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 (Themnissen, 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
Persichilli et al. : GENOME-WIDE DIVERSITY OF HOLSTEIN CATTLE

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 β-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; Marsusi 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; Abondì 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 (FST) 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 FST 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 Moscaresi 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 FST approaches were used: (i) a genome-wide FST outlier method using the software BayeScan (Foll and Gaggiotti, 2008) and (ii) a genome window-based analysis using weighted FST 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 FST outlier analysis we retained SNPs based on the top 0.05% FST value (99.95th percentile). We then considered a region of ± 60 kb around the significant SNPs to retain genes and explore their function.

For the FST window-based approach, after retaining windows with at least 4 SNPs, the top 0.05% FST (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 FST 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
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 FST 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` < 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 FST-based or EHH-falling within significantly enriched QTL as revealed and pathways. As a gene list we included all genes 2009) was used to explore potential biological functions between genes, the software ToppGene (Chen et al., 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 FST-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). 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 (`ZNF45`, `ZNF226`, `ZNF404`, `3C2B`, `PIK3C2B`, `VPS8`, `P2RY1`, `BTG2`, `CHIT1`, `MYBPH`, `CHI3L1`, `PPP1R15B`, `PIK3C2B`, `ZNF404`, `ZNF45`, and `ZNF226`).

### 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))

<table>
<thead>
<tr>
<th>Population</th>
<th>Code</th>
<th>No.</th>
<th><code>Ho ± SD</code></th>
<th><code>He ± SD</code></th>
<th><code>F_Roh</code></th>
<th><code>Ne</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian Holstein</td>
<td>itHOL</td>
<td>74</td>
<td>0.373 ± 0.160</td>
<td>0.354 ± 0.144</td>
<td>0.102</td>
<td>43.39</td>
</tr>
<tr>
<td>Parmigiano Reggiano Holstein</td>
<td>prHOL</td>
<td>105</td>
<td>0.363 ± 0.138</td>
<td>0.368 ± 0.135</td>
<td>0.125</td>
<td>89.05</td>
</tr>
<tr>
<td>North American Holstein</td>
<td>acHOL</td>
<td>101</td>
<td>0.342 ± 0.171</td>
<td>0.332 ± 0.160</td>
<td>0.168</td>
<td>39.97</td>
</tr>
</tbody>
</table>

1`F_Roh` = genomic inbreeding coefficient (`F`) derived from runs of homozygosity (ROH).
2`Ne` = effective population size.

### 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; Supplimental 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 FST 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.
In the BayeScan analysis the FST 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% FST) 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 FST window-based approach, we found 12 significant windows for each comparison (Supplemental Table S2). However, several windows were partially

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.

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.
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 \((\text{acHOL})\), a second including Italian Holstein samples not used to produce Parmigiano Reggiano \((\text{itHOL})\), and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano \((\text{prHOL})\)\(^1\)).

<table>
<thead>
<tr>
<th>BTA SNP</th>
<th>Gene</th>
<th>BTA SNP</th>
<th>Gene</th>
<th>BTA SNP</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BTA-38502-NO-RS</td>
<td>(\text{VPSS8}^*)</td>
<td>2 BOVINEHD020003369</td>
<td>(\text{IRSI}_1, \text{RHBDD1}^*)</td>
<td>3 BOVINEHD030004169</td>
<td>(\text{FCLR1}^*)</td>
</tr>
<tr>
<td>1 ARS-BFGL-NGS-14913</td>
<td>(\text{VPSS8}^*)</td>
<td>3 BTR-01071906</td>
<td>(\text{NEGR1}^*)</td>
<td>3 ARS-BFGL-NGS-113768</td>
<td>(\text{P2RY1}^*)</td>
</tr>
<tr>
<td>2 ARS-BFGL-NGS-94696</td>
<td>(\text{TMEFF2}^*)</td>
<td>3 BTA-10777-NO-RS</td>
<td>(\text{NEGR1}^*)</td>
<td>10 BOVINEHD1000020414</td>
<td>(\text{DACT1}^*)</td>
</tr>
<tr>
<td>2 BOVINEHD020003298</td>
<td>(\text{TMEFF2}^*)</td>
<td>3 ARS-BFGL-NGS-119369</td>
<td>(\text{NEGR1}^*)</td>
<td>11 ARS-BFGL-NGS-77753</td>
<td>(\text{CAPN1}^*)</td>
</tr>
<tr>
<td>4 BTA-83577-NO-RS</td>
<td>(\text{UMAD1}^*, \text{RF00001})</td>
<td>4 BTA-00171304</td>
<td>(\text{HDAC9})</td>
<td>12 BOVINEHD120006184</td>
<td>(\text{C12H1}^*)</td>
</tr>
<tr>
<td>4 BOVINEHD-08374-NO-RS</td>
<td>(\text{UMAD1}, \text{RF00001})</td>
<td>4 BTA-70407-NO-RS</td>
<td>(\text{RELN})</td>
<td>16 UA-IFASA-9087</td>
<td>(\text{CHIT1}, \text{CHIT1L}), (\text{MYBPH}), (\text{ADORAI})</td>
</tr>
<tr>
<td>4 HAPMAP50489-</td>
<td>(\text{NXPH1}^*, \text{TRNAC-GCA})</td>
<td>6 HAPMAP54653-</td>
<td>(\text{KLH}L5^*)</td>
<td>16 HAPMAP54719-</td>
<td>(\text{CHIT1})</td>
</tr>
<tr>
<td>BTA-83596</td>
<td>10 BTA-00430000</td>
<td>RS29025767</td>
<td>10 BOVINEHD10000206</td>
<td>(\text{ITPKC}^*)</td>
<td>(\text{BTG2}^*)</td>
</tr>
<tr>
<td>10 BTB-00430000</td>
<td>(\text{UNC13C}^*)</td>
<td>10 BOVINEHD1000016688</td>
<td>(\text{UNC13C}^*)</td>
<td>16 BOVINEHD10000201</td>
<td>(\text{BTG2}^*)</td>
</tr>
<tr>
<td>10 BOVINEHD1000016688</td>
<td>(\text{UNC13C}^*)</td>
<td>10 ARS-BFGL-NGS-55539</td>
<td>(\text{MYOSC}^*)</td>
<td>16 BOVINEHD100000513</td>
<td>(\text{PPPIR15B})</td>
</tr>
<tr>
<td>10 ARS-BFGL-NGS-55539</td>
<td>(\text{MAPK6}, \text{LEO1})</td>
<td>10 BOVINEHD100000531</td>
<td>(\text{MAPK6}, \text{LEO1})</td>
<td>16 BOVINEHD100000544</td>
<td>(\text{PK3C2B}^*)</td>
</tr>
<tr>
<td>12 BOVINEHD120006184</td>
<td>(\text{C12H1}^*)</td>
<td>12 BOVINEHD1200006184</td>
<td>(\text{C12H1}^*)</td>
<td>16 BOVINEHD1600000210</td>
<td>(\text{PK3C2B}^*)</td>
</tr>
<tr>
<td>14 BOVINEHD140000920</td>
<td>(\text{C18H19orf54})</td>
<td>14 BOVINEHD140000920</td>
<td>(\text{RF00100})</td>
<td>16 BOVINEHD1600000210</td>
<td>(\text{PK3C2B}^*)</td>
</tr>
<tr>
<td>15 BOVINEHD160000920</td>
<td>(\text{CFR7})</td>
<td>15 HAPMAP54387-</td>
<td>(\text{GLDN})</td>
<td>16 BOVINEHD170001720</td>
<td>(\text{FBXO33}^*)</td>
</tr>
<tr>
<td>16 BOVINEHD170001720</td>
<td>(\text{GLDN})</td>
<td>16 HAPMAP54719-</td>
<td>(\text{FBXW8})</td>
<td>17 BOVINEHD170001720</td>
<td>(\text{FBXO33}^*)</td>
</tr>
<tr>
<td>17 BOVINEHD170001720</td>
<td>(\text{GLDN})</td>
<td>17 HAPMAP54719-</td>
<td>(\text{FBXW8})</td>
<td>17 ARS-BFGL-NGS-101538</td>
<td>(\text{GSE1}^*)</td>
</tr>
<tr>
<td>18 BOVINEHD1800000452</td>
<td>(\text{ZFPF1}^*)</td>
<td>18 BOVINEHD1800000452</td>
<td>(\text{ZFPF1}^*)</td>
<td>18 ARS-BFGL-NGS-9872</td>
<td>(\text{BCAS3}^*)</td>
</tr>
<tr>
<td>19 BOVINEHD1800000452</td>
<td>(\text{ZFPF1}^*)</td>
<td>19 BOVINEHD1800000452</td>
<td>(\text{ZFPF1}^*)</td>
<td>18 ARS-BFGL-NGS-77925</td>
<td>(\text{TSPQA1}, \text{MIR142}, \text{SPT4H1}, \text{RECQL5}^*), (\text{SMIM5}), (\text{SMIM6}), (\text{MYO15B})</td>
</tr>
<tr>
<td>20 BOVINEHD1800000452</td>
<td>(\text{ZFPF1}^*)</td>
<td>20 BOVINEHD1800000452</td>
<td>(\text{ZFPF1}^*)</td>
<td>19 BOVINEHD1900016000</td>
<td>(\text{BCAS3}^*)</td>
</tr>
<tr>
<td>21 BOVINEHD1900016000</td>
<td>(\text{ZFPF1}^*)</td>
<td>21 BOVINEHD1900016000</td>
<td>(\text{ZFPF1}^*)</td>
<td>20 BOVINEHD1900016000</td>
<td>(\text{BCAS3}^*)</td>
</tr>
<tr>
<td>22 BOVINEHD2000007525</td>
<td>(\text{ZFPF1}^*)</td>
<td>22 BOVINEHD2000007525</td>
<td>(\text{ZFPF1}^*)</td>
<td>20 BOVINEHD2000007525</td>
<td>(\text{BCAS3}^*)</td>
</tr>
</tbody>
</table>

\(^1\)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.
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 BTA5 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
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 has 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 (Makanjula 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 indicate 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 though 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 ($VPS8$, $P2RY1$, $BTA2$, $CHIT1$, $MYBPH$, $CHI3L1$, $PPP1R15B$, $PIK3C2B$, $ZNF404$, $ZNF45$, and $ZF226$) 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 $ZF226$ 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 ($GSK3A$, $SUPT4H1$, $RNF43$, $VPS8$, $KSR2$, $ZFPM1$, $KCNMA1$, $VPS8$, $KSR2$, $ZFPM1$, $KCNMA1$, $VPS8$, $KSR2$, $ZFPM1$, $KCNMA1$).
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).

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). 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 putative and functional traits such as protein composition traits and coagulation (Zhou 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).

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). 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 putative and functional traits such as 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 ($ZNF112$, $CBLC$, $SHKBPI$, $PGLYRP1$, $RAB4B$, $TMEM145$, $FOX3$, $APOC2$, $BLVRB$, $TMEM98$, $DCDC1$, $SLC24A2$, $GPR20$, $PRL/miR-183/IRS1$, 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 $KCNA1$ has been predicted to have a functional interaction with the $\kappa$-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 ($ZFP71$, $KIAA0586$, $VPS8$, $C4A$, $CYP7B1$, 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).

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).

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). 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 putative and functional traits such as protein composition traits and coagulation (Zhou et al., 2019).
### 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 (ihHOL), and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano (phHOL))

<table>
<thead>
<tr>
<th>BTA</th>
<th>ihHOL vs. acHOL</th>
<th>Start (Mb)</th>
<th>End (Mb)</th>
<th>Gene (no.) Length (Mb)</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td>40.96</td>
<td>48.45</td>
<td>52</td>
<td>7.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MGAT2, DNAAF2, MIR6517, POLY5, KLHDGC1, KLHDGC2, NEMF, ARF6, VCPKMT, SOS2, L24GDH,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATPS3, CDKL1, MAPK5, AT8L, SAV1, NIN, ABHD12B, PYGL, TRIM9, TMX1, FRMD6, GNG2,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTARF, NID2, PTGDR, PLEKH02, PFF1, RBPM52, OAZ2, ZNF609, TRIP4, PCLAF, CSNK1G1, PPB,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SNX22, SNX1, FAM96A, DAPK2, HERC1, FBXL22, USP3, CA12, APH1B, RAB8B, RPS27L,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LACTB, TMI, TLN3, MIR190A, C2CD4A, VPS13C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ihHOL vs. phHOL</td>
<td>61.90</td>
<td>63.88</td>
<td>22</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NSD1, SYNPO, MYOZ3, RRM212, DCTN4, SMIM3, ZNF300, GPE3, TNNP1, ANXAX, CCDC69, GM2A,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SLC36A3, SLC36A2, SLC36A1, FAT2, BERV-K1, SPARC, ATOX1, G3BP1, GLRA1, NMUR2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>41.69</td>
<td>43.94</td>
<td>24</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RPS29, RN7SL1, LRRK1, RPL36A1, MGAT2, DNAAF2, MIR6517, POLY5, KLHDGC1, KLHDGC2, NEMF,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARF6, VCPKMT, SOS2, L24GDH, ATPS3, CDKL1, MAPK5, AT8L, SAV1, NIN, ABHD12B, PYGL,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TRIM9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>acHOL vs. phHOL</td>
<td>42.28</td>
<td>44.48</td>
<td>1</td>
<td>2.25</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>16.82</td>
<td>19.8</td>
<td>1</td>
<td>2.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TUSC1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>15.45</td>
<td>19.07</td>
<td>33</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DYL, SELENOV, EID2, DYRK1B, FBL, FCGBP, PSMC4, MAP3K10, TT2B, CNTD2, AKT2,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C18H19orf17, PLD3, HIPK4, FRX, SERTAD1, SERTAD3, BLHRV4, SPTBN4, SHKBP1, LTBP4,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NUMBL, CCOQ8B, IDKPC, C18H19orf54, SNRPB, MIA, RAB4B, EGLN3, CYP26B, CYP26, AXL,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HRNRPRUL1,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CDDC97, TGFB1, B9D2, TMEM91, EXOSC5, BCDH1A, B3GNT8, DAMAC2, ERIC1, CEACAM1,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LIPE, CNFN, MECGF3, TMEM145, PRR11, PAFH1B3, C1, ERC, GSK3A, ZNF526, DEDD2, POU2F2,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZNF574, GRIK3, ATP1A3, RABAC1, ARHGEF1, CD9A, RPS19, DMRTC2, LYPD4, CXCL7, CD177,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEX101, DSP3, BPS5, LYPD5, PHLDG3, ETHE1, ZNF573, XRC1, IRGQ, ZNF576, ZNF28,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRRM5, CADMI1, PLAUR, IRGC, SNCA, KCN4, LDADP5, ZNF283, ZNF104, ZNF45, ZNF226,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZNF229, ZNF235, ZNF112, ZNF285, ZNF380, CEACAM20, IGF2B2, PVR, CEACAM19, CEACAM16,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BCL3, BCLC, BCAM, NECTIN2, TOMM40, AP0E, AP0E, AP0C2, CLPTM1, RELB, CLASRP,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TM296, GEMI1N7, PPP1R37, NKD1, TRAPP56A, BLOC1SA, EXOCL2L, MARK1, CKM, KLC3,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ERCC2, P2P1R13L1, CD3EAP, ERCC1, FOSEB, RTN2, PPM1N, VASP, OPA3, GBP4, EML2, MIR330,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GIPR, SNRPD2, QPCTL, FBXO16, BMGRI, SIX5, DMPK, DMWD, RSP46A, SYMPK, FOXA3,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IFRSB1, MYP, NAXO52, NOVA2, CCDC61, MIR769, PGLYRP1, IGEL1, HIF3A,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TMEM133E, CCLI, CCL8, CCL11, CCL2, ASIC2, SPACA3, TMEM98, MYO1D, CSDK5R1, PSMD11,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZNF207, C19H19orf75, RHBDLD3, RHT1, RNFI35, ADAP2, TEFM, ATAD5, CRLF3, SUZ12, UTP6,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIR2322, COPRS, MIR325-2, MIR2332, MIR193A, RAB181FIP4, EVI2A, EVI2B, OMG, NF1, WSBI</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1In 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 FST windows-based and EHH-based approaches showed the highest fraction of QTL related to milk traits in all comparisons. For example, in the FST 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; Pérez-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

This study was supported by “Latteco2 project, sotomisura 10.2 of the PSRNBiodiversity 2020–2023” (MIPAAF. D.M. no. 465907 del 24/09/2021, project unique code J12C21004080005). The authors have not stated any conflicts of interest.

REFERENCES


Persichilli et al.: GENOME-WIDE DIVERSITY OF HOLSTEIN CATTLE


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

Journal of Dairy Science Vol. 106 No. 8, 2023