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

Christian Persichilli,[1] **Gabriele Senczuk,**[1]* **Salvatore Mastrangelo,**[2] **Maurizio Marusi,**[3]
Jan-Thijs van Kaam,[3] **Raffaella Finocchiaro,**[3] **Marika Di Civita,**[1] **Martino Cassandro,**[3,4] **and Fabio Pilla**[1]

[1]Department of Agriculture Environment and Food Science, University of Molise, Via A. De sanctis, 86100 Campobasso (CB), Italy
[2]Department of Agricultural, Food and Forest Sciences, University of Palermo, Viale delle Scienze, 90128 Palermo (PA), Italy
[3]National Association of Italian Holstein, Brown and Jersey Breeders, Via Bergamo, 292, 26100 Cremona (CR), Italy
[4]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy

ABSTRACT

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

Key words: cattle, Holstein, genetic structure, selection signature, candidate genes

INTRODUCTION

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

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

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

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

MATERIALS AND METHODS

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

Source of the Genotypes and Quality Control

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

To harmonize the data set, as they come from different SNPs panels, we performed an imputation approach using the software Pedimpute (Nicolazzi et al., 2013). We obtained 86,841 imputed SNPs that were then mapped against the ARS-UCD1.2 reference genome (Rosen et al., 2020). The software PLINK (Purcell et al., 2007; Chang et al., 2015) was used to filter for minor allele frequency (maf 0.05) and missing genotype call rate (geno 0.01) obtaining a final data set of 79,464 SNPs.

Analysis of Genetic Diversity and Population Structure

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

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

Selection Signature Through Genome-Wide Allele Frequency Approaches

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

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

For the single locus pairwise F_{ST} outlier analysis we retained SNPs based on the top 0.05% F_{ST} value

(99.95th percentile). We then considered a region of $\pm$ 60 kb around the significant SNPs to retain genes and explore their function.

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

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

Selection Signature Through Extended Haplotype Homozygosity

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

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

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

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

Quantitative Trait Loci and Functional Enrichment Analysis

For both F_{ST} and EHH analyses, we used the R package *Genomic Annotation in Livestock for positional candidate loci* (*GALLO*; Fonseca et al., 2020), to perform an enrichment analysis. The QTL annotation file (*.gff*) was imported from Animal QTL Database (Cattle QTLdb, 2021; Hu et al., 2013, 2022), using ARS_UCD1.2 as reference genome. Concerning EHH regions, all the subsequent analyses were conducted only on the shared haplotypes detected by the 2 approaches (XP-EHH and Rsb). The proportion of QTL

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

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

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

[2]Ne = effective population size.

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

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

RESULTS

Analysis of Genetic Relationships and Population Structure

Genetic diversity indices for the 3 Holstein populations are reported in Table 1. The itHOL showed the highest level of observed heterozygosity and the lowest level of inbreeding, the acHOL showed an opposite pattern, whereas the prHOL showed intermediate values. Concerning contemporary effective population size (***Ne***), the prHOL population showed the highest value

whereas the acHOL showed the lowest (Table 1). Trends in population size as shown by exploring *Ne* over the last 100 generations indicated a similar pattern with a general decrease of all populations starting between 25 and 20 generations ago. However, the Italian Holstein (itHOL and prHOL) showed higher values of *Ne* compared with the North American Holstein (acHOL; Supplemental Figure S2; https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023).

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

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

Selection Signature Through Genome-Wide Allele Frequency Approaches

Among the 3 sets of comparisons, a total of 120 outlier SNPs were detected using the single marker Bayesian approach whereas 140 SNPs were identified using the F_{ST} window-based approach (Figure 3). The list of all significant SNPs is reported in Supplemental Table S2 (https://doi.org/10.6084/m9.figshare.22786295.v1; Senczuk, 2023). Among these, we found 15 SNPs encompassing 6 chromosomes (BTA1, BTA3, BTA14, BTA16, BTA18, and BTA20) that were coincident between the 2 approaches. Within these candidate regions, 11 annotated genes were located (*VPS8, P2RY1, BTG2, CHIT1, MYBPH, CHI3L1, PPP1R15B, PIK-3C2B, ZNF404, ZNF45,* and *ZNF226*).
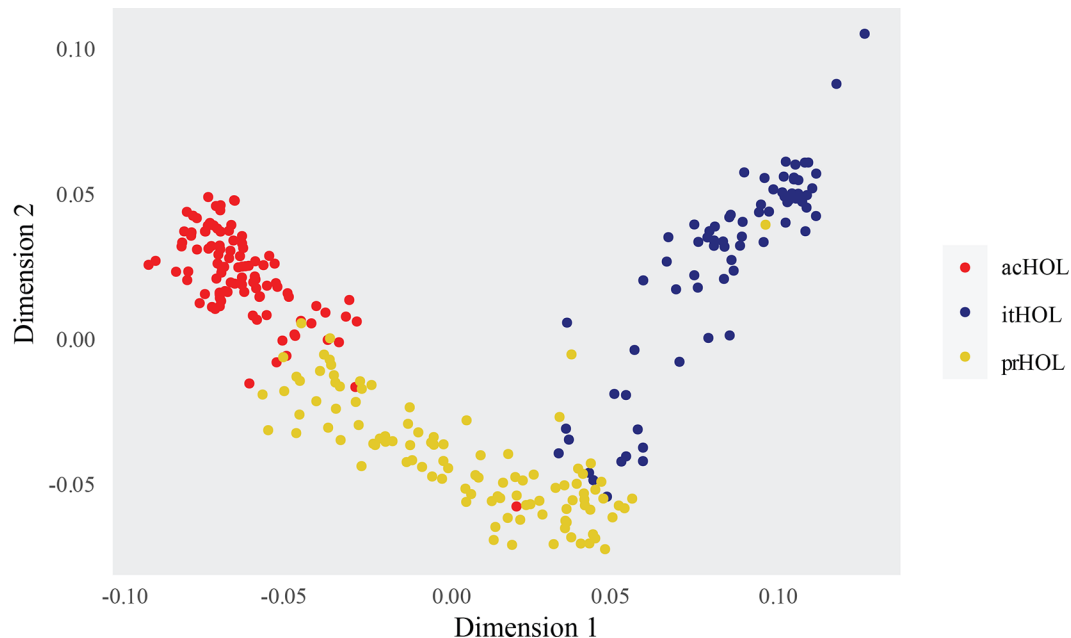
Figure 1. Multidimensional scaling plot showing the genetic relationships of the analyzed Holstein populations as reported in Table 1. acHOL = only individuals from the United States and Canada; itHOL = Italian Holstein samples not used to produce Parmigiano Reggiano; prHOL = only Italian Holstein cattle used to produce Parmigiano Reggiano..

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

Considering the F_{ST} window-based approach, we found 12 significant windows for each comparison (Supplemental Table S2). However, several windows were partially
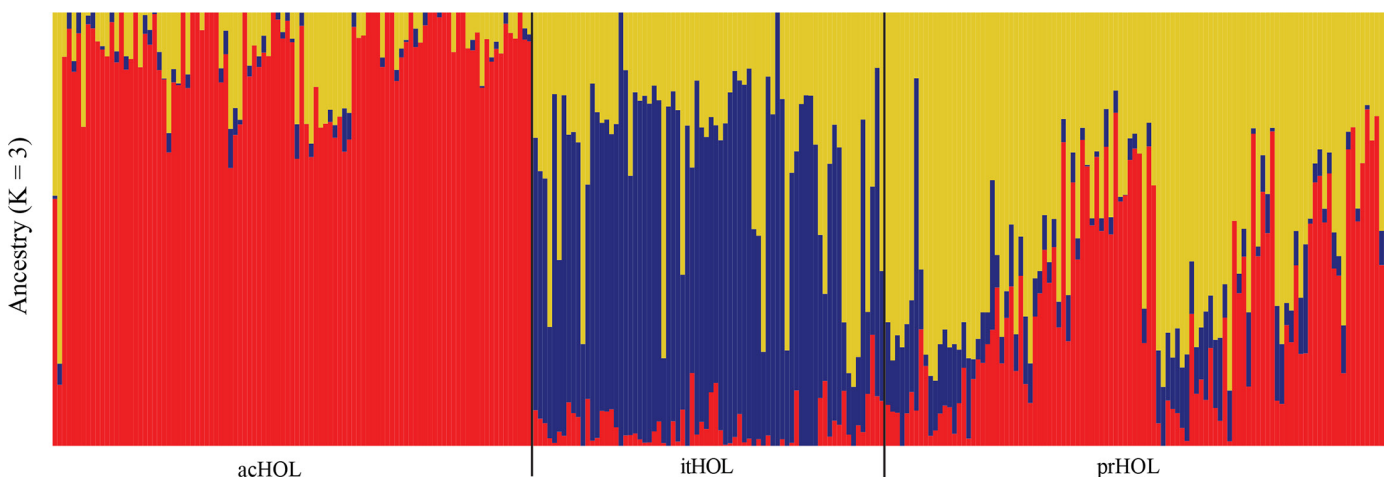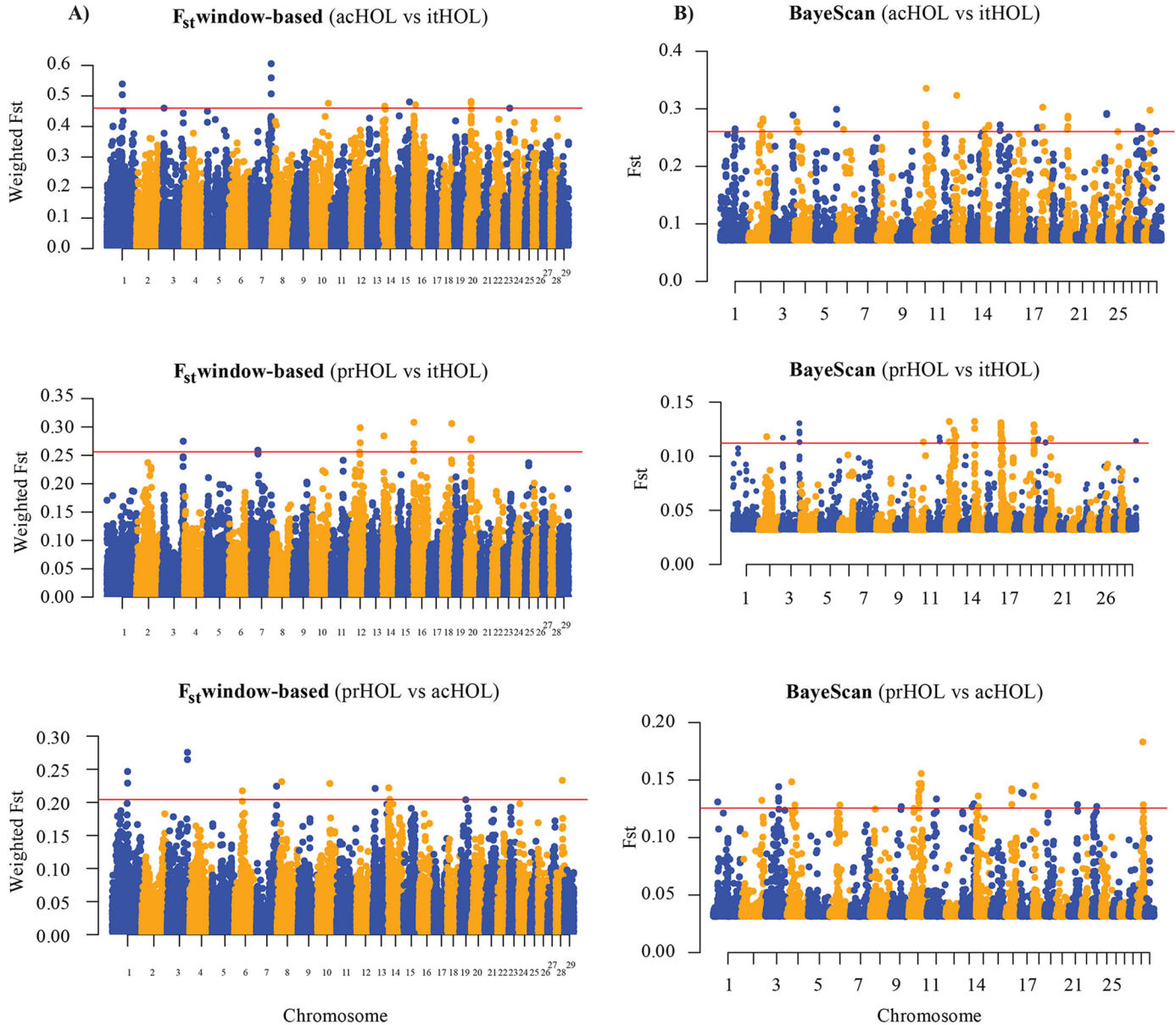

Figure 2. Bar plot of the ADMIXTURE analysis considering the most reliable number of given ancestral populations (K values; K = 3). acHOL = only individuals from the United States and Canada; itHOL = Italian Holstein samples not used to produce Parmigiano Reggiano; prHOL = only Italian Holstein cattle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

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

overlapped and were therefore merged (Table 3). In addition, only 7, 6, and 5 windows contained annotated genes in the itHOL versus acHOL, prHOL versus acHOL, and prHOL versus itHOL comparisons, respectively (Table 3). In the comparison between the itHOL and acHOL, prHOL and acHOL, and prHOL and itHOL, 9, 8, and 15 annotated genes located in 7, 5, and 4 autosomes were found. Also, in this case, overlapping significant SNPs in the comparisons itHOL versus acHOL and prHOL versus acHOL mapped within known genes (*VPS8* and

FER; Table 3). In addition, within these 2 comparisons, the genes *VPS8*, *FER*, and *GRIK3* were detected in more than one significant window.

Selection Signature Through Extended Haplotype Homozygosity

The 2 EHH-based analyses between populations revealed a total of 29 candidate regions putatively under selection (Figure 4). Among those, 11 were obtained

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

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

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

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

	itHOL vs. acHOL					prHOL vs. acHOL					prHOL vs. itHOL			
BTA	Start	End	no.	Gene		BTA	Start	End	no.	Gene	BTA	Start	End	no.	Gene
1	82260001	82440000	5	_VPS8*_		1	82260001	82440000	5	_VPS8*_	3	114900001	115020000	5	_P2RY1_
3	12240001	12360000	5	KIRREL1*, TRNAG-CCC		3	108840001	109020000	4	GRIK3*	12	48420001	48540000	5	KLF12*
7	107940001	108300000	4	FER*		7	107940001	108060000	4	FER*	16	960001	1080000	6	_CHI3L1*, MYBPH, CHIT1*, BTG2_
10	87780001	87900000	4	ESRRB*		8	24540001	24660000	4	SLC24A2*	16	2100001	2280000	5	_PPP1R15B*, PIK3C2B*, PLEKHA6, MDM4, TRNAK-UUU_
14	10560001	10680000	6	ASAP1*, FAM49B*		14	2520001	2640000	4	MROH5, PTP4A3, GPR20*	18	52020001	52140000	4	_ZNF404*, ZNF283*, ZNF45*, ZNF226*_
15	61620001	61740000	4	DCDC1*		14	7860001	7980000	6	ST3GAL1*					
16	10440001	10560000	4	HMGB1P1											

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

using the XP-EHH whereas 18 were obtained using the Rsb approach, respectively (Supplemental Table S2). We found a total of 7 overlapping genomic regions between the 2 EHH-based analyses, specifically 1 in the itHOL versus acHOL comparison, 3 in the prHOL versus acHOL comparison, and 3 in the prHOL versus itHOL comparison (Table 3).

Within the overlapping candidate regions identified by the 2 EHH-based analyses, a total of 266 annotated genes were found (Table 3). Among those, 52, 47, and 187 annotated genes were found within the itHOL versus acHOL, the itHOL versus prHOL, and the prHOL versus acHOL comparisons, respectively. In contrast, just a single genomic region located on the BTA10 was found in both the itHOL versus acHOL and the itHOL versus prHOL comparisons. In this region, we found 20 overlapping annotated genes (_MGAT2_, _DNAAF2_, _MIR6517_, _POLE2_, _KLHDC1_, _KLHDC2_, _NEMF_, _ARF6_, _VCPKMT_, _SOS2_, _L2HGDH_, _ATP5S_, _CDKL1_, _MAP4K5_, _ATL1_, _SAV1_, _NIN_, _ABHD12B_, _PYGL_, and _TRIM9_).

Quantitative Trait Loci and Functional Enrichment Analysis

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

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

Concerning the EHH-based approach, among the significantly enriched milk type QTL (all detected with _Padj_-value < 0.05) we found milk protein percentage, milk fat content, and milk caproic acid content on
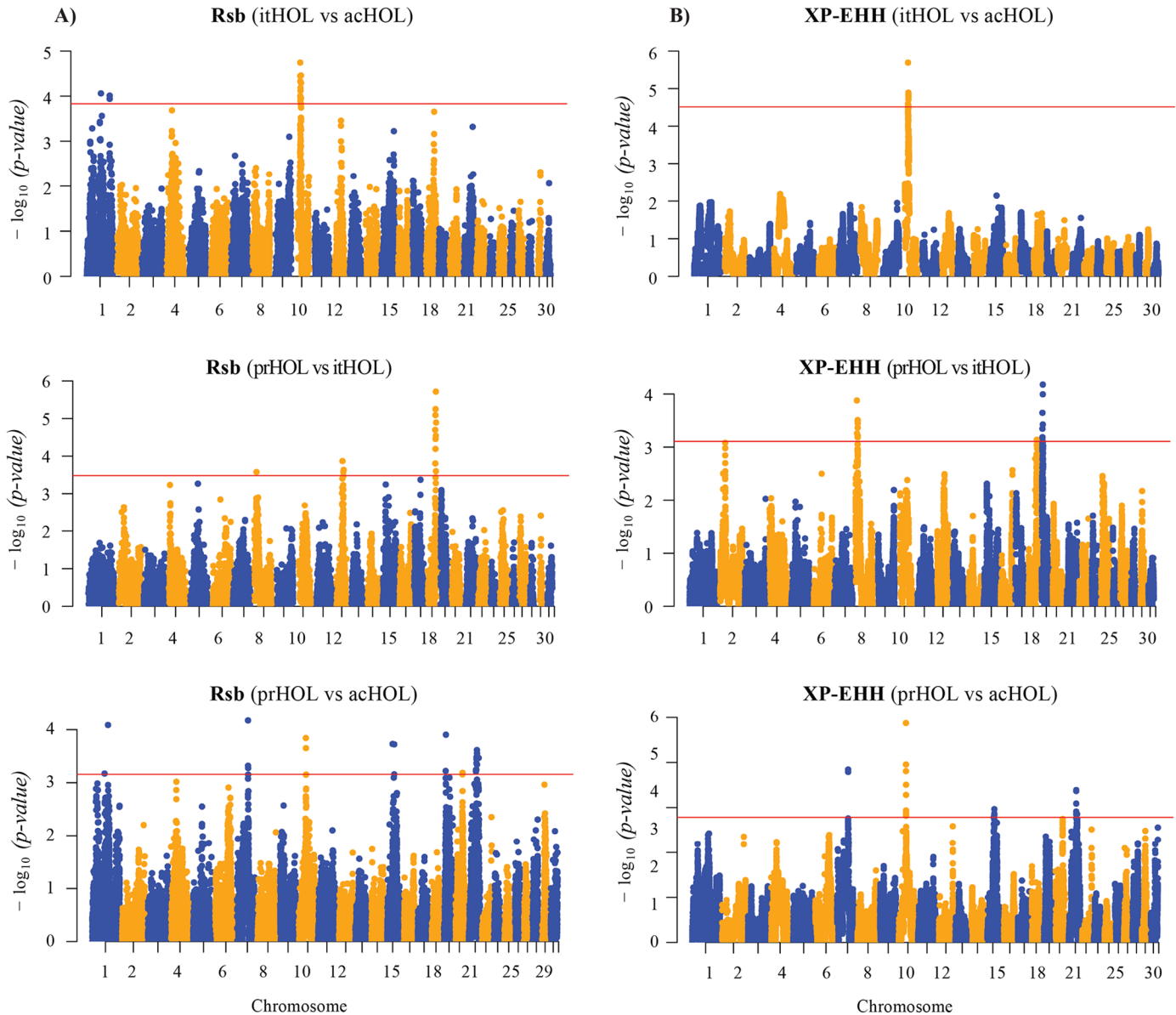
Figure 4. Manhattan plot of the $-\log_{10}$ (P-value) calculated between pairs of the 3 Holstein populations using the statistical method based on extended haplotype homozygosity [ratio of extended haplotype homozygosity segments between populations (Rsb; A) and the cross-population extended haplotype homozygosity (XP-EHH; B)] approaches. acHOL = only individuals from the United States and Canada; itHOL = Italian Holstein samples not used to produce Parmigiano Reggiano; prHOL = only Italian Holstein cattle used to produce Parmigiano Reggiano. The horizontal red line indicates the threshold based on the 0.05% $\log_{10}$(P-value) (99.95th percentile).

BTA10 in the itHOL versus acHOL comparison, milk linoleic acid content and milk casein percentage within the selection signals located on BTA19 and BTA18 in the prHOL versus acHOL comparison, and milk phosphorus content and milk α-S1-CN percentage within the selection signal on BTA7 in the itHOL versus prHOL comparison (Supplemental Figure S5).

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

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

were found significantly enriched in the itHOL versus acHOL and prHOL versus acHOL, respectively.

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

DISCUSSION

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

Our study shows that both the regular Italian Holstein and Italian Holstein used to produce Parmigiano Reggiano are clearly differentiated from the North American Holstein. In particular, we found that Italian Holstein used for the production of Parmigiano Reggiano seems to retain higher genetic diversity in terms of genomic admixture whereas both the North American and Italian groups seem to be characterized by a more homogeneous pattern. Such a pattern is probably due to stronger genetic drift caused by the early implementation of genomic selection in acHOL, and later in itHOL, as also detected in previous works (Makanjuola et al., 2020; Hulsegge et al., 2022). In contrast, the higher genomic admixture highlighted in the Holstein population used for the production of Parmigiano Reggiano might reflect the past use of the American stock to implement the breeding selection strategy linked to dairy productive traits. In addition, our results are also in line with other studies which found a certain distinctiveness of some Italian Holstein cows reared for other cheese production chains (Ablondi et al., 2021). These outcomes are also confirmed by all genetic diversity indices which indicates higher inbreeding and lower observed heterozygosity in the North American population which is well known to have been exposed to intense genomic selection. To strengthen these considerations are the effective population size estimates which underline a high contemporary value for the prHOL and lower values for both itHOL and acHOL. In general, the average inferred contemporary Ne for the 3 populations resulted comparable with those observed in others worldwide cosmopolitan breeds (Senczuk et al., 2020b). Finally, trends in Ne show considerable drops

from 25 generations ago in all the analyzed populations but in general, the Italian Holstein populations (itHOL and prHOL) show higher values. Even thoough the inferred sudden Ne decline may reflect a recent increase in the genetic drift due to the strong selection, these results should be taken with a certain caution as very recent drastic drops might be related to bias estimates either in structured populations characterized by low migration rate or in population with recent migrants (Santiago et al., 2020).

Selection Signature Through Genome-Wide Allele Frequency Approaches

The analytical approach using F_{ST} has been recently applied in cattle for the identification of genomic regions involved in phenotypic differences (Mastrangelo et al., 2019a,b; Senczuk et al., 2020a; Moscarelli et al., 2021; Bertolini et al., 2021). The genome-wide information in comparative F_{ST} was used to detect signatures of selection that diverge among the 3 groups of Holstein cattle considered in this study. Several genes with known roles in traits related to milk quality, resistance to disease and immune response, and fertility have been identified (Table 2). The fact that we do not find a great overlap in significant SNPs between the 2 approaches in almost all comparisons is not surprising considering the 2 different schemes in selecting SNPs we used. Indeed, for the F_{ST} windows-based approach we used a weighted F_{ST} average, retaining only windows with at least 4 SNPs (Weir and Cockerham, 1984). However, a total of 11 genes (*VPS8*, *P2RY1*, *BTG2*, *CHIT1*, *MYBPH*, *CHI3L1*, *PPP1R15B*, *PIK3C2B*, *ZNF404*, *ZNF45*, and *ZNF226*) still showed concordant results between the 2 approaches. Moreover, most of these genes have been shown to have a role in different productive traits in cattle (Pimentel et al., 2011; Du et al., 2019; Mohammadi et al., 2020; Illa et al., 2021; Soares et al., 2021; Yang et al., 2021). For example, the *VPS8* gene overlapped with the bovine QTL region of milk traits (Taye et al., 2017), the gene *ZNF226* is associated with growth and carcass quality QTL (MacHugh et al., 1997), and *CHI3L1* is related to the innate immune response during bacterial inflammation of the mammary gland (Breyne et al., 2018).

The BayeScan analysis detected 86 unique annotated genes falling within the upstream and downstream 60 kb from the significant SNPs. Several of these genes were detected in more than one comparison (*UNC13C*, *MAPK*, *LEO1*, *ZFPM1*, *KCNMA1*, and *C12H13orf42*). Among the 86 annotated genes, 55 have been shown to play a role in productive traits (in bold Tables 2 and 3) including genes related to milk production (*GSK3A*, *SUPT4H1*, *RNF43*, *VPS8*, *KSR2*, *ZFPM1*, *KCNMA1*,

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

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

Selection Signature Through Extended Haplotype Homozygosity

The genome-wide scan addressed to find putative genomic regions under selection using EHH approaches (XP-EHH and Rsb), identified a total of 29 genomic regions putatively under selection (Supplemental Table S2). However, in the haplotypic approach we did not find any common signal with the 2 F_{ST}-based approaches. This is a common observation in genome scans suggesting that detection of selection depends on the method used. In fact, F_{ST} and EHH-based approaches should be considered as complementary because of the different time scale resolution in which these methods are able to

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

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

Finally, in the comparison prHOL versus acHOL, we found 14 genes associated with milk production traits (*ZNF112*, *CBLC*, *SHKBP1*, *PGLYRP1*, *RAB4B*, *TMEM145*, *FOXA3*, *APOC2*, *BLVRB*, *TMEM98*,

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

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

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

MYO1D, *RHOT1*, *ADAP2*, and *UTP6*) which mapped on BTA8, BTA18, and BTA19. Among these, Sablik et al. (2020) demonstrated how a mutation in *PGLYRP1* increased milk and protein production during first lactation in Holstein Friesian cattle, whereas *FOXA3* was observed to be associated with milk and protein yield in several studies (Kolbehdari et al., 2009; Valdez-Torres et al., 2020).

Quantitative Trait Loci and Functional Enrichment Analysis

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

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

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

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

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

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

CONCLUSIONS

In this study we showed that the 3 Holstein populations are clearly differentiated; however, both the Italian and North American Holstein showed low levels of genetic diversity indices, higher inbreeding coefficients compatible with their history of isolation, drift, and strong genomic selection. Conversely, the Holstein population used to produce Parmigiano Reggiano, showed higher genetic diversity and lower inbreeding values. These results suggest that also in a cosmopolitan breed, often characterized by intense genetic exchange, the different breeding goals, environmental conditions, and farm systems can produce a genetic differentiation in a very short number of generations. Finally, the selec-

tion signature analyses detected the presence of several genes within QTL regions for important production traits associated with fertility, resistance to disease but especially with milk production traits. Overall, the results of the present study indicate that the 3 Holstein populations are genetically distinguishable from each other and that our selection signature analyses can be a starting point to identify causal mutations useful in future studies and lately in more practical applications.

ACKNOWLEDGMENTS

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Mastrangelo, S., G. Sottile, M. T. Sardina, A. M. Sutera, M. Tolone, R. Di Gerlando, and B. Portolano. 2019b. A combined genome-wide approach identifies a new potential candidate marker associ-

ated with the coat color sidedness in cattle. Livest. Sci. 225:91–95. https://doi.org/10.1016/j.livsci.2019.05.009.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Soares, R. A. N., G. Vargas, T. Duffield, F. Schenkel, and E. J. Squires. 2021. Genome-wide association study and functional analyses for clinical and subclinical ketosis in Holstein cattle. J. Dairy Sci. 104:10076–10089. https://doi.org/10.3168/jds.2020-20101.

Tang, K., K. R. Thornton, and M. Stoneking. 2007. A new approach for using genome scans to detect recent positive selection in the

human genome. PLoS Biol. 5:e171. https://doi.org/10.1371/journal.pbio.0050171.

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

Theunissen, B. 2012. Breeding for nobility or for production? Cultures of dairy cattle breeding in the Netherlands, 1945–1995. Isis 103:278–309. https://doi.org/10.1086/666356.

Turner, S. D. 2018. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. J. Open Source Softw. 3:731. https://doi.org/10.21105/joss.00731.

Vacek, M., M. Štípková, E. Němcová, and J. Bouška. 2006. Relationships between conformation traits and longevity of Holstein cows in the Czech Republic. Czech J. Anim. Sci. 51:327–333. https://doi.org/10.17221/3946-CJAS.

Valdez-Torres, J. M., J. A. Grado Ahuir, B. E. Castro-Valenzuela, and M. E. Burrola-Barraza. 2020. QTL analysis associated to single nucleotide polymorphisms (SNP) involved in the dairy phenotype of Holstein cattle. Rev. Mex. Cienc. Pecu. 11:1192–1207. https://doi.org/10.22319/rmcp.v11i4.5295.

Visentin, G., M. Marusi, and G. Civati. 2018. ICS-PR: Una nuova selezione per la Frisona italiana [ICS-PR: A new selection for the Italian Holstein]. Bianconero n.6 Novembre-Dicembre 2018:12–15. http://server01.anafi.it/Latteco/articoli/2018/07-BN-Nov-Dic-2018.pdf.

Vohra, V., S. Chhotaray, G. Gowane, R. Alex, A. Mukherjee, A. Verma, and S. M. Deb. 2021. Genome-wide association studies in Indian Buffalo revealed genomic regions for lactation and fertility. Front. Genet. 12:696109. https://doi.org/10.3389/fgene.2021.696109.

Vollema, A. R., and A. F. Groen. 1998. A comparison of breeding value predictors for longevity using a linear model and survival analysis. J. Dairy Sci. 81:3315–3320. https://doi.org/10.3168/jds.S0022-0302(98)75897-7.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370. https://doi.org/10.2307/2408641.

Wickham, H. 2011. ggplot2. Wiley Interdiscip. Rev. Comput. Stat. 3:180–185. https://doi.org/10.1002/wics.147.

Yang, Z., Z. Lian, G. Liu, M. Deng, B. Sun, Y. Guo, D. Liu, and Y. Li. 2021. Identification of genetic markers associated with milk production traits in Chinese Holstein cattle based on post genome-wide association studies. Anim. Biotechnol. 32:67–76. https://doi.org/10.1080/10495398.2019.1653901.

Yerex, R. P., C. W. Young, J. D. Donker, and G. D. Marx. 1988. Effects of selection for body size on feed efficiency and size of Holsteins. J. Dairy Sci. 71:1355–1360. https://doi.org/10.3168/jds.S0022-0302(88)79693-9.

Zhao, F., S. McParland, F. Kearney, L. Du, and D. P. Berry. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet. Sel. Evol. 47:49. https://doi.org/10.1186/s12711-015-0127-3.

Zhou, C., C. Li, W. Cai, S. Liu, H. Yin, S. Shi, Q. Zhang, and S. Zhang. 2019. Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. Front. Genet. 10:72. https://doi.org/10.3389/fgene.2019.00072.

ORCIDS

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