied the LRM to real data from three beef cattle breeds: Charolais (CHA), Limousine (LIM), and Simmental (SIM). For each breed, partial and whole EBV (EBV_p and EBV_w, respectively) were computed for a set of economically relevant traits (i.e., carcass traits, calving, and weaning weight). By applying the LRM to EBV_w and EBV_p derived from $data_w$ and $data_p$ datasets, this study aimed to evaluate accuracy, dispersion and potential bias of the genetic evaluations, thereby providing new evidence on the suitability of the LRM as a routine tool for genetic evaluation in beef cattle breeding programs.

Materials and Methods

The data used in this study were provided by Interbull Centre and originate from countries participating in Interbeef service. This data comprises national genetic evaluation information routinely submitted by collaborating countries within the Interbeef framework. The present study forms part of a broader international project aimed at developing and identifying a robust methodology for the validation of genetic evaluations in beef cattle. Within this context, the analyses presented here contribute to the assessment and comparison of alternative validation approaches for beef cattle genetic indices.

To characterize the amount of information available across generations, the number of bulls and total progeny per sire birth cohort were divided in partial ($data_p$) and whole ($data_w$) datasets. Bulls born from 1980 prior to 2000 were grouped into a single <2000 cohort, due to the small number of animals for each year. For each breed (CHA, LIM, and SIM), three economically important traits were considered in the LRM analysis: i) Carcass (carc), consisting of three sub-traits: carcass weight (cwe), carcass conformation (cco), and carcass fatness (cfa); ii) Calving, with two sub-traits: calving ease (cae) and birth weight (bwt); iii) Weaning weight (aww). Since LRM allow the inclusion of maternal traits, in addition to the direct animal effect (ANI) considered for each trait, the maternal effect (DAM) was also included for calving and aww. Only countries that submitted data were included in the analysis. Countries were defined as those in which each bull had recorded progeny. Country for each trait are reported in Table 1.

Table 1. Considered countries for each trait and cattle breed. Breed: Simmental, Limousine, Charolaise. Trait: Carcass (carc), Calving (calv), Weaning weight (adww). Sub-trait: Carcass weight (cwe), Carcass conformation (cco), Carcass fatness (cfa), Birth weight (bwt), Calving ease (cae), Average weaning weight (aww).

Trait	Sub-trait	Breed	A	B	C	D	E	F	G	H
Carcass (carc)	Conformation (cco) Fatness (cfa) Weight (cwe)	Charolaise	✓	✓						
		Limousine	✓	✓						
		Simmental	✓	✓						
Calving (calv)	Calving ease (cae) Birth weight (bwt)	Charolaise		✓	✓	✓	✓*			
		Limousine	✓	✓	✓	✓	✓*			
		Simmental	✓		✓	✓				
Average weaning weight (aww)	Average weaning weight (aww)	Charolaise	✓		✓	✓			✓	
		Limousine	✓	✓	✓	✓	✓	✓	✓	✓
		Simmental	✓		✓	✓	✓	✓	✓	✓

* only cae

Data preparation for LRM

For each breed and for each trait (and sub-trait) combination, the LRM whole dataset ($data_w$) and the partial dataset ($data_p$) were created. The partial dataset was obtained by truncating the most recent years of information, in order to mimic an earlier national evaluation: for carc, the last five years were removed (records up to 2018), whereas for calv and aww the last four years were removed (records up to 2019). Supplementary Table 1 reports the distribution of bulls by year of birth in $data_w$ and $data_p$, along with the distribution of their progeny by bulls' birth-year cohort, for each breed/trait combination. Separate genetic evaluations were then run on $data_w$ and $data_p$. In particular, in the last dataset, Estimated Breeding Values (EBV) for animals born from 2018 or later (for carcass traits) and 2019 or later (for calv and aww), were estimated using only the parent average information. EBV and their associated reliability were obtained using the software MIX99 (Lidauer et al., 2022). Genetic evaluations were carried out fitting a multi-trait, multi-country model trait to obtain EBVs for both $data_p$ and $data_w$ where the phenotype in each population was treated as a distinct trait. Number of random and fixed effects in each national model are reported in Supplementary Table 2 for carc, Supplementary Table 3 for calv and Supplementary Table 4 for aww. Heritability estimates, derived from the (co)variance components obtained from the variance–covariance matrices produced by MIX99, are reported in Supplementary Table 5.

Linear Regression

LRM analysis was performed on bulls that had progeny records in $data_w$ but not in $data_p$. After matching sires across evaluations, we identified as "new proven" those bulls with no recorded progeny in $data_p$ ($Progeny_p = 0$) and at least one progeny record in dataw ($Progeny_w \geq 1$). Validation was conducted by birth-year cohorts corresponding to the cut-off years used to construct data$_p$ (from 2018 onwards for carc traits and from 2019 onwards for calv and aww traits). Bulls meeting these criteria, together with the total progeny per bull birth-year cohort, are reported in Table 2. Table 3 shows the progeny classes, the number of bulls within each class, and the corresponding percentage relative to the total number of bulls. LRM statistics (Bias, Dispersion, Correlation, and Relative gain in accuracy) were then computed by EBV$_p$ and EBV$_w$ within this group.

Table 2. Number of bulls in validation with progeny only in the whole dataset (data$_w$) per year of birth. Breed: (a) Simmental, (b) Limousine, (c) Charolaise. Traits: Carcass, Calving and Weaning weight.

(a)

	Simmental					
	Carcass		Calving		Average daily weight gain	
Year of birth	Bulls data$_w$	Progeny data$_w$	Bulls data$_w$	Progeny data$_w$	Bulls data$_w$	Progeny data$_w$
2018	223	1,488	-	-	-	-
2019	212	1,212	399	8,115	232	4,035
2020	83	386	346	5,089	200	2,455
2021	11	53	288	3,006	162	1,541
2022	3	11	208	1,366	84	432
2023	-	-	1	1	-	-
Total	529	3,150	1,242	17,576	678	8,463

(b)

	Limousine					
	Carcass		Calving		Average daily weight gain	
Year of birth	Bulls data$_w$	Progeny data$_w$	Bulls data$_w$	Progeny data$_w$	Bulls data$_w$	Progeny data$_w$
2018	531	2,406	-	-	-	-
2019	333	1,090	696	6,259	2,071	48,124
2020	56	139	896	9,669	1,832	30,150
2021	2	4	591	4,279	1,391	10,658
2022	-	-	232	1,250	101	454
2023	-	-	7	22	1	7
Total	922	3,639	2,422	21,479	5,396	89,393

(c)

	Charolaise					
	Carcass		Calving		Average daily weight gain	
Year of birth	Bulls data$_{w}$	Progeny data$_{w}$	Bulls data$_{w}$	Progeny data$_{w}$	Bulls data$_{w}$	Progeny data$_{w}$
2018	242	897	-	-	-	-
2019	126	363	684	9,290	2,630	71,886
2020	39	88	614	8,281	2,569	52,608
2021	5	17	439	4,311	2,231	20,509
2022	-	-	178	1,058	626	3,272
2023	-	-	1	1	-	-
Total	412	1,365	1,916	22,941	8,056	148,275

Table 3. Progeny classes, number of bulls within each class, and corresponding percentage relative to the total number of bulls.

Class of progeny	N°Bulls	% Bulls
<=10	12,470	58.88
From >=10 to <=50	7,852	37.08
From >=50 to <=100	741	3.50
From >=100 to <=500	97	0.46
From >=500 to <=1000	11	0.05
>=1000	7	0.03

LR statistics were estimated within each validation group, defined by the combination of year of birth (YoB), country (From A to H), breed (CHA, LIM and SIM), trait (Carcass: cco, cfa, cwe; calving: bwt, cae; weaning aww), and effect (ANI and DAM). Cohort-specific estimates were then summarized across birth years to obtain population-level mean values for each country–breed–trait combination. Population bias (Δp) was estimated following Legarra and Reverter (2018) as:

$$\Delta p = \bar{\hat{u}}_p - \bar{\hat{u}}_w \quad [1]$$

where $\bar{\hat{u}}_p$ and $\bar{\hat{u}}_w$ are the mean EBV from the partial (EBV_P) and whole (EBV_W) datasets, respectively. Under an unbiased evaluation, the expected value of Δp is 0 whereas negative or positive values indicate systematic underestimation and overestimation of genetic progress, respectively. For dispersion, the expected value is 1 when the variability of EBV_P correctly reflects that of EBV_W. Values greater or lower than 1 indicate over or under dispersion, corresponding to overestimation or underestimation of EBV variability in the partial evaluation, respectively. Population dispersion was estimated using a weighted approach, applying weights proportional to the number of progeny per bull using the following linear regression model:

$$EBVw_i = \beta_0 + \beta_1\, EBVp_i + \varepsilon_i \quad [2]$$

where $\mathrm{EBV}_{\mathrm{w},i}$and $\mathrm{EBV}_{\mathrm{p},i}$are the EBVs of individual i within the validation group from the whole and the partial datasets, respectively, $\beta_0$is the intercept, $\beta_1$the regression coefficient, $\varepsilon_i$the residual error. The model was weighted by the number of offspring, to account for heterogeneity in the precision of partial evaluations and to evaluate the sensitivity of dispersion estimates to differences in number of progeny. Regression parameters were tested under the null hypothesis that β_0=0 and $\beta_1 = 1$ (Mayer et al., 1994).

The accuracy of genetic evaluations and the gain in accuracy from the partial to the whole dataset were assessed following Macedo et al. (2020). The ratio of accuracies, $\frac{acc_p}{acc_w}$, was estimated as the correlation between EBV_p and EBV_w within each validation group:

$$\rho_{w,p} = \frac{cov(\bar{\hat{u}}_p,\bar{\hat{u}}_w)}{\sqrt{var(\bar{\hat{u}}_p)\, var(\bar{\hat{u}}_w)}} \qquad [3]$$

This estimator represents the inverse of the relative gain in accuracy when additional phenotypic information is incorporated into genetic evaluations. Values close to 1 indicate minimal improvement in accuracy, whereas lower values reflect substantial accuracy gains.

The relative increase in accuracy ratio provided by the additional phenotypes can be expressed as:

$$\frac{1}{\rho_{w,p}} = \frac{acc_w - acc_p}{acc_p} \qquad [4]$$

For instance, a $\rho_{w,p}$value of 0.7 suggests that EBV_p captures approximately 70% of the accuracy ratio achieved with EBV_w, indicating a moderate gain from the inclusion of additional phenotypic records (Macedo et al., 2020).

To investigate the factors affecting LRM quality metrics (Bias, Dispersion, and Accuracy), separate weighted, for the progeny number, linear models were fitted for each metric. Fixed effects included country, breed, Year of Birth (Yob), heritability estimate (h^2), trait (cwe, cco, cfa, cae, bwt and aww), and the interaction between traits and h^2. The significance of fixed effects was assessed using analysis of variance (ANOVA). The general form of the model was:

$$LR_i = \beta_1 Country_i + \beta_2 Breed_i + \beta_3 YoB_i + \beta_4 Trait_i + \beta_5 h_i^2 + \beta_6 (Trait_i \times h_1^2) + \varepsilon_i \qquad [5]$$

where $LR_i$is the cohort-specific estimate of Bias, Dispersion, or Accuracy ratio for the i-th validation group. Analyses were performed in R (R Core Team, 2023). Analysis of variance tables were obtained for each model, and the proportion of explained variance was quantified using the partial eta-squared coefficients (η^2), defined as the proportion of variance attributable to a given effect relative to the sum of that effect and the residual variance ($\eta^2 = \frac{SS_{effect}}{SS_{effect}+SS_{total}}$) (R "*effectsize*" package, Ben-Shachar et al., 2020). Since bias is a key indicator of systematic inflation or deflation in international genetic evaluations, additional analyses of temporal patterns were performed exclusively for bias, as changes over time may reflect shifts in evaluation comparability across countries (Bonaiti et al., 1993; Boichard et al., 1995; Palucci and Dürr, 2014). To investigate whether Bias exhibited systematic patterns across validation cohorts, Year of Birth (YoB) was included in the model as a categorical fixed effect. Adjusted marginal means for YoB were estimated using the "*emmeans*" package (Lenth, 2023), accounting for all other fixed effects in the model.

Results

Carcass traits

LRM results for bias are reported in Figure 1. Bias values showed substantial variation across traits and breeds. In CHA, moderate bias was observed for cco and cfa, whereas very large positive deviations were detected for cwe in one country. In LIM, bias ranged from moderate negative to very large positive values, again with the largest deviations observed for cwe. In SIM, similar patterns were observed, with moderate bias for cco and cfa but extremely large positive values for cwe in country A. Overall, cwe exhibited markedly higher bias than the other carc traits.

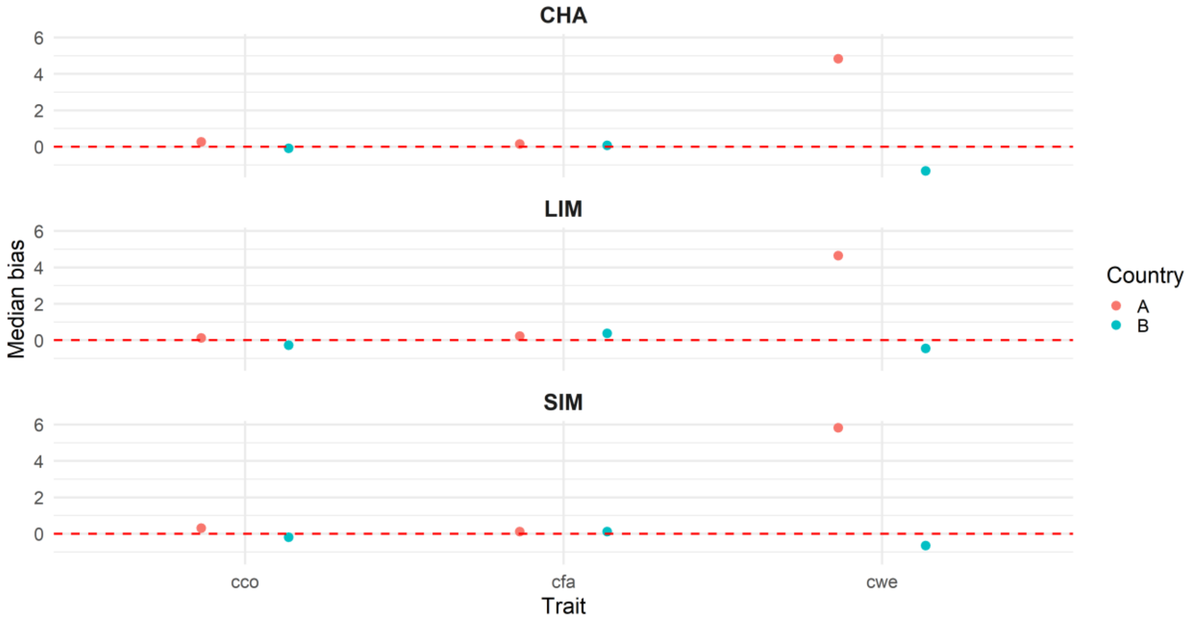


Figure 1. Median bias for carcass traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line denotes the expected bias under perfect calibration (bias = 0).

Dispersion varied considerably across traits (Figure 2). In CHA, dispersion ranged from extreme deflation (very low values for cwe in one population) to values close to 1 for cfa. In LIM, dispersion was generally below 1 across traits, indicating systematic deflation of breeding values under partial data. In SIM, dispersion ranged from moderate deflation to strong inflation, depending on trait and country. Across breeds, dispersion was mostly unstable for cwe.

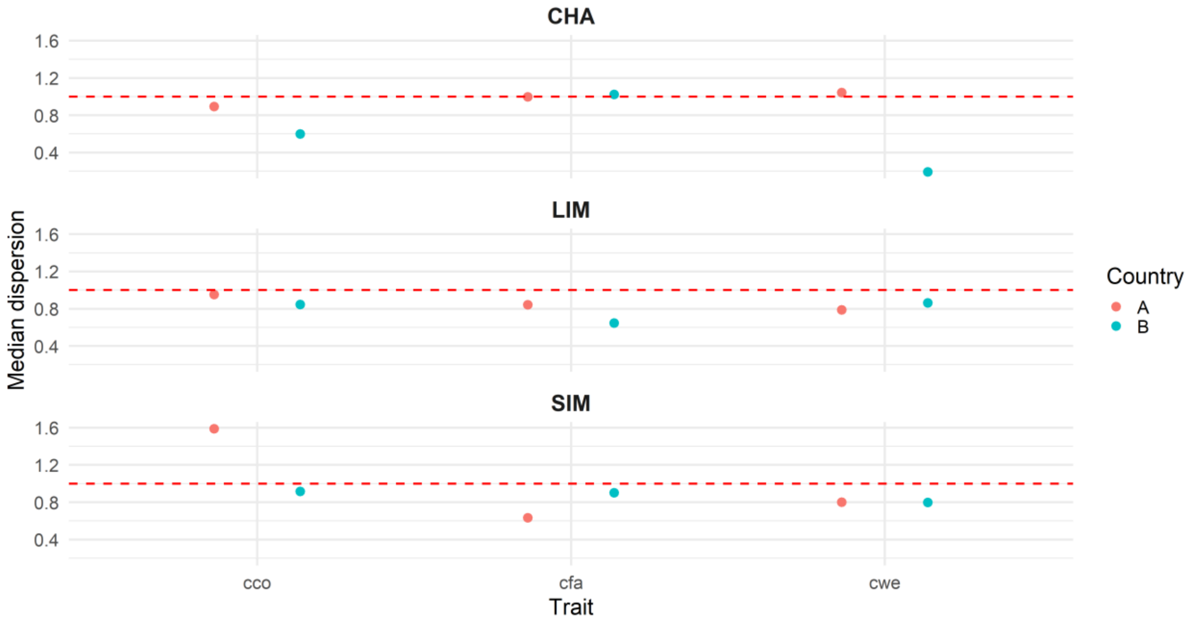


Figure 2. Median dispersion for carcass traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line denotes the expected bias under perfect calibration (dispersion = 1).

Accuracy ratio values ranged from low to high depending on trait and population (Figure 3). In CHA, accuracy ratio was moderate for cco and cfa but very low for cwe in some countries, indicating poor agreement between datasets for this trait. In LIM, accuracy ratio was generally moderate across traits, with no extremely high values. In SIM, accuracy ratio ranged from moderate to high, including strong agreement for cco in one population. Overall, accuracy ratio was consistently lowest for cwe (Figure 3).

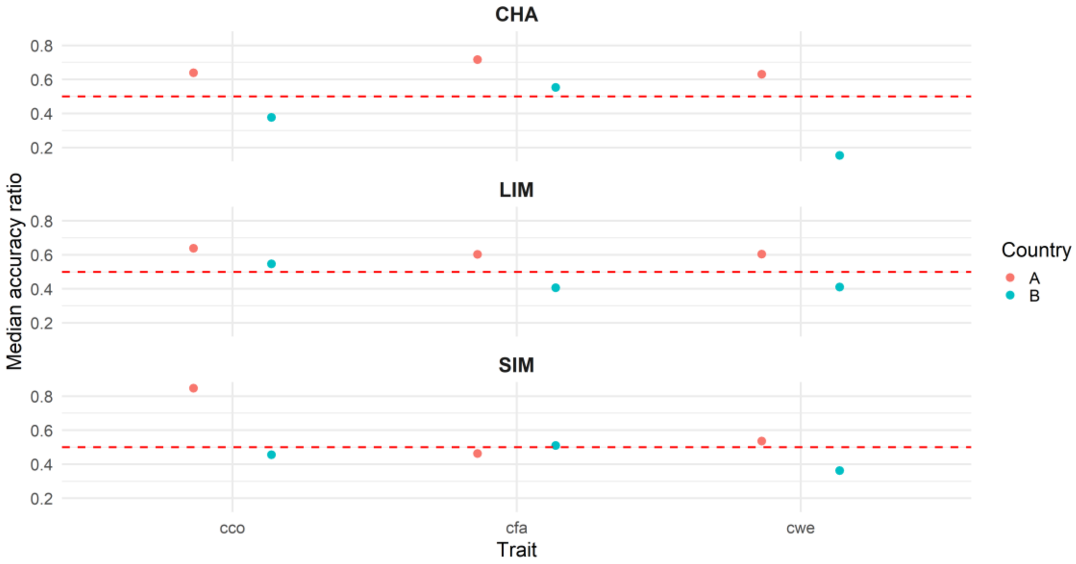


Figure 3. Median accuracy ratio for carcass traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line indicate an accuracy ratio level of 0.50.

Relative gain showed substantial heterogeneity across traits and populations (Figure 4). In CHA, values ranged from moderate to high, with elevated values for cco and particularly low values for some cfa evaluations. In LIM, relative gain exceeded 1 in several cases, especially for cfa and cwe, indicating strong sensitivity to data reduction. In SIM, relative gain varied widely, with values both below and well above 1 depending on trait and country. Across breeds, cwe frequently showed elevated relative gain, reflecting large changes in response when using partial data. (Figure 4).

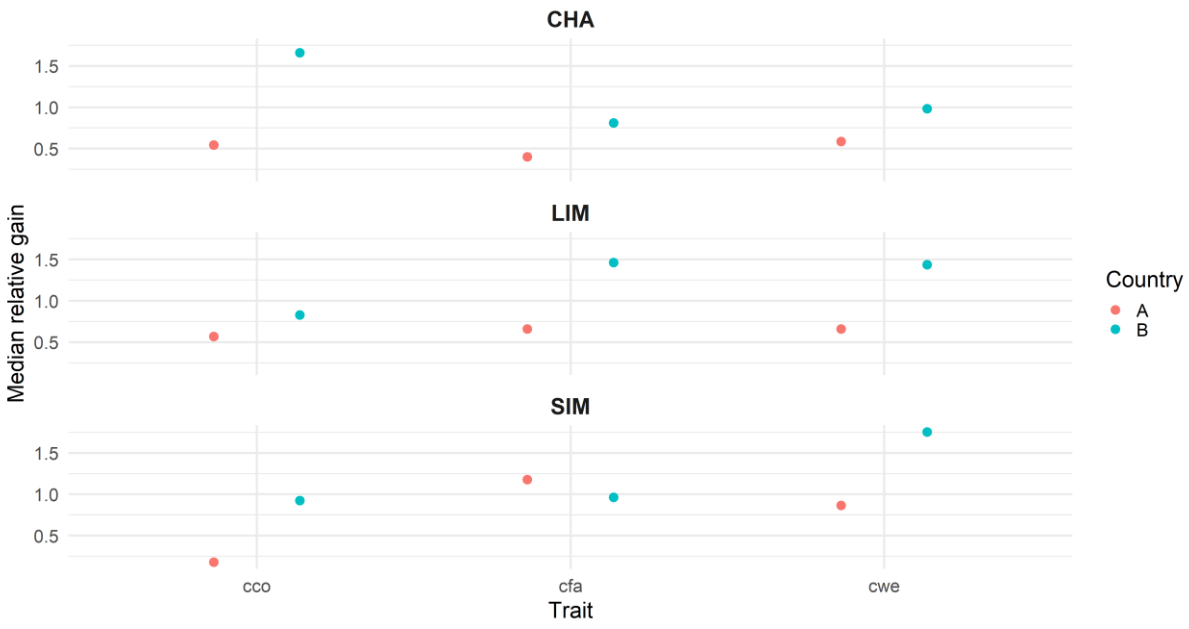


Figure 4. Median relative gain for carcass traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais).

Calving traits

Median bias values for calving traits (bwt and cae) are reported in Figure 5. Overall, median bias values were generally small across breeds, countries and effects, indicating good agreement between breeding values estimated from partial and complete datasets. Most estimates were close to zero for both traits. In CHA, bias was negligible in most cases, although moderate positive deviations were observed for bwt in some countries. In LIM, bias showed greater variability, including both positive and negative deviations. One notable exception was a large negative bias for bwt in country A , indicating substantial disagreement between evaluations. In SIM, bias remained small overall, with only moderate deviations observed for bwt in isolated cases. Across breeds, cae generally showed smaller and more stable bias than bwt.

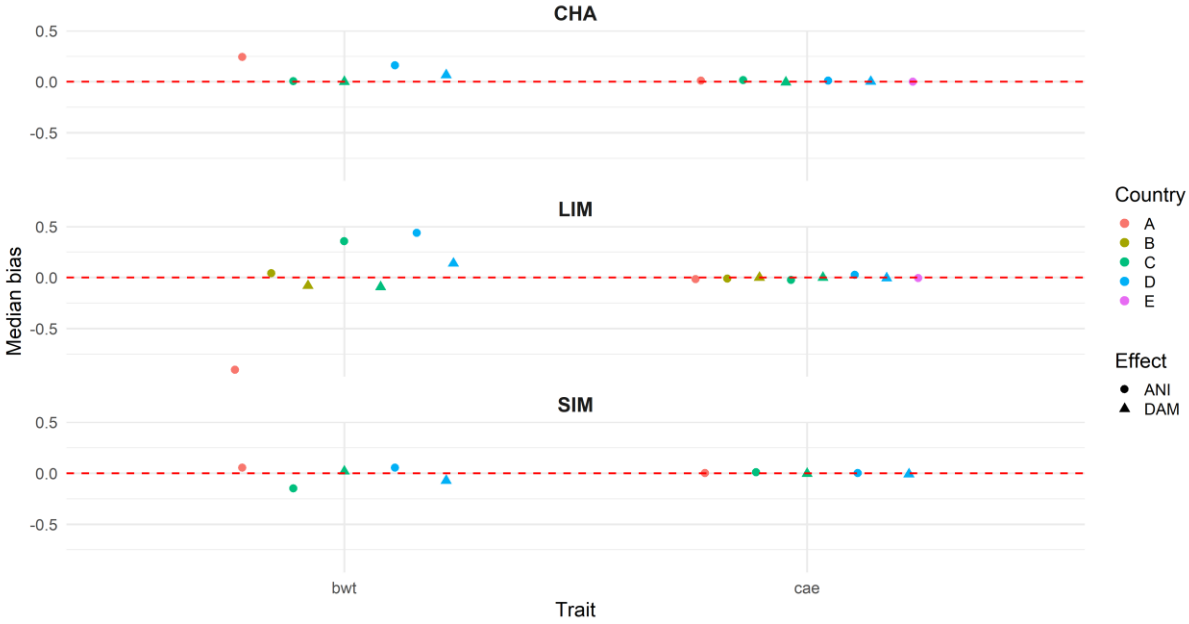


Figure 5. Median bias for calving traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line denotes the expected bias under perfect calibration (bias = 0).

Dispersion values were mostly close to 1, indicating comparable scaling of breeding values between partial and complete datasets in many scenarios. However, deflation (dispersion < 1) was frequent (Figure 6). In CHA, dispersion ranged widely, particularly for cae, with several values clearly below 1, indicating deflation under partial data. In LIM, dispersion was generally below 1, especially for cae, suggesting reduced variability in partial evaluations. In SIM, dispersion values were closer to unity overall, with some inflation (> 1) observed for bwt in specific country–effect combinations. overall, cae showed more consistent deflation than bwt across breeds.

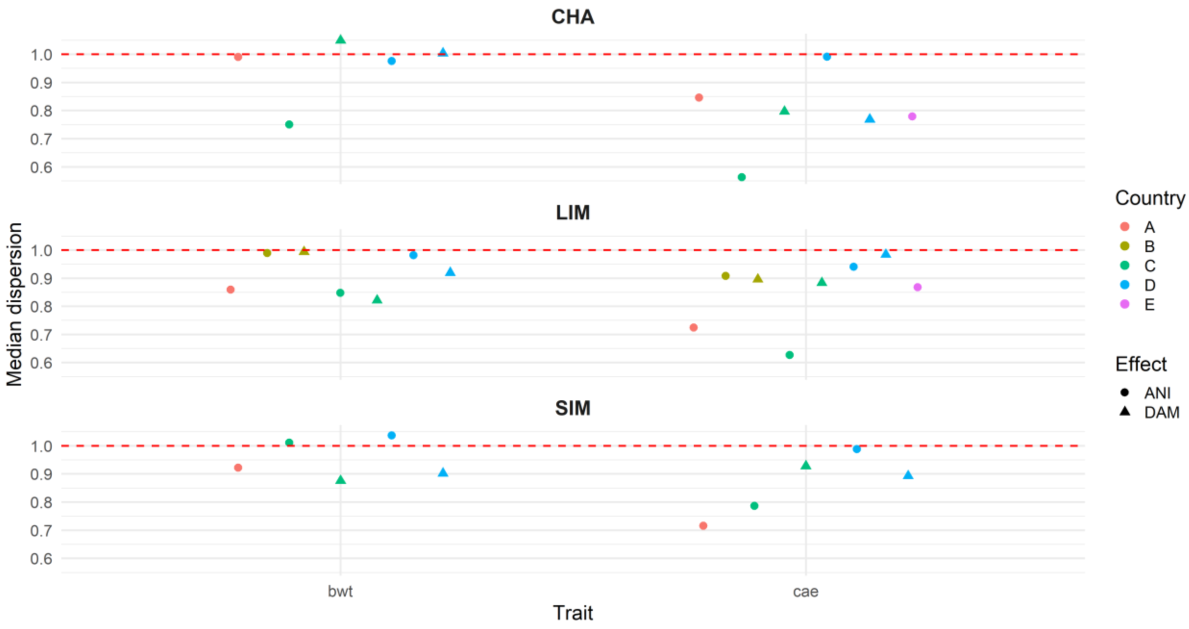


Figure 6. Median dispersion for calving traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line denotes the expected dispersion under perfect calibration (dispersion = 1).

Accuracy ratio, for calv ranged from moderate to high across breeds and traits (Figure 7). In CHA, accuracy ratio ranged from moderate to high, with bwt often showing stronger agreement than cae in some countries. In LIM, accuracy ratio was generally high for bwt in several populations, but more variable for cae. In SIM, accuracy ratio values were comparable to those observed in the other breeds, with high agreement in some country–trait combinations. Across breeds, bwt tended to show slightly higher accuracy ratio than cae in several national evaluations, although this pattern was not universal.

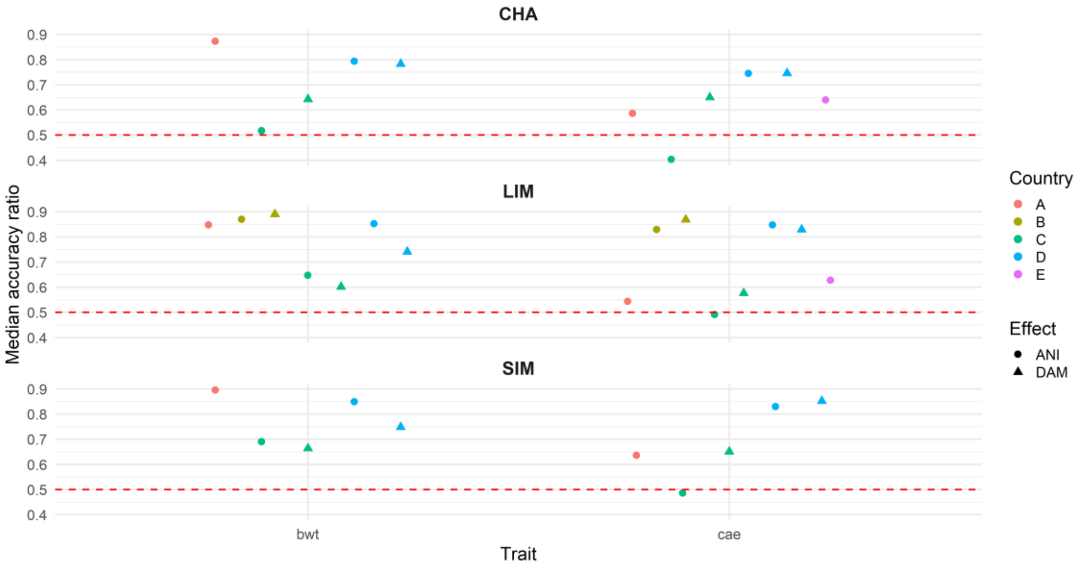


Figure 7. Median accuracy ratio for calving traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line indicate an accuracy level of 0.50.

Relative gain showed substantial variability across breeds, traits and countries, ranging from low values (< 0.2) to values exceeding 1 (Figure 8). In CHA, relative gain was generally low to moderate for bwt but more variable for cae, with some values above 1, indicating increased relative response under partial data. In LIM, relative gain varied widely, with moderate values in many scenarios but elevated values for cae in some populations. In SIM, relative gain was generally moderate, with occasional values close to or slightly above 1 for cae. Overall, cae tended to show greater variability in relative gain than bwt.

When distinguishing between direct (ANI) and maternal (DAM) effects, bias values were consistently close to zero for both effects across breeds and countries, indicating good agreement between partial and complete evaluations. However, dispersion revealed more frequent deflation for the direct effect, particularly for calving ease, whereas maternal effects generally showed scaling closer to unity. Accuracy ratio was moderate to high for both effects, with the maternal component showing comparable or occasionally higher stability than the direct effect in some populations. Relative gain in accuracy ratio was more variable for the direct effect, especially for calving ease.

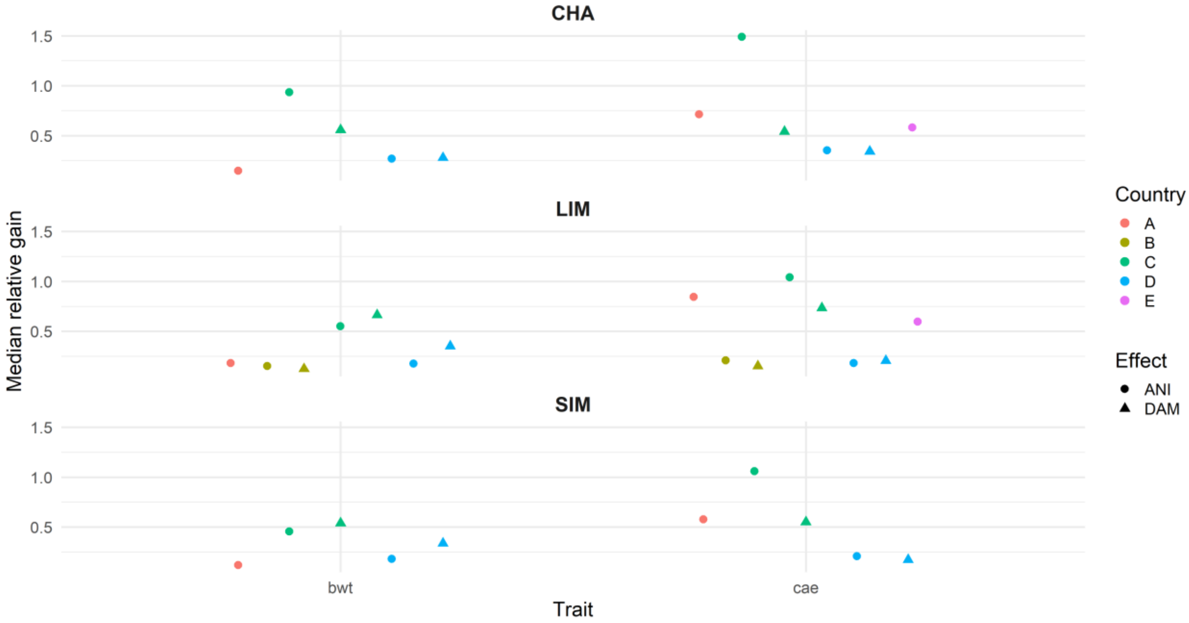


Figure 8. Median relative gain for calving traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais).

Average weaning weight

Median bias between the partial and the complete dataset for aww is presented in Figure 9 for CHA, LIM and SIM populations, separately for direct (ANI) and maternal (DAM) effects across countries.

Overall, median bias values were generally close to zero, indicating good agreement between breeding values estimated from the partial and complete datasets. No consistent systematic inflation or deflation was observed across breeds. In CHA, bias estimates were mostly small in magnitude, with slightly negative values predominating across several countries. Positive deviations were limited and of modest size, suggesting stable predictions when using the partial dataset. In LIM, a wider dispersion of bias values was observed. Both positive and negative deviations were detected, with some countries showing relatively large departures from zero for either ANI or DAM effects. This indicates greater variability in the impact of data reduction in this breed compared to CHA.

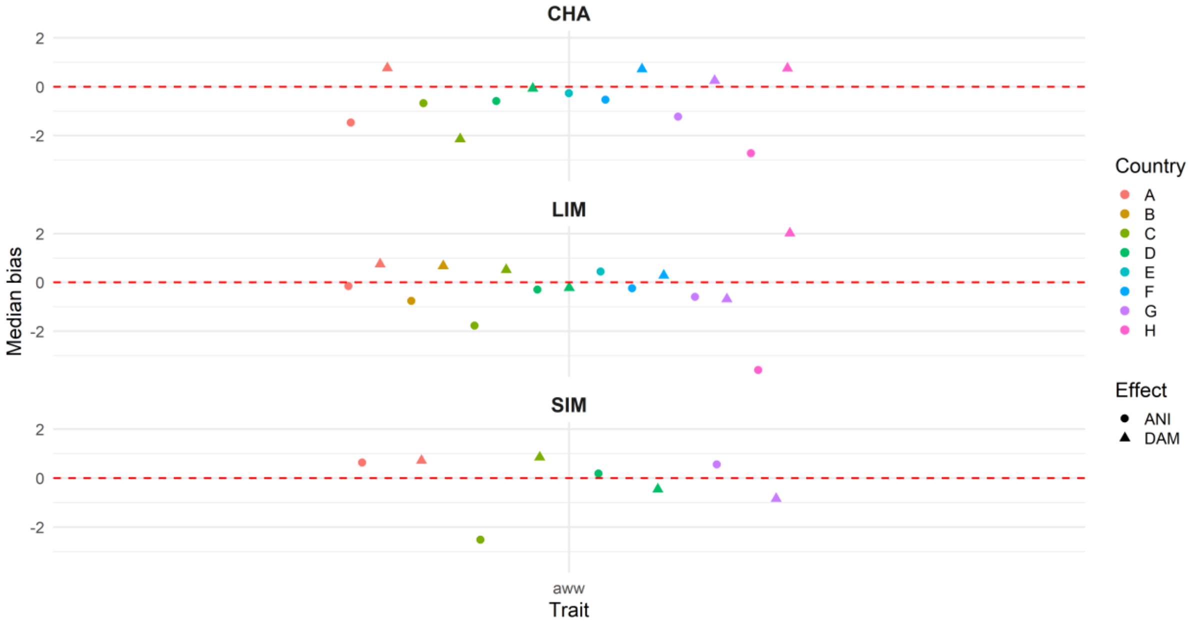


Figure 9. Median bias for average weaning weight by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line denotes the expected bias under perfect calibration (bias = 0).

Median dispersion values for average weaning weight (AWW) across breeds, countries and effects are reported in Figure 10. Overall, dispersion values were generally close to 1, indicating a comparable scale of estimated breeding values between the partial and complete datasets in most scenarios. However, breed-specific differences were evident. In CHA, dispersion values ranged from 0.76 to 1.00 for both ANI and DAM effects. Most estimates were slightly below 1, suggesting a mild deflation of breeding values when using the partial dataset. The variability across countries was limited, and no marked inflation was observed. In LIM, a wider range of dispersion values was detected (0.52–1.08). For ANI, several countries showed clear deflation (e.g. values around 0.52–0.70), whereas others were close to or slightly above 1. For DAM, dispersion was generally closer to 1, although one country showed noticeable inflation (1.08). This indicates greater heterogeneity in scale consistency across national evaluations compared to CHA. In SIM, the contrast between ANI and DAM effects was more pronounced. ANI dispersion values were consistently below 1 (0.70–0.90), indicating systematic deflation under the partial dataset. In contrast, DAM dispersion frequently exceeded 1 (1.08–1.39), with the highest values observed in specific countries. This suggests inflation of maternal breeding values in the reduced dataset for this breed. Across breeds, no uniform pattern was observed between ANI and DAM effects. However, SIM showed a tendency toward deflation for direct effects and inflation for maternal effects, whereas CHA displayed the most stable behavior across countries.

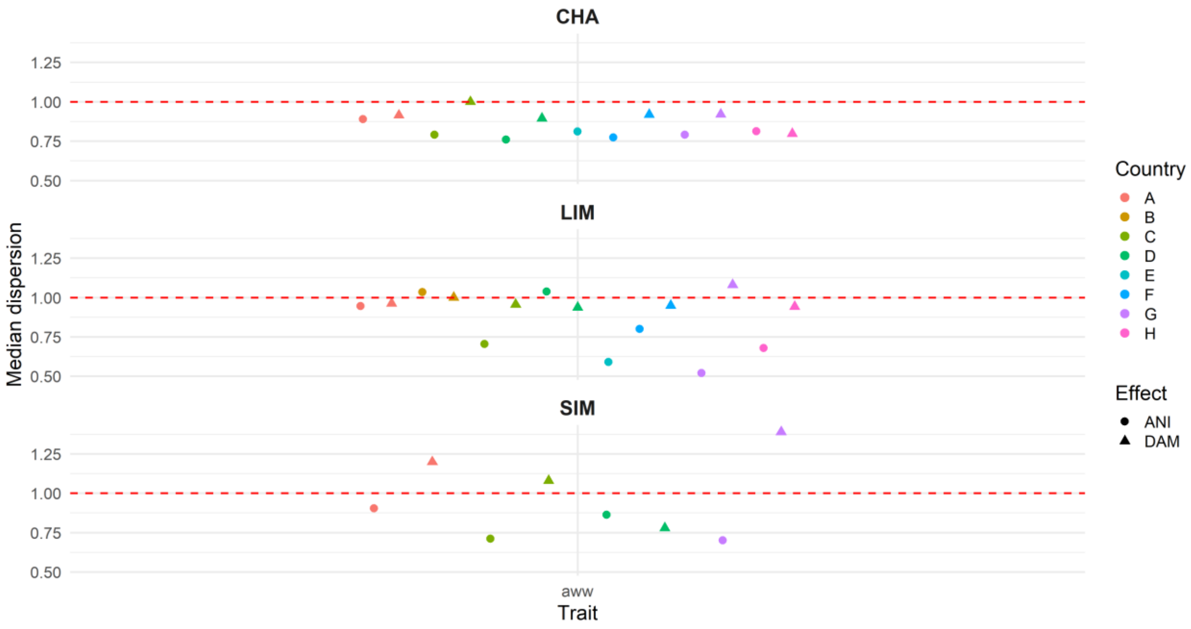


Figure 10. Median dispersion for average weaning weight by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line denotes the expected bias under perfect calibration (dispersion = 0).

Median accuracy ratio values for aww are presented in Figure 11. Overall, accuracy ratio was moderate to high across breeds, with values ranging from 0.44 to 0.89 depending on breed, country and effect. Maternal effects generally showed higher accuracy ratio than direct effects. In CHA, accuracy for ANI ranged from 0.54 to 0.72, indicating moderate agreement between evaluations. Higher values were consistently observed for DAM (0.67–0.85), suggesting that maternal breeding values were more stable under data reduction. Variation across countries was limited, and no extremely low values were observed.

In LIM, a wider range of accuracy ratio values was observed (0.44–0.89), indicating greater heterogeneity among national evaluations. Several country–effect combinations showed high agreement between datasets, particularly for DAM, which reached values above 0.85 in some cases. However, lower accuracy values were also detected, particularly for ANI in specific countries, indicating reduced stability of direct breeding values under partial data. In SIM, ANI accuracy ratio values were moderate (0.53–0.66), whereas DAM accuracy was consistently higher (0.70–0.86). This pattern indicates a clearer separation between direct and maternal effects compared to the other breeds, with maternal evaluations showing greater robustness to data reduction. Across breeds, maternal effects tended to maintain higher agreement between partial and complete evaluations, whereas direct effects showed greater variability across countries.

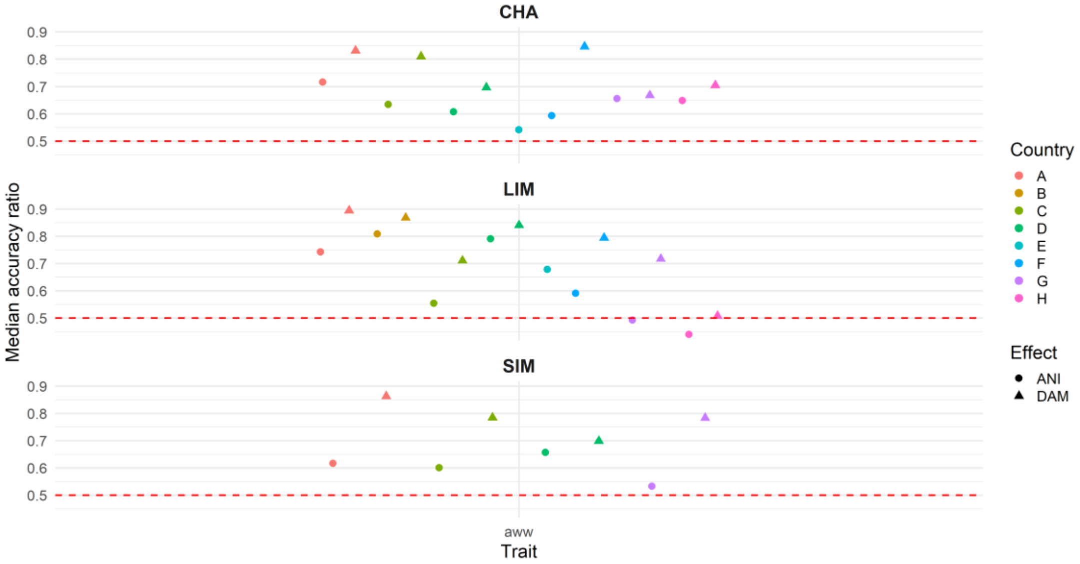


Figure 11. Median accuracy ratio for average weaning weight traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais). The red horizontal line indicate an accuracy level of 0.50.

Median relative gain values for aww are reported in Figure 12. Overall, relative gain varied substantially across breeds and countries, ranging from approximately 0.12 to 2.40. In most cases, values were below 1, indicating reduced response or information when using the partial dataset. However, some country–effect combinations showed values close to or above 1. In CHA, relative gain for ANI was generally moderate to high (0.40–0.85), whereas DAM values were consistently lower (0.18–0.50). This indicates that the partial dataset preserved a larger proportion of information for maternal than for direct effects. In LIM, relative gain showed the greatest variability among breeds. While many country–effect combinations displayed moderate or low values (≈0.12–0.81), some cases showed markedly elevated values, particularly for ANI in specific countries (up to 2.40). DAM relative gain was generally lower and more stable, although one country showed values close to 1. In SIM, relative gain for ANI ranged from moderate to high (0.53–1.00), whereas DAM values were consistently lower (0.16–0.43). This pattern mirrors that observed in CHA, with maternal effects retaining a larger proportion of information under partial data. Overall, relative gain showed stronger variability across countries than accuracy, with the largest deviations observed in LIM. Across breeds, direct effects generally exhibited higher relative gain than maternal effects.

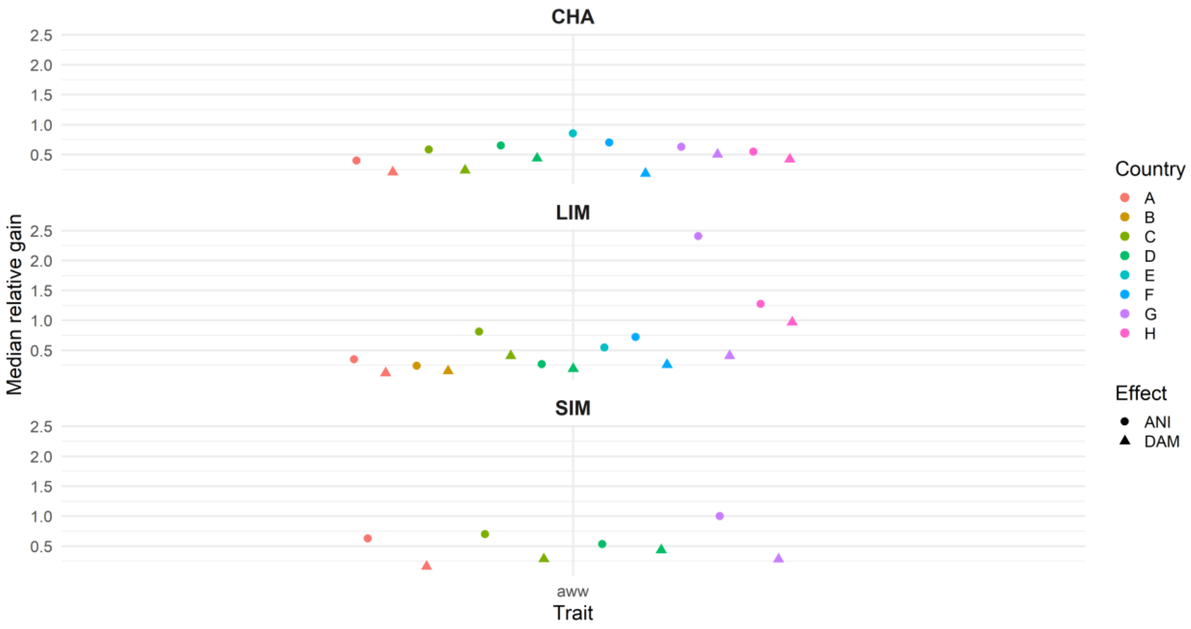


Figure 12. Median relative gain for average weaning weight traits by breed (LIM = Limousine, SIM = Simmental and CHA = Charolais).

The median values underlying the figures are reported in Supplementary Table 6.

Statistical analysis

The linear regression between the mean EBVs from $data_P$ and $data_W$ [2] showed a strong correspondence, with a slope of 0.963 (SE = 0.005, $p < 0.001$), slightly below one, indicating minor deflation in the EBV scale. The intercept was not significantly different from zero (intercept = -0.123, SE = 0.069, $p = 0.078$), suggesting minimal systematic bias. Residuals were generally small relative to the range of EBVs, and the model explained 98.9% of the variance ($R^2 = 0.989$).

The weighted linear models [5] revealed that several fixed effects significantly influenced LRM quality metrics (Table 4). For accuracy, all considered factors (Country, Trait, heritability, Breed, Year of Birth, and the interaction between Trait and heritability) showed highly significant effects ($P \leq 0.02$), indicating that the predictive ability of EBV validation is strongly dependent on both population structure and genetic architecture. The model explained a substantial proportion of the variability in accuracy ($R^2 = 64.13\%$). For bias, significant effects were detected for country, heritability, year of birth, and the trait × heritability interaction ($P \leq 0.01$), whereas trait and breed did not significantly affect this metric. The proportion of explained variance was moderate ($R^2 = 35.56\%$). Regarding dispersion, all fixed effects except for cohort (YoB) significantly affected the LRM estimates ($P \leq 0.01$). However, the overall proportion of explained variance was low ($R^2 = 26.72\%$).

Table 4. P-value and coefficient of determination (R^2) of the effect of country, trait, heritability (h^2), validation cohort (YoB), and interaction of the trait with the heritability (Trait:h^2) on EBV quality metrics, including accuracy, bias and dispersion obtained with the LRM.

	Accuracy ratio	Bias	Dispersion
Country	0.0001	0.0001	0.0001
Trait	0.0001	0.73	0.01
h^2	0.0001	0.01	0.0001
Breed	0.02	0.98	0.01
YoB	0.0001	0.0001	0.89
Trait x h^2	0.0001	0.0001	0.0001
R^2,%	64.13	35.56	26.72

The relative contributions of the examined factors to the variance of LRM quality metrics were assessed using the eta squared (η^2) index. As shown in Table 5, for accuracy, country, trait and heritability were the major contributors, each explaining approximately one third of the variance, whereas breed accounted for a minimal proportion. For bias, the year of birth explained a substantial fraction of the variance, while the effect of trait was negligible. Dispersion was mainly influenced by heritability and the interaction between trait and heritability.

Table 5. Eta squared (η^2) values indicating the relative contribution of each factor to the variance of EBV validation metrics, including accuracy, bias and dispersion obtained with the LRM.

	Accuracy ratio	Bias	Dispersion
Country	0.33	0.14	0.07
Trait	0.25	0.01	0.04
h2	0.33	0.02	0.12
Breed	0.02	0.0001	0.03
YoB	0.06	0.15	0.0001
Trait x h^2	0.28	0.08	0.08

Estimated marginal means (EMMeans) for the effect of YoB on bias varied across years, with the 2019 cohort exhibiting the highest positive bias (1.14 ± 0.32; 95% CI: 0.50–1.78), followed by 2020 (1.03 ± 0.33; 95% CI: 0.38–1.67) and 2018 (0.80 ± 0.48; 95% CI: -0.15–1.76). The 2021 and 2022 cohorts showed near-zero or slightly negative bias (-0.08 ± 0.35; 95% CI: -0.75–0.60 and -0.04 ± 0.42; 95% CI: -0.88–0.79, respectively). However, no statistical difference was observed across years.

Discussion

The LRM is a semi-parametric approach developed to validate and compare genetic evaluation models by assessing how EBVs change as additional information becomes available (Legarra and Reverter, 2018). The theoretical foundation of the method is rooted in Henderson's mixed model theory (Henderson, 1982): if a model is correctly specified and properly accounts for the selection process, the prediction error associated with EBV estimated from partial data should be uncorrelated with the information added at a later stage, thereby ensuring the absence of bias. Operationally, the method compares EBV obtained from a "partial" dataset with those derived from a "whole" dataset that includes additional information (e.g., new performance records or progeny data), and allows the estimation of three key indicators of evaluation quality: bias, dispersion, and accuracy ratio (Legarra and Reverter, 2018). Unlike traditional validation tests, the LRM does not require knowledge of the true breeding values (TBV) or the use of pre-adjusted phenotypes, making it particularly suitable for complex applied contexts such as beef cattle populations, which are often characterized by low-heritability traits, maternal effects, or limited progeny group sizes (Legarra and Reverter, 2018; Macedo et al., 2020b). Simulation studies have further demonstrated that the method is robust to mild model misspecifications (e.g., slightly incorrect heritability assumptions), maintaining its ability to detect both the presence and direction of bias and to discriminate among alternative evaluation models (Macedo et al., 2020b). More recently, the LR method has also been applied to real international beef cattle evaluations, where it proved useful for quantifying changes in population accuracy and dispersion under alternative model specifications, particularly for maternally influenced traits (Bonifazi et al., 2021). In the present study, we evaluated the performance of the LRM in beef cattle using real data from three breeds (CHA, LIM, and SIM) and three economically relevant traits (carc, calv, and aww).

Although the modern formulation of the LRM primarily focuses on mean bias and regression slope as indicators of systematic distortion and dispersion, respectively (Legarra and Reverter, 2018; Macedo et al., 2020b), additional regression parameters such as the intercept and the coefficient of determination (R^2) can provide complementary information regarding the global stability of EBV predictions. Historically, intercept and slope have played a central role in international validation studies and conversion equations among national genetic evaluations (e.g., Boichard et al., 1992; 1995; Banos et al., 1992; Bonaiti et al., 1993), where the objective was to align evaluation scales across populations. In contrast, the LRM framework is designed to test theoretical properties of prediction under increasing information and therefore emphasizes bias and dispersion rather than the regression intercept per se (Legarra and Reverter, 2018). In the present study, the intercept value of 0.123 and the extremely high R^2 value ($R^2 = 0.989$) indicate that the relationship between partial and whole EBV was highly stable, reinforcing the conclusion that the observed LRM statistics reflect structural and biological differences rather than inconsistencies in scale definition. Heritability values found in the present study were generally consistent with those commonly reported in beef cattle populations, showing low estimates for calving traits, moderate estimates for growth traits, and moderate to relatively higher estimates for carcass traits (Supplementary Table 5) (Eriksson et al., 2004; Bucek et al., 2019; Osorio et al., 2021; Interbull Centre, 2026). As expected, calving traits exhibited lower heritability compared to growth and carcass traits, supporting the biological plausibility of the dataset (Gutierréz et al., 2007; Chud et al., 2014; Englishby et al., 2016).

Determinants of LRM indicators

Accuracy ratio of EBV predictions estimated through the LRM showed substantial variability across countries and traits, and was strongly influenced by heritability. As reported in Table 4, country, trait, heritability (h^2), validation cohort (YoB), and their interaction significantly affected accuracy ratio ($P < 0.001$), with the overall model explaining 64.13% of the observed variability. The η^2 values (Table 5) further highlighted that country ($\eta^2 = 0.33$) and heritability ($\eta^2 = 0.33$) were the main contributors to accuracy ratio variation, followed by the interaction between trait and heritability ($\eta^2 = 0.28$) and trait itself ($\eta^2 = 0.25$). In contrast, breed and YoB showed comparatively smaller effects. In a related Interbeef study on Limousin cattle, Bonifazi et al. (2021) evaluated the impact of alternative model assumptions on LR validation statistics and reported marked heterogeneity across countries. In particular, smaller countries showed larger gains in accuracy ratio from international evaluations, whereas responses were more limited in larger and better-connected populations. Although the relative contribution of fixed effects was not formally quantified, these results could support the major role of country structure, connectedness, and data size also observed in the present study. The strong dependence of accuracy ratio on heritability is consistent with quantitative genetic theory, where prediction accuracy is expected to increase with additive genetic variance and the signal-to-noise ratio of the trait (Goddard and Hayes, 2009; Claus eta l., 2009; Daetwyler et al., 2013; Macedo et al., 2020a). In fact, accuracy is expected to increase for high heritable traits (such as production and morphological traits) compared to lower heritable ones (such as fertility traits). This relationship has been documented in genomic prediction studies. In longitudinal analyses of feed efficiency traits in pigs, accuracy ratio was closely related to the genetic variance captured by the model (David et al., 2022). Although caution is required when comparing different validation metrics, similar associations between heritability and predictive performance have also been reported when prediction accuracy, computed from prediction error variance and additive genetic variance, was used as the validation criterion. For example, in Nellore cattle, absolute accuracy of genomic prediction for carcass traits increased with heritability, reaching values of 0.46 for rib eye area and 0.37 for hot carcass weight when adjusted phenotypes were used (Fernandes et al., 2016). The large contribution of countries to accuracy variation likely reflects differences in population structure, connectedness, and effective training population size across national datasets (Bonaiti et al., 1993; Calus et al., 2009).

Bias of EBV predictions, assessed within the LRM framework as the difference between evaluations based on partial and complete datasets, was significantly affected by country, heritability, validation cohort (YoB), and the interaction between trait and heritability ($P \leq 0.01$; Table 4). The overall model explained 35.56% of the variability in bias, indicating a moderate capacity to account for systematic differences across scenarios. According to the η^2 estimates (Table 5), YoB ($\eta^2 = 0.15$) and country ($\eta^2 = 0.14$) were the main contributors to bias variation, followed by the trait × h^2 interaction ($\eta^2 = 0.08$). In contrast, trait and breed showed negligible effects ($\eta^2 = 0.01$ and 0.0001, respectively), suggesting that systematic inflation or deflation of EBV was less trait-specific and more related to structural characteristics of the dataset. Within the LRM, unbiased predictions are expected to have a mean difference equal to zero between

EBV obtained from partial and complete information. Deviations from this expectation may reflect changes in connectedness between training and validation populations, selective recording, or differences in the amount of information available across cohorts (Legarra and Reverter, 2017; Macedo et al., 2020b). Although low-heritability traits are theoretically more sensitive to changes in data completeness due to the larger contribution of residual variance (Macedo et al., 2020a; Macedo et al., 2020b), cae showed bias values consistently close to zero across all breeds and countries in this study. This suggests that, despite its low genetic signal, the evaluations for this trait were stable and not systematically inflated or deflated under partial data scenarios.

Dispersion of EBV predictions, evaluated as the regression coefficient of complete EBV on partial EBV within the LRM framework, was significantly influenced by country, trait, heritability, breed, and their interaction ($P \leq 0.01$; Table 4). However, the proportion of explained variance was lower compared both to accuracy and bias ($R^2 = 26.72\%$), indicating that dispersion is more weakly structured by the considered fixed effects. The η^2 values (Table 5) showed that heritability had the largest relative contribution to dispersion ($\eta^2 = 0.12$), followed by the trait × h^2 interaction ($\eta^2 = 0.08$) and country ($\eta^2 = 0.07$). Trait and breed had smaller but non-negligible effects ($\eta^2 =$ 0.04 and 0.03, respectively), whereas YoB had virtually no impact ($\eta^2 \approx 0.0001$). In the LRM framework, a regression coefficient equal to 1 indicates correctly scaled predictions, whereas values below or above 1 indicate over- or under-dispersion, respectively (Legarra and Reverter, 2018). The strong influence of heritability on dispersion suggests that scaling of EBV is closely related to the signal-to-noise ratio of the trait (Calus et al., 2009). For traits with lower h^2, partial evaluations are expected to shrink more strongly towards the mean due to the lower signal-to-noise ratio, potentially leading to over-dispersion when compared to complete evaluations (Macedo et al., 2020b).

These findings suggest that LRM statistics should be interpreted within homogeneous groups defined by country, trait and heritability, which showed the largest impact on accuracy, bias and dispersion. Mixing heterogeneous populations may mask or inflate validation metrics. Therefore, defining comparable validation cohorts and explicitly evaluating the effect of key fixed factors represent important best practices when applying the LRM.

ANI vs DAM effect differences

The integration of maternal genetic effects into breeding value validation through the Linear Regression (LRM) method is particularly challenging, as maternal traits are not directly expressed on the individual animal but are inferred from progeny performance and pedigree relationships (Legarra and Reverter, 2018). Consequently, the estimation of bias and absolute accuracy for maternal EBV is intrinsically more complex. Studies have shown that maternal EBV are generally more sensitive than direct EBV to model specifications, particularly to assumptions regarding the direct–maternal genetic correlation and covariance structure, especially when data are limited, and may require larger amounts of information to achieve comparable gains in accuracy (Meyer et al., 1992; David et al., 2015; Bonifazi et al., 2021). This difference arises because direct effects are supported by the animal’s own performance, whereas maternal effects depend entirely on relatives and progeny information (Willham, 1972). These difficulties are particularly evident for pre-weaning growth traits such as aww or pre-weaning average daily gain (Wolf et al., 2011). In these traits, maternal heritability is typically low, and negative direct–maternal genetic correlations are frequently reported, which complicate parameter estimation and may introduce bias (Eriksson et al., 2004). Such correlations may arise from identification errors, limited connectedness among herds, or unmodelled environmental covariances between dam and offspring (Eriksson et al., 2004). In addition, aww is strongly influenced by permanent environmental effects of the dam and other non-genetic maternal factors, reducing the effective signal-to-noise ratio. As a result, traits with a weak genetic signal are more sensitive to data completeness, and LRM dispersion parameters may show stronger shrinkage under partial evaluations (Macedo et al., 2020a; Macedo et al., 2020b). In contrast, calving traits such as birth weight and calving ease generally present moderate direct heritability and a non-negligible maternal component, providing a clearer genetic signal to the model (Eriksson et al., 2004; Macedo et al., 2020b). Moreover, although maternal effects are involved, these traits are less affected by prolonged postnatal maternal environmental influences (David et al., 2015). This structural difference may contribute to the relatively greater stability observed for maternal effects in calving traits compared with aww under LRM validation. In the case of maternal effects, the LRM accuracy ratio is particularly informative because prediction accuracy is difficult to assess in real populations, as maternal genetic effects are not directly expressed by the individual but are inferred from progeny and pedigree information (Legarra and Reverter, 2018; Macedo et al., 2020b; Bonifazi et al., 2021). In this context, the correlation between partial and whole evaluations provides an indicator of how maternal EBV respond when additional data become available, without requiring knowledge of the true breeding value (Legarra and Reverter, 2018).

Linking statistical and biological patterns

The median LRM statistics reported in Supplementary Table 6 highlight clear biological and structural differences across traits, breeds, and evaluation effects (ANI vs DAM), which are consistent with the statistical patterns observed in the global models. Across breeds, carcass traits (cco, cfa, cwe) generally showed moderate to high accuracy ratio (up to 0.85 in SIM, country A). In contrast, calving traits (bwt, cae) displayed more variable dispersion and slightly lower accuracy ratio, which is biologically plausible given their typically lower heritability and greater environmental influence (Gutierréz et al., 2007; Chud et al., 2014). Average weaning weight (aww), although moderately heritable, exhibited substantial variability in bias across countries, especially for ANI evaluations, suggesting sensitivity to data structure and cohort effects. These biological differences align with theoretical expectations: traits with higher additive genetic variance and more consistent recording schemes (e.g., carcass traits) tend to show better-scale and more accurate predictions, whereas traits influenced by maternal environment, management, or heterogeneous recording (e.g., calving traits and weaning weight) are more prone to instability in bias and dispersion (Macedo et al., 2020b).

The patterns observed in Supplementary Table 6 are consistent with the results of the global statistical analyses. Substantial variation in accuracy ratio, bias, and dispersion across countries confirms the strong and significant effect of Country detected in the linear models, particularly for accuracy ratio. Similarly, differences among trait categories in the magnitude and direction of bias and dispersion align with the significant Trait and Trait × h^2 effects identified in the overall analysis. Moreover, the variability observed between ANI and DAM evaluations, especially for bias and dispersion in aww, reflects the moderate explanatory power of the global models for these metrics. This indicates that, while systematic effects are detectable at the aggregate level, important heterogeneity remains at the level of specific genetic components and trait-by-country combinations.

The results reported by Alexandre et al. (2021) for carcass weight in Australian Angus provide an informative benchmark. In their study, accuracy was estimated as the Pearson correlation between GEBV and adjusted phenotypes divided by the square root of heritability, yielding an average value of 0.64, with dispersion close to unity (mean 0.99), indicating well-scaled genomic predictions. It should be noted, however, that this definition of accuracy differs from the LRM-based accuracy ratio used in the present study, which measures the consistency between EBV obtained from partial and complete evaluations rather than the correlation with adjusted phenotypes. In addition, Alexandre et al. evaluated genomic breeding values (GEBV), whereas the present study focuses on pedigree-based EBV. Despite these methodological differences, both studies identified heritability and trait as major drivers of accuracy. Alexandre et al. reported a high proportion of explained variance for LRM accuracy (R^2 = 84.9%), whereas in the present study the explained variance was lower ($R^2 \approx$ 64%) and dispersion showed greater variability across countries and evaluation effects. Moreover, bias was not significantly structured in their analysis (R^2 = 14.5%), while in the present study bias exhibited stronger dependence on structural factors such as country and cohort. Because EBV_p and EBV_w were derived from successive international evaluations, residual differences in evaluation base or scaling cannot be completely excluded as contributing factors.

From an operational perspective, in agreement with previous literature (Misztal et al., 2020; Macedo et al., 2020b; Bonifazi et al., 2021) adoption of the LRM as a standard validation metric within Interbull/Interbeef could provide several advantages. First, it would harmonize validation procedures across countries and traits by using a common framework based on bias, dispersion and accuracy ratio. Second, it could improve routine quality control by detecting inflation/deflation of EBV/GEBV, poor connectedness, or inconsistencies in model assumptions. Third, it would be particularly informative for smaller populations, which often obtain the largest relative gains from international evaluations. However, the present results also indicate that LRM statistics are strongly context-dependent and should be interpreted within homogeneous groups defined by country, trait and data structure. Therefore, successful implementation of the LRM would require carefully defined validation cohorts and regular stratified monitoring rather than reliance on pooled summary indicators alone.

Conclusion

This study evaluated the quality of EBV predictions using the LRM across multiple breeds, traits, and national evaluation contexts. By jointly analyzing accuracy, bias, and dispersion, we assessed the stability and consistency of genetic evaluations under different validation scenarios. The results showed that prediction accuracy is strongly context-dependent, with national evaluation frameworks representing a major source of variation. In addition, the genetic architecture of the trait—particularly its heritability—emerged as a key determinant of predictive performance. Traits with a stronger additive genetic signal consistently showed higher and more stable accuracy ratio. Differences between direct (ANI) and maternal (DAM) genetic effects highlight the need for specific attention when evaluating traits influenced by maternal components. Among the LRM quality metrics, dispersion appeared to be the least predictable and most variable across scenarios. This suggests that scaling of EBV should be interpreted with caution, particularly when comparing results across countries, traits, or genetic components. Overall, these findings confirm that the LRM provides a valuable framework for assessing the robustness of genetic evaluations.

Acknowledgments

The authors gratefully acknowledge the Interbull Centre for providing access to the data used in this study and for proposing the original idea that motivated this research.

References

Alexandre, P.A., Li, Y., Hine, B.C.*et al* Bias, dispersion, and accuracy of genomic predictions for feedlot and carcass traits in Australian Angus steers. *Genet Sel Evol* 53, 77 (2021). doi: 10.1186/s12711-021-00673-8

Banos, G., Wiggans, G. R., & Robinson, J. A. B. (1992). Comparison of methods to rank bulls across countries. *Journal of dairy science*, *75*(9), 2560-2568. doi: 10.3168/jds.S0022-0302(92)78018-7

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Estimation of effect size indices and standardized parameters. *Journal of open source software*, *5*(56), 2815. doi: 10.21105/joss.02815

Boichard, D., & Lee, A. J. (1992). Approximate accuracy of genetic evaluation under a single-trait animal model. *Journal of dairy science*, *75*(3), 868-877. doi: 10.3168/jds.S0022-0302(92)77827-8

Boichard, D., Bonaiti, B., Barbat, A., & Mattalia, S. (1995). Three methods to validate the estimation of genetic trend for dairy cattle. *Journal of dairy science*, *78*(2), 431-437. doi: 10.3168/jds.S0022-0302(95)76652-8

Bonaiti, B., Boichard, D., Barbat, A., & Mattalia, S. (1993). Problems arising with genetic trend estimation in dairy cattle. *Interbull Bulletin*, (8).

Bonifazi, R., Vandenplas, J., Ten Napel, J., Veerkamp, R. F., & Calus, M. P. (2021). The impact of direct-maternal genetic correlations on international beef cattle evaluations for Limousin weaning weight. *Journal of Animal Science*, *99*(9), skab222. doi: 10.1093/jas/skab222

Bucek, P. (2013). International genetic evaluation of calving traits in beef cattle. *Interbull Bulletin*, (47) Calus, M. P. (2010). Genomic breeding value prediction: methods and procedures. *animal*, *4*(2), 157-164. doi: 10.1017/S1751731109991352

Chud, T. C., Caetano, S. L., Buzanskas, M. E., Grossi, D. A., Guidolin, D. G., Nascimento, G. B., ... & Munari, D. P. (2014). Genetic analysis for gestation length, birth weight, weaning weight, and accumulated productivity in Nellore beef cattle. *Livestock Science*, *170*, 16-21. doi: 10.1016/j.livsci.2014.09.024

Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM. Genomic prediction in animals and plants: simulation of data, validation, reporting, and bench marking. Genetics. 2013;193:347–65. doi: 10.1534/genetics.112.143313

David, I., Bouvier, F., Banville, M., Canario, L., Flatres-Grall, L., Balmisse, E., & Garreau, H. (2015). The direct-maternal genetic correlation has little impact on genetic evaluations. *Journal of Animal Science*, *93*(12), 5639-5647. doi: 10.2527/jas.2015-9548

Englishby, T. M., Banos, G., Moore, K. L., Coffey, M. P., Evans, R. D., & Berry, D. P. (2016). Genetic analysis of carcass traits in beef cattle using random regression models. *Journal of Animal Science*, *94*(4), 1354-1364. doi: 10.2527/jas.2015-0246

Eriksson, S., Nasholm, A., Johansson, K., & Philipsson, J. (2004). Genetic relationships between calving and carcass traits for Charolais and Hereford cattle in Sweden. *Journal of Animal Science*, *82*(8), 2269-2276. doi: 10.2527/2004.8282269x

Fernandes Júnior, G. A., Rosa, G. J., Valente, B. D., Carvalheiro, R., Baldi, F., Garcia, D. A., ... & De Albuquerque, L. G. (2016). Genomic prediction of breeding values for carcass traits in Nellore cattle. *Genetics Selection Evolution*, *48*(1), 7. doi: 10.1186/s12711-016-0188-y

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programs. *Nature Reviews Genetics*, *10*(6), 381-391. doi: 10.1038/nrg2575

Henderson, C. R. (1982). Best linear unbiased prediction in populations that have undergone selection. In: Proceedings of the world congress on sheep and beef cattle breeding: 28 October–13 November 1980; Palmerston North and Christchurch. Palmerston North: Dunmore Press; 1982.p.191–201.

Interbull Centre. 2026. PREPdb - Performance Recording, Evaluation and Publication database [Online]. Interbull Centre- Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden. Available: https://prep.interbull.org/ [Accessed date 2026/04/22].

J. P. Gutiérrez, F. Goyache, I. Fernández, I. Alvarez, L. J. Royo, Genetic relationships among calving ease, calving interval, birth weight, and weaning weight in the Asturiana de los Valles beef cattle breed, *Journal of Animal Science*, Volume 85, Issue 1, January 2007, Pages 69–75. doi:10.2527/jas.2006-168

Kennedy, B. W., Schaeffer, L. R., & Sorensen, D. A. (1988). Genetic properties of animal models. *Journal of Dairy science*, *71*, 17-26. doi: 10.1016/S0022-0302(88)79975-0

Legarra, A., & Reverter, A. (2017). Can we frame and understand cross-validation results in animal breeding?.

Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution*, *50*(1), 53. doi: 10.1186/s12711-018-0426-6

Macedo, F. L., Christensen, O. F., Astruc, J. M., Aguilar, I., Masuda, Y., & Legarra, A. (2020a). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genetics Selection Evolution*, *52*(1), 47. doi: 10.1186/s12711-020-00567-1

Macedo, F. L., Reverter, A., & Legarra, A. (2020b). Behavior of the linear regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *Journal of Dairy Science*, *103*(1), 529-544. doi: 10.3168/jds.2019-16603

Mayer, D.G., Stuart, M.A., Swain, A.J., 1994. Regression of real-world data on model output: an appropriate overall test of validity. Agric. Syst. 45, 93–104. doi: 10.1016/S0308-521X(94)90282-8

Meyer, K. (1992). Bias and sampling covariances of estimates of variance components due to maternal effects. *Genetics Selection Evolution*, *24*(6), 487-509. doi:10.1186/1297-9686-24-6-487

Osorio Gallardo, T. (2021). Assessing the benefits from joining the International beef cattle genetic evaluation (Interbeef) at SLU’s Interbull Centre.

Palucci, V., & Dürr, J. (2014). Trend validation procedures applied by Interbull–historical overview. *population*, *1*(2), 5

Ramirez-Valverde, R., Misztal, I., & Bertrand, J. K. (2001). Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle. *Journal of Animal Science*, *79*(2), 333-338. doi: 10.2527/2001.792333x

Sullivan, P. G., & VanRaden, P. M. (2009). Development of genomic GMACE. *Interbull bulletin*, (40), 157-157.

Tyrisevä, A. M., Mäntysaari, E. A., Fiske, F., & Lindauer, M. H. (2011). Simulation study on Mendelian sampling variance tests. *Interbull Bulletin*, (44).

VanRaden, P. M., Tooker, M. E., Wright, J. R., Sun, C., & Hutchison, J. L. (2014). Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *Journal of dairy science*, *97*(12), 7952-7962. doi: 10.3168/jds.2014-8489

Willham, R. L. (1972). The role of maternal effects in animal breeding: III. Biometrical aspects of maternal effects in animals. *Journal of Animal Science*, *35*(6), 1288-1293. doi: 10.2527/jas1972.3561288x

Wolf, J. B., Leamy, L. J., Roseman, C. C., & Cheverud, J. M. (2011). Disentangling prenatal and postnatal maternal genetic effects reveals persistent prenatal effects on offspring growth in mice. *Genetics*, *189*(3), 1069-1082. Doi: 10.1534/genetics.111.130591

Appendix Chapter 5

Application of the Legarra & Reverter Method to Evaluate Carcass, Reproductive, and Weaning Traits in Three beef Cattle Breeds

Supplementary Table 1. Distribution of bulls by year of birth in data$_W$ and data$_P$, together with the distribution of their progeny, for each breed (Simmental, Limousine e Charolaise) and trait (Carcass, Calving and Weaning weight) combination. (a) Simmental, (b) Limousine, (c) Charolaise. Traits: Carcass, Calving and Average weaning weight.

Supplementary Table 2. List of random and fixed effects in each national model for carcass traits. Breeds: Charolaise, Limousine and Simmental. Country: A and B.

Supplementary Table 3. List of random and fixed effects in each national model for calving traits. Breeds: Charolaise, Limousine and Simmental. Country: A, B, C, D and E.

Supplementary Table 4. List of random and fixed effects in each national model for calving traits. Breeds: Charolaise, Limousine and Simmental. Country: A, B, C, D, E, F, G and H.

Supplementary Table 5. Heritability, estimates derived from the (co)variance components obtained from the variance–covariance matrices produced by MIX99. Breeds: Charolaise, Limousine and Simmental. Country: A, B, C, D, E, F, G and H. Traits: carcass weight (cwe), carcass conformation (cco), carcass fatness (cfa), calving ease (cae), birth weight (bwt) and average weaning weight (aww).

Supplementary Table 6. Median LRM statistics (Bias, Dispersion, Accuracy, and Relative Gain) for the three traits (Carcass, Calving, and Average Weaning Weight) across the three breeds (Limousin, Charolais, and Simmental).

Supplementary Table 1. Distribution of bulls by year of birth in data$_w$ and data$_p$, together with the distribution of their progeny, for each breed (Simmental, Limousine e Charolaise) and trait (Carcass, Calving and Weaning weight) combination. (a) Simmental, (b) Limousine, (c) Charolaise. Traits: Carcass, Calving and Average weaning weight.

a)

	Simmental											
	Carcass				Calving				Weaning weight			
Year of birth	Bulls data$_w$	Progeny data$_w$	Bulls data$_p$	Progeny data$_p$	Bulls data$_w$	Progeny data$_w$	Bulls data$_p$	Progeny data$_p$	Bulls data$_w$	Progeny data$_w$	Bulls data$_p$	Progeny data$_p$
<2000	7	53	5	44	36	3,941	36	3,839	20	1405	20	1377
2000	1	1	0	0	11	3,413	10	3,339	6	1,245	6	1,189
2001	2	6	2	3	6	1,882	6	1,871	6	637	6	630
2002	2	55	2	38	7	1,164	7	1,151	1	37	1	35
2003	1	3	1	1	11	4,973	10	4,862	7	1,732	7	1,671
2004	4	51	4	41	8	3,023	8	2,909	8	1,008	8	975
2005	3	75	3	67	10	2,407	9	2,298	7	701	6	669
2006	8	80	8	64	15	1,219	14	1,150	10	641	9	595
2007	10	204	8	181	14	3,261	13	3,055	11	1,593	11	1,438
2008	27	708	21	601	10	1,289	10	1,049	12	735	11	562

2009	36	581	29	450	16	2,461	15	2,045	16	991	14	793
2010	62	882	46	595	24	1,469	24	1,376	18	741	18	681
2011	96	1,914	65	1,159	34	4,315	32	3,750	28	2,141	24	1,914
2012	140	2,093	109	1,261	57	3,571	52	2,866	35	1,284	31	1,059
2013	209	4,055	143	1,884	87	6,640	73	5,453	58	3,093	46	2,491
2014	223	2,834	120	1, 001	99	5,224	87	3,544	47	2,741	39	1,913
2015	307	3,185	137	768	176	8,066	155	4,635	91	4,025	73	2,244
2016	299	3,173	41	157	235	6,926	179	3,726	119	3,495	95	1,929
2017	338	2,829	1	7	296	6,935	223	2,399	140	3,412	100	1,188
2018	223	1,488	0	0	313	6,336	141	991	178	3,576	79	597
2019	212	1,212	0	0	404	8,186	5	5	233	4,079	1	1
2020	83	386	0	0	346	5,089	0	0	200	2,455	0	0
2021	11	53	0	0	288	3,006	0	0	162	1,541	0	0
2022	3	11	0	0	208	1,366	0	0	84	432	0	0
2023	-	-	-	-	1	1	0	0	-	-	-	-

(b)

	Limousine											
	Carcass				Calving				Weaning weight			
Year of birth	Bulls data$_w$	Progeny data$_w$	Bulls data$_p$	Progeny data$_p$	Bulls data$_w$	Progeny data$_w$	Bulls data$_p$	Progeny data$_p$	Bulls data$_w$	Progeny data$_w$	Bulls data$_p$	Progeny data$_p$
<2000	18	548	18	523	69	15,737	67	15643	119	239,054	118	238,042
2000	3	118	2	112	13	4,246	13	4,228	17	18,707	17	18,643
2001	3	34	3	28	10	3,730	10	3,695	18	8,890	18	8,813
2002	6	115	6	110	14	3,410	14	3,380	28	18,609	27	18,517
2003	6	141	6	112	21	4,332	21	4,252	29	29,134	29	28,823
2004	7	35	6	24	17	1,759	17	1,715	27	13,848	27	13,709
2005	21	207	16	155	31	5,628	31	5,484	31	23,706	31	23,192
2006	24	281	22	248	28	3,008	27	2,883	46	23,341	46	21,328
2007	64	698	50	543	31	4,630	30	4,371	56	24,839	55	23,605
2008	96	1,461	73	1,092	41	5,257	39	5,061	65	17,561	62	16,843
2009	135	1,663	96	1,089	57	4,403	56	3,965	89	23,232	84	19,618
2010	237	2,287	155	1,367	68	5,766	65	5,358	143	21,011	138	18,298
2011	344	4,239	241	2,497	135	12,324	131	11,002	273	34,982	246	27,757
2012	385	4,591	255	2,226	194	7,150	181	6,423	432	38,758	394	29,942

2013	495	4,897	293	2,128	259	12,915	237	10,670	599	51,382	542	36,563
2014	673	5,776	365	1,917	389	15,358	357	11,396	808	65,948	710	41,143
2015	692	5,246	252	905	497	14,625	433	9,828	1,102	68,250	971	39,695
2016	803	5,085	85	230	681	18,267	564	11,559	1,514	78,747	1,294	36,287
2017	700	3,723	4	7	768	18,302	588	9,228	1,862	82,465	1,509	25,747
2018	531	2,406	0	0	854	14,582	519	4,662	1,981	64,258	827	6,099
2019	333	1,090	0	0	921	12,804	225	1,447	2,076	48,178	5	6
2020	56	139	0	0	907	9,821	11	32	1,832	30,150	0	0
2021	2	4	0	0	591	4,279	0	0	1,391	10,658	0	0
2022	-	-	-	-	232	1,251	0	0	102	472	1	1
2023	-	-	-	-	7	22	0	0	1	7	0	0

(c)

	Charolaise											
	Carcass				Calving				Weaning weight			
Year of birth	Bulls data$_{w}$	Progeny data$_{w}$	Bulls data$_{p}$	Progeny data$_{p}$	Bulls data$_{w}$	Progeny data$_{w}$	Bulls data$_{p}$	Progeny data$_{p}$	Bulls data$_{w}$	Progeny data$_{w}$	Bulls data$_{p}$	Progeny data$_{p}$
<2000	25	509	23	484	108	15,744	108	15,505	193	334,706	190	333,862
2000	4	38	3	34	16	2793	16	2741	39	68,527	39	68,378
2001	7	80	7	64	21	2504	20	2427	41	51,650	40	51,458
2002	3	29	3	19	21	2249	21	2158	35	39,899	35	39,678
2003	4	10	4	9	26	1491	26	1417	43	53,836	43	53,302
2004	8	85	8	71	23	3520	23	3353	49	58,745	49	58,202
2005	9	34	6	25	20	2214	19	2074	50	37,778	50	37,157
2006	17	177	15	145	28	2471	27	2201	68	49,392	66	48,456
2007	20	157	19	145	32	3315	31	3076	90	67,970	89	65,181
2008	52	428	50	332	42	2066	39	1904	95	32,920	92	30,665
2009	62	399	53	323	61	3115	59	2775	118	24,415	115	22,572
2010	131	1187	111	838	106	6873	100	5937	191	33,775	181	29,487
2011	161	921	125	619	141	4453	131	3631	300	55,817	290	42,237
2012	208	1,370	158	907	213	5406	200	4426	449	48,638	433	38,272

2013	268	1,365	165	663	278	5794	252	4579	697	62,590	669	52,258
2014	297	1,358	135	487	375	7782	342	5624	958	81,062	898	59,188
2015	326	1,683	134	386	507	12895	447	8195	1,379	101,960	1,281	65,780
2016	327	1,204	21	48	555	13767	445	6761	1,854	118,355	1,673	56,604
2017	292	1,053	0	0	614	10888	391	3689	2,128	107,601	1,773	35,659
2018	242	897	0	0	697	11315	214	1399	2,600	100,594	1,309	10,070
2019	126	363	0	0	688	9406	4	8	2,787	77,885	157	541
2020	39	88	0	0	614	8281	0	0	2,569	52,608	0	0
2021	5	17	0	0	439	4311	0	0	2,213	20,509	0	0
2022	-	-	-	-	178	1058	0	0	626	3,272	0	0
2023	-	-	-	-	1	1	0	0	-	-	-	-

Supplementary Table 2. List of random and fixed effects in each national model for carcass traits. Breeds: Charolaise (CHA), Limousine (LIM) and Simmental (SIM). Country: A and B. Effects: ANI = animal effect, MAT = maternal effect, CG = contemporary group , MPE = maternal permanent environment, HYS = herd-year-season.

Breed	Country	Fixed	Covariates	Random
CHA	A	3	16	ANI; MPE; HYS (3)
	B	5	14	ANI (1)
LIM	A	3	16	ANI; MPE; HYS (3)
SIM	B	5	14	ANI (1)
	A	3	16	ANI; MPE; HYS (3)

Supplementary Table 3. List of random and fixed effects in each national model for calving traits. Breeds: Charolaise (CHA), Limousine (LIM) and Simmental (SIM). Country: A, B, C, D and E. Effects: ANI = animal effect, MAT = maternal effect, CG = contemporary group , MPE = maternal permanent environment, HYS = herd-year-season.

Breed	Country	Fixed	Covariates	Random
CHA	C	2	2	ANI; MAT; MPE; CG (4)
	D	5	N/A	ANI; MAT; MPE (3)
	E	5	N/A	ANI (1)
	F	4	N/A	ANI; MAT; MPE (3)
	A	3	21	ANI; MAT; MPE; HYS (4)
LIM	C	2	2	ANI; MAT; MPE; CG (4)
	D	5	2	ANI; MAT; MPE (3)
	E	5	N/A	ANI (1)
	F	4	N/A	ANI; MAT; MPE (3)
	A	3	21	ANI; MAT; MPE; HYS (4)
	B (bwt)	6	2	ANI; MAT; MPE (3)
	B (cae)	7	2	ANI; MAT; MPE (3)
SIM	C	2	2	ANI; MAT; MPE; CG (4)
	D	5	2	ANI; MAT; MPE (3)
	A	3	21	ANI; MAT; MPE; HYS (4)

Supplementary Table 4. List of random and fixed effects in each national model for calving traits. Breeds: Charolaise (CHA), Limousine (LIM) and Simmental (SIM). Country: A, B, C, D, E, F, G and H. Effects: ANI = animal effect, MAT = maternal effect, CG = contemporary group , MPE = maternal permanent environment, HYS = herd-year-season.

Breed	Country	Fixed	Covariates	Random
CHA	C	2	2	ANI; MAT; MPE; CG (4)
	D	5	2	ANI; MAT; MPE (3)
	E	5	N/A	ANI (1)
	F	4	N/A	ANI; MAT; MPE (3)
	A	3	6	ANI; MAT; MPE (3)
	H	2	N/A	ANI; MAT; HY (3)
	G	4	1	ANI; MAT; MPE (3)
LIM	C	2	2	ANI; MAT; MPE; CG (4)
	D	5	2	ANI; MAT; MPE (3)
	E	5	N/A	ANI (1)
	F	4	N/A	ANI; MAT; MPE (3)
	A	3	21	ANI; MAT; MPE; HYS (4)
	H	2	N/A	ANI; MAT; HY (3)
	G	4	1	ANI; MAT; MPE (3)
SIM	C	2	2	ANI; MAT; MPE; CG (3)
	D	5	2	ANI; MAT; MPE (3)
	A	3	6	ANI; MAT; MPE (3)
	G	4	1	ANI; MAT; MPE (3)

Supplementary Table 5. Heritability, estimates derived from the (co)variance components obtained from the variance–covariance matrices produced by MIX99. Breeds: Charolaise, Limousine and Simmental. Country: A, B, C, D, E, F, G and H. Traits: carcass weight (cwe), carcass conformation (cco), carcass fatness (cfa), calving ease (cae), birth weight (bwt) and average weaning weight (aww).

Breed	Trait	Country	h^2	Effect
CHA	aww	A	0.201	ANI
		C	0.263	
		D	0.100	
		F	0.138	
		H	0.087	
		G	0.161	
		E	0.262	
		A	0.113	DAM
		C	0.113	
		D	0.081	
		F	0.067	
		H	0.018	
		G	0.057	
	bwt	C	0.148	ANI
		D	0.217	
		F	0.217	
		A	0.108	
		C	0.042	DAM
		D	0.084	
		F	0.082	
	cae	C	0.097	ANI
		D	0.044	
		F	0.099	
		A	0.081	
		E	0.050	
		C	0.023	DAM
		D	0.023	
		F	0.035	
	cwe	A	0.096	ANI
		B	0.210	
	cco	A	0.147	
		B	0.210	
	cfa	A	0.121	
		B	0.230	

LIM	aww	C	0.262	ANI
		D	0.104	
		B	0.190	
		A	0.202	
		F	0.195	
		H	0.083	
		G	0.161	
		E	0.255	
		C	0.125	DAM
		D	0.082	
		B	0.059	
		A	0.115	
		F	0.077	
		H	0.019	
		G	0.057	
	bwt	C	0.148	ANI
		D	0.217	
		F	0.242	
		B	0.162	
		A	0.108	
		C	0.042	DAM
		D	0.084	
		F	0.083	
		B	0.041	
	cae	C	0.097	ANI
		D	0.024	
		F	0.053	
		B	0.062	
		A	0.081	
		E	0.050	
		C	0.023	DAM
		D	0.024	
		F	0.023	
		B	0.033	
	cwe	A	0.096	ANI
		B	0.233	
	cco	A	0.147	
		B	0.292	
	cfa	A	0.121	
		B	0.236	
SIM	aww	D	0.099	ANI
		A	0.201	
		C	0.261	
		G	0.161	

	D	0.103	DAM
	A	0.062	
	C	0.107	
	G	0.057	
bwt	C	0.148	ANI
	D	0.217	
	A	0.108	
	C	0.042	DAM
	D	0.084	
cae	C	0.097	ANI
	D	0.044	
	A	0.081	
	C	0.023	DAM
	D	0.023	
cwe	A	0.096	ANI
	B	0.210	
cco	A	0.147	
	B	0.210	
cfa	A	0.121	
	B	0.230	

Supplementary Table 6. Median LRM statistics (Bias, Dispersion, Accuracy ratio, and Relative Gain) for the three traits (Carcass, Calving, and Average Weaning Weight) across the three breeds (Limousine - LIM, Charolais - CHA, and Simmental - SIM).

Breed	Trait	Country	Effect	Bias	Dispersion	Accuracy ratio	Relative gain
CHA	cco	B	ANI	-0.08	0.67	0.38	1.66
	cfa			0.07	1.02	0.55	0.81
	cwe			-1.33	0.19	0.15	0.98
	cco	A	ANI	0.27	0.91	0.64	0.55
	cfa			0.14	1.01	0.72	0.40
	cwe			4.84	1.09	0.63	0.58
LIM	cco	B	ANI	-0.28	0.82	0.55	0.83
	cfa			0.38	0.61	0.41	1.46
	cwe			-0.45	0.81	0.41	1.44
	cco	A	ANI	0.13	0.96	0.64	0.57
	cfa			0.23	0.85	0.60	0.66
	cwe			4.65	0.82	0.60	0.66
SIM	cco	B	ANI	-0.20	0.87	0.46	0.92
	cfa			0.11	0.95	0.51	0.96
	cwe			-0.66	0.68	0.36	1.75
	cco	A	ANI	0.31	1.16	0.85	0.18
	cfa			0.12	0.70	0.46	1.18
	cwe			5.83	0.52	0.54	0.87
CHA	bwt	C	ANI	0.00	0.72	0.52	0.93
	cae			0.02	0.57	0.40	1.49
	bwt		DAM	0.00	1.10	0.64	0.56
	cae			0.00	0.80	0.65	0.54
	bwt	D	ANI	0.16	0.94	0.79	0.27
	cae			0.01	0.98	0.74	0.35
	bwt		DAM	0.06	0.99	0.78	0.28
	cae			0.00	0.77	0.75	0.34
	cae	E	ANI	0.00	0.77	0.64	0.58
	bwt	A	ANI	0.24	0.98	0.87	0.15
	cae			0.01	0.86	0.59	0.71
LIM	bwt	C	ANI	0.35	0.83	0.65	0.55
	cae			-0.03	0.60	0.49	1.04
	bwt		DAM	-0.09	0.80	0.60	0.66
	cae			0.00	0.91	0.58	0.73
LIM	bwt	D	ANI	0.44	0.97	0.85	0.18
	cae			0.02	0.94	0.85	0.18
	bwt		DAM	0.14	0.91	0.74	0.35

	cae			-0.01	0.99	0.83	0.21
LIM	cae	E	ANI	-0.01	0.84	0.63	0.60
LIM	bwt	B	ANI	0.04	1.01	0.87	0.15
	cae			-0.01	0.92	0.83	0.21
	bwt		DAM	-0.08	0.98	0.89	0.12
	cae			0.00	0.94	0.87	0.15
LIM	bwt	A	ANI	-0.91	0.87	0.85	0.18
	cae			-0.02	0.74	0.54	0.84
SIM	bwt	C	ANI	-0.15	1.01	0.69	0.46
	cae			0.01	0.74	0.49	1.06
	bwt		DAM	0.02	0.87	0.66	0.54
	cae			0.00	0.92	0.65	0.55
SIM	bwt	D	ANI	0.05	1.02	0.85	0.18
	cae			0.00	0.97	0.83	0.21
	bwt		DAM	-0.07	0.89	0.75	0.34
	cae			-0.01	0.89	0.85	0.17
SIM	bwt	A	ANI	0.05	0.94	0.90	0.12
	cae			0.00	0.72	0.64	0.58
CHA	aww	C	ANI	-0.69	0.79	0.63	0.58
			DAM	-2.15	0.96	0.81	0.23
CHA	aww	D	ANI	-0.60	0.77	0.61	0.65
			DAM	-0.08	0.90	0.70	0.44
CHA	aww	E	ANI	-0.28	0.80	0.54	0.85
CHA	aww	F	ANI	-0.54	0.78	0.59	0.70
			DAM	0.71	0.92	0.85	0.18
CHA	aww	A	ANI	-1.48	0.89	0.72	0.40
			DAM	0.76	0.93	0.83	0.20
CHA	aww	H	ANI	-2.73	0.81	0.65	0.54
			DAM	0.74	0.81	0.70	0.42
CHA	aww	G	ANI	-1.23	0.85	0.65	0.62
		G	DAM	0.24	0.88	0.67	0.50
LIM	aww	C	ANI	-1.77	0.73	0.55	0.81
		C	DAM	0.52	0.97	0.71	0.41
LIM	aww	D	ANI	-0.31	1.05	0.79	0.27
		D	DAM	-0.22	0.94	0.84	0.19
LIM	aww	E	ANI	0.44	0.63	0.68	0.54
LIM	aww	F	ANI	-0.25	0.80	0.59	0.72
		F	DAM	0.28	0.95	0.79	0.26
LIM	aww	B	ANI	-0.77	1.02	0.81	0.24
		B	DAM	0.67	0.96	0.87	0.15
LIM	aww	A	ANI	-0.16	0.95	0.74	0.35
		A	DAM	0.74	0.97	0.89	0.12
LIM	aww	H	ANI	-3.59	0.67	0.44	1.27
		H	DAM	2.01	1.00	0.51	0.97

LIM	aww	G	ANI	-0.60	0.56	0.49	2.40
		G	DAM	-0.69	1.08	0.72	0.41
SIM	aww	C	ANI	-2.52	0.72	0.60	0.70
		C	DAM	0.85	1.10	0.78	0.28
SIM	aww	D	ANI	0.18	0.88	0.66	0.53
		D	DAM	-0.45	0.79	0.70	0.43
SIM	aww	A	ANI	0.63	0.98	0.62	0.62
		A	DAM	0.71	1.16	0.86	0.16
SIM	aww	G	ANI	0.55	0.68	0.53	1.00
		G	DAM	-0.84	1.02	0.78	0.28

CHAPTER 6

General conclusion

This thesis explored two complementary dimensions of cattle genetics: (i) the characterization and conservation of local breeds, shaped by historical selection and environmental adaptation and (ii) the evaluation and validation of genetic improvement strategies in cosmopolitan breeds sustaining modern livestock production systems. Chapter 2 and Chapter 4 highlighted genetic diversity, selective signatures, and adaptive patterns reflecting centuries of interaction between genotype, environment, and management in local cattle breeds.

Understanding the relationships among breeds and subpopulations is essential, as genetic proximity and differentiation have direct consequences for both conservation and breeding decisions. In the Spanish cattle breeds analyzed, the identification of divergent genomic regions helped clarify both historical divergence and more recent differentiation driven by breeding goals, conservation programs, and local management systems. Such information is relevant for matching breed resources with specific production environments, allowing farmers and breeders to choose populations according to adaptive capacity, management conditions, and market objectives. Likewise, the study of shared genomic regions among Sardinian local cattle breeds provided insight into the common ancestral background still preserved across these populations, while also highlighting the risk of losing unique adaptive variation through demographic decline or genetic homogenization. Recognizing what is shared and what is breed-specific is fundamental for designing conservation priorities and for integrating valuable traits of niche breeds, such as robustness, fertility, product quality, and environmental adaptation, into future breeding strategies. The identification of genomic regions associated with morphology, behaviour, adaptation, immunity, and productive traits confirmed that local breeds represent structured and functionally meaningful reservoirs of genetic variation. In the context of increasing environmental uncertainty, maintaining such diversity is not merely a cultural or historical priority, but a biological safeguard that preserves evolutionary options for future livestock systems. Beyond identifying genomic distinctiveness and similarities, the conservation of local breeds also requires monitoring their demographic and genetic sustainability. In this context, estimates of effective population size (Ne) are particularly informative. Although LD-based estimates of Ne are widely regarded as more informative for describing realized genetic diversity, census-based estimates still provide valuable complementary information. In particular, census-derived Ne reflects the current demographic structure of the breeding population, including the number and balance of reproductive males and females, and can be obtained even when genomic data are incomplete or unavailable. Therefore, combining census- and LD-based approaches may offer a more comprehensive view of population status, integrating demographic sustainability with realized genomic diversity. This is especially relevant in local breeds, where both management practices and genetic variability should be considered jointly.

At the same time, Chapter 4 and Chapter 5 focused on cosmopolitan cattle breeds highlighted the complexity of achieving sustainable genetic progress under intensive selection. The antagonistic relationship between production and fertility traits confirmed that genetic improvement cannot rely only on productivity-driven objectives. Nevertheless, the persistence of genetic variability for functional traits demonstrates that balanced selection remains achievable. The evaluation of EBV accuracy, bias, and dispersion across contexts further emphasized that genetic progress must be supported by robust, validated, and transparent assessment frameworks.

Taken together, these findings suggest that conservation and selection should not be viewed as opposing paradigms. Rather, they represent complementary components of a unified genetic strategy where local breeds provide reservoirs of adaptive potential, while cosmopolitan breeds ensure efficiency and food security. In this perspective, the role of animal genetics extends beyond improving performance, as it becomes central to designing resilient agricultural systems capable of responding to both present demands and uncertain future scenarios. The integration of biodiversity conservation and advanced quantitative methodologies thus emerges not as a theoretical ideal, but as a practical necessity for the next generation of breeding programs.

Future developments emerging from this thesis concern both biodiversity conservation and quantitative genetic improvement. For local Sardinian breeds, a natural next step is the use of whole-genome sequencing to uncover additional shared and breed-specific variants linked to adaptation, product quality, robustness, and environmental resilience. Beyond scientific interest, such knowledge could support the economic valorization of these populations within sustainable production systems. Although their production may be limited in quantity, these breeds may offer high organoleptic value and strong links to local identity, biodiversity-rich grazing systems, and low-input management. For modern breeding programs, future progress will increasingly depend on the integration of genomic information into selection indices, particularly for complex traits such as fertility, health, and resilience. In parallel, the validation of breeding values will remain an evolving challenge. Further research should focus on defining more informative validation cohorts based on the actual amount of available information for each sire, using indicators such as reliability or effective record contribution. This would allow a more realistic comparison of prediction quality across countries, traits, and levels of data availability.

From a farmer’s perspective, the practical value of this thesis lies in supporting more informed and forward-looking decisions. Livestock systems and market demands are continuously evolving, making it increasingly important to match breed choice, production goals, and environmental conditions. The results on local breeds highlight opportunities for farms operating in marginal areas or targeting quality-oriented niche markets, where adaptation, robustness, and product identity may represent competitive advantages. At the same time, the results on selection indexes and validation methods emphasize the importance of using reliable genetic tools to define breeding objectives and monitor real progress over time. Whether the goal is improving fertility, growth, calving performance, or efficiency, validated breeding values help farmers understand where their animals stand at national and international levels and make more effective selection decisions. Ultimately, genetics is not only a scientific discipline, but a practical management tool for improving profitability, resilience, and long-term sustainability.