Animal 18 (2024) 101142

Contents lists available at ScienceDirect

Animal
The international journal of animal biosciences

Investigation of cosmopolitan and local Italian beef cattle breeds uncover common patterns of heterozygosity

M.C. Fabbri [a,*], F. Tiezzi [a], A. Crovetti [a], C. Maltecca [a,b], R. Bozzi [a]

[a] Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università di Firenze, Firenze, Italy
[b] Department of Animal Science, North Carolina State University, Raleigh, NC 27695, United States

ARTICLE INFO

Article history:
Received 5 December 2023
Revised 15 March 2024
Accepted 18 March 2024
Available online 25 March 2024

Keywords:
Autochthonous breeds
Balancing selection
Beef Cattle
Fertility traits
Heterozygosity-Rich Regions

ABSTRACT

The analysis of livestock heterozygosity is less common compared to the study of homozygous patterns. Heterozygous-Rich Regions (**HRR**s) may harbor significant loci for functional traits such as immune response, survival rate, and fertility. For this reason, this study was conducted to investigate and characterize the heterozygosity patterns of four beef cattle breeds, which included two cosmopolitan breeds (Limousine and Charolaise) and two local breeds (Sarda and Sardo Bruna). Our analysis identified regions with a high degree of heterozygosity using a consecutive runs approach, the Tajima D test, nucleotide diversity estimation, and Hardy Weinberg equilibrium test. These regions exhibited recurrent heterozygosity peaks and were consistently found on specific chromosomes across all breeds, specifically autosomes 15, 16, 20, and 23. The cosmopolitan and Sardo Bruna breeds also displayed peaks on autosomes 2 and 21, respectively. Thirty-five top runs shared by more than 25% of the populations were identified. These genomic fragments encompassed 18 genes, two of which are directly linked to male fertility, while four are associated with lactation. Two other genes play roles in survival and immune response. Our study also detected a region related to growth and carcass traits in Limousine breed. Our analysis of heterozygosity-rich regions revealed particular segments of the cattle genome linked to various functional traits. It appears that balancing selection is occurring in specific regions within the four examined breeds, and unexpectedly, they are common across cosmopolitan and local breeds. The genes identified hold potential for applications in breeding programs and conservation studies to investigate the phenotypes associated with these heterozygous genotypes. In addition, Tajima D test, Nucleotide diversity, and Hardy Weinberg equilibrium test confirmed the presence of heterozygous fragments found with Heterozygous-Rich Regions analysis.

Implications

Directional and disruptive selection could affect population structure and homozygosity, but heterozygosity and balancing selection also define population history. Nucleotide diversity, Tajima D test and heterozygosity-rich regions are proposed to identify segments showing the impact of balancing selection, either by gene introgression or admixture. Regions resulting in high heterozygosity contained genes associated with fertility, growth, and adaptation, and the balancing selection footprint has been evidenced. Breeding programs and conservation schemes that use genomic information should alleviate the pressure generated by directional selection specifically on these regions, allowing the conservation of diversity and heterozygosity in the population.

Introduction

Natural selection plays a crucial role in promoting the survival of animals that exhibit improved adaptation to changing environmental conditions. Moreover, artificial selection has been widely applied in livestock species to attain more beneficial and economically valuable traits (Biscarini et al., 2015). Natural and artificial selection both apply selective pressure to specific genomic regions that control various traits associated with productivity and adaptability in diverse environments. These traits encompass fitness, body conformation, behavior, as well as resistance to diseases and parasites.

The availability of high-throughput sequencing and genotyping data for single nucleotide polymorphisms (**SNPs**) has made it possible to characterize genome segments based on their homozygosity and/or heterozygosity. One of the most widely used measures to describe and investigate homozygosity patterns is the runs of

* Corresponding author.
E-mail address: mariachiara.fabbri@unifi.it (M.C. Fabbri).

https://doi.org/10.1016/j.animal.2024.101142

homozygosity (**ROH**). ROH are contiguous segments of homozygous genotypes that are present in an individual due to parents transmitting identical haplotypes to their offspring (Ceballos et al., 2018). More recently, heterozygosity-rich regions (**HRRs**) have been proposed to complement homozygosity measures and investigate the significance of heterozygosity patterns. HRRs have been used to identify genomic regions that are under gene introgression or admixture (Tsartsianidou et al., 2021). HRRs can also provide insights into population structure and demographic history (Bizarria Dos Santos et al., 2021), and may harbor important loci for key functional traits such as immune response, survival rate, fertility, and other fitness traits (Mc Parland et al., 2009). Evidence suggests that increased heterozygosity occurs in regions under balancing selection or with a high recombination rate, as low linkage disequilibrium leads to high region diversity (Fijarczyk and Babik, 2015; Ferrer et al., 2016; Mulim et al., 2022). It has been extensively studied as directional and disruptive selection could affect population structure and homozygosity (Saravanan et al., 2020), but less attention has been given to how heterozygosity and balancing selection could define population history and vice versa.

Balancing selection is a common term that includes three types of mechanism (heterozygote advantage, negative frequency-dependent, or fluctuating selection) that maintain higher than expected levels of heterozygosity and allelic diversity within populations (Fijarczyk and Babik, 2015).

In assessing patterns of heterozygosity and the potential influence of balancing selection on a population, several parameters can be considered, including nucleotide diversity, the Tajima D test, and expected and observed heterozygosity derived from the Hardy-Weinberg equilibrium test.

Nucleotide diversity, often denoted as π, quantifies the average number of nucleotide differences per site when comparing two randomly selected DNA (Nei and Li, 1979). This metric contains information about a DNA region's functional significance, estimated lower in intergenic regions (Tatarinova et al., 2016), as well as about the population's demographic history (Neumann et al., 2023).

The Tajima D test statistic distinguishes between DNA sequences evolving randomly and those undergoing non-random processes, which may involve directional or balancing selection, demographic changes such as expansion or contraction, genetic hitchhiking, or introgression (Tajima, 1989).

Using different approaches (e.g., Tajima D test, HRR investigation) to estimate heterozygosity is important to evaluate the degree of agreement of results in order to make them as reproducible as possible. HRR approach is becoming popular in population genetics studies. HRRs are easy to obtain -the methodology is almost the same as ROH (Biscarini et al., 2019) and can be seen as the complementary measure of ROH. Additionally, HRR, compared to other approaches, offers some additional advantages since the length of the runs identified allows to better characterize the demographic history of populations: long HRRs occur from recent matings, as recombination has not had the possibility of breaking up the segment. At the same time, short HRR could indicate balancing selection or fixed haplotypes in the population as a result of selection. Finally, compared to more rigorous tests such as Tajima D, HRRs offer a more lenient approach in identifying regions of interest since, in selected populations, conditions close to mutation drift are not achievable, and measures of demography and selective history need to be investigated considering the population standing variation. An explicative study was a relatively recent work (Williams et al., 2016) where the heterozygosity in the Chillingham cattle, a local breed of northern England, not subject to selection, that remained a closed herd for at least 350 years, was analyzed. Results confirmed the lack of the population genetic variability, even if fertility remained high in the breed. Intriguingly, the authors identified regions that were maintained at a high level of heterozygosity. These regions could contain loci that contribute to the survival rate, fertility, and other fitness traits and can be segments of the genome where diversity could be very beneficial. Other authors (Fijarczyk and Babik, 2015) studied HRR in Pinzgauer cattle and identified genes important in biological processes, in particular genes that play a role in the Krebs cycle, as well as the innate immune and inflammatory response. Recently, a study examining heterozygosity-rich regions in the Purunã breed was conducted (Mulim et al., 2022). This breed, originating from Southern Brazil, is a composite population formed through the crossbreeding of Angus, Charolais, Canchim, and Caracu. The Purunã breed enjoys the benefits of heterosis resulting from these crossbreeds, characterized by superior growth performance. Their investigation revealed HRR, encompassing genes previously linked to carcass weight, meat, carcass quality, and marbling deposition. HRR could give new insights into understanding the relationship between genetic diversity and fitness, assessing the population's evolutionary potential and simultaneously predicting the reduction of breeds' genetic variability (Ferenčaković et al., 2017). In this study, four Italian beef breeds were considered: Limousine (**LIM**), Charolaise (**CHA**), Sarda (**SAR**), and Sardo Bruna (**SAB**), which have not been previously characterized in terms of patterns of heterozygosity. Two of them are cosmopolitan breeds under selection program, i.e., Limousine and Charolaise, reared in the Italian country and originating from France, from which animals and semen are still bought. Currently, Italian Limousine counts ~80 000 animals kept in more than 2 000 farms, while Charolaise in Italy consists of ~ 21 000 individuals distributed in approximately 1 000 farms. Limousine and Charolaise are in continuous population size expansion and are becoming two of the most important beef breeds reared in Italy.

Sarda and Sardo Bruna are local breeds enrolled in the register of Italian breeds at limited diffusion and are not subjected to a breeding program; they are reared exclusively on the Italian island of Sardinia and are originating from several past crosses between local and specialized breeds, such as Bruna Alpina, Maremmana, Piemontese, Simmental. Sarda and Sardo Bruna amount to ~ 21 000 and ~ 25 000 animals, respectively. Given the similar numerosity but the different selective history, we chose to compare Sarda and Sardo Bruna along with Limousine and Charolaise to identify patterns of heterozygosity based on a combination of HRR, Tajima D test, and nucleotide diversity, with the aim to elicit particular selection patterns present in the local breeds that might differentiate them from the cosmopolitan ones.

Material and methods

Samples, genotyping, and data editing

A total of 13 577 individuals were included in the analysis belonging to four breeds, which will be named LIM, CHA, SAR, and SAB hereafter. All individuals were genotyped with the GeneSeek GGP-LDv4 33 k (Illumina Inc., San Diego, California, USA) SNP chip. Sampled animals were 8 348 LIM, 2 331 CHA, 941 SAR, and 1 957 SAB. Genotype quality control and data filtering were performed with PLINK v1.9 (Chang et al., 2015): only SNPs located on the 29 autosomes were included (n = 28 289), and all individuals and SNPs with more than 10% of missing values were removed from the dataset. For the subsequent analyses, 27 091 SNPs and 12 611 individuals were retained (7 775 LIM, 2 041 CHA, 1 867 SAB, 928 SAR).

Multidimensional scaling analysis, admixture analysis, and editing

A multidimensional scaling analysis (**MDS**) was performed to investigate the population structure between the four breeds based on genetic distances, particularly to establish how phylogenetically distant local breeds were compared to the cosmopolitan ones. The first two dimensions were obtained with PLINK v1.9 (Chang et al., 2015) using the *--mds-plot* flag, which estimated distances based on the genome-wide pairwise Identical by State matrix. The proportion of mixed ancestry was assessed using the ADMIXTURE 1.22 software (Alexander et al., 2009), in order to identify potential recent crosses, which were then removed from the dataset to avoid overestimating heterozygosity due to admixture. The number of ancestries (K) to be retained in admixture (K = 2–4) was evaluated via a fivefold cross-validation (CV). The model with minimum CV error was selected to discard all individuals, which presented a probability of membership to one breed less than 80%. Several thresholds of admixture proportions of individuals (Q) were tested (i.e., Q >= 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9), and 0.8 resulted as the most acceptable threshold in terms of animals retained (∼11 500) and accuracy of breed membership (Fig. S1).

Detection of heterozygosity-rich regions and gene annotation

Analysis of runs of Heterozygosity-Rich Regions (HRRs) (also known as "runs of heterozygosity," ROHet) was conducted with the R package detectRUNS v. 0.9.5 (Biscarini et al., 2019). The consecutive method was used, which means that a scan of the genome SNP by SNP was directly carried out. A sensitivity analysis was performed on specific parameters that have been found to affect results in HRR detection: namely, the minimum number of SNPs in a run and the number of opposite and/or missing genotypes allowed (Biscarini et al., 2020; Chen et al., 2022). Consequently, three different scenarios were applied to the reduced dataset obtained after admixture and individual editing (n = 11 442). The three scenarios were ordered from the most liberal to the most conservative. Scenario 1 (**SC1**) provided the minimum number of SNP in an ROHet set to 10, and missing and homozygous genotypes were allowed, more precisely, one missing SNP and two homozygous SNPs. Scenario 2 (**SC2**) had intermediate restrictions: missing and homozygous SNPs remained as SC1, but the minimum number of SNPs in the run was increased (15 SNPs). Scenario 3 (**SC3**) was the most conservative and did not allow missing and homozygous SNPs, but the minimum number of SNPs was left as in the previous scenario (SNPs = 15). The minimum length of an ROHet was set to 10 Kbp, and the maximum gap between consecutive homozygous SNPs was 1 Mbp for all scenarios.

The HRR class of lengths were calculated considering five intervals: from 0 to 0.5 Mbp, > 0.5 Mbp and ≤ 1 Mbp, > 1 Mbp and ≤ 2 Mbp, > 2 Mbp and ≤ 4 Mbp and, > 4 Mbp.

Highly heterozygous genomic regions (HRR islands) were identified by selecting SNPs with an in-HRR frequency > 0.25 (Biscarini et al., 2020). The top HRRs were investigated using Genome Data Viewer of NCBI (https://www.ncbi.nlm.nih.gov/genome/gdv/), and Bos taurus genome assembly Bos taurus UMD 3.1.1 was used as reference (https://www.ncbi.nlm.nih.gov/data-hub/assembly/GCF_000003055.6/). For the Gene set enrichment analysis, the lists of protein-coding genes were uploaded to STRING 11.5 (Szklarczyk et al., 2017).

Observed and expected heterozygosity, nucleotide diversity, and Tajima's D statistic test

Genetic diversity of a population can be measured with several approaches. Here, three parameters were tested: nucleotide diversity, Tajima's D test values, and observed and expected heterozygosity, estimates that can be integrated and compared with HRR approach information. VCFtools (Danecek et al., 2011) was used to estimate each marker nucleotide diversity and to perform the Tajima's D test. The genomic windows analyzed were set to 10 kb for both methods. Tajima's D values can offer insights into the population's evolutionary trajectory: a value of 0 suggests that the population is evolving in line with mutation-drift equilibrium; values less than 0 indicate recent selective sweeps or population expansion following a bottleneck event; in contrast, values greater than 0 point to populations experiencing balancing selection or abrupt population contractions.

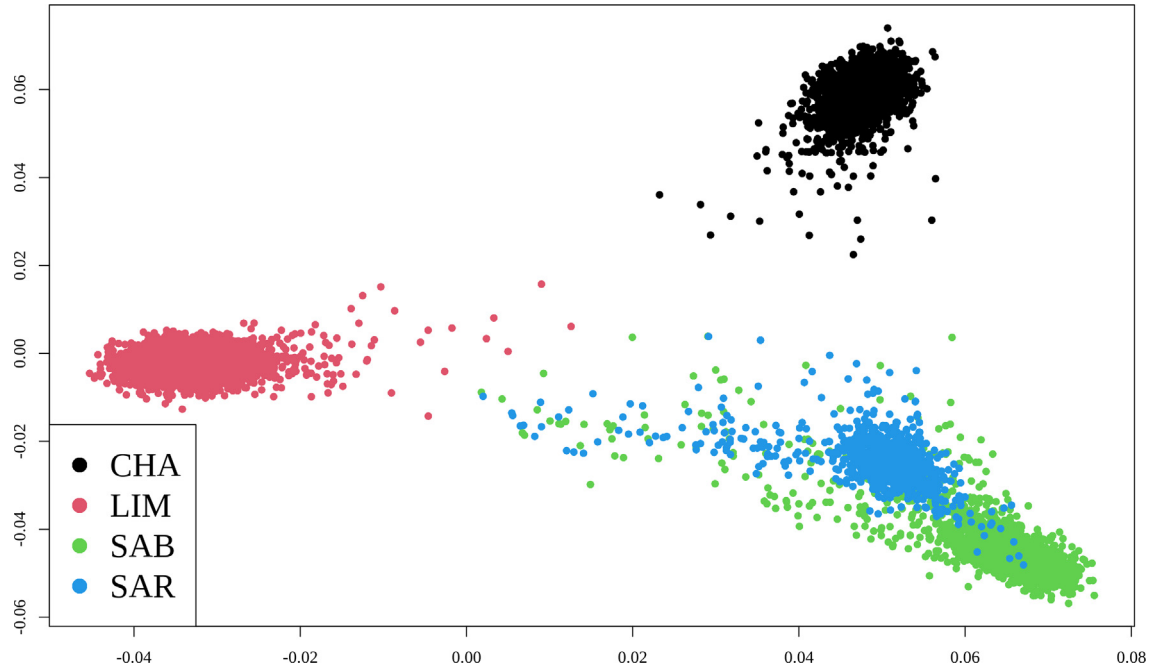

Fig. 1. Multidimensional-scaling plot of the four cattle breeds (CHA = Charolaise, LIM = Limousine, SAB = Sardo Bruna, SAR = Sarda).

Observed and expected heterozygosity level was obtained for each SNP using the Hardy-Weinberg test statistics implemented in the PLINK v.1.09 software (Chang et al., 2015).

Results

Multidimensional scaling analysis, admixture analysis, and editing

Fig. 1 depicts the MDS plot for the four breeds examined. A clear distinction between local and commercial populations emerged. CHA and LIM breeds displayed lower variability within breed, forming distinct clusters separated from SAR and SAB. However, some individuals within these clusters plotted at intermediate distances, potentially indicating crossbreeding. Conversely, Sardinian breeds exhibited noticeable introgression or genetic similarity among themselves, making it challenging to distinguish between the two breeds. Many animals appeared to be crossbreds, particularly SAR vs SAB and these hybrids vs LIM. CHA remained distinctly separate from the other three breeds.

To address this issue, we conducted an admixture analysis to assess the likelihood of individual membership to a specific breed and to exclude crossbred individuals that might introduce bias into the analysis. As illustrated in Fig. 2A, the admixture analysis clearly revealed shared ancestry proportions between SAB and SAR breeds, and some individuals within the LIM and CHA breeds displayed evidence of admixture. Therefore, a threshold of Q, i.e., the degree of assignment of each individual to one of the four breeds, was set to $\geq$ 80% to identify potential crossbreeds (Fig. S1). Subsequently, the MDS analysis was repeated and presented in Fig. 2B, with only samples that displayed high assignment ($\geq$ 80% probability) to one of the four breeds retained for the reduced dataset.

The reduced dataset comprised 11 442 individuals (LIM = 7 733; CHA = 1 988, SAB = 1 487 and SAR = 234).

Detection of heterozygosity-rich regions and gene annotation

To detect HRR, we assessed three scenarios in which three parameters were adjusted: the minimum number of SNPs in an HRR (changed from 10 in SC1 to 15 in SC2 and SC3), the number

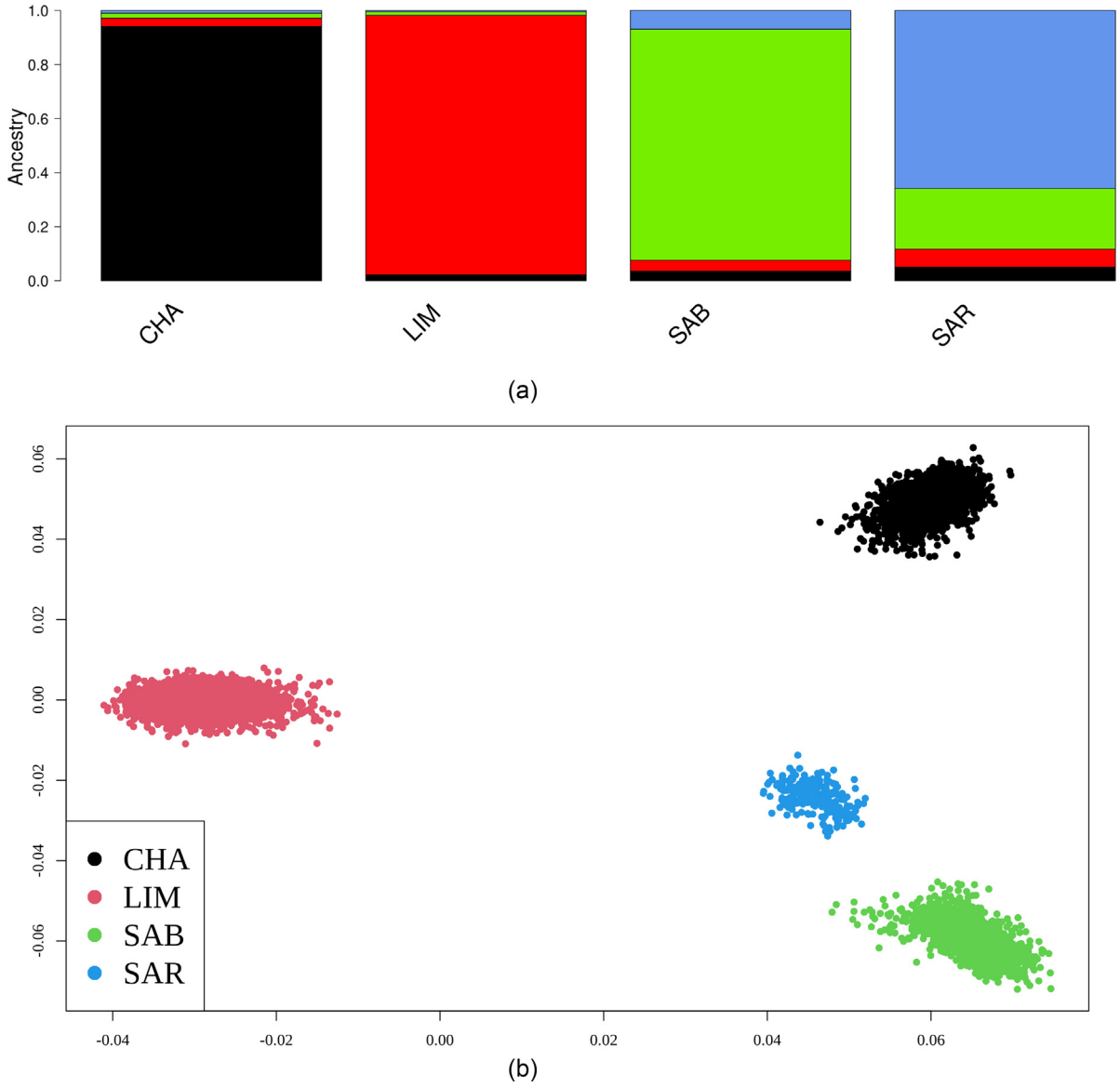

Fig. 2. (A) Admixture plot with the number of ancestries K = 4. (B) Multidimensional scaling plot using the reduced dataset. Only animals correctly assigned to one cattle breed with more than 80% of probabilities were retained and plotted (CHA = Charolaise, LIM = Limousine, SAB = Sardo Bruna, SAR = Sarda).

of missing and homozygous genotypes was set from 1 (SC1 and SC2) to 0 in SC3 and from 2 (SC1 and SC2) to 0 in SC3, respectively. The Venn diagram (Fig. 3) highlights SC2 as the most interesting scenario since all HRRs in SC2 were identified in the other two scenarios. Furthermore, when the minimum number of SNPs in an HRR was set to 10 (SC1), it resulted in a significant increase in the total number of HRRs, exceeding that of SC2, nearly tenfold. Finally, SC3 identified 220 unique HRRs and 694 runs shared with SC1 and SC2. The total number of HRR for each scenario, separated by breed, is provided in Table S1.

Results obtained from SC2 have been chosen for further analyses. The total number of HRR per breed is presented in Table 1.

The average number of HRRs per individual was higher for LIM (~14 HRRs). Results in SAB and CHA were similar, while SAR had the lowest number of HRR per individual (~12 runs). The length intervals chosen to analyze the HRR resulted in the distribution

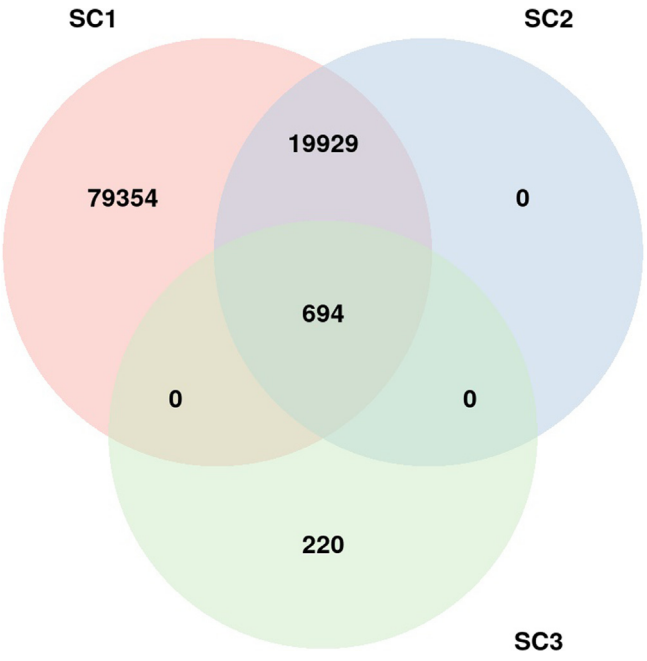

Fig. 3. Venn Diagram of the total number of Heterozygous-Rich Regions detected in each scenario (SC) for the cattle breeds.

Table 1
Total amount of Heterozygous-Rich Regions (HRRs) identified per cattle beef breed and the average number of HRRs per individual (between parentheses).

Breed[1]	No. HRR	No. of animals
CHA	27 066 (13.61)	1 988
LIM	110 970 (14.35)	7 733
SAB	20 049 (13.48)	1 487
SAR	2 780 (11.88)	234
Total	160 865	11 442

[1] CHA = Charolaise, LIM = Limousine, SAB = Sardo Bruna, SAR = Sarda.

Table 2
Number of Heterozygous-Rich Regions (HRRs) for class of length for each cattle beef breed, where CHA = Charolaise, LIM = Limousine, SAB = Sardo Bruna, SAR = Sarda.

Classes	CHA	LIM	SAB	SAR
0–0.5	14 110	63 686	10 946	1 531
0.5–1	3 625	13 825	2 476	327
1–2	7 471	26 234	5 228	714
2–4	1 858	7 206	1 397	208
> 4	2	19	2	0

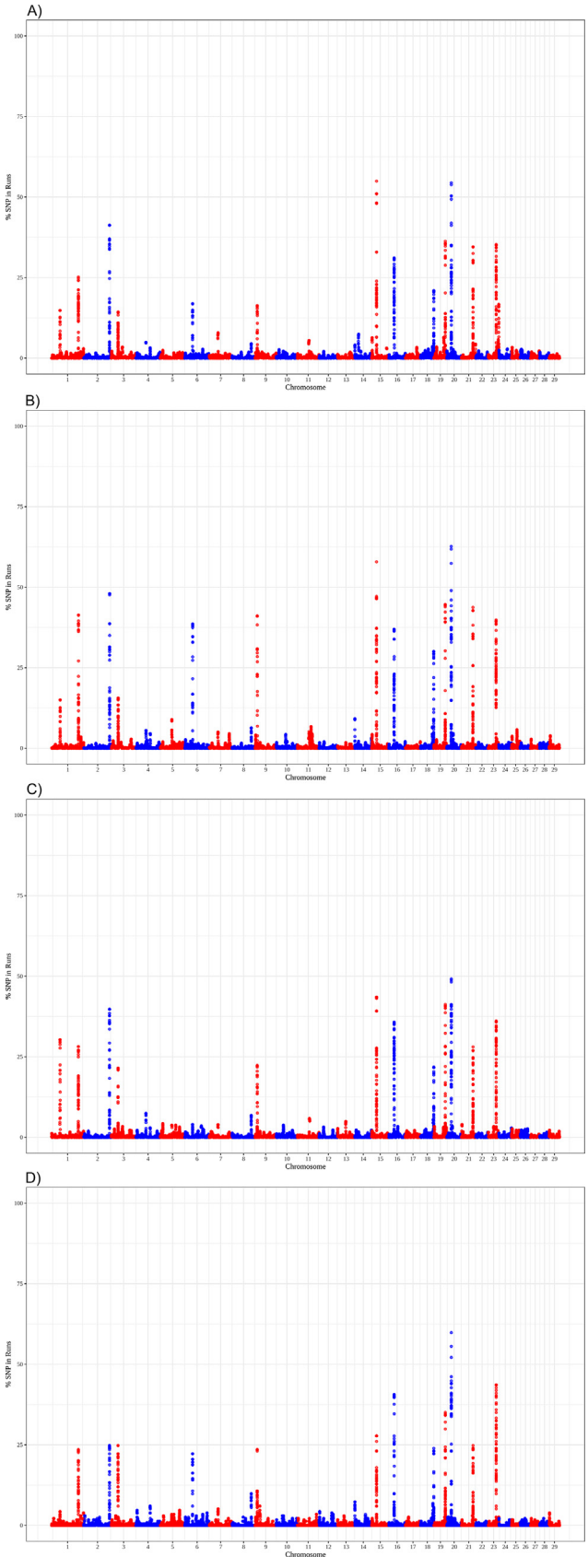

Fig. 4. Manhattan plots for each cattle breed, where A = Charolaise, B = Limousine, C = Sardo Bruna, D = Sarda.

Table 3
List of Heterozygous-Rich Regions present in more than 25% of each cattle population and genes identified within regions, where CHA = Charolaise, LIM = Limousine, SAB = Sardo Bruna, SAR = Sarda, SNP = single nucleotide polymorphism and BTA = Bos taurus autosome.

Breed	Start SNP	End SNP	BTA	from	To	Genes
SAB	BovineHD0100037693	BovineHD0100037709	1	132 339 365	132 398 347	
LIM	BovineHD0100037693	BovineHD0100037712	1	132 339 365	132 409 877	
LIM	BovineHD0200037002	BovineHD0200037019	2	127 476 802	127 540 073	*CATSPER4,CNKSR1,*
CHA	BovineHD0200037002	BovineHD0200037019	2	127 476 802	127 540 073	*ZNF593,FAM110D*
SAB	BovineHD0200037002	BovineHD0200037021	2	127 476 802	127 549 465	
CHA	BovineHD0200037056	BovineHD0200037075	2	127 644 712	127 680 352	*EXTL1*
SAB	BovineHD0200037056	BovineHD0200037075	2	127 644 712	127 680 352	
LIM	BovineHD0200037056	BovineHD0200037076	2	127 644 712	127 683 153	*EXTL1, SLC30A2*
LIM	6_38668893	BovineHD0600010764	6	38 668 893	38 950 070	*FAM184B, NCAPG, DCAF16, LCORL*
LIM	BovineHD0900002717	BovineHD0900031354	9	10 949 242	10 991 125	
LIM	BovineHD1500006301	BovineHD1500006338	15	24 175 920	24 238 959	*TTC12*
SAB	BovineHD1500006317	BovineHD1500006335	15	24 212 885	24 233 148	
SAR	BovineHD1500006317	BovineHD1500006330	15	24 212 885	24 224 422	
CHA	BovineHD1500006317	BovineHD1500006338	15	24 212 885	24 238 959	
SAB	BovineHD1600007151	BovineHD1600007172	16	25 709 923	25 774 613	
CHA	BovineHD1600024421	Hapmap45728-BTA-38205	16	25 903 792	26 455 382	*DUSP10*
SAB	BovineHD1600024421	BTA-88802-no-rs	16	25 903 792	26 613 598	
SAR	BovineHD1600007201	BTA-88802-no-rs	16	25 911 857	26 613 598	
LIM	BovineHD1600007207	BTA-88802-no-rs	16	25 933 132	26 613 598	
SAB	BovineHD1900015979	BovineHD1900015997	19	56 530 998	56 560 675	*SAP30BP,SMIM6, RECQL5*
LIM	BovineHD1900015983	BovineHD1900015998	19	56 537 361	56 564 441	
CHA	BovineHD1900015984	BovineHD1900015998	19	56 539 142	56 564 441	
SAR	BovineHD1900015984	BovineHD1900015990	19	56 539 142	56 552 581	*SAP30BP, RECQL5*
SAB	BovineHD2000006630	BovineHD2000006646	20	22 151 938	22 171 796	
LIM	BovineHD2000006630	BovineHD2000006681	20	22 151 938	22 234 812	
CHA	BovineHD2000006630	BovineHD2000006685	20	22 151 938	22 258 146	
SAR	BovineHD2000006630	BovineHD2000006691	20	22 151 938	22 288 482	
SAB	BovineHD2000006669	BovineHD2000006701	20	22 216 004	22 316 326	*MIER3*
LIM	BovineHD2000006690	BovineHD2000006721	20	22 285 687	22 375 058	
CHA	BovineHD2100016595	BovineHD2100016614	21	57 719 367	57 750 099	*MIR2284F, SLC24A4*
LIM	BovineHD2100016595	BovineHD2100016622	21	57 719 367	57 763 783	
SAR	BovineHD2300011345	BovineHD2300011382	23	39 269 822	39 377 385	
LIM	BovineHD4100016194	BovineHD2300011366	23	39 285 499	39 340 244	
CHA	BovineHD4100016194	BovineHD2300011366	23	39 285 499	39 340 244	
SAB	ARS-BFGL-NGS-65768	BovineHD2300011382	23	39 303 680	39 377 385	

reported in Table 2. As anticipated, the greatest number of HRR was observed in the shortest length class (HRR < 0.5 Mbp). The number of HRR in this category ranged from approximately 63 000 in the LIM breed to around 1 500 in the SAR breed. However, the intermediate length categories (1–2 Mbp and 2–4 Mbp) also showed elevated values, with the lowest count of less than one hundred HRRs only in the SAR breed. Additionally, two HRRs larger than 4 Mbp were identified in CHA and SAB, while LIM had 19 long runs of heterozygosity. It is worth noting that the Limousine breed displayed the highest number of HRR in all length classes, emphasizing the importance of sample size in such analyses to prevent potential bias and misinterpretations, considering that results are often reported as the total amount of segments.

Manhattan plots of HRR regions are reported in Fig. 4. Thirteen peaks were defined for cosmopolitan breeds and 12 peaks for local ones. Peaks on Bos taurus autosome 15 (**BTA15**), BTA16, BTA20, and BTA23 were identified in all breeds. Peaks on BTA2 and BTA21 were identified on cosmopolitan and SAB breeds, while a peak on BTA 1 was identified only in LIM and SAB.

To identify the more prevalent HRR in the studied populations, we selected the HRR represented in more than 25% of all individuals for each breed. This threshold retained 35 HRRs that collectively contained 18 unique genes (Table 3). As shown in the Manhattan plots, four genomic regions were found to be common across all four breeds, although they did not overlap precisely. These recurrent HRRs were observed in the following regions: (i) BTA15, which included the *TTC12* gene; (ii) BTA16, featuring the *DUSP10* gene; (iii) BTA19, shared regions across breeds containing the *SAP30BP* and *RECQL5* genes; (iv) the top HRR on BTA20 and BTA23 did not contain annotated genes. Additionally, on BTA2, two HRRs were detected in the cosmopolitan and Sardo Bruna breeds. The first one included four genes (*CATSPER4*, *CNKSR1*, *ZNF593*, *FAM110D*), and the second one harbored the *EXTL1* gene. The cosmopolitan breeds exhibited a unique HRR not found in the local breeds, located on BTA21 and containing the *MIR2284F* and *SLC24A4* genes.

The LIM breed exhibited a distinct breed-specific HRR on BTA6, which included four annotated genes (*FAM184B*, *NCAPG*, *DCAF16*,

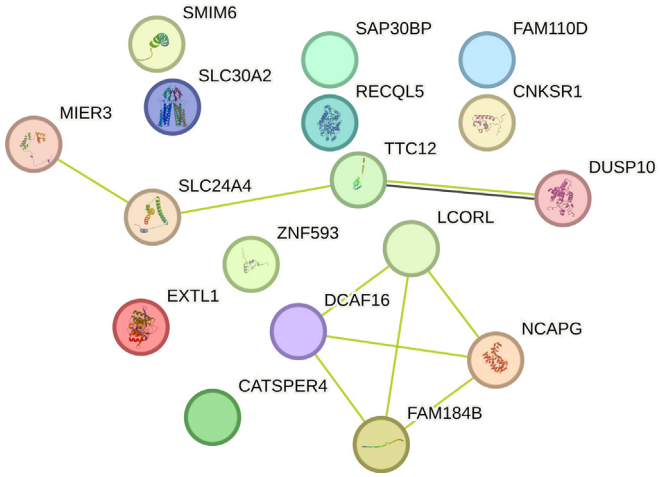

Fig. 5. Network of proteins expressed by genes identified in significant Heterozygous-Rich Regions (HRRs) in the four cattle breeds. Proteins included in all significant HRRs identified were reported.

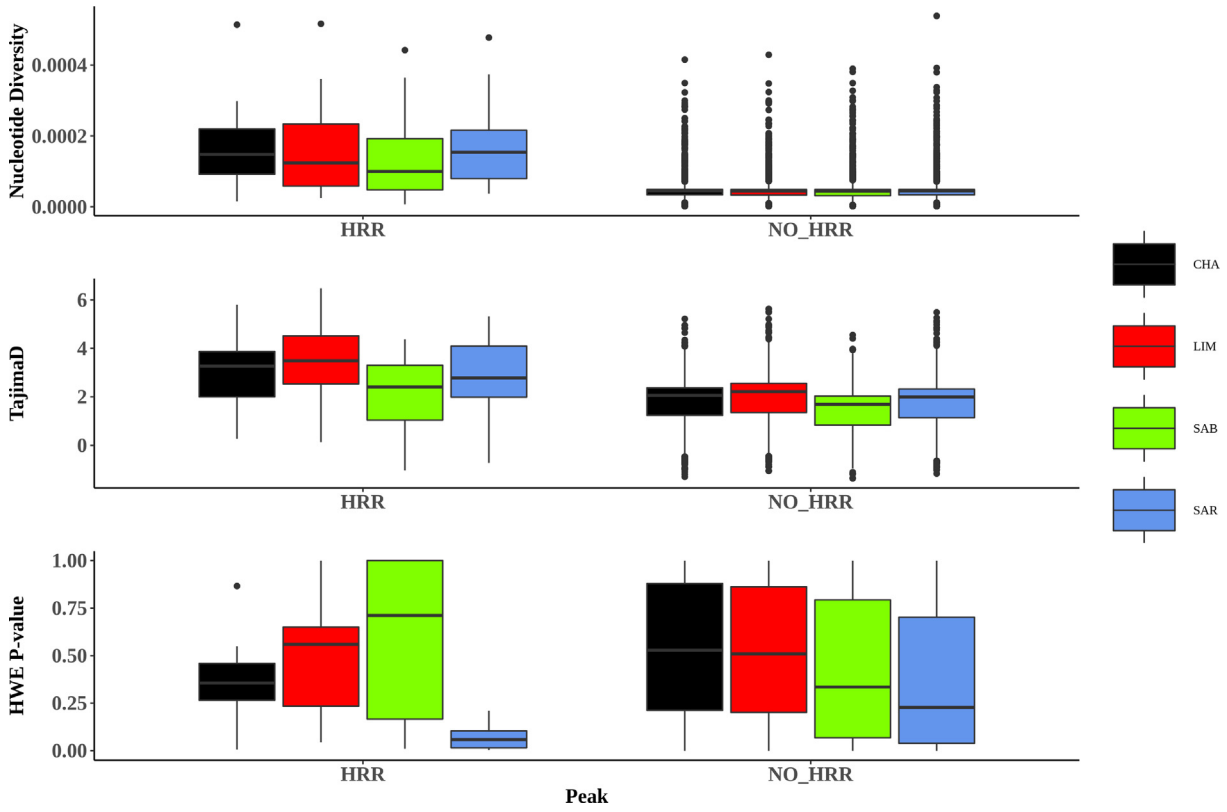

Fig. 6. Cattle beef breeds heterozygous estimates (Nucleotide Diversity, Tajima D test estimate, Hardy Weinberg (HWE) P value) for Single Nucleotide Polymorphisms included in Heterozygosity-Rich Regions (HRRs) island compared to Single Nucleotide Polymorphisms and intervals not included (CHA = Charolaise, LIM = Limousine, SAB = Sardo Bruna, SAR = Sarda).

and *LCORL*). A unique HRR on BTA20 containing the *MIER3* gene was identified in the SAB breed. Notably, this gene was located downstream from a common HRR shared by all four breeds, although it lacked any annotated genes.

The potential interactions among the 18 genes are illustrated in Fig. 5. This analysis revealed potential connections: the *TTC12* and *DUSP10* genes showed co-expression, further supporting the significance of the HRR found on BTA15 and BTA16. In the same cluster, alongside *TTC12* and *DUSP10*, the *SLC24A4* gene was identified, and interestingly, *MIER3* (which was found in the SAB breed) was positioned downstream of a significant genomic region shared by all breeds.

The last noteworthy cluster related to the HRR was discovered in the LIM breed on BTA6, spanning from 38.6 to 38.9 million base pairs (Mbp). This cluster encompassed four genes: *FAM184B*, *NCAPG*, *DCAF16*, and *LCORL*.

Nucleotide diversity, Tajima's D statistic test, and observed and expected heterozygosity

The average nucleotide diversity varied slightly among the four breeds, with the highest diversity observed in CHA, followed by SAR, LIM, and SAB breeds. π values were of approximately 4.10×10^{-5}, 4.08×10^{-5}, 4.06×10^{-5}, and 3.95×10^{-5} for the four breeds, respectively. The autosomes exhibiting the highest diversity values were BTA15, BTA19, BTA20, BTA21, and BTA23 (Fig. S2), which were also the chromosomes featuring the most prominent heterozygosity islands.

Nucleotide diversity values within regions that include HRR islands were compared with values from other genomic intervals (Fig. 6), resulting in significantly higher and more distinctly distributed π values within HRR regions.

No detectable differences in π values within HRR regions were found between breeds, indicating very similar results across the breeds considered.

Average Tajima's D values ranged from 1.314 in the SAB breed to 1.819 in the LIM breed, with CHA and SAR displaying values close to 1.60. Variations in values were observed among chromosomes (Table S2), with higher average values on BTA18, BTA19, BTA23, and BTA28. When examining intervals containing significant HRR, Tajima's D estimates were notably higher than the population mean (Fig. 6, Table 4). Although SAB had lower estimates, they remained consistently above 0.

We assessed observed and expected heterozygosity using *P*-values from the exact test for Hardy-Weinberg equilibrium. The average *P*-values per breed were as follows: CHA = 0.53, LIM = 0.52, SAB = 0.43, and SAR = 0.37. As anticipated, the Sardinian breeds exhibited slightly lower average values compared to the cosmopolitan breeds. A more detailed examination within each breed (Fig. 6) revealed that CHA had lower HWE *P*-values with reduced variability for SNPs included in HRR. LIM maintained similar SNP estimates both inside and outside HRR. In contrast, SAB

Table 4
Tajima D test values for each cattle breed in significant Heterozygous-Rich Regions (HRRs), where CHA = Charolaise, LIM = Limousine, SAB = Sardo Bruna, SAR = Sarda and BTA = Bos taurus autosome.

BTA	CHA	LIM	SAB	SAR
2	3.056	3.244	2.248	3.265
2	2.535	3.334	1.825	2.771
15	3.920	4.474	2.458	3.369
16	3.022	3.197	2.851	2.802
20	2.065	3.067	1.813	2.417
23	3.303	3.707	3.027	3.280

displayed higher variability in SNP *P*-values within HRR compared to other genomic regions. SAR appeared to have SNPs generally not in equilibrium within HRR, a condition not evident when considering the average HWE *P*-values (Fig. 6).

Discussion

Previous studies have highlighted the reduced genetic variability in cosmopolitan breeds compared to their local counterparts (Makina et al., 2014; Senczuk et al., 2020). To design effective conservation programs and accurately assess the risk status of the breeds we are studying, it is essential to examine the demographic history of these populations. This study focused on the cosmopolitan breeds Limousine and Charolaise and the local Sarda and Sardo Bruna breeds. While previous studies have described genetic diversity parameters for these breeds, including inbreeding estimates, the length of runs of homozygosity, and population sizes (Fabbri et al., 2021), no studies have investigated and characterized these four breeds in terms of heterozygosity and selection patterns.

Multidimensional scaling analysis and admixture analysis

Admixture and introgression events were expected based on breed histories and their geographical distribution, particularly for the SAR and SAB breeds. Historically, the Sarda breed exhibited considerable variability in morphological, reproductive, and productive traits, stemming from 2-3 distinct subpopulations adapting to diverse environmental conditions in the rearing areas. This historical context supports the hypothesis of crossbreeding with Sardo Bruna and possibly other local breeds, such as Sardo Modicana, as well as more productive breeds like Limousine, which is extensively reared in Sardinia.

Similarly, SAB exhibits phenotypic similarities to Bruna Alpina, resulting from crossbreeding initiated in the 19th century. Since the black coat phenotype disqualifies animals from the national register, animals classified as crossbreds in this study likely resulted from less recent admixture events.

Genetic distances observed within and between CHA and LIM breeds recapitulate these cosmopolitan beef breeds' demographic history and breeding programs, each characterized by distinct phenotypic traits.

Detection of heterozygosity-rich regions and gene annotation

Determining the parameters for analyzing HRR is a challenging aspect of this type of investigation. Selecting values to define an HRR is often arbitrary, leading to non-overlapping and non-reproducible outcomes. This can make it difficult to compare results between different studies and among animals of the same breed. To address these challenges, we conducted a sensitivity analysis. Based on our findings, we made specific parameter choices for our study. We adopted a more stringent criterion for the minimum number of SNPs and a less stringent one for handling missing and homozygous genotypes. Since these studies are often sensitive to parameter selection, we have included information on all three scenarios in the paper to offer guidance for future studies. Ultimately, we used the results from Scenario 2 for gene enrichment analysis, where we determined that fifteen SNPs should be set as the minimum number of markers in an HRR, allowing for one missing value and two homozygous SNPs. We also increased the animal sample size and selected a higher threshold for identifying HRR islands compared to previous studies (Biscarini et al., 2020; Mulim et al., 2022) to decrease the chance of spurious results.

In our study, we have identified two distinct gene clusters. The first cluster includes regions shared by both local and cosmopolitan breeds, containing genes such as *TTC12*, *DUSP10*, *SLC24A4*, and *MIER3*. The second cluster is unique to the LIM breed, comprising genes like *FAM184B*, *NCAPG*, *DCAF16*, and *LCORL*. Notably, regions on BTA2, while not forming networks, exhibit significant overlap across three or more breeds.

The *TTC12* gene is particularly interesting due to its association with male fertility. It plays a crucial role in assembling dynein arms within motile cilia present in human respiratory cells and sperm flagella. Mutations in the *TTC12* gene can lead to alterations in both outer and inner dynein arm complexes (Thomas et al., 2020), which are essential for the motility of cilia and flagella in humans. These defects manifest as primary ciliary dyskinesia, a condition characterized by airway infections and male infertility (Thomas et al., 2020).

A human study has identified the *TTC12* gene as one of the 40 genes linked to male infertility, particularly due to asthenoteratozoospermia, a condition characterized by reduced sperm motility and increased morphological abnormalities in sperm (Meng et al., 2023).

In cattle, especially within the Holstein breed, *TTC12* has been proposed as a candidate gene associated with prefreeze semen quality traits (Abril-Parreño et al., 2023). Additionally, in Murrah buffalo, this gene has shown associations with the age at first calving (George et al., 2023), highlighting its significant role in both male and female fertility. Indeed, the negative effect of inbreeding depression in selected cattle (Holstein breed) was defined by a recent study (Ghoreishifar et al., 2023), which revealed that the increase of 1% in autozygosity led to a decrease in sperm motility and sperm concentration (0.28 and 0.42%, respectively). These results reinforced heterozygosity's importance in these genomic regions related to male fertility.

We identified the *DUSP10* (dual-specificity phosphatase 10) gene in the same cluster. *DUSP10's* role involves regulating inflammation - it can either increase or decrease inflammation in response to factors like infection, injury, disease, or certain treatments (Jiménez-Martínez et al., 2019). It plays a key part in both innate and adaptive immune responses by inhibiting inflammation (Zhang et al., 2016). In cattle, *DUSP10* has also been linked to fat metabolism and adipogenesis in Holstein and Wagyu cattle (Huang et al., 2017), as well as carcass traits in Charolaise cattle (Keogh et al., 2021). This connection likely arises from *DUSP10's* role in regulating the MAPK (Mitogen-activated protein kinase) signaling pathway, which, in turn, impacts growth and proliferative processes.

The third gene, clustering with *TTC12* and *DUSP10*, is the *SLC24A4* (solute carrier family 24 member 4) gene, part of the potassium-dependent sodium or calcium exchanger protein family. Research has linked *SLC24A4* to body conformation (Yan et al., 2021) and calving interval traits in dairy cattle (Nayeri et al., 2016). Notably, signatures of selection housing the *SLC24A4* gene were found in dairy cows (Maiorano et al., 2018). Our study detected *SLC24A4* in an HRR island shared exclusively by Limousine and Charolaise, two prevalent beef cattle breeds in Italy. Importantly, given its connection to pigmentation traits (Sulem et al., 2007), this gene could have implications for health and adaptation, being associated in humans with hair, eye color, and sun sensitivity (Gerstenblith,Shi,and Landi, 2010).

The last gene within this cluster is *MIER3* (Mesoderm Induction Early Response 1, Family Member 3). In a genome-wide association study across multiple cattle breeds focused on milk production, the *MIER3* gene was found to be located in proximity to the most significant SNP associated with survival (Illa et al., 2021). Collectively, these genes constitute the first cluster identified in our study.

The second gene cluster was exclusive to the Limousine breed and encompassed significant HRRs housing the *FAM184B* (Family With Sequence Similarity 184 Member B), *NCAPG* (Non-SMC Condensin I Complex Subunit G), *DCAF16* (DDB1 And CUL4 Associated Factor 16), and *LCORL* (Ligand Dependent Nuclear Receptor Corepressor Like) genes. This region on BTA6 has been the focus of several studies on cattle carcass and growth traits (Lindholm-Perry et al., 2011; Zhang et al., 2016; Xia et al., 2017). Notably, multi-strategy genome-wide association studies have identified the DCAF16-NCAPG region as a susceptibility locus for average daily gain in Simmental beef cattle (Zhang et al., 2016). Additionally, *NCAPG* has been recognized as a significant modulator of cattle's postnatal growth and lipid deposition (Weikard et al., 2010). Downstream of *NCAPG* and *DCAF16* is *LCORL*, which has been associated with height and body size in various domestic animal species, including cattle (Signer-Hasler et al., 2012; Takasuga, 2016; Plassais et al., 2019; Saif et al., 2020). Notably, significant SNPs near the *FAM184B* and *LCORL* genes have also been linked to cattle's BW and pH value traits (Xia et al., 2017). Nevertheless, it is important to consider increasing the sample size for CHA and local breeds in future studies to ensure that the identification of this cluster solely in LIM (~8 000 samples) is not due to statistical power limitations and/or the sample collection strategy.

The HRR on BTA19 contained the SAP30BP (SAP30 Binding Protein), *SMIM6* (Small Integral Membrane Protein 6), and *RECQL5* (RecQ Like Helicase 5) genes, and these genes were common to all breeds. *SAP30BP* plays a role in modulating symbiont transcription in humans and positively regulates histone deacetylation. *SMIM6* is predicted to negatively regulate calcium ion binding activity, while *RECQL5* is an important helicase for maintaining human genome stability. While no direct associations have been established between these genes and production or fitness traits in livestock, the region is intriguing, as it appears in both cosmopolitan and local breeds, hinting at a potential role in survival traits.

The two regions on BTA2 were identified in LIM, CHA, and SAB but not in SAR. It is possible that the sample size in SAR, after data pruning, hindered their detection in that breed. Nonetheless, these regions contained two groups of genes: (*CATSPER4*, *CNKSR1*, *ZNF593*, and *FAM110D*) and (*EXTL1* and *SLC30A2*) genes.

Some of these genes have been linked to fertility or lactation traits. *CATSPER4*, which belongs to the CatSper family, plays a crucial role in sperm motility and sodium ion transport, making it related to male infertility (Sivakumar et al., 2018), similar to the previously discussed *TTC12* gene. *ZNF593* (Zinc Finger Protein 593) and *FAM110D* (Family With Sequence Similarity 110 Member D) were identified in a genomic region associated with lactation persistency in Holstein cattle (Do et al., 2017). Notably, *SLC30A2* (Solute Carrier Family 30 Member 2) is linked to human lactation, with eight missense mutations causing pathologically low breast milk zinc concentrations (approximately 50–95% reduction) and severe zinc deficiency in infants (Itsumura et al., 2016; Lee et al., 2018).

The gene annotation performed in the current study for the identified HRR highlights how these heterozygosity regions could potentially harbor significant loci related to key functional traits such as fertility, survival rate, and growth. An association study could be performed to complete and extend these findings.

Nucleotide diversity, Tajima's D test observed and expected heterozygosity

The comparison of diversity estimates between breeds and chromosomes confirms the significance of regions identified through the HRR approach, as we indeed found high agreement between HRR analysis and Tajima D test values, nucleotide diversity estimates, and HWE test *P*-values. Notably, the areas with the highest nucleotide diversity closely align with the most frequently occurring heterozygous regions, as indicated by elevated Tajima D test values in these peaks. All breeds showed almost similar and coherent estimates in HRR and out of HRR for the parameters as mentioned above. HWE *P*-value had different trends in SAB and SAR within heterozygous regions: SAB had a high variability within group, while SAR seemed to be characterized by loci, not in equilibrium. The latter could be for the reduced sample size of SAR or could harbor a specificity of this breed. Further analyses are needed on the Sarda breed. These findings suggest that these regions might be subject to balancing selection, a process wherein multiple alleles are maintained within a population, potentially leading to their long-term preservation throughout evolution.

Furthermore, from our analysis, it emerged that several genes found in these regions are associated with fertility, growth, and lactation. Maintaining heterozygosity within these regions could be linked to a selective advantage, albeit further studies would be needed to confirm these suggestive results.

Conclusion

This study is the first to investigate heterozygosity patterns in Italian Limousine, Charolaise, and Sarda and Sardo Bruna local beef breeds. While the genetic distance confirms the divergence between selectively bred and autochthonous unselected populations, the trends and distribution of heterozygosity paint a different picture. More recently, heterozygosity-rich regions (HRRs) have been proposed to complement homozygosity measures and investigate the significance of heterozygosity patterns. Despite distinct population structures and demographic histories, almost identical heterozygous genomic regions across all four breeds were found. These regions with high heterozygosity contained genes associated with fertility, survival, growth, and adaptation traits. It could be interesting to associate HRR found in this study with phenotypic records, such as fertility and survival. The results of this study allow us better to understand the structure and pattern of Heterozygosity-rich regions, to highlight genomic regions under balancing selection, as well as to define potential candidate genes of fertility and fitness traits in Italian beef cattle breeds, which could result in fundamental both in breeding programs and in conservation schemes. It could be interesting to extend this study to other cattle breeds to verify if these heterozygous regions are characteristic of Italian beef breeds, or specific beef breeds, or in more general, in cattle population.

Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.animal.2024.101142.

Ethics approval

Not applicable.

Data and model availability statement

The datasets generated and/or analyzed during the current study are not publicly available due owned by a third party, ANACLI, Associazione Nazionale degli Allevatori delle razze bovine Charolaise e Limousine Italiane (https://www.anacli.it/), but are available from the corresponding author on reasonable request.

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

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

Author ORCIDs

M.C. Fabbri: https://orcid.org/0000-0002-3224-745X.
F. Tiezzi: https://orcid.org/0000-0002-4358-9236.
C. Maltecca: https://orcid.org/0000-0002-9996-4680.
R. Bozzi: https://orcid.org/0000-0001-8854-0834.

CRediT authorship contribution statement

M.C. Fabbri: Software, Methodology, Formal analysis, Data curation, Conceptualization. **F. Tiezzi:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **A. Crovetti:** Writing – original draft, Investigation. **C. Maltecca:** Writing – review & editing, Supervision, Methodology, Conceptualization. **R. Bozzi:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of interest

None.

Acknowledgments

We acknowledge Associazione Nazionale Allevatori delle razze bovine Charolaise e Limousine (ANACLI) for providing the data.

Financial support statement

This work was financially supported by Associazione Nazionale degli Allevatori delle razze bovine Charolaise e Limousine Italiane (Cup:J89H18000010005, URL: http://www.anacli.it/ibeef). Grant was received by RB. The funders had no role in study design and analysis, decision to publish, or manuscript preparation.

References

Abril-Parreño, L., Carthy, T.R., Keogh, K., Štiavnická, M., O'Meara, C., Lonergan, P., Kenny, D.A., Fair, S., 2023. Genome-wide association study reveals candidate markers related to field fertility and semen quality traits in Holstein-Friesian bulls. Animal 17, 100841.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19, 1655–1664.

Biscarini F, Cozzi P, Gaspa G and Marras G 2019. detectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes. Accessed on 15 March 2024 from https://rdrr.io/cran/detectRUNS/f/vignettes/detectRUNS.vignette.Rmd.

Biscarini, F., Nicolazzi, E.L., Stella, A., Boettcher, P.J., Gandini, G., 2015. Challenges and opportunities in genetic improvement of local livestock breeds. Frontiers in Genetics 6, 33.

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G., Ciampolini, R., 2020. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. Animals 10, 2285.

Bizarria Dos Santos, W., Pimenta Schettini, G., Fonseca, M.G., Pereira, G.L., Loyola Chardulo, L.A., Rodrigues Machado Neto, O., Baldassini, W.A., Nunes De Oliveira, H., Abdallah Curi, R., 2021. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. Journal of Animal Breeding and Genetics 138, 161–173.

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., Wilson, J.F., 2018. Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics 19, 220–234.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM and Lee JJ 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, s13742-015-0047–8.

Chen, Z., Zhang, Z., Wang, Z., Zhang, Z., Wang, Q., Pan, Y., 2022. Heterozygosity and homozygosity regions affect reproductive success and the loss of reproduction: a case study with litter traits in pigs. Computational and Structural Biotechnology Journal 20, 4060–4071.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, and 1000 Genomes Project Analysis Group 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Do, D.N., Bissonnette, N., Lacasse, P., Miglior, F., Sargolzaei, M., Zhao, X., Ibeagha-Awemu, E.M., 2017. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. Journal of Dairy Science 100, 1955–1970.

Fabbri, M.C., Dadousis, C., Tiezzi, F., Maltecca, C., Lozada-Soto, E., Biffani, S., Bozzi, R., 2021. Genetic diversity and population history of eight Italian beef cattle breeds using measures of autozygosity. PLOS ONE 16, e0248087.

Ferenčaković, M., Banadinović, M., Mercvajler, M., Khayatzadeh, N., Čurik, V.C., Curik, I., Sölkner, J., Mészáros, G., 2017. Mapping of heterozygosity rich regions in Austrian pinzgauer cattle. Acta Agriculturae Slovenica 5, 42.

Ferrer, E.S., García-Navas, V., Sanz, J.J., Ortego, J., 2016. The strength of the association between heterozygosity and probability of interannual local recruitment increases with environmental harshness in blue tits. Ecology and Evolution 6, 8857–8869.

Fijarczyk, A., Babik, W., 2015. Detecting balancing selection in genomes: limits and prospects. Molecular Ecology 24, 3529–3545.

George, L., Alex, R., Sukhija, N., Jaglan, K., Vohra, V., Kumar, R., Verma, A., 2023. Genetic improvement of economic traits in Murrah buffalo using significant SNPs from genome wide association study. Tropical Animal Health and Production 55, 199.

Gerstenblith, M.R., Shi, J., Landi, M.T., 2010. Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. Pigment Cell & Melanoma Research 23, 587–606.

Ghoreishifar, M., Vahedi, S.M., Salek Ardestani, S., Khansefid, M., Pryce, J.E., 2023. Genome-wide assessment and mapping of inbreeding depression identifies candidate genes associated with semen traits in Holstein bulls. BMC Genomics 24, 230.

Huang, W., Guo, Y., Du, W., Zhang, X., Li, A., Miao, X., 2017. Global transcriptome analysis identifies differentially expressed genes related to lipid metabolism in Wagyu and Holstein cattle. Scientific Reports 7, 5278.

Illa, S.K., Mukherjee, S., Nath, S., Mukherjee, A., 2021. Genome-wide scanning for signatures of selection revealed the putative genomic regions and candidate genes controlling milk composition and coat color traits in Sahiwal cattle. Frontiers in Genetics 12, 699422.

Itsumura, N., Kibihara, Y., Fukue, K., Miyata, A., Fukushima, K., Tamagawa-Mineoka, R., Katoh, N., Nishito, Y., Ishida, R., Narita, H., Kodama, H., Kambe, T., 2016. Novel mutations in SLC30A2 involved in the pathogenesis of transient neonatal zinc deficiency. Pediatric Research 80, 586–594.

Jiménez-Martínez, M., Stamatakis, K., Fresno, M., 2019. The dual-specificity phosphatase 10 (DUSP10): its role in cancer, inflammation, and immunity. International Journal of Molecular Sciences 20, 1626.

Keogh, K., Carthy, T.R., McClure, M.C., Waters, S.M., Kenny, D.A., 2021. Genome-wide association study of economically important traits in Charolais and Limousin beef cows. Animal 15, 100011.

Lee, S., Zhou, Y., Gill, D.L., Kelleher, S.L., 2018. A genetic variant in SLC30A2 causes breast dysfunction during lactation by inducing ER stress, oxidative stress and epithelial barrier defects. Scientific Reports 8, 3542.

Lindholm-Perry, A.K., Sexten, A.K., Kuehn, L.A., Smith, T.P.L., King, D.A., Shackelford, S.D., Wheeler, T.L., Ferrell, C.L., Jenkins, T.G., Snelling, W.M., Freetly, H.C., 2011. Association, effects and validation of polymorphisms within the NCAPG - LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. BMC genetics 12, 103.

Maiorano, A.M., Lourenco, D.L., Tsuruta, S., Ospina, A.M.T., Stafuzza, N.B., Masuda, Y., Filho, A.E.V., Cyrillo, J.N.D.S.G., Curi, R.A., Silva, J.A.I.D.V., 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLOS ONE 13, e0200694.

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

Mc Parland, S., Kearney, F., Berry, D.P., 2009. Purging of inbreeding depression within the Irish Holstein-Friesian population. Genetics Selection Evolution 41, 16.

Meng, L., Liu, Q., Tan, C., Xu, X., He, W., Hu, T., Tu, C., Li, Y., Du, J., Zhang, Q., Lu, G., Fan, L.-Q., Lin, G., Nie, H., Zhang, H., Tan, Y.-Q., 2023. Novel homozygous variants in TTC12 cause male infertility with asthenoteratozoospermia owing to dynein arm complex and mitochondrial sheath defects in flagella. Frontiers in Cell and Developmental Biology 11, 1184331.

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

Nayeri, S., Sargolzaei, M., Abo-Ismail, M.K., May, N., Miller, S.P., Schenkel, F., Moore, S.S., Stothard, P., 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genetics 17, 75.

Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76, 5269–5273.

Neumann, G.B., Korkuć, P., Arends, D., Wolf, M.J., May, K., König, S., Brockmann, G.A., 2023. Genomic diversity and relationship analyses of endangered German Black Pied cattle (DSN) to 68 other taurine breeds based on whole-genome sequencing. Frontiers in Genetics 13, 993959.
Plassais, J., Kim, J., Davis, B.W., Karyadi, D.M., Hogan, A.N., Harris, A.C., Decker, B., Parker, H.G., Ostrander, E.A., 2019. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nature Communications 10, 1489.
Saif, R., Henkel, J., Jagannathan, V., Drögemüller, C., Flury, C., Leeb, T., 2020. The LCORL locus is under selection in large-sized Pakistani Goat breeds. Genes 11, 168.
Saravanan, K.A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., Mishra, B.P., 2020. Selection signatures in livestock genome: a review of concepts, approaches and applications. Livestock Science 241, 104257.
Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., Crepaldi, P., Mantovani, R., Bongioni, G., Pagnacco, G., Portolano, B., Rossoni, A., Pilla, F., Cassandro, M., 2020. The genetic heritage of Alpine local cattle breeds using genomic SNP data. Genetics Selection Evolution 52, 40.
Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T., Rieder, S., 2012. A genome-wide association study reveals loci influencing height and other conformation traits in horses. PloS One 7, e37282.
Sivakumar, A., Kumar, S., Yathish, H.M., Mishra, C., Modi, R.P., Chaudhary, R., Khan, S., Sivamani, B., Ghosh, S.K., Sarkar, M., 2018. Expression profiling and Identification of novel SNPs in CatSper2 Gene and their influence on sperm motility parameters in bovines. Animal Biotechnology 29, 34–40.
Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., Jakobsdottir, M., Steinberg, S., Pálsson, S., Jonasson, F., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K.R., Aben, K.K., Kiemeney, L.A., Olafsson, J.H., Gulcher, J., Kong, A., Thorsteinsdottir, U., Stefansson, K., 2007. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nature Genetics 39, 1443–1452.
Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., Jensen, L.J., von Mering, C., 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research 45, D362–D368.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.
Takasuga, A., 2016. PLAG1 and NCAPG-LCORL in livestock. Animal Science Journal = Nihon Chikusan Gakkaiho 87, 159–167.
Tatarinova, T.V., Chekalin, E., Nikolsky, Y., Bruskin, S., Chebotarov, D., McNally, K.L., Alexandrov, N., 2016. Nucleotide diversity analysis highlights functionally important genomic regions. Scientific Reports 6, 35730.
Thomas, L., Bouhouche, K., Whitfield, M., Thouvenin, G., Coste, A., Louis, B., Szymanski, C., Bequignon, E., Papon, J.-F., Castelli, M., Lemullois, M., Dhalluin, X., Drouin-Garraud, V., Montantin, G., Tissier, S., Duquesnoy, P., Copin, B., Dastot, F., Couvet, S., Barbotin, A.-L., Faucon, C., Honore, I., Maitre, B., Beydon, N., Tamalet, A., Rives, N., Koll, F., Escudier, E., Tassin, A.-M., Touré, A., Mitchell, V., Amselem, S., Legendre, M., 2020. TTC12 loss-of-function mutations cause primary ciliary dyskinesia and unveil distinct dynein assembly mechanisms in motile cilia versus flagella. The American Journal of Human Genetics 106, 153–169.
Tsartsianidou, V., Sánchez-Molano, E., Kapsona, V.V., Basdagianni, Z., Chatziplis, D., Arsenos, G., Triantafyllidis, A., Banos, G., 2021. A comprehensive genome-wide scan detects genomic regions related to local adaptation and climate resilience in Mediterranean domestic sheep. Genetics Selection Evolution 53, 90.
Weikard, R., Altmaier, E., Suhre, K., Weinberger, K.M., Hammon, H.M., Albrecht, E., Setoguchi, K., Takasuga, A., Kühn, C., 2010. Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. Physiological Genomics 42A, 79–88.
Williams, J.L., Hall, S.J.G., Del Corvo, M., Ballingall, K.T., Colli, L., Ajmone Marsan, P., Biscarini, F., 2016. Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. Animal Genetics 47, 19–27.
Xia, J., Fan, H., Chang, T., Xu, L., Zhang, W., Song, Y., Zhu, B., Zhang, L., Gao, X., Chen, Y., Li, J., Gao, H., 2017. Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle. Scientific Reports 7, 42048.
Yan, G., Liu, K., Hao, Z., Shi, Z., Li, H., 2021. The effects of cow-related factors on rectal temperature, respiration rate, and temperature-humidity index thresholds for lactating cows exposed to heat stress. Journal of Thermal Biology 100, 103041.
Zhang, W., Li, J., Guo, Y., Zhang, L., Xu, L., Gao, X., Zhu, B., Gao, H., Ni, H., Chen, Y., 2016. Multi-strategy genome-wide association studies identify the DCAF16-NCAPG region as a susceptibility locus for average daily gain in cattle. Scientific Reports 6, 38073.