ORIGINAL ARTICLE

A genome-wide comparison between selected and unselected Valle del Belìce sheep reveals differences in population structure and footprints of recent selection

Slim Ben Jemaa[1] | Marco Tolone[2] | Maria Teresa Sardina[2] |
Rosalia Di Gerlando[2] | Giorgio Chessari[3] | Andrea Criscione[3] |
Christian Persichilli[4] | Baldassare Portolano[2] | Salvatore Mastrangelo[2]

[1]Laboratoire des Productions Animales et Fourragères, Institut National de la Recherche Agronomique de Tunisie, Université de Carthage, Ariana, Tunisia

[2]Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, Palermo, Italy

[3]Dipartimento Agricoltura, Alimentazione e Ambiente, University of Catania, Catania, Italy

[4]Dipartimento di Agraria, Ambientale e Scienze dell'alimentazione, University of Molise, Campobasso, Italy

Correspondence
Salvatore Mastrangelo, Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, Palermo, Italy.
Email: salvatore.mastrangelo@unipa.it

Funding information
Livestock Environment Opendata, Grant/Award Number: PSRN 2014-2020, PRJ-0185 and J84I18000090007

Abstract

About three decades of breeding and selection in the Valle del Belìce sheep are expected to have left several genomic footprints related to milk production traits. In this study, we have assembled a dataset with 451 individuals of the Valle del Belìce sheep breed: 184 animals that underwent directional selection for milk production and 267 unselected animals, genotyped for 40,660 single-nucleotide polymorphisms (SNPs). Three different statistical approaches, both within (iHS and ROH) and between (Rsb) groups, were used to identify genomic regions potentially under selection. Population structure analyses separated all individuals according to their belonging to the two groups. A total of four genomic regions on two chromosomes were jointly identified by at least two statistical approaches. Several candidate genes for milk production were identified, corroborating the polygenic nature of this trait and which may provide clues to potential new selection targets. We also found candidate genes for growth and reproductive traits. Overall, the identified genes may explain the effect of selection to improve the performances related to milk production traits in the breed. Further studies using high-density array data, would be particularly relevant to refine and validate these results.

KEYWORDS
candidate genes, *Ovis Aries*, population genomics, selection signatures, SNP markers

1 | INTRODUCTION

Valle del Belìce is a sheep breed reared in Sicily for milk production. The breeding history of this breed is relatively recent (Portolano, 1987). In the 1990s, the data collected on production and morphological traits allowed the development of Valle del Belìce breed standard. In 1997, the breed was given official recognition. Since the beginning of the 21st century, the Valle del Belìce breed has been subjected to selection, and great efforts have been made by local farmers to keep its breeding sustainable, mainly by increasing milk production through better management

—WILEY⏐559

and selection. Nowadays, it is one of the most economically important dairy sheep breeds in Southern Italy, with about 200,000 animals. Three decades of breeding and selection in Valle del Belìce sheep are expected to have left several genomic footprints related to milk production traits.

From the end of the 1990s until today, the genetic material of this sheep has been collected and stored by the University of Palermo. This provided the opportunity to map genomic regions affected by selection pressure in Valle del Belìce sheep through the analysis of genomic differences between selected (more recently sampled) and unselected animals.

Advances in genomic methodologies have enabled the development and combination of methods to explore the effects of selection in the genome of livestock species (e.g. Avila et al., 2018; Brito et al., 2017; Fariello et al., 2014). Most of them rely on linkage disequilibrium, site frequency spectrum, reduced local variability and population differentiation (Qanbari & Simianer, 2014). Each of these methods benefits from different sources of information on genome variation and is sensitive to different categories of selection signatures. For instance, the iHS test has advantages in exploring selective sweeps with variants at moderate frequencies (Sabeti et al., 2007), whereas Rsb is better suited for detecting recent selective sweeps that resulted in near or complete fixation of an allele in a population (Tang et al., 2007). The selection also leads to reduced genetic diversity in some regions of the genome, which results in stretches of consecutive homozygous genotypes, known as runs of homozygosity (ROH) islands (Mastrangelo et al., 2017).

In sheep, several studies have been carried out using the above-mentioned methods to identify the genomic regions under selection (e.g. Ahbara et al., 2019; Eydivandi et al., 2021; Kijas et al., 2012), even within genetically close populations (e.g. Cesarani et al., 2019; Mastrangelo et al., 2019). However, changes in the genomic architecture over time due to within-breed selection processes are rarely studied (Hulsegge et al., 2022), especially in sheep.

In this study, we used the Illumina OvineSNP50 BeadChip genotype data and the comparison between two groups of Valle del Belìce ewes formed, respectively, by individuals that underwent directional selection (recent sampling) and unselected animals (old sampling), with the aim to identify genomic regions potentially linked to the selection process.

2 | MATERIALS AND METHODS

Animal Care and Use Committee approval was not needed as data were obtained from pre-existing databases.

2.1 | Animals and genotyping

We have assembled a dataset with 451 individuals of the Valle del Belìce sheep breed: 267 old samples (VDB-old, collected before the 2000s) and 184 new samples (VDB-new, collected between 2018 and 2020). The Department SAAF, University of Palermo, has preserved the genetic material of VDB-old. As regards recent samples, genomic DNA was extracted from blood using the commercial Illustra blood genomic Prep Mini Spin kit (GE Healthcare).

All animals were genotyped for 54,241 single-nucleotide polymorphisms (SNPs) using the Illumina OvineSNP50K array. Chromosomal coordinates for each SNP were obtained from the Oar_v4.0 ovine (*Ovis aries*) genome sequence assembly. Only SNPs located on autosomes were considered. After filtering for minor allele frequency (maf 0.05), missing genotype call rate (geno 0.01) and missingness per individual (mind 0.1), a final dataset consisting of 40,660 SNPs was obtained. All animals had high-quality genotyping and were included in the analysis.

2.2 | Population structure

To investigate the genetic relationships among individuals and between groups, multidimensional scaling (MDS) analysis based on the matrix of identity-by-state pairwise distances was performed using PLINK 1.7 (Purcell et al., 2007). Population structure has been inspected using the maximum likelihood clustering approach implemented in the ADMIXTURE software v1.3.0 (Alexander et al., 2009), with K values ranging from 2 to 4.

2.3 | Identifying selection signatures

In this study, we used two statistical approaches designed to quantify the excess of haplotype homozygosity within (iHS) or among groups (Rsb) and the analysis of consecutive regions of homozygosity (ROH islands).

2.3.1 | iHS and Rsb approaches

The two EHH-based metrics were computed within 1.5-Mb sliding windows with 50-kb overlapping step using the *rehh* package v. 3.2.2 in R (Gautier & Vitalis, 2012). Haplotypes were reconstructed from the SNP data using fastPHASE v. 1.4 (Scheet & Stephens, 2006).

For each SNP, iHS and Rsb tests are constructed to have an approximately standard Gaussian distribution (Tang et al., 2007; Voight et al., 2006). The scores, calculated for each SNP, were transformed into *p*-values using the *rehh*

package (Gautier & Vitalis, 2012). Candidate regions were then defined by the windows containing at least three neighbouring SNPs with a−log10(p-value)>=2.5. We have chosen to focus on clusters of neighbouring SNPs because it has been demonstrated that it is more powerful to look for windows of consecutive SNPs that contain numerous extreme scores rather than treating each SNP separately (Voight et al., 2006).

2.3.2 | Runs of homozygosity islands

Runs of homozygosity were investigated using PLINK 1.7 (Purcell et al., 2007). The minimum length that constituted the ROH was set to 1 Mb. The following criteria were also used: (i) one missing and one possible heterozygous genotype were allowed in the ROH; (ii) the minimum number of SNPs that constituted the ROH was set to 30; (iii) the minimum SNP density per ROH was set to one SNP every 100 kb; (iv) the maximum gap allowed between consecutive homozygous SNPs was 1000 kb. The percentage of SNP within an ROH was calculated by counting the number of times each SNP was in an ROH and by dividing that number by the number of animals in each group, thus obtaining homozygosity estimates per locus. To identify ROH islands, the top 0.2% SNPs were selected. Finally, the inbreeding coefficient (F) based on ROH (F_{ROH}) for each animal was also estimated as follows:

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

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

2.4 | Gene annotation

Annotated genes within the identified genomic regions were obtained from Genome Data Viewer provided by the National Center for Biotechnology Information (NCBI) using the ovine assembly OAR_v4.0. Gene Ontology (GO) and the enrichment analysis of annotated genes identified by at least two approaches (Table 1) were conducted using the open-source Database for Annotation, Visualization, and Integrated Discovery ver. 2021 package (https://david-d.ncifcrf.gov), imposing a p-value <0.05. The associated QTLs were investigated using the animal QTL database (QTLdb; https://www.animalgenome.org/cgi-bin/QTLdb/index). Finally, we conducted a comprehensive literature search to investigate the biological function and the phenotypes that are known to be affected by each annotated gene.

3 | RESULTS

3.1 | Population structure and inbreeding

Multidimensional scaling analysis revealed a separation between the two investigated groups along the first dimension (PC1 with 12.34% of the variance; Figure 1). In addition, the VDB-new group formed two clusters with unequal sizes. Conversely, VDB-old individuals did not form a cluster but were scattered. This distribution pattern was confirmed by the results of the Admixture analysis observed at $K=2$ in which VDB-old individuals are already distinguishable from their selected counterparts The increase of K at 3 and 4 highlighted the two subgroups in VDB-new sampling (Figure S1), also revealed by MDS. The ROH-based inbreeding coefficient was lower in the

TABLE 1 Genomic regions identified by two statistical approaches in the Valle del Belice sheep.

Test	OAR	Star bp	End bp	Genes
iHS, ROH	4	99,200,128	103,490,698	*NUP205, SLC13A4, FAM180A, MTPN, LUZP6, CHRM2, PTN, DGKI, CREB3L2, AKR1D1, TRIM24, SVOPL, ATP6V0A4, TMEM213, KIIA1549, ZC3HAV1L, ZC3HAV1, TTC26, UBN2, KLRG2, CLEC2L, HIPK2, TBXAS1, PARP12, KDM7A, SLC37A3*
iHS, ROH	4	104,579,997	104,883,987	*AGK, KIAA1147, WEE2, SSBP1, PRSS37*
Rsb, iHS	4	104,950,000	106,450,000	*MGAM, MGAM2, EPHB6, TRPV6, C4H7orf34, KEL, GSTK1, CASP2, CLCN1, FAM131B, ZYX, EPHA1*
iHS, ROH	11	20,360,396	21,976,316	*TAOK1, ABHD15, TP53I13, GIT1, ANKRD13B, CORO6, SSH2, EFCAB5, NSRP1, SLC6A4, BLMH, TMIGD1, CPD, GOSR1, TUSC5, BHLHA9, ABR, TIMM22, NXN, RNMTL1, GLOD4, GEMIN4, FAM57A, VPS53, FAM101B, RPH3AL*

Abbreviation: OAR, *Ovis aries* chromosome number.

—WILEY—

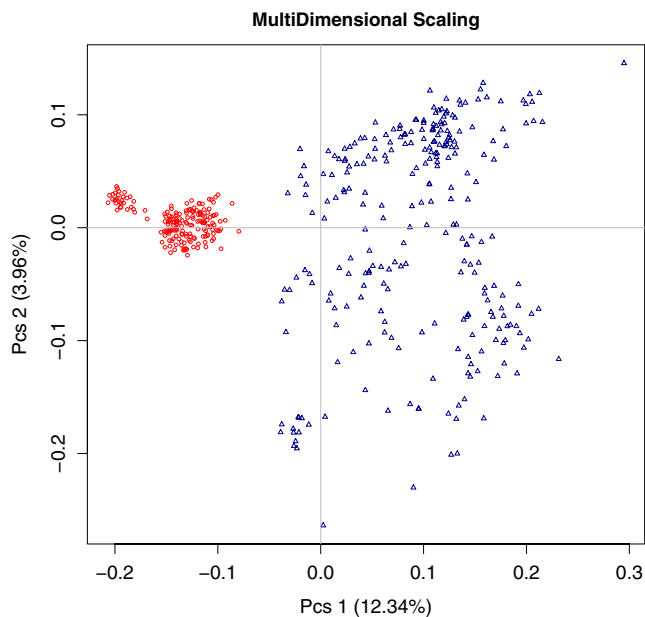**MultiDimensional Scaling**

FIGURE 1 Genetic relationship defined with multidimensional scaling analysis between unselected (blue) and selected (red) group in Valle del Belice sheep breed. [Colour figure can be viewed at wileyonlinelibrary.com]

VDB-new group ($F_{ROH} = 0.062 \pm 0.055$) than in the VDB-old group ($F_{ROH} = 0.097 \pm 0.064$).

3.2 | Identifying selection signatures between groups using Rsb

We applied the Rsb test to detect potential selective sweeps fixed (or nearly fixed) in the VDB-new but which remain segregating in the VDB-old (or vice versa; Figure 2). In total, 115 outlier SNPs were revealed (Table S1). The approach defined a total of three candidate regions, spanning 45 known protein coding genes.

3.3 | Identifying selection signatures within the VDB-new group

3.3.1 | iHS approach

The VDB-new group showed a total of 1577 outlier SNPs, represented by the Manhattan plot in Figure 3. These outlier markers defined a total of 23 candidate genomic regions encompassing 511 known genes (Table S2). The highest number of regions identified by iHs were located on the first four chromosomes. The results also showed five strong iHS signals, defined as regions with a high number of consecutive outliers (>=10): one on OAR03 (174,300,000–181,300,000 bp), three on OAR04

(84,550,000–88,550,000 bp; 90,650,000–96,100,000 bp and 99,200,000–110,650,000 bp) and one on OAR11 (19,050, 000–23,750,000 bp; Table S2).

3.3.2 | Runs of homozygosity approach

A total of four genomic regions that frequently appeared within ROH were identified. Table S3 provides the chromosome, the start and end positions of ROH islands found in the VDB-new group. These regions ranged from 4.30 Mb (on OAR04) to 0.03 Mb in length (on OAR04). We identified a total of 57 known genes within the ROH islands.

3.4 | Congruence among selection signals

A total of four genomic regions were jointly identified by at least two statistical approaches (Table 1). Two adjacent regions on OAR04 were jointly identified by two within population approaches (iHS and ROH). The first region containing 26 genes, was ~4.3 Mb in size (at position: 99,200,128 bp–103,490,690 bp) while the second was a ~300 kb region (at position: 104,579,997–104,883,987 bp) harbouring only five genes. Another region on OAR04 (104,950,000–106,450,000 bp) containing 12 genes was identified by Rsb and iHS. It is worthy of note that the regions on OAR04 also showed the strongest signals in iHS analysis with a total of 44 SNPs above the significance threshold. Finally, iHS and ROH registered another common selection signal on OAR11 (20,360,396–21,297,316 bp) containing 26 genes.

3.5 | Gene identification and functional enrichment analysis

To better understand the biological implications of the selection signals, enrichment investigations were performed for annotated genes that were identified by at least two approaches. No significantly enriched annotation functions (Benjamini-corrected p-value <0.05) were observed. The search on the sheep QTLdb revealed 22 QTLs belonging to different traits (Table S4). In particular, we identified QTL associated with milk production and growth traits.

4 | DISCUSSION

Valle del Belìce is the most reared dairy sheep in Sicily. This breed has undergone selection during the last 30 years. Such selective pressure is expected to cause
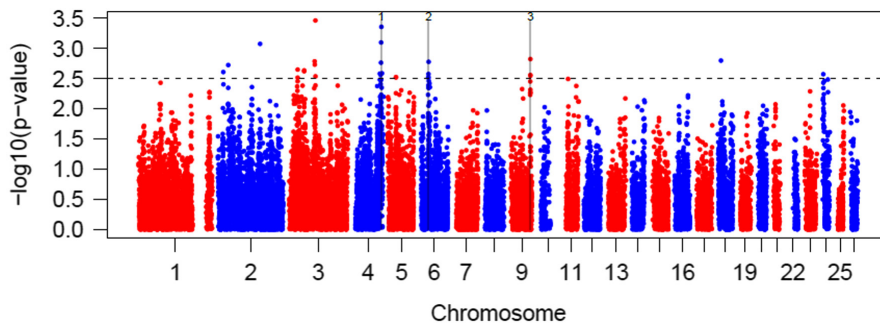
FIGURE 2 Manhattan plots of *Rsb* test in the comparison between unselected and selected groups in Valle del Belice sheep breed. [Colour figure can be viewed at wileyonlinelibrary.com]
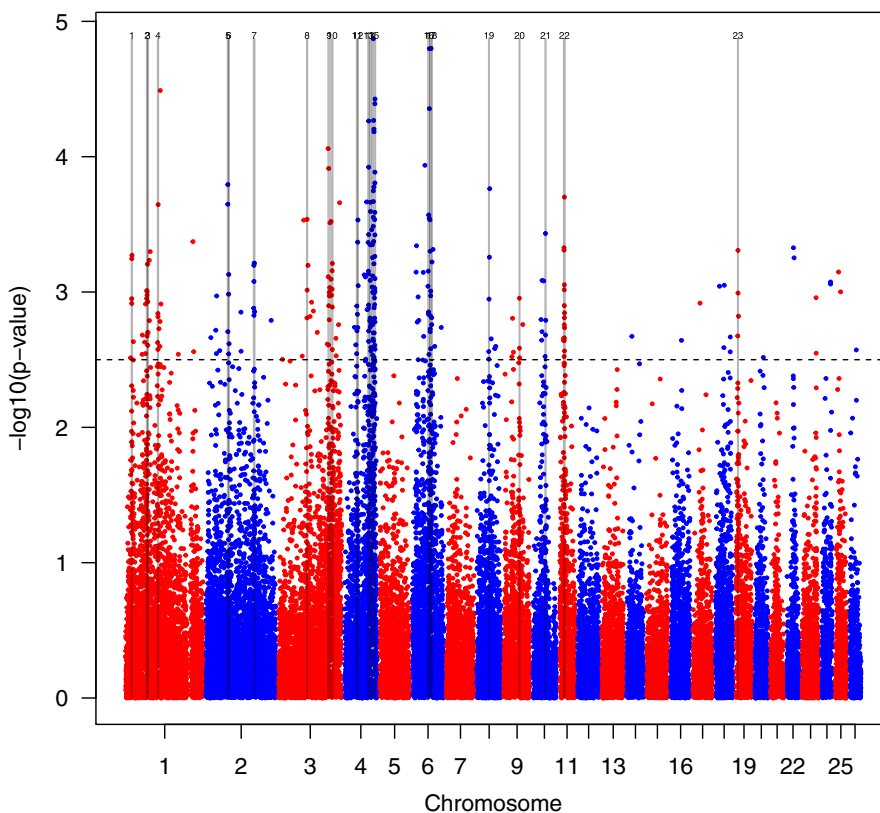

FIGURE 3 Manhattan plot of the genome-wide *iHS* analysis for selected group. [Colour figure can be viewed at wileyonlinelibrary.com]

changes in the genetic structure and diversity of this population. Accordingly, both aspects were recognizable in our population structure results, with VDB-new occupying a distinct space from VDB-old in MDS, having different genomic pattern in ADMIXTURE, and a different ROH-based inbreeding coefficient. Regarding this latter aspect, we found a lower F_{ROH} value for VDB-new compared to the unselected population, which is an interesting result considering that selection often results in an inevitable increase in inbreeding. This can be explained by the widespread mating between close relatives, a typical feature of the old Sicilian farming system (Mastrangelo et al., 2014). Therefore, the lowest F_{ROH} value in VDB-new indicates that the breed has benefited from proper mating management. Likewise, the small group of VDB-new individuals revealed by MDS and ADMIXTURE is likely the result of higher levels of relatedness in this group compared to the

whole VDB-new sample. The genetic divergences among groups of the same population but undergoing different selection pressures have been already reported in cattle (Cardoso et al., 2018; Hulsegge et al., 2022). Presumably, differences in the genetic structure between groups could be driven by the combined action of selection and genetic drift (Bertolini et al., 2020).

The main aim of the present study was to investigate selection signatures linked to the production traits in Valle del Belìce sheep. This was performed by comparing the genomes of a group of animals subjected to a directional selection process with an unselected population, using three statistical approaches that have been successfully applied in other studies on various livestock species (e.g. Brito et al., 2017; Mastrangelo et al., 2023; Moscarelli et al., 2021; Nolte et al., 2019). Genome scans comparing breeds under artificial selection with their unselected

—WILEY—

founder population have been previously applied in cattle (e.g. Cardoso et al., 2018; Hulsegge et al., 2022; Seo et al., 2022) and could provide valuable insight into the genes associated with the traits that have been under artificial selection. The main drawback of such strategy is the likelihood that genetic drift can lead to divergence between selected and control populations for reasons related to randomness rather than specific genes playing a role in the variability of selected traits. We believe that the population size in the present study was sufficiently large to prevent genetic drift from overriding the effect of selection for the identified loci in the VDB-new population. However, since Valle del Belìce sheep can be considered a recent population resulting from crossbreeding practices, some assumptions might be violated due to population expansion, subdivision and bottleneck; these factors, although the candidate genes were identified using different statistical methods, could lead to the identification of false positive signals.

In the present study, we found a few number of selective sweeps with the Rsb test compared to iHS (3 vs. 23, respectively) which is in line with selective pressure on beneficial alleles that is still ongoing in the VDB-new in which most of the variants under selection are still present with intermediate frequency. It is worth noting that the iHS statistic achieves its maximal power when selected alleles segregate at intermediate frequencies in the population (Zhao et al., 2015), while Rsb can detect selected alleles that have risen to near-fixation or a point of fixation in a specific population but remain polymorphic in the other population (Tang et al., 2007).

No genomic regions were jointly identified by the three statistic methods, since each approach tends to detect different signals (Liu et al., 2013), and most of the shared regions have been identified between complementary approaches (Table 1). In fact, three genomic regions detected by iHS were also located in ROH, lending further support that these regions are likely true candidates. Many studies on livestock species have shown a positive correlation between the frequency of ROH and iHS (Gautason et al., 2021; Liu et al., 2021; Mastrangelo et al., 2020; Nolte et al., 2019).

Due to the large number of genes included in the outlier windows, we chose to focus on two types of candidate regions to discuss the biological implication of our findings with regard to genes whose functions have been characterized: (i) the five candidate regions with more than 10 SNPs exceeding the significance threshold identified by iHS. This is because selective sweeps detected by iHS tend to produce clusters of extreme scores across the sweep region, while under a neutral model, extreme scores are scattered more uniformly (Voight et al., 2006) and (ii) the candidate regions jointly detected by the two methods

(Table 1). The rationale is that selection signals supported across different methodologies can be interpreted as reliable signals (Saravanan et al., 2020).

We reviewed the literature to obtain a list of genes known to be associated with milk production traits included in the aforementioned candidate regions.

Three out of the five relevant candidate regions detected by iHS are located on OAR4 (Table S2). The most relevant region on this chromosome is located at position 99,200,000–110,650,000 bp with 44 SNPs exceeding the significance threshold (22% of the total number of SNPs in this region). This large region included the three signatures of selection identified on OAR04 combining the results between approaches (Table 1). The SNP with the highest p-value in this region (rs414908437, $-\log10(p\text{-value}) = 4.87$, position: 103,490,698 bp) is located within *SLC37A3* gene. Importantly, this SNP is located at the end of the first of the two ROHs detected on OAR4 (Table S3). *SLC37A3* belongs to the *SLC37* gene family, an almost unexplored group of transmembrane sugar transporters (Bartoloni & Antonarakis, 2004). The gene *SLC37A1* is highly expressed in the mammary gland (Kemper et al., 2015; Raven et al., 2016). Previous studies found the *SLC37A1* gene associated with several milk production traits in cattle including milk yield, fat and protein percentages (Kemper et al., 2015; Pausch et al., 2017; Raven et al., 2016). Furthermore, Sanchez et al. (2019) reported *SLC37A1* associated with milk's cheese-making properties and milk mineral content in Montbéliarde cows. Another gene of interest is *KDM7A*, which is only 152 kb downstream of the rs414908437 SNP and is a key regulator of *DGAT2* gene (Kim et al., 2021). The latter is a member of the diacylglycerol acyltransferases (DGAT), a vital group of enzymes in catalysing triacylglycerol biosynthesis. DGAT genes have been identified as functional candidate genes affecting milk production traits, especially for fat content in milk (Liu et al., 2020). The outlier window on OAR11 spans a 4.7-Mb region (from 19,050,000 to 23,750,000 bp) with more than 22% of the SNPs having p-values exceeding the significance threshold. The highest signal in this region spans from 20,360,396 to 21,680,960 bp. Within this window, there is a shared region with a ROH encompassing *TAOK1*, a gene modulating IgG and IgM concentration in colostrum and serum (Lin et al., 2022). Furthermore, among the cluster of outliers in this region (seven SNPs), two highly significant SNPs (rs417806498 and rs428702588) are located in *SSH2*, a gene that encodes a protein tyrosine phosphatase which plays a key role in the regulation of actin filaments, and that is linked to intra-muscular fat in sheep (Daetwyler et al., 2012) and height at withers in cattle (Vanvanhossou et al., 2020). This is in line with the fact that Valle del Belìce sheep

was secondarily selected for growth traits which led to the identification of candidate genes that are rather associated with such trait. The peak area on chromosome 11, at position 21,680,960 bp with a −log10 (p-value = 3.7), is located only ~8 kb downstream of *FAM57A* gene. There is no clear trade-off between this gene and milk production traits although a previous study reported that *FAM57A* is among the genes that are related to 305 milk yield in Guzerá cattle (Paiva et al., 2020). *FAM57A* is rather thought to play a crucial role in spermatogenesis (Wang et al., 2016) and is among the differentially expressed genes when comparing high- and low-quality ejaculates in Australian sheep (Hodge et al., 2021). Hence, it seems likely that the selection process in the VDB breed has incidentally impacted some male reproductive traits. In line with this hypothesis, on the another candidate region on OAR4, the area bounded by the two most significant SNPs (rs401685703, −log10(p-value) = 4.26, position: 87,097,602 bp and rs398805194, −log10(p-value) = 3.92, position: 87,153,327 bp) encompasses two genes coding for ring finger proteins, *RNF133* and *RNF148*, two evolutionarily conserved testis-specific E3 ubiquitin-protein ligases. At least one of these genes (*RNF133*) plays a critical role in sperm function during spermatogenesis in mice (Nozawa et al., 2022).

Although we have mainly focused attention on the common candidate regions identified by at least two approaches, it is important to reiterate that each method has specific features implying that non-overlapping regions of selection signatures across different methods should be treated as complementary information to better understand the different patterns of variation in a population (Santos et al., 2021). In this regard, the Rsb approach identified a region on OAR06 with well-known candidate genes. This test is suitable for detecting signatures of recent selection (Bahbahani et al., 2015; Epstein et al., 2016; Oleksyk et al., 2010), such as those present in the Valle del Belìce breed. Within this region, we mapped important genes (*ABCG2*, *PKD2*, *SPP1*, *MEPE*, *IBSP*, *LAP3*, *MED28*, *FAM184B*, *NCAPG*, *LCORL*) previously reported within a QTL related to milk production traits in cattle (e.g. Olsen et al., 2008; Zhang et al., 1998) and that overlapped with a candidate genomic window for milk traits identified in a previous study in Valle del Belìce sheep (Mohammadi et al., 2022). However, limited overlap was observed in the candidate genomic regions identified in the present study. This could be explained by at least two potential factors. Firstly, potential artefacts due to population stratification or cryptic relatedness could lead to spurious associations in trait-associated genetics, especially in the field of polygenic analysis (Novembre & Barton, 2018). For instance, the polygenic selection signals observed for the height GWAS meta-analysis in humans are now considered to be partly driven by residual population structure (Berg et al., 2019; Sohail et al., 2019). Second, the selection signals detected in the present study are likely to be associated, in addition to milk production, with other traits. This is exemplified by the selection signatures on chromosomes 4 and 11 where the selection signal peaks fall within (or are close to) genes associated with growth and reproduction.

5 | CONCLUSIONS

About three decades of artificial selection have led to a significant change in the population structure of Valle del Belìce sheep. Our findings demonstrate that changes in genetic structure and diversity have occurred in the group of animals subjected to directional selection, compared with the group of old unselected animals. There were no overlapped regions jointly identified by the three statistic methods, since each approach tends to detect different signals. We found a congruence among the EHH-based methods (iHS and Rsb), but also betweeen iHS and ROH approaches. A total of four genomic regions were jointly identified by at least two statistical approaches, most of which overlapped known candidate genes. We highlighted several interesting sweeps underlying candidate genes associated with milk production, which may provide clues to potential new selection targets. Our results also show that selection in the Valle del Belice breed led to the targeting of genes associated with growth and reproductive traits. Further studies using higher-density array data, would be particularly relevant to refine and validate these candidate regions.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

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

ORCID
Slim Ben Jemaa https://orcid.org/0000-0002-7103-3315
Marco Tolone https://orcid.org/0000-0002-6043-9782
Maria Teresa Sardina https://orcid.org/0000-0001-9527-6197
Rosalia Di Gerlando https://orcid.org/0000-0001-5135-6531

Giorgio Chessari https://orcid.
org/0000-0002-9841-4820
Andrea Criscione https://orcid.
org/0000-0002-6701-1906
Christian Persichilli https://orcid.
org/0000-0003-2664-136X
Baldassare Portolano https://orcid.
org/0000-0003-0792-9405
Salvatore Mastrangelo https://orcid.
org/0000-0001-6511-1981

REFERENCES

Ahbara, A., Bahbahani, H., Almathen, F., Al Abri, M., Agoub, M. O., Abeba, A., Kebede, A., Musa, H. H., Mastrangelo, S., Pilla, F., Ciani, E., Hanotte, O., & Mwacharo, J. M. (2019). Genome-wide variation, candidate regions and genes associated with fat deposition and tail morphology in ethiopian indigenous sheep. *Frontiers in genetics*, 9, 699. https://doi.org/10.3389/fgene.2018.00699

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

Avila, F., Mickelson, J. R., Schaefer, R. J., & McCue, M. E. (2018). Genome-wide signatures of selection reveal genes associated with performance in American Quarter Horse subpopulations. *Frontiers in Genetics*, 9, 249. https://doi.org/10.3389/fgene.2018.00249

Bahbahani, H., Clifford, H., Wragg, D., Mbole-Kariuki, M. N., Van Tassell, C., Sonstegard, T., Woolhouse, M., & Hanotte, O. (2015). Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. *Scientific Reports*, 5(1), 11729. https://doi.org/10.1038/srep11729

Bartoloni, L., & Antonarakis, S. E. (2004). The human sugar-phosphate/phosphate exchanger family SLC37. *Pflügers Archiv*, 447(5), 780–783. https://doi.org/10.1007/s00424-003-1105-0

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., Boyle, E. A., Zhang, X., Racimo, F., Pritchard, J. K., & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK biobank. *eLife*, 8, e39725. https://doi.org/10.7554/eLife.39725

Bertolini, F., Schiavo, G., Bovo, S., Sardina, M. T., Mastrangelo, S., Dall'Olio, S., Portolano, B., & Fontanesi, L. (2020). Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano-Reggiano cheese production system. *Animal*, 14, 921–932. https://doi.org/10.1017/S1751731119003318

Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L. R., Cánovas, A., Feng, Z., Jafarikia, M., & Schenkel, F. S. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics*, 18(1), 229. https://doi.org/10.1186/s12864-017-3610-0

Cardoso, D. F., de Albuquerque, L. G., Reimer, C., Qanbari, S., Erbe, M., do Nascimento, A. V., Venturini, G. C., Becker Scalez, D. C., Baldi, F., Ferreira de Camargo, G. M., Zerlotti Mercadante, M.

E., do Santos Gonçalves Cyrillo, J. N., Simianer, H., & Tonhati, H. (2018). Genome-wide scan reveals population stratification and footprints of recent selection in Nelore cattle. *Genetics Selection Evolution*, 50(1), 1–12. https://doi.org/10.1186/s12711-018-0381-2

Cesarani, A., Sechi, T., Gaspa, G., Usai, M. G., Sorbolini, S., Macciotta, N. P. P., & 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 Research*, 177, 68–75. https://doi.org/10.1016/j.smallrumres.2019.06.014

Daetwyler, H. D., Swan, A. A., van der Werf, J. H., & Hayes, B. J. (2012). Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genetics Selection Evolution*, 44(1), 1–11. https://doi.org/10.1186/1297-9686-44-33

Epstein, B., Jones, M., Hamede, R., Hendricks, S., McCallum, H., Murchison, E. P., Schönfeld, B., Wiench, C., Hohenlohe, P., & Storfer, A. (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nature Communications*, 7(1), 1–7. https://doi.org/10.1038/ncomms12684

Eydivandi, S., Roudbar, M. A., Ardestani, S. S., Momen, M., & Sahana, G. (2021). A selection signatures study among middle eastern and European sheep breeds. *Journal of Animal Breeding and Genetics*, 138(5), 574–588. https://doi.org/10.1111/jbg.12536

Fariello, M. I., Servin, B., Tosser-Klopp, G., Rupp, R., Moreno, C., International Sheep Genomics Consortium, Cristobal, M. S., & Boitard, S. (2014). Selection signatures in worldwide sheep populations. *PLoS One*, 9(8), e103813. https://doi.org/10.1371/journal.pone.0103813

Gautason, E., Schönherz, A. A., Sahana, G., & Guldbrandtsen, B. (2021). Genomic inbreeding and selection signatures in the local dairy breed Icelandic cattle. *Animal Genetics*, 52(3), 251–262. https://doi.org/10.1111/age.13058

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

Hodge, M. J., de Las Heras-Saldana, S., Rindfleish, S. J., Stephen, C. P., & Pant, S. D. (2021). Characterization of breed specific differences in spermatozoal transcriptomes of sheep in Australia. *Genes*, 12(2), 203. https://doi.org/10.3390/genes12020203

Hulsegge, I., Oldenbroek, K., Bouwman, A., Veerkamp, R., & Windig, J. (2022). Selection and drift: A comparison between historic and recent Dutch Friesian cattle and recent Holstein Friesian using WGS data. *Animals*, 12(3), 329. https://doi.org/10.3390/ani12030329

Kemper, K. E., Reich, C. M., Bowman, P. J., Vander Jagt, C. J., Chamberlain, A. J., Mason, B. A., Hayes, B. J., & Goddard, M. E. (2015). Improved precision of QTL mapping using a non-linear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genetics Selection Evolution*, 47(1), 1–17. https://doi.org/10.1186/s12711-014-0074-4

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raasma, H., McEwan, J., Dalrymple, B., & International Sheep Genomics Consortium.

(2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, *10*(2), e1001258. https://doi.org/10.1371/journal.pbio.1001258

Kim, J. H., Nagappan, A., Jung, D. Y., Suh, N., & Jung, M. H. (2021). Histone demethylase KDM7A contributes to the development of hepatic steatosis by targeting diacylglycerol acyltransferase 2. *International Journal of Molecular Sciences*, *22*(20), 11085. https://doi.org/10.3390/ijms222011085

Lin, S., Ke, C., Liu, L., Gao, Y., Xu, L., Han, B., Zhao, Y., Zhang, S., & Sun, D. (2022). Genome-wide association studies for immunoglobulin concentrations in colostrum and serum in Chinese Holstein. *BMC Genomics*, *23*(1), 1–13. https://doi.org/10.1186/s12864-021-08250-5

Liu, D., Chen, Z., Zhao, W., Guo, L., Sun, H., Zhu, K., Liu, G., Shen, X., Zhao, X., Wang, Q., Ma, P., & Pan, Y. (2021). Genome-wide selection signatures detection in Shanghai Holstein cattle population identified genes related to adaption, health and reproduction traits. *BMC Genomics*, *22*(1), 1–19. https://doi.org/10.1186/s12864-021-08042-x

Liu, J., Wang, Z., Li, J., Li, H., & Yang, L. (2020). Genome-wide identification of diacylglycerol acyltransferases (DGAT) family genes influencing milk production in Buffalo. *BMC Genetics*, *21*(1), 1–14. https://doi.org/10.1186/s12863-020-0832-y

Liu, X., Ong, R. T. H., Pillai, E. N., Elzein, A. M., Small, K. S., Clark, T. G., Kwiatkowski, D. P., & Teo, Y. Y. (2013). Detecting and characterizing genomic signatures of positive selection in global populations. *The American Journal of Human Genetics*, *92*(6), 866–881. https://doi.org/10.1016/j.ajhg.2013.04.021

Mastrangelo, S., Ben Jemaa, S., Ciani, E., Sottile, G., Moscarelli, A., Boussaha, M., Montedoro, M., Pilla, F., & Cassandro, M. (2020). Genome-wide detection of signatures of selection in three Valdostana cattle populations. *Journal of Animal Breeding and Genetics*, *137*(6), 609–621. https://doi.org/10.1111/jbg.12476

Mastrangelo, S., Ben Jemaa, S., Sottile, G., Casu, S., Portolano, B., Ciani, E., & Pilla, F. (2019). Combined approaches to identify genomic regions involved in phenotypic differentiation between low divergent breeds: Application in Sardinian sheep populations. *Journal of Animal Breeding and Genetics*, *136*(6), 526–534. https://doi.org/10.1111/jbg.12422

Mastrangelo, S., Ben-Jemaa, S., Perini, F., Cendron, F., Biscarini, F., Lasagna, E., Penasa, M., & Cassandro, M. (2023). Genome-wide mapping of signatures of selection using a high-density array identified candidate genes for growth traits and local adaptation in chickens. *Genetics Selection Evolution*, *55*(1), 1–16. https://doi.org/10.1186/s12711-023-00790-6

Mastrangelo, S., Di Gerlando, R., Tolone, M., Tortorici, L., Sardina, M. T., & Portolano, B. (2014). Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. *BMC Genetics*, *15*(1), 1–10. https://doi.org/10.1186/s12863-014-0108-5

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

Mohammadi, H., Farahani, A. H. K., Moradi, M. H., Mastrangelo, S., Di Gerlando, R., Sardina, M. T., Scatassa, M. L., Portolano, B., & Tolone, M. (2022). Weighted single-step genome-wide association study uncovers known and novel candidate genomic regions for Milk production traits and somatic cell score in Valle del Belice dairy sheep. *Animals*, *12*(9), 1155. https://doi.org/10.3390/ani12091155

Moscarelli, A., Sardina, M. T., Cassandro, M., Ciani, E., Pilla, F., Senczuk, G., Portolano, B., & Mastrangelo, S. (2021). Genome-wide assessment of diversity and differentiation between original and modern Brown cattle populations. *Animal Genetics*, *52*(1), 21–31. https://doi.org/10.1111/age.13019

Nolte, W., Thaller, G., & Kuehn, C. (2019). Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. *PLoS One*, *14*(4), e0215913. https://doi.org/10.1371/journal.pone.0215913

Novembre, J., & Barton, N. H. (2018). Tread lightly interpreting polygenic tests of selection. *Genetics*, *208*(4), 1351–1355. https://doi.org/10.1534/genetics.118.300786

Nozawa, K., Fujihara, Y., Devlin, D. J., Deras, R. E., Kent, K., Larina, I. V., Umezu, K., Yu, Z., Sutton, C. M., Ye, Q., Dean, L. K., Emori, C., Ikawa, M., Garcia, T. X., & Matzuk, M. M. (2022). The testis-specific E3 ubiquitin ligase RNF133 is required for fecundity in mice. *BMC Biology*, *20*(1), 1–19. doi:10.1186/s12915-022-01368-2

Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 185–205. 10.1098%2Frstb.2009.0219

Olsen, H. G., Meuwissen, T. H. E., Nilsen, H., Svendsen, M., & Lien, S. (2008). Fine mapping of quantitative trait loci on bovine chromosome 6 affecting calving difficulty. *Journal of Dairy Science*, *91*, 4312–4322. https://doi.org/10.3168/jds.2008-1000

Paiva, J. T., Peixoto, M. G. C. D., Bruneli, F. A. T., Alvarenga, A. B., Oliveira, H. R., Silva, A. A., Silva, D. A., Veroneze, R., Silva, F. F., & Lopes, P. S. (2020). Genetic parameters, genome-wide association and gene networks for milk and reproductive traits in Guzerá cattle. *Livestock Science*, *242*, 104273. https://doi.org/10.1016/j.livsci.2020.104273

Pausch, H., Emmerling, R., Gredler-Grandl, B., Fries, R., Daetwyler, H. D., & Goddard, M. E. (2017). Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. *BMC Genomics*, *18*(1), 1–11. https://doi.org/10.1186/s12864-017-4263-8

Portolano, N. (1987). La pecora della Valle del Belice. In Edagricole (Ed.), *Pecore e capre italiane* (pp. 117–124). Edagricole.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (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

Qanbari, S., & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock Science*, *166*, 133–143. https://doi.org/10.1016/j.livsci.2014.05.003

Raven, L. A., Cocks, B. G., Kemper, K. E., Chamberlain, A. J., Vander Jagt, C. J., Goddard, M. E., & Hayes, B. J. (2016). Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle. *Mammalian Genome*, *27*(1), 81–97. https://doi.org/10.1007/s00335-015-9613-8

Sabeti, P. C., Vasrilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., & The International HapMap Consortium. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*, 913–918. https://doi.org/10.1038/nature06250

Sanchez, M. P., Ramayo-Caldas, Y., Wolf, V., Laithier, C., El Jabri, M., Michenet, A., Boussaha, M., Taussat, S., Fritz, S., Celacroix-Buchet, A., Brochard, M., & Boichard, D. (2019). Sequence-based GWAS, network and pathway analyses reveal genes co-associated with milk cheese-making properties and milk composition in Montbéliarde cows. *Genetics Selection Evolution*, *51*(1), 1–19. https://doi.org/10.1186/s12711-019-0473-7

Santos, W. B., Schettini, G. P., Maiorano, A. M., Bussiman, F. O., Balieiro, J. C., Ferraz, G. C., Pereira, G. L., Baldassini, W. A., Neto, O. R. M., Oliveira, H. N., & Curi, R. A. (2021). Genome-wide scans for signatures of selection in Mangalarga Marchador horses using high-throughput SNP genotyping. *BMC Genomics*, *22*(1), 1–17. https://doi.org/10.1186/s12864-021-08053-8

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

Scheet, P., & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics*, *78*(4), 629–644. https://doi.org/10.1086/502802

Seo, D., Lee, D. H., Jin, S., Won, J. I., Lim, D., Park, M., Kim, T. H., Lee, H. K., Kim, S., Choi, I., Lee, J. H., Gondro, C., & Lee, S. H. (2022). Long-term artificial selection of Hanwoo (Korean) cattle left genetic signatures for the breeding traits and has altered the genomic structure. *Scientific Reports*, *12*(1), 1–15. https://doi.org/10.1038/s41598-022-09425-0

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., Chiang, C. W. K., Hirschorn, J., Daly, M. J., Patterson, N., Neale, B., Mathieson, I., Reich, D., & Sunyaev, S. R. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, *8*, e39702. https://doi.org/10.7554/eLife.3970

Tang, K., Thornton, K. R., & Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology.*, *5*, e171. https://doi.org/10.1371/journal.pbio.0050171

Vanvanhossou, S. F. U., Scheper, C., Dossa, L. H., Yin, T., Brügemann, K., & König, S. (2020). A multi-breed GWAS for morphometric traits in four Beninese indigenous cattle breeds reveals loci associated with conformation, carcass and adaptive traits. *BMC Genomics*, *21*(1), 1–16. https://doi.org/10.1186/s12864-020-07170-0

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*, e72. https://doi.org/10.1371/journal.pbio.0040154

Wang, S., Wang, X., Ma, L., Lin, X., Zhang, D., Li, Z., Wu, Y., Zheng, C., Feng, X., Liao, S., Feng, Y., Chen, J., Hu, X., Wang, M., & Han, C. (2016). Retinoic acid is sufficient for the in vitro induction of mouse spermatocytes. *Stem Cell Reports*, *7*(1), 80–94. https://doi.org/10.1016/j.stemcr.2016.05.013

Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister-Genskow, M., Witte, L. A., Grignola, F. E., & Uimari, P. (1998). Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics*, *149*, 1959–1973. https://doi.org/10.1093/genetics/149.4.1959

Zhao, F., McParland, S., Kearney, F., Du, L., & Berry, D. P. (2015). Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics Selection Evolution*, *47*(1), 1–12. https://doi.org/10.1186/s12711-015-0127-3

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.