

## Genomic approach to manage genetic variability in dairy farms

Chiara Punturiero, Raffaella Milanese, Francesca Bernini, Andrea Delledonne, Alessandro Bagnato and Maria Giuseppina Strillacci

Department of Veterinary Medicine and Animal Production, Università degli Studi di Milano, Lodi, Italy

### ABSTRACT

In this study we investigated the genetic variability, the inbreeding and allele frequencies of monogenic traits in seven herds of Holstein breed and provided insight to farmers on the value of genomic management of reproduction in their herds. A total of 3,953 Holstein cows were sampled and genotyped with the Neogen GGP Bovine 100K SNP chip within the activities of the Regione Lombardia funded GO-PEI project 'GENOMIC tool for the management of reproduction in dairy cattle and for the control of inbreeding – GENORIP'. Principal component analysis was applied for analysing the genetic variability within and among farms using the SVS software of Golden Helix. Run of Homozygosity (ROH) and the genomic inbreeding were obtained with the detectRUNS package of the R software. Genotype frequencies for mendelian disease, fertility and production traits were also obtained. A total of 458,267 ROH were identified and ROH were distributed on all autosomes with an average length of 2,703,811 bp covering 12.7% of the genome. Several genomic regions appear under selection, while a specific region on BTA4 was identified in one herd, harbouring genes mainly related to the specific selection strategy of the farmer. The  $F_{ROH}$  values obtained considering ROH greater than 16Mb, varied from 0.004 to 0.325, with the highest  $F_{ROH}$  average value of 0.136. Among mendelian heritable diseases, the Haplotype Cholesterol Deficiency was the one with the largest proportion of carrier animals, i.e. 5.6%. A herd-tailored process to assist farmers in genomic management of reproduction was released. The ROH distribution within herd, together with the genotype frequencies for disease, fertility and production mendelian traits, suggest that similar directional selection is occurring across herds. This study released to each farmer the genomic make-up of their herd used jointly with the gEBV estimated by their national breeders' association (ANAFIBJ) for herd reproductive management.

### HIGHLIGHTS

- All females of 7 herds have been genotyped with the GGP 100K SNP chip.
- Genomic information on all females can be used by farmers in the process to manage reproduction, selection and genetic herd variability.
- The availability of genomic information on the whole herd allowed to release to each farmer the genomic make-up of their herd.
- The ROH distribution together with the genotype frequencies for disease, fertility and production mendelian traits, made it possible to identify genomic regions under selection according to farmer strategies.

### ARTICLE HISTORY

Received 23 April 2023

Revised 26 July 2023

Accepted 27 July 2023

### KEYWORDS

ROH; genomic variability; genomic inbreeding; Holstein cattle breed


## Introduction

The Holstein breed is widely recognised as the most productive dairy cattle in the world. The Holstein is the most wide-spread dairy cattle breed in Italy with 9,552 farms, 1,130,734 lactating cows and an average production of 10,710 kg of milk. In particular, in the Lombardy region located in the northern flat, which is suited to intensive dairy farming, there are 2,759 farms

with a total of 566,583 animals (ANAFIBJ – National statistics 2022).

The selection index of the Holstein breed has evolved through the years, with an initial emphasis on increasing milk yield per cow, followed by a shift towards milk components and functional and health traits (Egger-Danner et al. 2015). However, as in all selection programs, the intense selection practiced

**CONTACT** Maria Giuseppina Strillacci  [maria.strillacci@unimi.it](mailto:maria.strillacci@unimi.it)

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/1828051X.2023.2243977>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

over the years may led to a loss of genetic variability and to an increase in inbreeding in the population. The need to control the increase in inbreeding even in large populations under selection has been discussed for a long time (McDaniel 2001; Weigel 2001). In the second half of the last century, there was a motivation to introduce new molecular tools to integrate traditional phenotypic selection programs (Henderson 1975) with the use of information of loci and QTL regions that contain genes capable of influencing economically important traits in animal production (Georges et al. 1995; Andersson 2001).

In the genomics era occurred in the last decade, the paradigm of animal breeding has changed significantly. Current genotyping techniques make it possible to determine the genotype of an animal at hundreds of thousands of markers known to be associated with phenotypic variability at very low cost and use this information to select animals even without any performance available. This is the principle of genomic selection proposed by Meuwissen et al. (2001) based on the use of Single Nucleotide Polymorphisms (SNP) as markers.

The SNP are biallelic genomic markers very frequent in the genome of any individual (approximately one per 100 base pairs) (Collins et al. 1997). The SNP genotyping technology is nowadays a routine process in cattle breeding, both for males and females, producing a large number of genotyping information, allowing as such to implement efficient selection programs also for traits with low heritability (Boichard et al. 2015), and to develop comprehensive mating plans that make use of all the genomic information available to the breeder female herd.

One of the most important elements necessary to perform selection in livestock populations is the existence of genetic variability. Indeed, the occurred selection of superior animals over time has resulted in a loss of genetic diversity. This may cause a reduced response to selection and an increase of the frequency of homozygous loci (Dickerson and Hazel 1944).

Before the advent of genomics, the study of inbreeding was based on pedigree information which, however, has limitations: 1) the value of inbreeding depends on the quality and completeness of the pedigree data (Oliehoek and Bijma 2009); 2) it does not consider the genetic variability between full siblings due to recombination during meiosis leading to an underestimation and/or overestimation of inbreeding (Hill and Weir 2011); 3) the comparison between genomic and pedigree information, showed that the frequency of misidentified bulls can be as high as 13.9% (Wiggans et al. 2012), affecting as such the inbreeding

values based on this information. With the genomic selection and the development of high-density SNP arrays, it has become possible to obtain more accurate estimates of genome-wide inbreeding and relatedness (Engelsma et al. 2012). Genomic inbreeding can be calculated from a genome-wide relationship matrix (GRM) between individuals (Hayes et al. 2009), or as ratio between the length of the genome where homozygous markers form Run Of Homozygosity (ROH) (McQuillan et al. 2008) and the total genome length analysed (FROH). The length of the ROH provides also information on whether a ROH segment is the result of recent (long ROH) versus more distant (short ROH) autozygosity events (Pemberton et al. 2012). Additionally, identification of genes annotated in the ROH can provide insights on the selection occurring in the population (MacLeod et al. 2009; Purfield et al. 2012).

Due to intense selection, in dairy populations the inbreeding rates (Charlesworth and Willis 2009) and the frequency of deleterious alleles (Ouborg et al. 2010) have increased significantly over the years with a possible effect on productivity due to inbreeding depression (Falconer and Mackay 1983; Keller and Waller 2002). In Friesian cattle deleterious effects due to this phenomenon have been described on productive and functional traits (Martikainen et al. 2018; Doekes et al. 2019).

For these reasons, fostering genetic variability in the herd and controlling inbreeding is considered a priority in dairy farms and in management of cattle populations under selection. A farm-driven project funded by the Regione Lombardia in the EC EIP-AGRI Rural Development Program 2014–2020 framework is bringing genomics into the management of female replacements through the genotyping of all the animals in the herds. The project is named GENORIP: 'GENOmics' tool for the management of reproduction in dairy cattle and for the control of inbreeding. The project aimed to release a process to integrate the genomic management of herd reproduction and to manage inbreeding and genetic variability using dense SNP genotyping data.

This study is part of the activities of the GENORIP and aimed to investigate the genetic variability, the genomic inbreeding and the allele frequencies of hereditary monogenic traits in the females of seven Italian Holstein large dairy cattle herds.

## Materials and methods

### Ethics statement

The study was approved by the OPBA of the University of Milan (Protocol number 160\_2019), in

accordance with the Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010, updating Directive 86/609/EEC of 1986 on the protection of animals used for scientific purposes.

### ***Animal sampling, DNA extraction, quality control and genotyping***

A total of 3,953 Holstein cows were sampled from seven different herds located in the Lombardy region of Italy. These herds were chosen for their different sizes and management practices, in order to provide some examples of dairy cattle farms in the Lombardy region. The farms differ, in fact, in their structures and available technologies (such as milking parlours vs. milking robots) and size (120 milking heads vs 600 milking cows).

Animals were sampled using ear Tissue Sampling Units (TSU) for adult individuals and bioptic ear tags for newborn calves. The collected samples were then classified in a project structured database and stored at the University of Milan tissue repository Animal Bio-Arkive (Longeri et al. 2021).

The DNA was extracted from ear tissue using the Quick-DNA™ Miniprep Kit of Zymo Research according to the manufacturer's protocol (Zymo Research Corporation). Cows were genotyped with NEOGEN's GGP Bovine 100K, consisting of approximately 100,000 SNPs, with an average SNP spacing of about 29 kb.

All samples had a call rate value >95%. Only SNPs located on the 29 autosomes annotated according to the ARS-UCD1.2 bovine genome assembly (n. 89,762) were considered in this study to perform all analyses. In order to avoid bias, we excluded SNPs detecting the same mutation: they were more than 600 SNPs on the autosomes.

### ***Analysis of population structure***

Principal component analysis (PCA) was used to determine genetic diversity within and among herds and has been performed using SNP & Variation Suite (SVS) v8.9 (Golden Helix Inc., Bozeman, MT, USA). The 2-D graphical visualisation of PCA results was obtained using the 'ggplot2' R library (Wickham 2016).

### ***Genotype frequencies for health, phenotypic and productive traits***

The GGP Bovine 100K chip releases a large number of SNPs genotypes associated to mendelian hereditary

traits, such as genetic disorders and mutations related to phenotypic and productive traits or haplotypes linked to fertility traits. The allele and genotypic frequencies for these loci were estimated using an in-house R script.

### ***Runs of homozygosity detection***

Runs of Homozygosity (ROH) were obtained for each individual using the consecutive run method of the 'detectRUNS' library of R software (Biscarini et al. 2019). The parameters used were: (i) minimum number of 30 SNPs/ROH; (ii) a minimum length of 1 Mb for the identified ROH, to avoid the detection of short and common ROH across the genome due to Linkage Disequilibrium; (iii) a maximum distance of 1 Mb between consecutive SNPs to eliminate the bias in detection due to the density occurrence of SNPs; (iv) no missing SNPs as well as no heterozygous genotypes presence in ROH definition.

The ROH distribution per herd was evaluated separately using five classes of ROH length (<2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb and >16 Mb). Descriptive statistics relative to the total number of ROH, the ROH average number per individuals and, the average length of ROH were calculated.

The 'detectRUNS' library was also used to obtain: i) the graphical representations (Manhattan plots) for the percentage of occurrence of SNPs in ROH, estimated by counting the number of times that each SNP falls inside a ROH over the total number of individuals; ii) the ROH\_islands, identified as peaks in Manhattan plot where SNPs are inside a ROH in more than 50% (chosen threshold) of the cows as discussed and suggested by Schiavo et al. (2021).

### ***Gene annotation of ROH\_islands and functional analyses***

All ROH\_islands were annotated with the genes downloaded from the NCBI online Database (NCBI Annotation Release: 106). Only genes with an official gene name were considered. Database for Annotation, Visualisation, and Integrated Discovery (DAVID) v6.8 (DAVID online Database) was used to perform a gene ontology (GO) functional annotation and KEGG pathway analyses.

Additionally, the CattleQTLdb database (AnimalQTLdb) was used to identify – using the 'Search by associated gene' option – the QTL overlapping the ROH\_islands.

### Inbreeding coefficient

The genomic inbreeding coefficients ( $F_{ROH}$ ) based on ROH were calculated for each cow as:

$$F_{ROH} = \sum L_{ROH}/L_{AUTO} \quad (1)$$

where  $L_{ROH}$  is the total length of ROH proper of each individual genome, and  $L_{AUTO}$  is the total genome length covered by the used SNP dataset (2,487,916,500 bp in this study).  $F_{ROH}$  were calculated for each of the five classes of ROH length ( $L_{ROH}$ ) previously defined.

### Results and discussion

Table 1 reports the number of individuals sampled and genotyped in each farm, and the average gEBV for several traits released by ANAFIBJ for all genotyped females. All paternity and maternity consistency was verified based on the genome data by ANAFIBJ to solve inconsistencies due to incorrect genealogy, i.e. errors in sire or in maternal grandsire registration. The proportion of these inconsistencies varied from 8% to 45%.

#### Population structure

A first sight of the genetic variability of the 7 herds is provided by the graphical representation of PCA shown in Figure 1. As visualised in Figure 1, within herd PCAs, cows cluster clearly in separate groups in Herd\_3 (PCA1 = 15.88%, PCA2 = 13.94%), Herd\_4 (PCA1 = 15.02%, PCA2 = 12.63%), Herd\_6 (PCA1 = 19.67%, PCA2 = 14.44%), and to some extent also in Herd\_2 (PCA1 = 16.81%, PCA2 = 16.16%) and Herd\_5 (PCA1 = 13.11%, PCA2 = 12.40%), while it appears to exist more homogeneity among cows for Herd\_7 (PCA1 = 14.79%, PCA2 = 13.08%) and Herd\_1 (PCA1 = 15.24%, PCA2 = 12.22%). The cow clustering is expected to reflect the choices made by farmers in terms of use of sires for reproduction. More specifically it is likely that the variability shown by PCAs depends on the sire origin as system/country of selection scheme (e.g. USA, CAN, NLD, DEU, FRA, ITA) is mediated by the AI centres selling the semen to farmers. Discussion with farmers (partners of the project) on this topic disclosed a different approach in sire choice: some farmers rely on the technical advice (and semen) from a unique AI centre, some others select sires personally across all available on the market, also taking advantage of the information deriving from the mating plans offered by the farmer association. Only one, Herd\_4 is selecting sires based on a herd genomic selection on females already applied for some years. All gEBVs here reported

(Table 1) are based on the breeders' reproductive choices that were made without taking into account the genomic information of the females in the herd, with the exception of Herd\_4.

To provide a rationale for the cows' clustering in PCAs, we investigated the variability of sires used in farms as number of daughters from same sire/maternal grandsire (i.e. a bull being both sire and maternal grandsire in the same herd), within herd and among herds (Table 2).

In relation to the herd size, Herd\_3 and Herd\_6 use the lowest number of bulls: each sire has, in fact, an average 10.5 and 8.1 daughters, respectively. Herd\_7 is the one with the largest number of sires (3.2 daughters per sire, on average) accounting for the herd size. The large size of Herd\_3 somehow affects the possibility to use a large number of sires in the breeding plans, maintaining a high genetic level of the group of males: using a large number of sires to decrease the number of daughters per male, would in fact diminish the average genetic value of the reproducers. To avoid decreasing too much the genetic level of the bulls Herd\_3 accept to have larger groups of daughters per sire, if compared to other herds. Nevertheless, the 5 clusters visible for Herd\_3 in Figure 1 indicate that the farmer in selecting sires, in addition to the selection goal, was also paying attention to the genetic variability: the EBVs of the cows for PFT (selection index for Productivity, Functionality and Type) are in fact comparable to those of other herds, with a higher average value for milk gEBV (1,052 kg). On the contrary, the large spectrum of sires used (173 sires and 186 maternal grand sires, the largest in all herds), was widening the genetic variability, but only at a very low extent also reducing the average genetic values of used sires. It is noteworthy that Herd\_7 is the one with the lowest proportion of bulls being at the same time sires and maternal grand sires of females in the herd. This indicates a fast change in sires used in reproduction and a great attention to reduce genetic inbreeding. Herd\_4 has even fewer sires being also maternal grand sires, only 38%, an indication consistent with the applied selection plan of the farmer, based on genomic selection of females for several years. Herd\_4 farmer is using young sires of the most recent generations as much as possible; the selection is addressed to prioritise improvement of functional traits, with particular emphasis on fertility and longevity. The mating plan based on genomic information used by Herd\_4 appears very successful, both for the very good gEBVs of females, greater than all other herds for functional and production traits (Somatic Cell Count, Udder

**Table 1.** Sampling and milk destination of each considered herd together with descriptive statistics (minimum, maximum and mean) of gEBV (genomic EBV) values calculated by ANAFBJ.

	HERD_1	HERD_2	HERD_3	HERD_4	HERD_5	HERD_6	HERD_7
N. samples	455	263	1494	329	220	638	554
GPFT	1559/4474 (3045.87)	1373/3875 (2689.16)	1726/4543 (3436.17)	1726/4543 (3436.2)	1760/4346 (3093.47)	1608/4072 (3002.92)	1847/4324 (3181.58)
IES€	-387/1228 (388.59)	-451/944 (257)	-345/875 (288.62)	-199/1266 (668.22)	-310/1124 (416.91)	-447/1007 (312.27)	-256/1083 (421.58)
ICS-PR	-990/1469 (108)	-1117/870 (-69.05)	-1235/888 (-55.85)	-684/1411 (519.04)	-824/1510 (168.35)	-1122/1135 (70.77)	-749/1386 (207.93)
Milk (kg)	-1168/2514 (605.81)	-896/1872 (424.01)	-1211/2437 (725.55)	-870/2055 (751.66)	-1265/2060 (574.73)	-1144/2021 (580.94)	-964/2061 (634.58)
Fat (kg)	-46/107 (32.78)	-48/99 (12.27)	-64/110 (25.12)	-44/122 (45.6)	-45/116 (31.69)	-55/94 (26.62)	-33/114 (30.05)
Protein (kg)	-30/85 (27.77)	-33/70 (16.52)	-27/84 (29.54)	-41/83 (36.28)	-37/78 (29.44)	-36/75 (23.06)	-22/86 (30.07)
Fat (%)	-0.57/0.84 (0.09)	-0.53/0.7 (-0.04)	-0.58/0.65 (-0.02)	-0.72/0.8 (0.16)	-0.53/0.66 (0.09)	-0.61/0.82 (0.05)	-0.61/0.76 (0.06)
Protein (%)	-0.32/0.39 (0.06)	-0.19/0.32 (0.02)	-0.29/0.34 (0.04)	-0.23/0.46 (0.09)	-0.18/0.39 (0.09)	-0.39/0.41 (0.03)	-0.28/0.36 (0.07)
Feed efficiency	94/106 (99.61)	94/105 (98.92)	94/108 (101.33)	90/104 (98.52)	94/105 (99.39)	89/104 (97.23)	88/106 (99.02)
Type	-1.61/1.63 (0.11)	-1.67/1.63 (0.16)	-1.67/2.27 (0.51)	-1.52/1.73 (0.51)	-0.91/1.93 (0.59)	-1.29/2.75 (1.17)	-0.56/3.01 (0.95)
ICM	-1.91/2.82 (0.79)	-2.38/2.89 (0.28)	-2.22/2.82 (0.18)	-2.24/3.23 (0.89)	-1.67/2.77 (0.78)	-1.95/3.78 (1.47)	-1.15/4.32 (1.3)
Feet and length	-2.74/3.7 (0.44)	-3.26/2.64 (-0.11)	-3.51/2.81 (-0.18)	-2.91/3.69 (0.27)	-2.48/3.48 (0.3)	-2.07/3.83 (0.72)	-2.82/3.77 (0.89)
Somatic cell	92/113 (102.5)	93/112 (102.14)	89/115 (100.84)	91/121 (104.39)	90/116 (101.86)	91/116 (103.34)	88/120 (103.41)
Udder health	92/113 (102.38)	92/113 (102.34)	86/115 (103.91)	91/118 (103.91)	89/115 (102.58)	90/116 (102.55)	91/118 (103.21)
Longevity	91/118 (105.24)	94/117 (104.97)	91/117 (102.96)	93/121 (110.23)	95/118 (105.66)	94/117 (105.42)	94/118 (105.42)
Fertility	88/114 (102.06)	93/113 (101.92)	85/116 (100.37)	89/116 (105.03)	94/116 (102.54)	90/112 (101.01)	91/113 (102.4)
BCS	86/114 (98.18)	88/113 (99.65)	84/114 (98.81)	87/116 (99.28)	78/109 (97.28)	82/109 (96)	78/113 (97.67)

GPFT: Production, Functionality and Type genomic index; IES€: Indice economico salute (Health Economic Index); ICS-PR: Indice Caseificazione e Sostenibilità-Parmigiano Reggiano (Cheesemaking and Sustainability Index - Parmigiano Reggiano); ICM: Indice Composto Mammella (Udder Composite Index); BCS: Body Condition Score.

Health, Longevity and Fertility), the selection indexes (PFT, IES€, ICS-PR), and for the maintenance of genetic variability. Herd\_2 is under a very successful introduction of precision farming (Automatic Milking Systems), never used genomic selection at farm female level and is systematically applying the same breeding goal in the herd for the past several years. The impact of GENORIP on this farm was positive as it allowed the farmer a fast step forward in matching the technology available in the herd with the genomic information to manage female reproduction in the herd.

When we compared the number of common bulls across herds, this can contribute to the explanation why Herd\_6 is clustering separately from others: the number of sires in Herd\_6, when compared to others, is a maximum 15 in common with maximum 2 other farms, while others have in common 21 to 59 sires up to 4 herds. The mating plan in Herd\_6 is in fact fully relying on the technical advice of a unique AI centre with all bulls deriving from its selection program, while other herds acquire bulls from various AI centres.

### Genotype frequencies for productive traits

Regarding genes linked to milk production and quality (Table 3), the AA SNP mutation of  $\alpha$ -S1-Casein is close to 100% in all herds, while the GG SNP mutation wasn't identified: as a key to compare to other studies the SNP allele A mutation correspond to the B variant and the G SNP mutation correspond to the C variant as usually reported (Sanchez et al. 2020). In Holstein the effect of B- and C-variants for  $\alpha$ -S1-Casein were identified (Poulsen et al. 2013), with the B variant linked to increase in milk yield and the positive effect of C variant on curd coagulation time and curd firmness rate (Bovenhuis et al. 1992). The frequency of the B variant indicates that the genetic potential of our herds is for milk production as it is closely to be fixed in the population.

Also for the  $\beta$ -Casein locus the most frequent genotype is AA in all herds (ranging from about 84% to 95%), whilst the BB genotype was found with a low genotype frequency only in Herd\_1 (1.07%) and in Herd\_7 (1.10%).  $\beta$ -Caseins show numerous genetic variants that result in different quality characteristics in milk. The most common variants are A1 and A2. A1A1 is the less frequent genotype variant of  $\beta$ -Casein in all herds; while the A1A2 and A2A2 vary according with herds, ranging from 37.4% to 51.7% for A1A2 and from 28.8% to 58.0% for A2A2. The molecular difference between the two proteins is related to a



**Figure 1.** Graphical representation of PCA results both for each herd and for all individual together.

**Table 2.** Number of sire and maternal grandsires used as reproducers in each herd and shared among them.

Herd_n	N. Sires	N. min -max of daughters per sire in the herd (average)	N. maternal grandsires	N. sires also maternal grand sire* (%)	N. of common sires (%) in N. of Herds (min -max)	N. of daughters of common sires in other herds (average)	N. daughters of common sires across all herds (average)
		Values within herds			Values among herds		
Herd_1	101	1–16 (4.5)	121	75 (74)	59 (58.0) – (1–4)	1–81 (7.2)	2–85 (14.9)
Herd_2	50	1–35 (5.3)	67	34 (68)	23 (46.0) – (1–3)	1–31 (7.8)	3–52 (17.9)
Herd_3	142	1–81 (10.5)	159	104 (73)	38 (26.8) – (1–4)	1–21 (4.4)	3–85 (22.4)
Herd_4	53	1–30 (6.2)	82	20 (38)	22 (41.0) – (1–4)	1–63 (7.6)	2–65 (18.5)
Herd_5	56	1–19 (3.9)	72	41 (73)	21 (37.5) – (1–4)	1–48 (7.8)	2–53 (16.8)
Herd_6	79	1–117 (8.1)	98	59 (75)	15 (18.8) – (1–2)	1–16 (4.4)	2–31 (11.3)
Herd_7	173	1–14 (3.2)	186	90 (52)	56 (32.4) – (1–4)	1–81 (8.0)	2–85 (14.6)

\*N. of sires that are maternal grandsire of another female. Proportion calculated as (n. sires also maternal grandsires)/(n. sires).

mutation resulting in an amino acid change (proline vs histidine) at position 67 of  $\beta$ -Casein (Ginger and Grigor 1999). The amino acid change was associated with a different gastric digestion of caseins. Indeed, during the enzymatic digestion of A1 casein, an opioid peptide (BCM-7) is released, which is not released in the digestion of A2 variants (Brooke-Taylor et al. 2017). In recent years, there has been an increased focus on the  $\beta$ -Casein A2 allele as some studies have

suggested that the  $\beta$ -Casein A2 allele is better tolerated by the human population (He et al. 2017). To date, however, no relationship has been found between the consumption of cow's milk with the A1 allele for  $\beta$ -Casein and disease incidence. In addition, the A1 variant improves rennet coagulation properties compared to the A2 variant (Dinc et al. 2013; Ketto et al. 2017). Interest in marketing dairy products, with improved health impact, has opened the market to

**Table 3.** Genotype frequencies for productive traits (i.e. milk protein and meat variant of monogenic related loci) in each herd.

Gene	Genotypes/mutation at SNP level	Herd_1	Herd_2	Herd_3	Herd_4	Herd_5	Herd_6	Herd_7
<b>Milk trait related loci</b>								
$\alpha$ -S1-Casein	BB (AA)	98.90	100	99.87	100	99.50	100	99.80
	BC (AG)	1.10	–	0.13	–	0.50	–	0.20
	CC (GG)	–	–	–	–	–	–	–
$\beta$ -Casein BB	AA	93.41	84.03	96.79	93.92	95.90	93.26	93.50
	AB	6.37	15.97	3.21	6.08	4.10	6.74	6.30
	BB	1.07	–	–	–	–	–	1.10
$\beta$ -Casein variants*	A1A1	13.41	20.15	12.58	15.20	6.82	4.70	13.92
	A1A2	44.62	51.71	48.33	51.37	39.09	37.77	48.64
	A2A2	41.98	28.14	39.09	33.43	54.09	57.52	37.43
$\beta$ -Lactoglobulin	AA	35.82	22.81	33.22	40.73	50.00	37.80	43.86
	AB	51.21	50.95	50.03	46.81	40.00	47.72	42.42
	BB	12.97	26.24	16.74	12.46	10.00	14.49	13.72
K-Casein	AA	17.58	36.12	29.74	14.29	16.82	10.03	18.95
	AB	42.86	39.16	36.97	39.82	50.45	48.12	39.35
	AE	10.99	6.46	11.52	6.99	5.45	5.17	10.83
	BB	17.14	12.93	12.12	26.75	25.00	29.00	19.49
	BE	9.89	5.32	8.51	10.33	2.27	7.21	10.65
Lactoferrin	EE	1.54	–	1.14	1.82	–	0.47	0.72
	AA (AA)	54.73	33.84	50.94	67.48	42.73	46.39	55.78
	AB (AG)	39.12	53.61	40.70	28.27	48.18	44.36	38.45
YellowFat	BB (GG)	6.15	12.55	8.37	4.26	9.09	9.25	5.78
	GG	100	100	99.53	99.09	100	100	99.64
	AG	–	–	0.47	0.91	–	–	0.36
	AA	–	–	–	–	–	–	–
<b>Meat traits related loci</b>								
Calpain_316	CC	36.92	19.77	28.67	34.95	34.09	26.33	37.36
	CG	49.45	51.33	51.11	46.20	46.36	50.47	47.83
	GG	13.63	28.90	20.23	18.84	19.55	23.20	14.80
Calpain_4751	CC	52.97	56.65	53.68	51.67	61.82	56.43	56.32
	CT	41.10	36.50	39.02	39.82	34.55	38.24	38.99
	TT	5.93	6.84	7.30	8.51	3.64	5.33	4.69
CAST_2870	AA (AA)	20.66	9.89	19.08	22.19	22.37	28.68	14.08
	AB (AG)	49.89	44.11	50.67	49.85	47.95	49.06	45.85
	BB (GG)	29.45	46.01	30.25	27.96	29.68	22.26	40.07
CAST_2959	AA (AA)	34.73	50.57	43.11	42.55	39.27	27.74	46.57
	AB (AG)	48.57	41.44	44.91	43.47	45.66	49.06	43.68
	BB (GG)	16.70	7.98	11.98	13.98	15.07	23.04	9.75
UoGCAST1	CC	27.21	35.88	35.37	29.18	28.90	23.38	37.64
	CG	49.12	51.53	47.11	48.63	48.17	47.87	48.18
	GG	23.67	12.60	17.52	22.19	22.94	28.75	14.18

\*Genotype frequencies at marker CSN2\_7 (the same frequencies were obtained for CSN2\_X14711\_8101 locus).

milk selected for its  $\beta$ -Casein A2 content only (Mendes et al. 2019). As Table 2 shows, it is evident that two herds, in particular Herd\_5 and Herd\_6, have a proportion of A2A2 genotypes >50%. These breeders, in fact, select for this genotype while the others generally have higher values for the heterozygous A1A2 genotype. The higher proportion of heterozygous genotype variants is in agreement with those reported by other authors for Holstein breed (Massella et al. 2017).

For  $\beta$ -Lactoglobulin, depending on the herd, the AA or AB genotypes are the most frequent genotypes: i.e. the AA genotype frequencies is higher in Herd\_5 and Herd\_7 (50% and 43.86%, respectively); in all other herds the higher frequency has been found for the AB genotype (ranging from 46.81% in Herd\_4 to 50.95% in Herd\_2). The  $\beta$ -Lactoglobulin is the major serum protein in cow's milk, accounting for about 50% the total amount of milk proteins and the B variant has been associated with a higher casein content,

resulting in a higher cheese yield (van den Berg et al. 1992; Stasio and Mariani 2000).

The frequencies of the six K-Casein genotypes (AA, AB, AE, BB, BE, and EE) had the same pattern in all herds: the most represented genotype was AB, ranging from about 36.97% in Herd\_3 to 50.45% in Herd\_5. Instead, the EE genotype has been registered in a very low number of cows (not one EE was found in Herd\_2 and Herd\_5) (Table 3). Studies in the literature show that in Holstein both A and B alleles are the most frequent (Prinzenberg et al. 1999; Farrell et al. 2004) and the E allele the least frequent (Caroli et al. 2000). In fact, negative effects on coagulum formation during cheesemaking have been observed in milk produced by individuals carrying the E allele variant (Caroli et al. 2000; Comin et al. 2008). As Table 3 shows, heterozygous are generally the most widespread, while the BB variant, which has intermediate values, positively influences the production of cheeses,

such as Parmigiano Reggiano and Grana Padano, by increasing cheese yields, as shown in the study of Mariani et al. (Mariani et al. 1976).

The heterozygous cows (AG) at marker linked to milk YellowFat feature, are still present with very low frequencies in Herd\_3 (0.47%), Herd\_4 (0.91%) and Herd\_7 (0.36%). The AA genotypes causes a characteristic yellow colour of fat in tissues and milk, due to carotenoids depositions in adipose tissue. (Yang et al. 1992).

Finally, for the Holstein cows here analysed, at Lactoferrin locus, we found a similar distribution of the three genotype frequencies across herds, with a higher frequency of AA, mainly in Herd\_4 (about 67%). AA genotype was associated with a low milk SCC values (Wojdak-Maksymiec et al. 2006).

Regarding meat traits, we observed high variability at all analysed loci (Table 3), as the Holstein breed is not selected for these traits.

### **Genotype frequencies for reproductive traits and disease**

Table 4 shows the genotype frequencies of genes and haplotypes influencing bovine fertility. Animals bearing mutations affecting reproduction efficiency were found for all haplotypes, with different frequencies in the seven herds. The haplotype HH5 was the one with the largest number of carriers in all herds, with about 10% of females being carriers in Herd\_7. Instead, high carrier haplotype frequencies were found in some herds only for specific HH. HH4 carriers are counted only for Herd\_3 and Herd\_7. The genotype frequencies for COQ9-rs109301586, STAT3 and 5, Leptin\_2F, and PKP2\_988 markers loci (all mainly involved in embryo development) are similar across all herds, counting a higher proportion of one or both homozygous genotypes, except for STAT3\_25402 and STAT5\_13319, for which the most frequent genotype was BB (Table 4).

In this study, the highest carrier frequency was found for HCD (haplotype cholesterol deficiency), that represents an economic loss for the farmer. Animals homozygous for this disease show the first clinical signs between 1 and 5 months of age with decreased appetite, weight loss, diarrhoea and subsequent death with a frequency in the German Friesian population of 4.2% (Kipp et al. 2016). In our study only one animal from Herd\_3 was affected. In previous study, it was shown that animals carrying HCD had significantly higher protein yields than non-carriers, but it is still unclear how HCD affects cheese yields (Cole et al. 2016).

The presence of a high proportion of carrier haplotypes/SNPs may be linked to the breeders'

choice to concentrate on the selection of productive traits, which are negatively correlated with certain reproductive traits, or the use of bulls carrying some haplotype. Some of these haplotypes influence heifer conception rate, cow conception rate, milk, and protein (Cole et al. 2016). The increased consideration of these haplotypes in cattle selection criteria could lead breeders to a gradual improvement in herd fertility, reducing the losses associated with it.

Regarding genetic diseases, this study has shown the presence of carriers with low frequencies (ranging from 0.07 to 4.28, Table 3) for BLAD, GSDV and RP1 in all herds. Higher carrier frequencies were observed for Brachyspina (up to 7.22% in Herd\_2).

The proportion of BLAD-carrier animals ranged from 0.07 to 0.66 which, compared with other studies such as in Brazilian Holsteins that found a carrier frequency of 5.7% (Ribeiro et al. 2000), is very low.

Avoiding carrier-to-carrier mating, and thus identifying heterozygous cows, would be a way of managing the reproduction and presence of female carriers in the herd.

### **Runs of homozygosity detection**

A total of 458,267 ROH was identified in all cows of the seven herds. The count of ROH (per herd) reflects the size of herd sampling (correlation = 0.997). At the individual level, the average number of ROH ranged from 7 (Herd\_1) to 251 (Herd\_3), with a similar total mean ROH length close to 2.6 Mb, except for Herd\_5 that had on average longer ROH, close to 3 Mb (Table 5, Figure 2A). Herd\_2 showed both the lowest mean number of ROH per individual (106.8) and lowest total genome length (average value) covered by ROH (11.3%) (Table 5). Differences among cows were identified also considering the total length of the genome covered by the ROH (sum of all ROH per animal, Figure S1).

Same selection occurring across herds may have affected same regions of the genome; at genomic level some evidence may be related to the fact that across herds ROH are found in largely overlapping genomic regions among females of different herds. Over the years selection may have affected same regions where genes involved in expression of traits under selection are annotated (Zhang et al. 2015).

The ROH were found for all classes of length (Figure 2B), with shorter regions (<2 Mb), being the most frequent classes of length (about 50%), even if this proportion may be slightly overestimated according to results from (Ferenčaković et al. 2013), study however based on a 50K SNP chip and not on a 100K SNP chip array. Contrariwise, a small number of ROH



**Table 4.** Genotype frequencies calculated for genes (haplotype) affecting fertility traits and genetic diseases in each herd.

Gene haplotype (HH)*	Genotypes at SNP level	Herd_1	Herd_2	Herd_3	Herd_4	Herd_5	Herd_6	Herd_7
<b>Fertility related traits loci</b>								
COQ9 rs109301586	AA (AA)	24.84	27.76	16.80	24.01	17.27	10.50	19.89
	AB (AG)	45.05	52.85	48.46	51.37	50.91	47.02	45.21
	BB (GG)	30.11	19.39	34.74	24.62	31.82	42.48	34.90
HH1	Free	96.26	100	97.59	100	93.18	97.34	93.50
	Carrier	3.74	–	2.41	–	6.82	2.66	6.50
HH3	Free	98.24	92.78	93.17	96.96	97.73	99.37	94.58
	Carrier	1.76	7.22	6.83	3.04	2.27	0.63	5.42
HH4	Free	100	100	96.72	100	100	100	99.09
	Carrier	–	–	3.28	–	–	–	0.91
HH5	Free	93.61	94.30	93.37	92.40	94.55	91.22	89.71
	Carrier	6.39	5.70	6.63	7.60	5.45	8.78	10.29
HH6	AA (AA)	99.12	99.62	97.39	99.39	93.64	100	99.10
	AB (AG)	0.88	0.38	2.61	0.61	6.36	–	0.90
HH7 CENPU	AA (DD)	–	–	–	–	–	–	–
	AB (DI)	0.20	0.40	0.40	–	–	0.20	0.40
	BB (II)	99.80	99.60	99.60	100	100	99.80	99.60
Leptin_2FB	CC	50.77	39.54	33.27	36.17	37.27	41.07	38.45
	CT	40.44	47.91	48.59	52.58	42.73	47.18	46.39
	TT	8.79	12.55	18.14	11.25	20.00	11.76	15.16
STAT3_19069	AA (TT)	17.36	23.95	16.33	14.29	15.91	15.99	13.18
	AB (TC)	48.13	52.85	48.39	49.54	52.73	48.75	44.95
	BB (CC)	34.51	23.19	35.27	36.17	31.36	35.27	41.88
STAT3_25402	AA (TT)	7.69	8.37	7.76	3.34	4.09	7.05	5.60
	AB (TG)	37.36	44.11	37.62	34.65	29.55	45.61	35.74
	BB (GG)	54.95	47.53	54.62	62.01	66.36	47.34	58.66
STAT5_13244	AA (TT)	14.07	14.83	18.14	7.29	7.27	19.75	12.45
	AB (TC)	47.25	50.19	47.86	44.68	41.82	54.55	45.85
	BB (CC)	38.68	34.98	34.00	48.02	50.91	25.71	41.70
STAT5_13319	AA (TT)	2.20	3.42	2.01	2.43	5.45	1.25	1.62
	AB (TG)	25.33	33.08	23.90	31.61	35.91	18.50	21.12
	BB (GG)	72.47	63.50	74.10	65.96	58.64	80.25	77.26
PKP2_988	AA (I/I)	41.32	31.68	37.51	31.00	45.91	57.77	43.14
	AB (I/D)	44.84	51.15	48.96	52.89	46.82	35.95	45.85
	BB (D/D)	13.85	17.18	13.53	16.11	7.27	6.28	11.01
<b>Mendelian heritable disease loci</b>								
BLAD	Free	99.34	100	99.93	100	100	99.84	100
	Carrier	0.66	–	0.07	–	–	0.16	–
	Affected	–	–	–	–	–	–	–
Brachyspina	Free	97.58	92.78	97.59	98.78	100	96.55	99.64
	Carrier	2.42	7.22	2.41	1.22	–	3.45	0.36
	Affected	–	–	–	–	–	–	–
GSDV	Free	99.78	100	99.8	100	100	100	100
	Carrier	0.22	–	0.2	–	–	–	–
	Affected	–	–	–	–	–	–	–
HCD	Free	90.99	98.10	92.97	98.48	92.73	92.95	94.22
	Carrier	9.01	1.90	6.96	1.52	7.27	7.05	5.78
	Affected	–	–	1	–	–	–	–
RP1	AA (I/I)	–	–	–	–	–	–	–
	AB (I/D)	1.76	0.38	4.28	–	0.50	–	–
	BB (D/D)	98.24	99.62	95.72	100	99.50	100	100

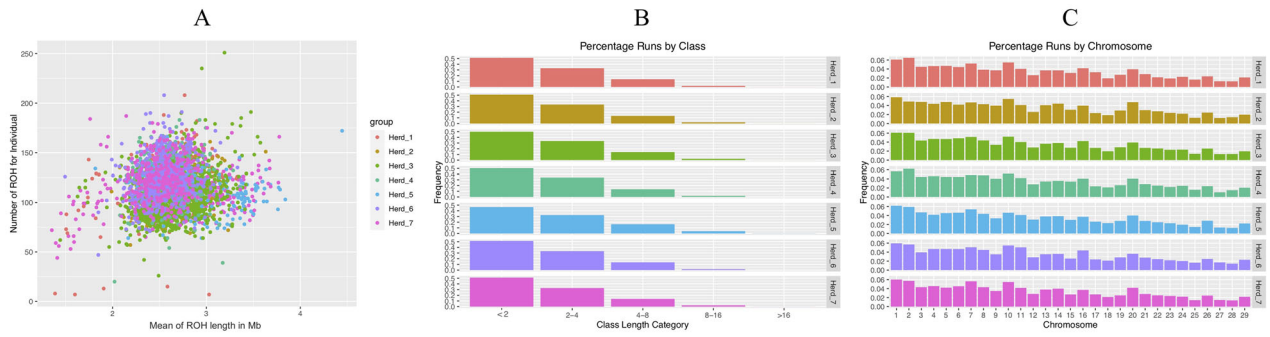
\*Coenzyme Q9 (COQ9) (G→A; rs109301586); Holstein Haplotype 1 (HH1); Holstein Haplotype 3 (HH3); Holstein Haplotype 4 (HH4); Holstein Haplotype 5 (HH5); Holstein Haplotype 6 (HH6); Holstein Haplotype 7 (HH7 CENPU); Leptin\_2FB; Brachyspina; Bovine Leukocyte Adhesion Deficiency (BLAD); Holstein Cholesterol Disorder (HCD); Glycogen storage disease type V (GSDV); Progressive retinal degeneration (RP1); Signal transducer and activator of transcription (STAT); plakophilin 2 (PKP2\_988). I = insertion; D = Deletion.

**Table 5.** Runs of Homozygosity (ROH) descriptive statistics.

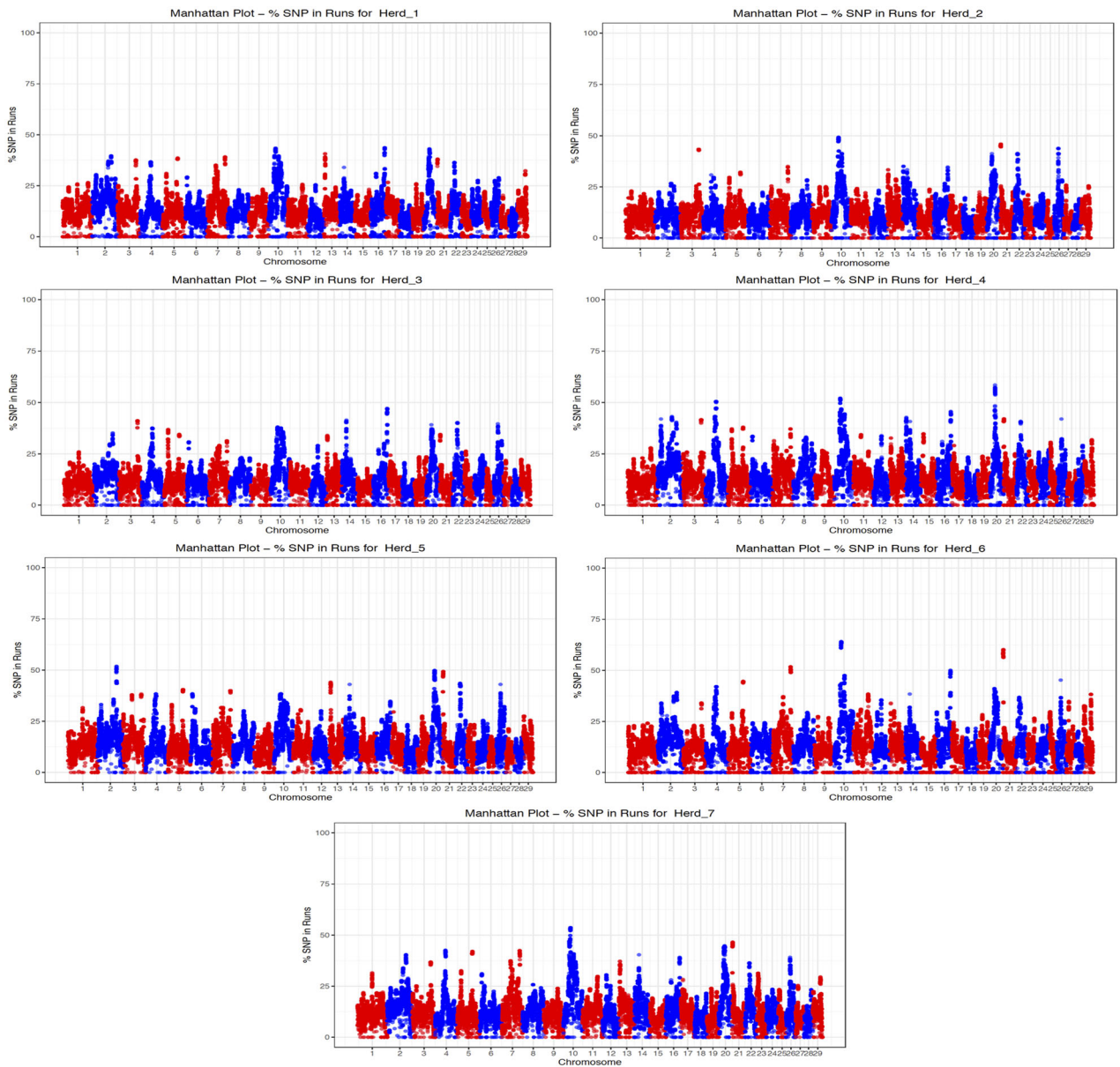
Herd	Total ROH	Min-Max (mean) n. ROH	Min ROH length	Max ROH length	Mean ROH length	Mean coverage <sup>1</sup> (%) <sup>2</sup>
Herd_1	53,513	7–208 (117.6)	1,000,172	33,248,436	2,659,912	312,834,920 (12.6)
Herd_2	28,094	57–174 (106.8)	1,000,172	23,591,055	2,630,000	280,940,043 (11.3)
Herd_3	164,064	26–251 (109.8)	1,000,014	45,522,015	2,685,399	294,89,7811 (11.8)
Herd_4	41,637	20–188 (126.6)	1,000,172	26,347,829	2,677,836	338,896,839 (13.6)
Herd_5	24,913	64–172 (113.2)	1,000,172	33,218,333	2,967,912	336,089,051 (13.5)
Herd_6	79,827	47–208 (125.1)	1,000,124	26,861,540	2,586,512	323,626,185 (13.0)
Herd_7	66,219	53–187 (119.5)	1,000,039	29,616,394	2,719,103	325,011,350 (13.1)

<sup>1</sup>Mean Coverage = average calculated by Herds on total ROH length cover each sample's genome.

<sup>2</sup>Proportion of genome covered by ROH (Mean Coverage) on 2,487,916,500 bp (Genome length covered by the 89,762 SNP). ROH lengths are expressed in base pairs (bps).



**Figure 2.** Graphical representation of ROH statistics per herd: (A) relationship between number and averaged total length (Mb) of ROH proper of each cow; (B) frequencies of ROH for each class of length together with details on the > 16 Mb ROH class of length.



**Figure 3.** Manhattan plot of the proportion of SNPs in identified Runs of Homozygosity (ROH), along all the autosomes, for all analysed Herds.

**Table 6.** Runs of Homozygosity (ROH)\_Islands detected and shared in at least 50% of cows together with the annotated genes annotated and associated traits.

Herd_n	chrom	nSNP	From	To	Genes	Cattle QTL database
Herd_5	2	7	100503013	100668079	ERBB4	
Herd_4	4	29	54647672	55649011	GPR85, BMT2, TMEM168	
Herd_6	7	33	94623898	95679696	FAM81B, TTC37, ARSK, GPR150, RFESD, SPATA9, RHOBTB3, GLRX, ELL2	
Herd_6	7	3	95832919	95890433	–	
Herd_4	10	58	31539600	32997552	C10H15orf41*, MEIS2	Bone quality, feet and leg conformation, foot angle
Herd_6	10	77	36261135	38520889	RAD51*, RMDN3,	Body depth, bovine
Herd_7	10	72	36420875	38520889	GCHFR, DNAJC17,	respiratory disease
Herd_2	10	28	36459819	37304540	C10H15orf62, ZFYVE19, PPP1R14D, SPINT1, RHOV, VPS18, DLL4, CHAC1, INO80, EXD1, CHP1, OIP5, NUSAP1*, NDUFAF1, RTF1*, ITPKA, RPAP1*, TYRO3, MGA, MAPKBP1*, JMJD7, PLA2G4B, SPTBN5, EHD4, PLA2G4E, PLA2G4D, PLA2G4F*, VPS39, TMEM87A, GANC*, CAPN3*, ZNF106, SNAP23, LRRC57, HAUS2, STARD9, CDAN1, TTBK2, UBR1*, TMEM62, CCNDBP1, EPB42	ease, dairy form, daughter pregnancy rate, feet and leg conformation, first service conception, inseminations per conception, length of productive life, milk protein percentage, net merit, PTA type, quality grade, rear leg placement - rear view, rump width, shear force, somatic cell score, stature, stillbirth, strength, teat length, teat placement, teat placement – rear, udder attachment, udder cleft, udder depth, udder height
Herd_4	20	42	26292282	27490668	PELO	Milk protein percentage
Herd_5	20	30	26591866	27490668		
Herd_6	21	46	497917	2416354	MKRN3, MAGEL2, NDN, SNRPN, SNURF, UBE3A	

\*Genes associated with traits reported in 'Cattle QTL Database' column.

longer than 16 Mb were mapped in all herds (observed frequencies ranging from 0.05% – Herd\_5 to 0.30% – Herd\_5) with a maximum of two ROH longer than 16 Mb per individual.

Finally, ROHs were found over all chromosomes: there was no evidence of a relationship between the chromosome's length and mean ROH length, as shown in Figure 2C.

The graphs in Figure 3 shows the proportion of SNP in ROH segments across all the autosomes (Manhattan Plots) for all Herds.

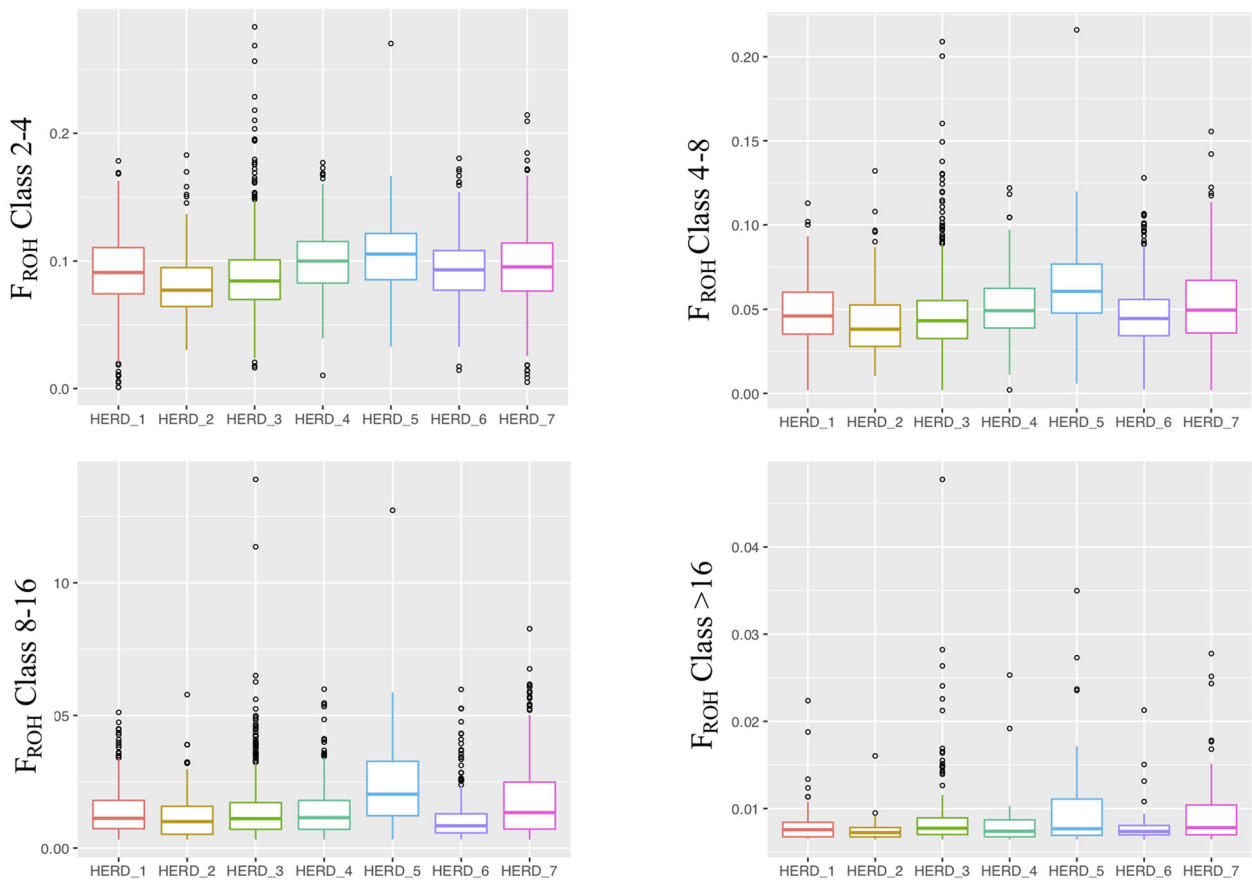
ROH\_islands (i.e. SNP within ROH with frequency value greater than 50% as herein before defined) were detected for the Herds 2, 4, 5, 6, and 7 and are listed in Table 6. On chr 7 two very close regions were identified for Herd\_6. ROH\_islands identified on chr 10 and 20 are identified in three and two herds, respectively. These two genomic regions overlap to that identified in other Holstein cows bred in Italy (Mastrangelo et al.

2018). All ROH\_islands except one mapping on chr 7, harboured annotated genes (n. 68).

According to the Animal QTL database, the genes lying within the ROH\_Island located on chr 10 (region shared by cows of three different herds) are mainly associated with reproduction traits (e.g. fertility traits) and morphology traits (e.g. Udder and Conformation traits). Among the genes annotated within the ROH\_Island reported in Table 6, the *ERBB4* and *MKRN3* genes affects udder and fertility traits. In detail, *ERBB4* was identified as the hub gene of the network that regulates udder growth and development and seems to affect the genes' expression that are involved in the udder involution and that promote mammary gland remodelling (Xuan et al. 2022), whereas *MKRN3* controls the initiation of puberty (Abreu et al. 2015) and inhibits the reproductive axis (Abreu et al. 2020). The ROH region on chr 20 is under selection in Herd\_4 and Herd\_5 and includes the *PELO* gene. This region

Table 7. Descriptive statistics Total  $F_{ROH}$  inbreeding coefficient.

Herd_n	$F_{ROH}$ Min	$F_{ROH}$ Max	$F_{ROH}$ Mean	$F_{ROH}$ SD
Herd_1	0.004	0.231	0.126	0.032
Herd_2	0.050	0.220	0.113	0.027
Herd_3	0.026	0.325	0.119	0.029
Herd_4	0.016	0.218	0.136	0.029
Herd_5	0.056	0.307	0.135	0.028
Herd_6	0.035	0.214	0.130	0.027
Herd_7	0.033	0.252	0.131	0.030



**Figure 4.** Descriptive statistics of total  $F_{ROH}$  (Table 7) and graphical representations (Boxplots) of  $F_{ROH}$  calculated in concordance with the five Runs of Homozygosity (ROH) classes of length.

has been recently found under selection also in US Holstein and Jersey by (Lozada-Soto et al. 2022) who characterised the ROH in several dairy cattle populations. In their study this ROH region resulted wider respect to the one here identified.

### Inbreeding coefficient

As reported in Table 7, the  $F_{ROH}$  values varied from 0.004 (Herd\_1) to 0.325 (Herd\_3), with an overall average value of  $F_{ROH}$  ranging from 0.113 to 0.136. These

values are comparable with the genomic inbreeding calculated in the US Holstein by Lozada-Soto et al. (2022) and in Italian Holstein as reported by Dadousis et al. (2022). As shown in Figure 4, the distribution of the  $F_{ROH}$  values calculated per each class of ROH length differed among the herds, ranging from 0.113 (Herd\_1) to 0.136 (Herd\_4). For the two greater classes of ROH, representing the most recent genomic inbreeding, the average values (per farm) were between 0.012 and 0.023 for ROH of 8–16 Mb and 0.008 and 0.010 for ROH > 16 Mb (Table S2). We want to highlight that the

maximum proportion of cows with ROH > 16 Mb was identified in Herd\_5 (25.5%). In other herds this proportion is lower than 17% with a minimum value in Herd\_3 of 5.7%. The inbreeding coefficients here obtained for each class of length were lower (except for the class of length <2 Mb) than calculated in Italian Holstein breed (Ablondi et al. 2022). Considering this overview, we can easily deduct that the inbreeding is taken under control in all farms in the last decades by the farmers, applying breeding strategies aimed to maintain genetic diversity among cows.

## Conclusions

In the last decade, genomic selection has been very successful and rapidly adopted in the genetic improvement plans of large dairy cattle populations, such as the Italian Friesian breed. The introduction of genomic selection in selection schemes made the improvement for low heritable traits, as functional or health traits, much more efficient. In addition, the availability of SNP genotypes also on females is making it possible for farmers to customise herd breeding goals by implementing efficient selection, especially for functional and health traits, and develop comprehensive mating plans that exploit all the information available to the breeder.

The use of SNP genotypes on females can be extended to optimise the herd mating plans also to manage herd genetic variability, control inbreeding at a genomic level and for specific selection for mendelian monogenic traits.

In this study, the analysis of genotypes produced by the GENORIP project provided a snapshot of the genetic variability, of the genomic inbreeding, as well as the presence of mendelian genetic variants linked to traits of interest in seven Holstein dairy herds for a total of 3,953 animals. A particularity of the project was to genotype all females present in the herds and the newborn female calves along the three years of the project duration, allowing as such to validate all the genealogical information. In addition, the availability of gEBV for all females makes it possible to evaluate the selection program adopted by farmers based on the sire side.

The knowledge of both milk properties and carriers of unfavourable traits is useful to farmers who process their milk into cheese or regarding payment according to the protein composition of the milk, and to implement mating plan, respectively.

## Authors' contributions

CP: collected the samples; MGS and AB: Conceived and designed the study; RM: performed the experiments, CP and

MGS: data analysis; AD and FB: participated to data analysis; CP, MGS and AB: wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This project was funded by EAFRD Rural Development Program 2014–2020, Management Authority Regione Lombardia - OP. 16.1.01 Project ID n. 201801062430 – 'Operational Group EIP AGRI' <https://ec.europa.eu/eip/agriculture/en/eip-agriprojects/projects/operational-groups>.

## Data availability statement

The data supporting the conclusions of this manuscript are included in the [Supplementary Materials](#).

## References

- Ablondi M, Sabbioni A, Stocco G, Cipolat-Gotet C, Dadousis C, Kaam J-Tv, Finocchiaro R, Summer A. 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. doi: [10.3389/fvets.2021.773985](https://doi.org/10.3389/fvets.2021.773985).
- Abreu AP, Macedo DB, Brito VN, Kaiser UB, Latronico AC. 2015. A new pathway in the control of the initiation of puberty: the MKRN3 gene. *J Mol Endocrinol.* 54(3):R131–R139. doi: [10.1530/JME-14-0315](https://doi.org/10.1530/JME-14-0315).
- Abreu AP, Toro CA, Song YB, Navarro VM, Bosch MA, Eren A, Liang JN, Carroll RS, Latronico AC, Rønnekleiv OK. 2020. MKRN3 inhibits the reproductive axis through actions in kisspeptin-expressing neurons. *J Clin Invest.* 130(8):4486–4500.
- ANAFIBJ – National statistics. 2022. Accessed Oct 18. <http://www.anafi.it/it/pubblicazioni-statistiche/statistiche-nazionali>.
- Andersson L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet.* 2(2):130–138. doi: [10.1038/35052563](https://doi.org/10.1038/35052563).
- Biscarini F, Cozzi P, Gaspa G, Marras G. 2019. detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. R package version 0.9.5. <https://CRAN.R-project.org/package=detectRUNS>.
- Boichard D, Ducrocq V, Fritz S. 2015. Sustainable dairy cattle selection in the genomic era. *J Anim Breed Genet.* 132(2): 135–143. doi: [10.1111/jbg.12150](https://doi.org/10.1111/jbg.12150).
- Bovenhuis H, Van Arendonk JAM, Korver S. 1992. Associations between milk protein polymorphisms and milk production traits. *J Dairy Sci.* 75(9):2549–2559. doi: [10.3168/jds.S0022-0302\(92\)78017-5](https://doi.org/10.3168/jds.S0022-0302(92)78017-5).
- Brooke-Taylor S, Dwyer K, Woodford K, Kost N. 2017. Systematic review of the gastrointestinal effects of A1 compared with A2  $\beta$ -casein. *Adv Nutr.* 8(5):739–748. doi: [10.3945/an.116.013953](https://doi.org/10.3945/an.116.013953).
- Caroli A, Bolla P, Budelli E, Barbieri G, Leone P. 2000. Effect of  $\kappa$ -casein E allele on clotting aptitude of Italian Friesian milk. *Zootecnica e Nutrizione Animale.* 26(3):127–130.

- Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nat Rev Genet.* 10(11):783–796. doi: [10.1038/nrg2664](https://doi.org/10.1038/nrg2664).
- Cole JB, Null DJ, VanRaden PM. 2016. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. *J Dairy Sci.* 99(9):7274–7288. doi: [10.3168/jds.2015-10777](https://doi.org/10.3168/jds.2015-10777).
- Collins FS, Guyer MS, Charkravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. *Science.* 278(5343):1580–1581. doi: [10.1126/science.278.5343.1580](https://doi.org/10.1126/science.278.5343.1580).
- Comin A, Cassandro M, Chessa S, Ojala M, Dal Zotto R, de Marchi M, Carnier P, Gallo L, Pagnacco G, Bittante G. 2008. Effects of composite  $\beta$ - and  $\kappa$ -casein genotypes on milk coagulation, quality, and yield traits in Italian Holstein cows. *J Dairy Sci.* 91(10):4022–4027. doi: [10.3168/jds.2007-0546](https://doi.org/10.3168/jds.2007-0546).
- Dadousis C, Ablondi M, Cipolat-Gotet C, van Kaam J-T, Marusi M, Cassandro M, Sabbioni A, Summer A. 2022. Genomic inbreeding coefficients using imputed genotypes: assessing different estimators in Holstein-Friesian dairy cows. *J Dairy Sci.* 105(7):5926–5945. doi: [10.3168/jds.2021-21125](https://doi.org/10.3168/jds.2021-21125).
- Dickerson G, Hazel LN. 1944. Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. *J Agric Res.* 69(2):459–476.
- Dinc H, Ozkan E, Koban E, Togan I. 2013. Beta-casein A1/A2, kappa-casein and beta-lactoglobulin polymorphisms in Turkish cattle breeds. *Arch Anim Breed.* 56(1):650–657. doi: [10.7482/0003-9438-56-065](https://doi.org/10.7482/0003-9438-56-065).
- Doekes HP, Veerkamp RF, Bijma P, De Jong G, Hiemstra SJ, Windig JJ. 2019. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. *Genet Sel Evol.* 51(1):54. doi: [10.1186/s12711-019-0497-z](https://doi.org/10.1186/s12711-019-0497-z).
- Egger-Danner C, Cole JB, Pryce JE, Gengler N, Heringstad B, Bradley A, Stock KF. 2015. Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal.* 9(2):191–207. doi: [10.1017/S1751731114002614](https://doi.org/10.1017/S1751731114002614).
- Engelsma KA, Veerkamp RF, Calus MPL, Bijma P, Windig JJ. 2012. Pedigree-and marker-based methods in the estimation of genetic diversity in small groups of Holstein cattle. *J Anim Breed Genet.* 129(3):195–205. doi: [10.1111/j.1439-0388.2012.00987.x](https://doi.org/10.1111/j.1439-0388.2012.00987.x).
- Falconer DS, Mackay TFC. 1983. *Quantitative genetics*. London: Longman.
- Farrell HM, Jr, Jimenez-Flores R, Bleck GT, Brown EM, Butler JE, Creamer LK, Hicks CL, Hollar CM, Ng-Kwai-Hang KF, Swaisgood HE. 2004. Nomenclature of the proteins of cows' milk—Sixth revision. *J Dairy Sci.* 87(6):1641–1674. doi: [10.3168/jds.S0022-0302\(04\)73319-6](https://doi.org/10.3168/jds.S0022-0302(04)73319-6).
- Ferenčaković M, Sölkner J, Curik I. 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol.* 45(1):42. doi: [10.1186/1297-9686-45-42](https://doi.org/10.1186/1297-9686-45-42).
- Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics.* 139(2):907–920. doi: [10.1093/genetics/139.2.907](https://doi.org/10.1093/genetics/139.2.907).
- Ginger MR, Grigor MR. 1999. Comparative aspects of milk caseins. *Comp Biochem Physiol B Biochem Mol Biol.* 124(2):133–145. doi: [10.1016/S0305-0491\(99\)00110-8](https://doi.org/10.1016/S0305-0491(99)00110-8).
- Hayes BJ, Visscher PM, Goddard ME. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res (Camb).* 91(1):47–60. doi: [10.1017/S0016672308009981](https://doi.org/10.1017/S0016672308009981).
- He M, Sun J, Jiang ZQ, Yang YX. 2017. Effects of cow's milk beta-casein variants on symptoms of milk intolerance in Chinese adults: a multicentre, randomised controlled study. *Nutr J.* 16(1):72. doi: [10.1186/s12937-017-0275-0](https://doi.org/10.1186/s12937-017-0275-0).
- Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics.* 31(2):423–447. doi: [10.2307/2529430](https://doi.org/10.2307/2529430).
- Hill WG, Weir BS. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet Res (Camb).* 93(1):47–64. doi: [10.1017/S0016672310000480](https://doi.org/10.1017/S0016672310000480).
- Keller LF, Waller DM. 2002. Inbreeding effects in wild populations. *Trends Ecol Evol.* 17(5):230–241. doi: [10.1016/S0169-5347\(02\)02489-8](https://doi.org/10.1016/S0169-5347(02)02489-8).
- Ketto IA, Knutsen TM, Øyaas J, Heringstad B, Ådnøy T, Devold TG, Skeie SB. 2017. Effects of milk protein polymorphism and composition, casein micelle size and salt distribution on the milk coagulation properties in Norwegian Red cattle. *Int Dairy J.* 70:55–64. doi: [10.1016/j.idairyj.2016.10.010](https://doi.org/10.1016/j.idairyj.2016.10.010).
- Kipp S, Segelke D, Schierenbeck S, Reinhardt F, Reents R, Wurmser C, Pausch H, Fries R, Thaller G, Tetens J, et al. 2016. Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. *J Dairy Sci.* 99(11):8915–8931. doi: [10.3168/jds.2016-11118](https://doi.org/10.3168/jds.2016-11118).
- Longeri ML, Zaniboni L, Cozzi MC, Milanese R, Bagnato A. 2021. *Animal Bio Arkivi: establishment of a phenotype and tissue repository for farm animals and pets at the University of Milan*. In ASPA Congress. Vol. 20. Taylor & Francis; p. 136.
- Lozada-Soto EA, Tiezzi F, Jiang J, Cole JB, VanRaden PM, Maltecca C. 2022. Genomic characterization of autozygosity and recent inbreeding trends in all major breeds of US dairy cattle. *J Dairy Sci.* 105(11):8956–8971. doi: [10.3168/jds.2022-22116](https://doi.org/10.3168/jds.2022-22116).
- MacLeod IM, Meuwissen THE, Hayes BJ, Goddard ME. 2009. A novel predictor of multilocus haplotype homozygosity: comparison with existing predictors. *Genet Res (Camb).* 91(6):413–426. doi: [10.1017/S0016672309990358](https://doi.org/10.1017/S0016672309990358).
- Mariani P, Losi G, Russo V, Castagnetti GB, Grazia L, Morini D, Fossa E. 1976. Prove di caseificazione con latte caratterizzato dalle varianti A e B della k-caseina nella produzione del formaggio Parmigiano-Reggiano. *Sci Tecn Latt-Cas.* 27: 208–27.
- Martikainen K, Sironen A, Uimari P. 2018. Estimation of intra-chromosomal inbreeding depression on female fertility using runs of homozygosity in Finnish Ayrshire cattle. *J Dairy Sci.* 101(12):11097–11107. doi: [10.3168/jds.2018-14805](https://doi.org/10.3168/jds.2018-14805).
- Massella E, Piva S, Giacometti F, Liuzzo G, Zambrini AV, Serraino A. 2017. Evaluation of bovine beta casein polymorphism in two dairy farms located in northern Italy. *Ital J Food Saf.* 6(3):6904. doi: [10.4081/ijfs.2017.6904](https://doi.org/10.4081/ijfs.2017.6904).
- Mastrangelo S, Sardina MT, Tolone M, Di Gerlando R, Sutera AM, Fontanesi L, Portolano B. 2018. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *Animal.* 12(12):2480–2488. doi: [10.1017/S1751731118000629](https://doi.org/10.1017/S1751731118000629).

- McDaniel BT. 2001. Uncontrolled Inbreeding. *J Dairy Sci.* 84: e 185–E186. doi: [10.3168/jds.S0022-0302\(01\)70214-7](https://doi.org/10.3168/jds.S0022-0302(01)70214-7).
- McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, et al. 2008. Runs of homozygosity in European populations. *Am J Hum Genet.* 83(3):359–372. doi: [10.1016/j.ajhg.2008.08.007](https://doi.org/10.1016/j.ajhg.2008.08.007).
- Mendes MO, de Morais MF, Rodrigues JF. 2019. A2A2 milk: brazilian consumers' opinions and effect on sensory characteristics of Petit Suisse and Minas cheeses. *LWT.* 108: 207–213.
- Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157(4):1819–1829. doi: [10.1093/genetics/157.4.1819](https://doi.org/10.1093/genetics/157.4.1819).
- Oliehoek PA, Bijma P. 2009. Effects of pedigree errors on the efficiency of conservation decisions. *Genet Sel Evol.* 41(1): 9. doi: [10.1186/1297-9686-41-9](https://doi.org/10.1186/1297-9686-41-9).
- Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. 2010. Conservation genetics in transition to conservation genomics. *Trends Genet.* 26(4):177–187. doi: [10.1016/j.tig.2010.01.001](https://doi.org/10.1016/j.tig.2010.01.001).
- Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. 2012. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet.* 91(2):275–292. doi: [10.1016/j.ajhg.2012.06.014](https://doi.org/10.1016/j.ajhg.2012.06.014).
- Poulsen NA, Bertelsen HP, Jensen HB, Gustavsson F, Glantz M, Månsson HL, Andrén A, Paulsson M, Bendixen C, Buitenhuis AJ, et al. 2013. The occurrence of noncoagulating milk and the association of bovine milk coagulation properties with genetic variants of the caseins in 3 Scandinavian dairy breeds. *J Dairy Sci.* 96(8):4830–4842. doi: [10.3168/jds.2012-6422](https://doi.org/10.3168/jds.2012-6422).
- Prinzenberg E, Krause I, Erhardt G. 1999. SSCP analysis at the bovine CSN3 locus discriminates six alleles corresponding to known protein variants (A, B, C, E, F, G) and three new DNA polymorphisms (H, I, A1). *Anim Biotechnol.* 10(1-2): 49–62. doi: [10.1080/10495399909525921](https://doi.org/10.1080/10495399909525921).
- Purfield DC, Berry DP, McParland S, Bradley DG. 2012. Runs of homozygosity and population history in cattle. *BMC Genet.* 13(1):70. doi: [10.1186/1471-2156-13-70](https://doi.org/10.1186/1471-2156-13-70).
- Ribeiro LA, Baron EE, Martinez ML, Coutinho LL. 2000. PCR screening and allele frequency estimation of bovine leukocyte adhesion deficiency in Holstein and Gir cattle in Brazil. *Genet Mol Biol.* 23(4):831–834. doi: [10.1590/S1415-4757200000400021](https://doi.org/10.1590/S1415-4757200000400021).
- Sanchez M-P, Fritz S, Patry C, Delacroix-Buchet A, Boichard D. 2020. Frequencies of milk protein variants and haplotypes estimated from genotypes of more than 1 million bulls and cows of 12 French cattle breeds. *J Dairy Sci.* 103(10):9124–9141. doi: [10.3168/jds.2020-18492](https://doi.org/10.3168/jds.2020-18492).
- Schiavo G, Bovo S, Muñoz M, Ribani A, Alves E, Araújo JP, Bozzi R, Čandek-Potokar M, Charneca R, Fernandez AI, et al. 2021. Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds. *Anim Genet.* 52(2):155–170. doi: [10.1111/age.13045](https://doi.org/10.1111/age.13045).
- Stasio L, Mariani P. 2000. The role of protein polymorphism in the genetic improvement of milk production. *Zootec Nutr Anim.* 26(3):69–90.
- van den Berg G, Jtm E, de Koning PJ, Bovenhuis H. 1992. Genetic polymorphism of  $\kappa$ -casein and  $\beta$ -lactoglobulin in relation to milk composition and processing properties. *Nederlands Melk En Zuiveltijdschrift.* 46(3–4):145–168.
- Weigel KA. 2001. Controlling inbreeding in modern breeding programs. *J Dairy Sci.* 84: e 177–E184. doi: [10.3168/jds.S0022-0302\(01\)70213-5](https://doi.org/10.3168/jds.S0022-0302(01)70213-5).
- Wickham H. 2016. *ggplot2: elegant graphics for data analysis* [Internet]. New York: Springer-Verlag. Accessed 2021 May 2. <https://ggplot2.tidyverse.org>.
- Wiggans GR, Cooper TA, VanRaden PM, Olson KM, Tooker ME. 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. *J Dairy Sci.* 95(3):1552–1558. doi: [10.3168/jds.2011-4985](https://doi.org/10.3168/jds.2011-4985).
- Wojdak-Maksymiec K. 2006. Associations between bovine lactoferrin gene polymorphism and somatic cell count in milk. *Vet Med Praha.* 51:14. doi: [10.17221/5512-VETMED](https://doi.org/10.17221/5512-VETMED).
- Xuan R, Chao T, Zhao X, Wang A, Chu Y, Li Q, Zhao Y, Ji Z, Wang J. 2022. Transcriptome profiling of the nonlactating mammary glands of dairy goats reveals the molecular genetic mechanism of mammary cell remodeling. *J Dairy Sci.* 105(6):5238–5260. doi: [10.3168/jds.2021-21039](https://doi.org/10.3168/jds.2021-21039).
- Yang A, Larsen T, Tume R. 1992. Carotenoid and retinol concentrations in serum, adipose tissue and liver and carotenoid transport in sheep, goats and cattle. *Aust J Agric Res.* 43(8):1809. doi: [10.1071/AR9921809](https://doi.org/10.1071/AR9921809).
- Zhang Q, Guldbbrandtsen B, Bosse M, Lund MS, Sahana G. 2015. Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics.* 16(1):542. doi: [10.1186/s12864-015-1715-x](https://doi.org/10.1186/s12864-015-1715-x).