Ann. Anim. Sci., Vol. 24, No. 2 (2024) 431–444 *DOI: 10.2478/aoas-2023-0100*

sciendo

COMPREHENSIVE ANALYSIS OF RUNS OF HOMOZYGOSITY AND HETEROZYGOSITY IN HOLSTEIN CATTLE ON THE BASIS OF MEDIUM AND HIGH DENSITY SNP PANELS AND LARGE POPULATION SAMPLE

Tomasz Szmatoła[1,2], Artur Gurgul[1], Igor Jasielczuk[1], Katarzyna Ropka-Molik[2♦]

[1]Center for Experimental and Innovative Medicine, University of Agriculture in Krakow, Rędzina 1c, 30-248 Kraków, Poland
[2]Department of Animal Molecular Biology, National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland
♦Corresponding author: katarzyna.ropka@iz.edu.pl

Abstract

This study reports runs of homozygosity (ROH) and heterozygosity (ROHet) distributed in a large population of Holstein cattle on the basis of two microarrays of medium (50k; 2163 animals; 54 609 SNPs) and high single nucleotide polymorphism (SNP) density (HD; 600 animals; 777 692 SNPs). To assess the inbreeding values of Holstein cattle, the ROH-based genomic inbreeding coefficient (F_{ROH}) was calculated. The comparison of SNP panels suggested that F_{ROH} values above 4 Mb should be considered for panels of medium densities as a relatively reliable measure of inbreeding. Moreover, ROH hotspots and coldspots were identified and compared between the HD and 50k SNP panels and were carefully examined for association with production and functional traits. The obtained results pinpointed genomic regions presumably under selection pressure in Holstein cattle. The regions overlapped with a large number of genes, including *GHR*, *GBF1*, *SUMF1*, *CCL28*, *NIM1K*, *U6*, *BTRC* and *FABP1,* many of which are involved in important Holstein cattle characteristics. We also found that some ROH hotspots and coldspots identified with the HD panel were not detected with the 50k panel, mainly because of insufficient SNP density in certain genomic regions. This suggests that using medium-density panels might not be the best choice when precise identification of ROH patterns is the main goal. In summary, in this work, we confirmed that a high-density SNP panel compared to a medium-density SNP panel allows for more precise identification of ROH patterns, especially in the case of short ROH that could be associated with ancestral inbreeding.

Key words: runs of homozygosity, microarray, Holstein cattle, inbreeding

Runs of homozygosity (ROH) are large continuous homozygous segments that are formed when both haplotypes inherited from the offspring's parents are identical (Gibson et al., 2006). The origin of ROH lies in a population demographic history and phenomena such as genetic drift, population bottlenecks or inbreeding. Short ROH are considered to be associated with ancestral inbreeding, while long ROH are associated with recent events (Keller et al., 2011). The distribution of ROH segments is non-random within the genome, and ROH create specific patterns of distribution in various populations that include sites with high ROH frequency, called ROH hotspots, and genomic regions with low ROH occurrence, known as ROH coldspots (Nothnagel et al., 2010; Pemberton et al., 2012; Wang et al., 2022). Natural or artificial selection acts as a strong pressure in shaping ROH distribution and characteristics across the genome, allowing for tracking of recent and ancient selection history by assessing the size and frequency of ROH regions (Howrigan et al., 2011; Mastrangelo et al., 2016; Peripolli et al., 2017). Several studies have pinpointed that ROH hotspots are associated with positive selection events in various animal species (Szmatoła et al., 2019; Biscarini et al., 2020; Santos et al., 2021; Li et al., 2022; Szmatoła et al. 2022) that in most scenarios increase homozygosity in certain genomic regions (Makino et al., 2018). Additionally, ROH can be easily utilized for estimating inbreeding via assessment of the ROH-based inbreeding coefficient (F_{ROH}), calculated as the ratio of the total sum of ROH per individual and the autosomal genome's length (McQuillan et al., 2008). This allows for the estimation of an individual's autozygosity without knowledge of its pedigree, which is especially important in regard to livestock species or individuals of unknown pedigree (Ferencakovic et al., 2011; Szmatoła et al., 2016, 2019). Thus, ROH-based methods of genomic inbreeding assessment may sup-

*The research was financed from the fund of Multiannual Program 17-2.10.9 – "Application and improvement of molecular analysis in cattle genotyping and maintaining genotype databases" and Statutory activities of MA PIB 01-5.02.1 – "Identification and characterization of the variability of the number of CNV in the genome of cattle and their impact on the level of production traits."

port standard estimates and be used in genomic selection (Topolski and Jagusiak, 2019).

Additionally, regions that are enriched in heterozygotes, known as runs of heterozygosity (ROHet), have also been of recent interest. These regions with high variability can provide information regarding population diversity and evolutionary history and pinpoint regions of the genomes in which maintaining greater genetic diversity may be more beneficial (Mulim et al., 2022).

Microarray density is an important factor in accurate ROH identification. In the study of Ferenčaković et al. (2013), it was shown that medium-density SNP panels tend to overestimate the number of short ROH (under 4 Mb) by calling artificial ROH due to the number of heterozygotes located within these fragments not being genotyped. In contrast, high-density panels offer a vastly higher number of genotypes at the expense of more genotyping errors and missing calls. Therefore, high-density panels offer more precise ROH identification, especially in the case of short ROH, and thus are considered more reliable for ROH identification. However, most studies of cattle are being performed using medium-density SNP panels, which are the gold standard in cattle genetics research, since these panels are used in most genomic-assisted breeding programs (Kim et al., 2015). Hence, it is necessary to compare high-density and medium-density SNP panels for a relatively large dataset and concentrate not only on standard ROH characteristics but also on ROH hotspots and coldspots.

In this study, we aim to determine the distribution and characteristics of runs of homozygosity and heterozygosity in a relatively large population of Holstein cattle on the basis of two microarrays of medium (50k) and high SNP density (HD). In addition, ROH hotspots and coldspots are identified and compared between the HD and 50k SNP panels. Moreover, the F_{ROH} (genomic inbreeding coefficient) are estimated, allowing an assessment of inbreeding in Holstein cattle populations. Finally, runs of heterozygosity hotspot regions are identified and compared between the two types of microarrays used.

The study assumption is that a high-density SNP panel will allow for a more detailed identification of ROH, especially short ROH, which are associated with ancestral inbreeding. Moreover, a high-density SNP panel should potentially allow for a more precise identification of ROH hotspot and cold spot regions, some of which could potentially be omitted in a medium-density SNP microarray.

Material and methods

Animal samples, DNA isolation and genotyping, filtering of genotypic data

The research material used in this study was tissue (ear punch, semen, blood, fur) stored in the Biological Material Bank of the National Research Institute of Animal Production. The material was obtained from 2736 animals that were further genotyped with Illumina BovineSNP50 (BovineSNP50 v2 DNA analysis BeadChip, later called 50k; n=2163) and Illumina BovineHD (BovineHD DNA analysis BeadChip, later called HD; n=600) microarrays. The animals were randomly selected Holstein cows and bulls maintained in Poland. No local animal care ethics committee approval was needed for this study since the material had been previously collected during standard breeding and veterinary testing procedures. The PCA for both microarray panels is presented in Supplementary Figure 1.

The DNA collected in the Biological Material Bank was isolated using a Sherlock AX (A & A Biotechnology, Gdynia, Poland) kit, followed by normalization to 50 ng/μL and analysis with the use of Illumina BovineSNP50 (50k) and Illumina BovineHD (HD) BeadChip microarrays (Illumina Inc., San Diego, CA, USA) scanned on a HiScanSQ system (Illumina). All procedures were carried out in accordance with the manufacturer's protocol.

Before filtering, the initial dataset contained 54 609 SNPs for 50k and 777 692 SNPs for HD BeadChip. The following filtering settings were applied: CallRate above 97%; Hardy-Weinberg equilibrium (HWE) test p value threshold set to 0.0001; minor allele frequency (MAF) greater than 0.0001; and SNPs on the X,Y and without a fixed genomic position removed. The final marker panels included 44 069 SNPs for 50k with a mean SNP density of 8.7 calculated in a 0.5 Mb window and 631 518 for HD panels with a mean SNP density of 125.4.

Identification of runs of homozygosity and heterozygosity; estimation of genomic inbreeding

Runs of homozygosity (ROH) and heterozygosity (ROHet) on both filtered SNP panels were identified with the use of detectRUNS software with the consecutive SNP-based run detection method (Biscarini et al., 2019). For ROH detection, the following criteria of identification were applied: a minimum number of 30 consecutive homozygous SNPs; a minimal ROH length set to 1 Mb; a maximum distance between SNPs equal to 1 Mb; a maximum of 1 SNP with a heterozygous genotype; and a maximum of 1 SNP with a missing genotype. For ROHet, since there are no standard criteria for their analysis and only a few studies concerning ROHet exist, we followed the criteria proposed by Biscarini et al. (2020), Santos et al. (2021) and Li et al. (2022): a minimum number of 15 consecutive homozygous SNPs; a minimal ROH length set to 1 Mb; a maximum distance between SNPs equal to 1 Mb; a maximum of 3 SNPs with homozygous genotypes; and a maximum of 2 SNPs with missing genotypes. The identified ROH were assigned to five length categories: >1 Mb, >2 Mb, >4 Mb, >8 Mb and above 16 Mb. For ROHet, however, the length categories were as follows: >0 Mb, >0.5 Mb, >1 Mb, >1.5 Mb, >2 Mb, and >2.5 Mb. The average sums of ROH and ROHet in selected categories were calculated by summing

all ROHs identified for each animal in each category and averaging the results.

The genomic inbreeding coefficient (F_{ROH}) was calculated with the use of detectRUNS software based on the research of McQuillan et al. (2008) by dividing the total sum of ROH lengths in the selected ROH length category for each individual by the sum of lengths of autosomal chromosomes covered by SNPs. ROH length categories were as follows: >1 Mb, >2 Mb, >4 Mb, >8 Mb and >16 Mb.

Identification of ROH hotspot, cold spot regions and ROHet hotspots

To detect runs of homozygosity and heterozygosity hotspot regions, we first calculated how often each SNP was present in ROH/ROHet in a given microarray panel. Then, 1% of the highest occurrences of SNPs in ROH/ROHet were merged into regions, and these regions were called ROH/ROHet hotspot regions. A similar procedure was maintained in the case of regions of low ROH frequency; however, this time, only 0.5% of the bottom occurrences of SNPs in ROH were taken into consideration, and these regions were called ROH coldspots. All of the ROH/ROHet hotspots and ROH coldspots were searched for overlapping genes with the use of Ensembl BioMart (https://www.ensembl.org/biomart) software based on Ensembl gene version 106. In addition, to assess the biological processes and molecular functions of the identified genes, the Panther Classification System (https://www.pantherdb.org/) was used.

Results

Distribution of ROH and ROHet

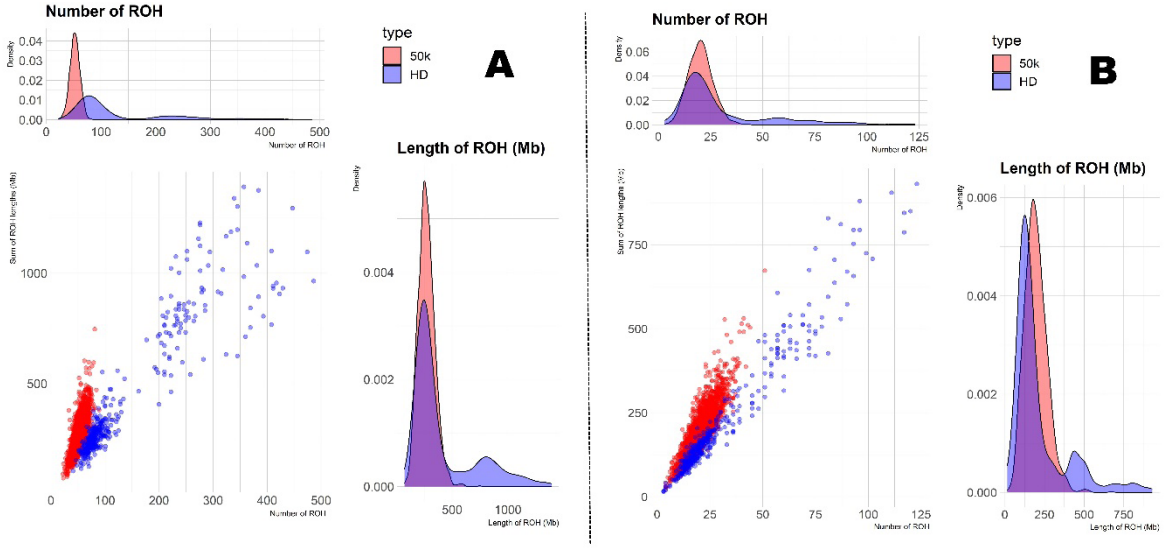
To assess the distribution and characteristics of identified ROH based on microarray panels differing in density, we concentrated on the number of ROH, sums of ROH lengths and assignment of ROH into specific length categories.

In the case of the 50k panel, 113,177 ROH were identified for 2163 animals, which averaged to 52.3 ROH per individual, while the HD microarray detected 52,347 ROH for 600 animals (130.8 on average per individual). In general, a higher number of short (1–4 Mb) and medium (4–8 Mb) ROH per individual was observed for the HD microarray, while similar numbers of ROH were observed in the length category of 8–16 Mb. However, in the case of long ROH, above 16 Mb, a higher number of ROH was observed in the case of the 50k SNP panel. This is partially because of the method used for ROH calculations that allowed only 1 heterozygote in ROH, which could lead to breaking of long ROH in the case of denser SNP HD microarray.

A mostly similar trend was observed in regard to total lengths of ROH per individual. Again, the HD microarray was characterized by higher sums of ROH lengths per animal in the category of short ROH (1–4 Mb), with somewhat similar values for the medium length ROH category and visibly lower values for the long ROH category. These results are presented in detail in Table 1 and Figure 1.

Table 1. Number and sum of lengths of ROH in the analysed microarray panels

		Stat	ROH length category (Mb)				
			>1	>2	>4	>8	>16
50k	Number/animal	Mean	52.3	39.0	20.2	8.8	2.6
		SD	9.1	8.2	6.0	3.6	1.9
		Min	22	11	4	0	0
		Max	89	71	51	32	13
	Length/animal (Mb)	Mean	270.5	249.0	196.0	131.6	64.2
		SD	74.4	74.0	71.3	62.2	48.8
		Min	72.3	57.8	23.6	0	0
		Max	746.2	728.8	673.4	556.4	396.6
HD	Number/animal	Mean	130.9	48.1	29.8	8.2	1
		SD	94.5	67.5	22.8	8.3	1.8
		Min	41	13	3	0	0
		Max	486	258	123	51	12
	Length/animal (Mb)	Mean	409.8	321.2	215.6	96.8	20.4
		SD	287.6	234.5	176.5	102.9	38.8
		Min	104.7	47.1	14.4	0	0
		Max	1391.3	1188.5	931.0	668.9	287.4
			ROH length category (Mb)				
			1–2	2–4	4–8	8–16	16+
50k	Number of ROH		28661	40820	24764	13203	5729
	Number of ROH/individual		13.25	18.87	11.45	6.10	2.65
HD	Number of ROH		25358	15072	8635	2885	397
	Number of ROH/individual		42.26	25.12	14.40	4.81	0.66



A – ROH length above 1 Mb, B – ROH length above 4 Mb.

Figure 1. Distribution of ROH lengths based on both SNP panels used and different length thresholds

Table 2. Number and sum of lengths of ROHet in the analysed populations and both microarray panels

		Stat	ROHet length category (Mb)					
			>0	>0.5	>1	>1.5	>2	>2.5
50k	Number/animal	Mean	37.9	37.9	8.5	1.2	0.2	0.02
		SD	6.1	6.1	2.8	1.1	0.4	0.14
		Min	19	19	1	0	0	0
		Max	62	62	20	5	2	1
	Length/animal (Mb)	Mean	32.0	32.0	10.5	2.1	0.5	0.05
		SD	5.4	5.4	3.6	1.9	1.0	0.36
		Min	15.8	15.8	1.2	0	0	0
		Max	51.7	51.7	24.0	9.0	4.9	2.64
HD	Number/animal	Mean	69.8	3.8	0.6	0.05	0	0
		SD	39.9	3.3	1.2	0.27	0	0
		Min	31	0	0	0	0	0
		Max	180	21	6	3	0	0
	Length/animal (Mb)	Mean	23.1	2.6	0.7	0.08	0	0
		SD	13.2	2.4	1.4	0.45	0	0
		Min	10.6	0	0	0	0	0
		Max	63.9	16.6	6.8	5.14	0	0
			ROHet length category (Mb)					
			0–0.5	0.5–1	1–2	2–4		
50k	Number of ROH		0	63,493	17,976	455		
	Number of ROHet/individual		0	29.4	8.3	0.2		
HD	Number of ROHet		26,402	1,400	107	0		
	Number of ROHet/individual		66.0	3.5	0.3	0		

A total of 81,924 ROHet were identified in the case of the 50k SNP panel, which averages 37.9 ROHet per animal. As many as 27,909 ROHet were detected for the HD panel, which is on average 46.5 ROHet per individual. ROHet were assigned into different categories than ROH and resulted in the identification of a generally higher number and shorter ROHet on the HD panel than in the case of the 50k SNP microarray. In addition, no ROHet above 2 Mb was identified for the HD microarray, while the 50k SNP chip was characterized by values of approximately 0.5 ROHet per individual.

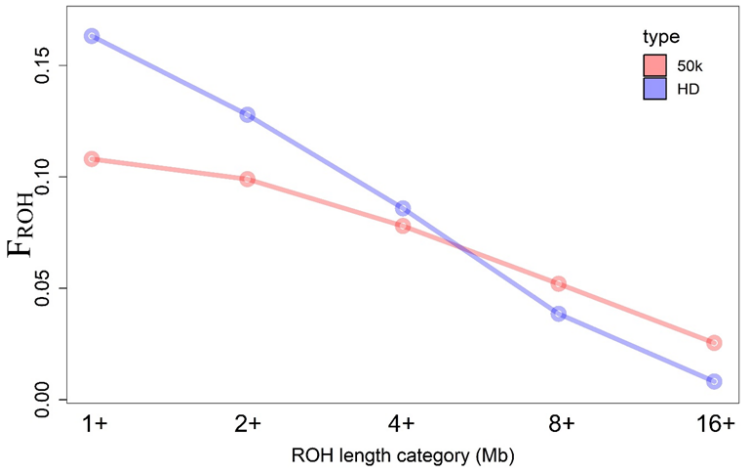


Figure 2. F_{ROH} values for the analysed animals presented for both microarray panels

The sum of ROHet lengths, however, showed a different trend. The 50k SNP chip, even though characterized by a lower number of ROHet, showed higher sums of ROHet length in every category. These results are presented in detail in Table 2.

As expected, F_{ROH} values presented the same trend in characteristics as the total sum of ROH lengths per individual, with generally higher F_{ROH} values in the case of the HD microarray for short ROH (above 1 Mb), somewhat similar values for medium length ROH (above 4 Mb) and generally lower values for ROH above 8 Mb. The results are graphically presented in Figure 2.

Identification of ROH hotspots, coldspots and ROHet hotspot regions

Runs of homozygosity hotspots, which are regions of high ROH frequency associated with selection pressure, were identified for both microarrays, forming 9 genomic regions (from 0.3 to 14.2 Mb in length) for the 50k SNP chip and 26 regions (from 0.1 Mb to 3.9 Mb) for the HD microarray. Eight of nine ROH hotspots identified for the 50k microarray were also detected by the HD panel, with the exclusion of one hotspot, localized on chromosome 8, in which there was a visible increase in SNP occurrences in ROH on the HD panel; however, it did not reach the adopted threshold value. Moreover, in the case of the 50k microarray, one long hotspot on chromosome 20 was broken into shorter segments in the case of the HD panel. The HD microarray allowed for the identification of 17 novel ROH hotspot regions that were not distinguishable at all in the case of the 50k SNP panel. Additionally, some hotspot regions observed in the case of the HD panel (for example, the ending part of BTA12, BT17 or BTA29) were not visible at all in the case of the 50k panel. This is due to low density of SNPs in the vicinity of these regions in case of 50k panel (BTA 12: 25 SNPs in total for 50k panel with a mean density of 1.6 in a 7 Mb region in contrast to 604 SNPs and mean density of 40.2 for HD panel; BTA 17: 59 SNPs in total for 50k panel with a mean density of 5.3 in a 5 Mb region in contrast to 1451 SNPs and mean density of 131.9 for HD panel; BTA 29: 18 SNPs in total for 50k panel with a mean density of 3.6 in a 2 Mb region in contrast to 339 SNPs and mean density of 67.8 for HD panel), which clearly shows that this panel omits important information captured by the HD panel. Figure 3 shows runs of homozygosity hotspots and coldspots on the selected chromosomes that were typed based on visual differences between microarray panels in the location of these regions, while Figure 4 presents three examples of hotspot regions that are absent in the 50k panel due to low SNP density. Additionally, all hotspot regions along with SNP density values are graphically presented in Supplementary Figure 2.

The identified regions of high ROH occurrences (ROH hotspots) overlapped with a number of genes: 104 for the 50k panel and 229 for the HD panel, of which 51 were common for both microarrays. The discussed ROH hotspot regions are graphically presented in Supplementary Figure 2, partially in Figure 3 and in detail in Supplementary File 1.

For ROH coldspots, we identified 20 regions (from 0.03 Mb to 10.5 Mb in length) with the 50k SNP panel and 32 (from 0.006 Mb to 2.66 Mb) with the HD panel. In the identified cold spot regions, we found 27 genes in the case of the 50k SNP panel and 82 genes in the HD panel, with 7 genes identified by both microarrays. The ROH cold spot regions are graphically presented in Supplementary Figure 2 and partially in Figure 3 and in detail in Supplementary File 2.

This figure shows only the selected chromosomes on which a visible difference in ROH frequency was shown. The results in detail, with SNP density values, for all chromosomes are presented in Supplementary Figure 2. Horizontal dashed lines represent the cold spot threshold, while solid lines represent the hotspot threshold.

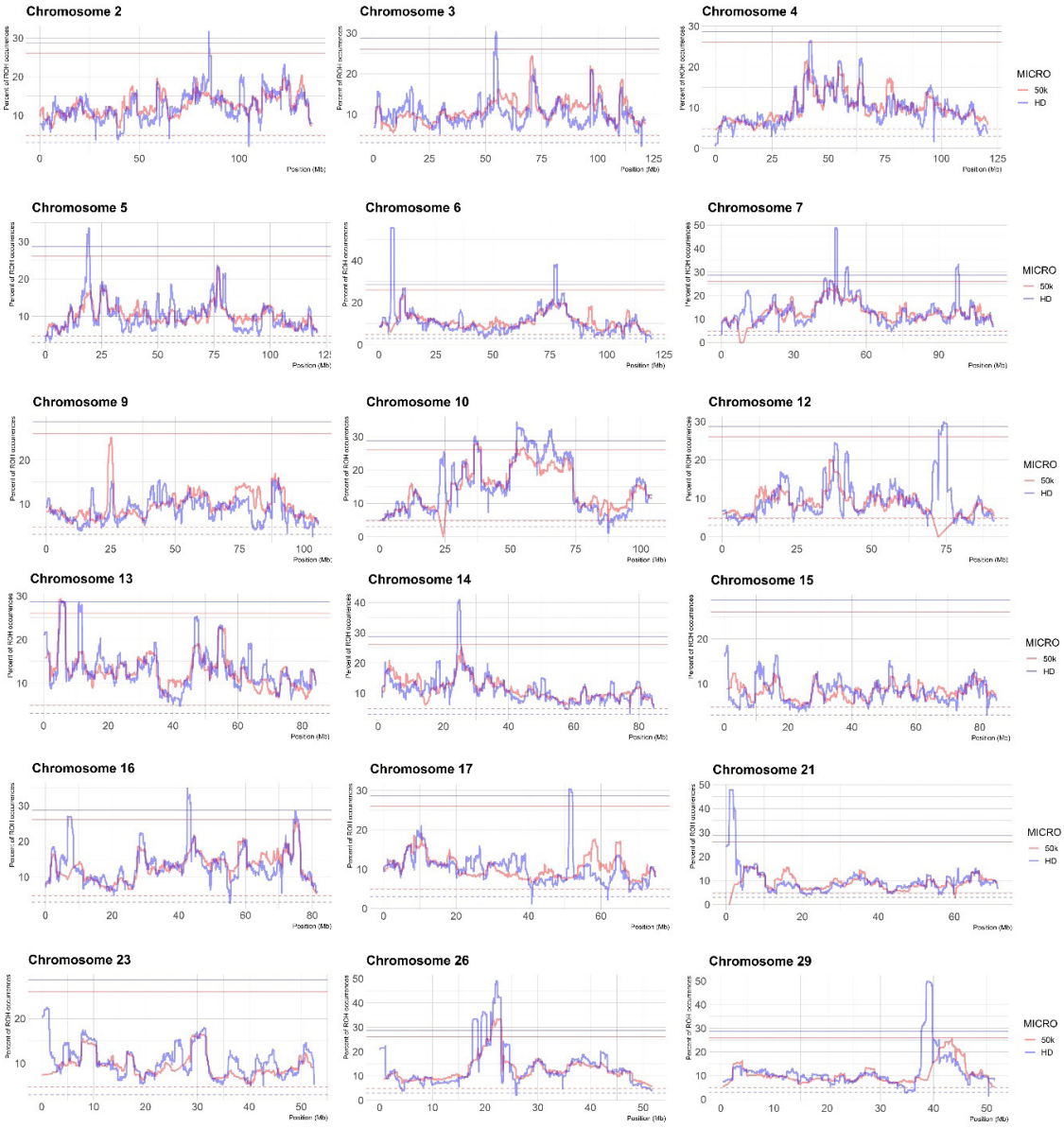


Figure 3. Runs of homozygosity hotspots and coldspots on the selected chromosomes of interest

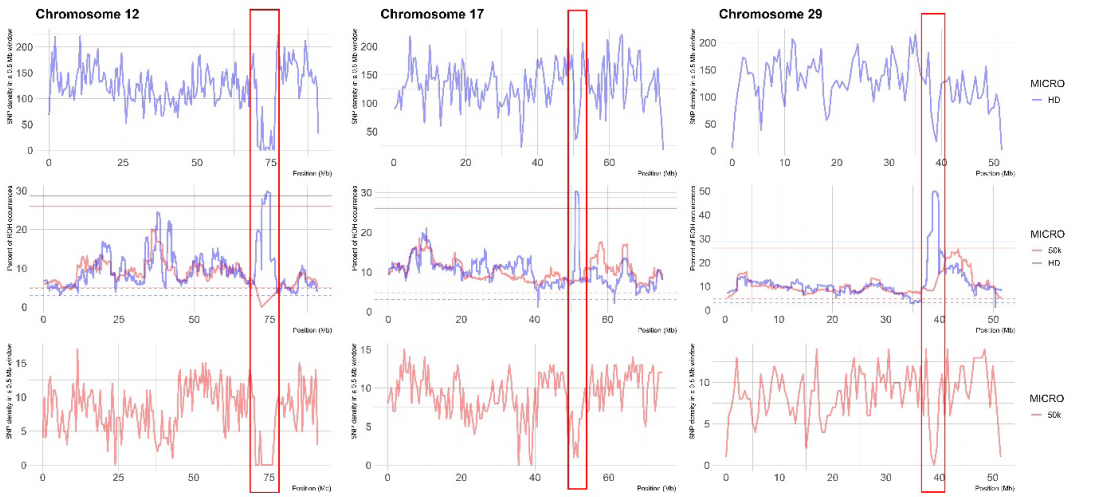
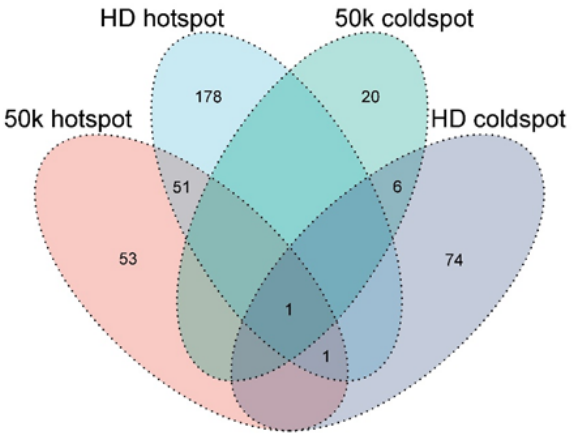


Figure 4. Examples of hotspot regions that were uniquely identified with the HD panel along with SNP density plots for both genotyping panels. The red frame represents hotspot regions that were not identified via the 50k panel due to low SNP density values

Figure 5. Comparison of heterozygosity hotspot regions for two types of microarrays.
Red represents the 50k SNP microarray, and blue represents the HD panel. Dark blue and dark red colours represent ROHet hotspot regions



50k – represents medium-density SNP panel, HD – represents high-density SNP panel.

Figure 6. Venn diagram representing common genes identified by two types of microarray panels in relation to ROH hotspots and coldspots

Identification of ROHet hotspot regions

ROHet hotspots were identified in a vastly larger number than ROH hotspot regions while being generally of a smaller size. For the 50k SNP chip, 30 regions were identified, with sizes ranging from 0.11 to 2.35 Mb. For the HD microarray, 129 such regions were identified, which were generally shorter than those identified on the 50k panel (from 0.02 to 1.046 Mb). Most of the regions were uniquely identified for each microarray panel; however, some overlapped between panels. These regions are graphically presented in Supplementary Figure 3 and Figure 5 and in detail in Supplementary file 3.

Discussion

Runs of homozygosity have been a hot topic in recent years, and many authors have contributed their research to this subject by comparing human populations and various animal species with respect to inbred assessment and recently to signatures of selection. Runs of homozygosity can be detected based on genotyping microarrays of various densities, as well as data obtained from next-generation sequencing (NGS) analysis. In this study, we used two microarrays of different SNP densities, medium-density BovineHD with approximately 50k SNPs (here abbreviated as 50k) and high-density BovineHD with approximately 700k SNPs (here abbreviated as HD), to comprehensively analyse ROH and ROHet. Similar comparisons of medium and high SNP density panels were performed previously by other authors (Purfield et al., 2012; Ferenčaković et al., 2013); however, in this study, we utilized a larger animal population (consisting of 2163 Holstein cattle for the 50k panel and 600 for the HD panel) and performed a more comprehensive analysis of hotspot/cold spot regions. Apart from comparative analysis of microarrays, analysis of a large dataset also allows verification of previously described ROH-associated regions and their possible association with production traits.

ROH and ROHet characteristics, distribution and F_{ROH} estimation

Runs of homozygosity and heterozygosity are commonly distributed across the genome and allow for recognition of the demographic history of populations. In addition, various authors have provided the insight that ROH and ROHet can be used to investigate genetic diversity and adaptive evolution (Ceballos et al., 2018).

In this study, the basic characteristics of ROH, which are the average sums of ROH lengths and number of ROH per individual described (for 50k the average sum of ROH lengths of 270.6 Mb and number of ROH of 52.3 and for HD 409.8 Mb and 130.9, respectively), correspond well to the results obtained by other researchers. In the research of Purfield et al. (2012), based on the HD panel, the authors showed that the average sum of ROH lengths for Holstein cattle was approximately 240 Mb, and for segments longer than 5 Mb, it was 115 Mb. North American Holstein cattle were characterized by a higher mean sum of ROH lengths that equalled 299.6 Mb and an average number of ROH of 82.3 (Forutan et al., 2018). Very similar results were reported for Italian Holstein cattle based on the SNP50 BeadChip by Marras et al. (2015), in which the mean sum of ROH lengths was estimated to be approximately 300 Mb and the average number of ROH was 81.7. In our previous research (Szmatoła et al., 2019), Polish Holstein cattle were characterized by a mean sum of ROH lengths of 295.1 Mb and a mean number of ROH per individual of 53.3.

It should be noted that the results of ROH identification can be strongly biased by the density of the microarray used, the parameters and the method of ROH identification. Recently, the most common method of ROH identification based on a sliding window was established to provide some form of analytical bias, which is why in this study, a consecutive approach that resolves around checking homozygosity status for each adjacent SNP was used (Santos et al., 2021). Compared to the sliding window method, this approach, however, is characterized by the identification of a smaller number of long ROH, especially in the case of high-density SNP panels, due to breaks in long ROH caused by a limited number of allowed heterozygotes.

The ROH identified with the use of the HD panel potentially better represent real ROH present in the genome due to the higher density of the markers. This is especially evident in a short ROH length category (1–4 Mb) in which the 50k panel identified a vastly lower number and total sum of lengths of ROH in comparison to the HD panel. It should be noted that due to the lower SNP density of the 50k panel, some short ROH identified by the HD panel were not visible at all (Figure 4 and Supplementary Figure 2). Moreover, some short ROH identified via the 50k panel could in fact be artificial ROH that were not identified at all with the HD panel, since medium-density panels tend to produce short artificial ROH due to their vastly lower SNP density (Ferenčaković et al., 2013).

It must be stated that in this study, due to the chosen method of ROH identification, longer ROH (above 8 Mb) tended to break, which is especially noticeable in the case of the HD panel. That is why we generally obtained a lower number of very long ROH when compared to studies based on the sliding window approach. The proposed method used in the detectRUNS software will provide some bias in the detected number and sum of ROH lengths when assigning ROH into ROH length categories; however, this will not have any effect on the total sum of ROH lengths, total F_{ROH} values, ROH hotspots or coldspots.

Runs of heterozygosity (ROHet) is a quite new term and refers to regions of high heterozygosity in the genome where most adjacent SNPs are in a heterozygous state. However, compared to ROH, more errors are generally allowed in ROHet; in most publications, up to three homozygous SNPs for short regions are allowed. ROHet

can be associated with heterozygous clusters (Williams et al., 2016), and their analysis can be used to identify balancing selection events or loci of lethal recessive mutations (Biscarini et al., 2020; Santos et al., 2021). Only a few reports on ROHet identification can be found. The numbers of identified ROHet per individual in livestock vary from 9.9 for cattle (Biscarini et al., 2020) and 28.3 for sheep (Tsartsianidou et al., 2021) to 57.8 for turkeys (Marras et al., 2018) and up to 52.2 for horses (Santos et al., 2021). The numbers have differed greatly due to the methods of ROHet identification and the density of the SNP panel used. Our results (mean number of ROHet of 37.9 for the 50k panel and 69.8 for the HD panel), however, are consistent in regard to ROHet length, showing ROHet under 1 Mb to be the most frequent, as well as in ROHet being generally scarce compared to ROH events. It should be noted that even though the detected number of all ROHet was much higher for the HD panel than for the 50k panel, the total length was vastly lower. This could be explained by the higher SNP density of the HD panel, which led to the identification of even very short ROHet fragments that could not be identified at all with the medium-density panel. Additionally, the 50k panel could lead to the identification of artificial ROHet, just as in the case of short ROH (Ferenčaković et al., 2013).

The genomic inbreeding coefficient calculated from ROH (F_{ROH}) is regarded as a very reliable measure of autozygosity that provides information about both recent and ancient relatedness of individuals (Ferenčaković et al., 2013; Curik et al., 2014). In the research of Ferenčaković et al. (2013), it was shown that F_{ROH} may be a better estimator of autozygosity levels than the traditionally used inbreeding coefficient, which is based on pedigree datasets (F_{PED}). It should be stated, however, that the density of the SNP panel used for F_{ROH} estimation is essential for the identification of short ROH, which contribute greatly to ancestral inbreeding (Szmatoła et al., 2016). This is clearly visible in this study, in which a vast number of short ROH identified with the use of the HD panel are not shown at all with the use of the 50k microarray. Additionally, since long ROH did tend to break in the case of the HD panel into shorter segments, the F_{ROH} was also generally lower for categories above 8 and 16 Mb for this panel. In general, the 50k panel underestimated F_{ROH} values for shorter ROH (>1 and >2 Mb categories) and overestimated F_{ROH} with lengths above 8 and 18 Mb when compared to the HD panel. The F_{ROH} values for ROH longer than 4 Mb were very similar for both panels. In general, it seems that the genomic inbreeding coefficient calculated based on medium-density arrays for ROH above 1 Mb could be underestimated by quite a large amount, suggesting that F_{ROH} values above 4 Mb should be utilized for a more precise analysis.

There are a number of manuscripts describing F_{ROH} for selected cattle breeds, and most of the findings in those studies do compare with the results obtained in this study. Purfield et al. (2012) showed that Holstein cattle kept in Europe were characterized by F_{ROH} coefficients in the range of 0.081 for ROH above 1 Mb to 0.046 for ROH longer than 5 Mb. Italian Holstein cattle in the report of Marras et al. (2015) had F_{ROH} values above 1 Mb of 0.116 and 0.073 for ROH above 4 Mb. Our previous report supported these results, with F_{ROH} values of 0.118 for ROH above 1 Mb and 0.088 for ROH above 4 Mb. The results presented in this study partially correspond to the results obtained by other authors, with mean F_{ROH} values above 1 Mb of 0.107 for the 50k panel and 0.162 for the HD panel and above 4 Mb of 0.078 for the 50k panel and 0.085 for the HD panel.

ROH hotspot and cold spot regions

This study aimed to comprehensively analyse ROH patterns in a relatively large dataset based on two types of microarrays, one of medium density and one of high density. It is easily noticed that some inconsistencies may be seen among the panels used, such as ROH hotspots originating from short ROH on BTA2, BTA3, BTA5, BTA6, BTA7, BTA13, BTA14, BTA16, BTA17, and BTA29 (Figure 3). Additionally, some ROH hotspots identified with the HD panel were not seen at all with the use of the 50k panel because of a lack of SNPs in these regions – for example, hotspots on BTA6, BTA10, BTA12, BTA21 or BTA29 (Figure 3, 4 and Supplementary Figure 2). Most of these differences arose due to the insufficient density of a 50k SNP panel, which did not allow for ROH identification at poorly covered genome regions. Some additional differences in ROH patterns can be related to the worse ability of the 50k panel to detect short ROH, representing ancestral inbreeding as described above.

ROH hotspots and their gene content

We observed that both microarray panels showed a high similarity of ROH patterns in the genome (Supplementary Figure 2 and Figure 3). However, the hotspot regions overlapped only to a certain extent and resulted in approximately 49% of common genes (when comparing 50k gene results to HD) from both panels (Figure 6). Nevertheless, most of these genes seem to be interesting because they have been previously described as being related to cattle production traits.

ROH hotspots identified within this study overlapped with a large and different number of genes depending on the microarray used; however, 51 genes were identified by both panels. These genes were previously proposed by other authors as important for production traits in cattle (Mohammadi et al., 2020). For example, ROH hotspots spanning the *GHR* and *GBF1* genes were identified. *GHR* is one of the major genes affecting milk production traits, including milk composition and yield (Blott et al., 2003; Strucken et al., 2015), lactogenesis and fertility, mammary gland development (Hadi et al., 2015) and growth performance (Zhao et al., 2007). *GHR* is localized on BTA20 within QTLs associated with milk production, and it has been confirmed that variation in the *GHR* locus determines dairy cattle production traits in different breeds (Kadri et al., 2015; Viitala et al., 2006) and even in

buffalos (EL-Komy et al., 2020). In turn, the *GBF1* gene encodes Golgi brefeldin A, which was shown to also be significantly associated with milk yield and fertility in dairy cattle (Nayeri and Stothard, 2016). Using GWAS, the authors pinpointed a SNP in the *GBF1* gene affecting milk production via modification of lipid and carbohydrate metabolism networks.

The recent research of Pedrosa et al. (2021) focused on the identification of genomic regions responsible for lactation persistency and milk production traits. The authors pinpointed *SUMF1*, *CCL28*, *NIM1K* and *U6* genes (Pedrosa et al., 2021) that were also detected in the present study by both microarrays. The *SUMF1* gene was associated with milk yield, *U6* with milk yield and fat content, and both *CCL28* and *NIM1K* with protein percentage in Holstein dairy cattle (Pedrosa et al., 2021). The *CCL28* gene was also identified as significantly related to protein content in a US population of Holstein cattle (Jiang et al., 2019). Moreover, the *SUMF1* gene was detected in indigenous dairy breeds (using HD microarray) as the selection signature region with a significant association with milk performance traits due to its involvement in lipid and lipoprotein metabolism (Dash et al., 2022).

Another interesting gene that was identified for both panels was *BTRC,* which determines mammary gland development in mice (Kudo et al., 2004). The *BTRC* gene was previously mapped to QTLs related to dairy cattle traits (van den Berg et al., 2014) and proposed as a candidate gene determining udder formation and development (Marete et al., 2018). Furthermore, the study of eQTL and QTL identification indicated that expression of the *BTRC* gene can be associated with the *TWNK* gene (twinkle mtDNA helicase), which was also detected in our data. The *TWNK* gene was noted to be associated with milk yield, fat content and the percent of fat in milk (van den Berg et al., 2019).

Our results confirmed that both microarrays were able to identify hotspots associated with various important cattle production traits; however, the HD panel allowed for the identification of a larger number of hotspots. The HD-based hotspots encompassed 178 unique genes that could be associated with regions of the genome not covered at all by the 50k SNP panel (or with insufficient SNP density) or identified short ROH that could be signs of ancient selection events that diminish with the SNP density of the selected microarray (as an example, three such regions are presented in Figure 4). These unique hotspots were localized on 17 different chromosomes, including the most numerous at BTA7 (3 regions and 27 genes). Within the regions, we observed genes involved inter alia in lactation and mammary gland function; milk production traits, implantation and embryo development; coat colour phenotype and heat stress.

The application of the HD panel also allowed the detection of 17 genes involved in anatomical structure morphogenesis and development processes (*FGF10, FGF8, CASZ1, TOX, TLX1, TMOD2, TMOD3, SEMA4G, WNT8B, HECW2, SLIT81, DUSP6, NGR2, SMAD5, TGFB1, UBE3A* and *HPS6*). One of these genes, *TGFB1,* has been previously proposed as a key factor controlling mammary gland remodelling and function in dairy cattle during both dry and lactation periods (Vries et al., 2011; Dai et al., 2018).

Another of the interesting genes was *FARP1,* which together with *HS6ST3* was localized on BTA12 and was detected only with the HD microarray. A recent study showed a possible significant association of the *FARP1* gene with milk urea nitrogen (MUN) content (Ma et al., 2023). The authors proposed a panel of candidate genes that affected MUN production or metabolism in Holstein cows. Moreover, *FARP1* was mapped to a QTL associated with reproductive traits, namely, the interval to first oestrus after calving. The other interesting gene related to MUN proposed in the study by Ma et al. (2023) was adhesion G protein-coupled receptor B1 (*ADGRB1*), which is involved in angiogenesis. In our study, we observed a hotspot region in which another gene belonging to the adhesion G protein-coupled receptor family was identified – Receptor L3 (*ADGRL3*), which suggests that this group of genes can be important in dairy cattle production traits. This gene was localized in the vast hotspot located on chromosome 12, which was not detected by the 50k microarray due to the very low number of SNPs in this region, which did not allow for ROH formation (12 SNPs in total in the 6.7 Mb region for the 50k panel compared to 310 SNPs for the HD panel); thus, this region was one of the regions that was lost in the ROH hotspot identification. This suggests that at least some of the ROH hotspots, not only short but also covering large genomic regions, can be omitted when using medium-density SNP panels.

In addition, our results confirmed a strong ROH hotspot in another genomic region associated with milk quality and the amount of milk produced. This hotspot region was localized on chromosome 6 and was uniquely identified with the HD panel. Within the hotspot, a *FABP2* gene was detected. The FABP family, especially *FABP4,* was shown to have a major effect on milk yield and protein content (Zhou et al., 2015; Ye et al., 2022). Our results also indicated that *FABP2*, which plays a role in the metabolism and transport of long-chain fatty acids, can be essential for milk composition in Holstein cows.

The next interesting ROH hotspot region identified only by the HD panel was localized on chromosome 26. This region overlapped with the *SCD* gene that encodes the stearoyl-CoA desaturase enzyme, which was proposed in numerous studies (Li et al., 2020; Tian et al., 2022) to regulate milk fat synthesis and influence milk composition. Again, this signal was not detected in the case of the medium-density SNP panel.

On chromosome 29, we observed a large region that included 14 genes representing pregnancy-associated glycoproteins – PAGs (*PAG5, PAG18, PAG7, PAG4, PAG16, PAG14, PAG19, PAG20, PAG21, PAG6, PAG17, PAG11, PAG5, PAG8)* and two other genes – *DDB1* and

VWEC. In the research of Buaban et al. (2022), utilizing the GWAS method, a very similar region on chromosome 29 was identified that was significantly associated with milk yield in Thai dairy cattle. In this genomic region (29: 37,949,466–40,507,958), the authors identified 20 PAG genes together with *DDB1*, *VWEC* and other provisional LOC genes and proposed them as candidate genes for milk and reproductive traits in Thai cows. Our results strongly support this finding and indicate that this region can also be important for Holstein dairy cattle production performance. PAG genes, expressed mainly in the placenta (placental trophoblasts), play a key role during implantation and pregnancy. In cows, PAG proteins are critical for pregnancy maintenance (Reese et al., 2019; Barbato et al., 2022). Considering PAG function, previously shown associations and ROH hotspot localization, we hypothesized that these genes should be investigated more thoroughly in terms of improving Holstein cattle reproductive traits. Once again, in our results for the 50k SNP panel, we did not identify any increase in ROH frequency for this region, probably because of a gap in SNPs on a microarray in the middle of the mentioned hotspot region (10 SNPs in the 2 Mb region in the 50k panel in comparison to 221 SNPs for the HD panel).

Other potentially interesting genes are *PITX3* and *GBF1* identified by both microarrays, together with *TOX* (identified only with HD) and *SUFU* (identified only with 50k). In the literature, these genes have been shown to be associated with pigmentation in cattle (Senczuk et al., 2020). The authors explained that these genes can be related to melanogenesis and melanocyte differentiation or early developmental events in cattle. Interestingly, the *TOX* gene was also analysed in terms of its association with reproductive traits in cattle (Camargo et al., 2015). Moreover, it has been established that the *MYO5A* gene, detected uniquely with an HD microarray in our study, determined a dilute coat colour phenotype in mice (Zhang et al., 2021) and rabbits (Fontanesi et al., 2012). Similarly, variation within the *HPS6* gene, uniquely identified with the HD panel, was associated with pigmentation patterns in humans (Karim et al., 2021).

A very interesting ROH hotspot region was observed on chromosome 21 using the HD approach. This hotspot region was not detected with the 50k panel, but the 50k panel identified an ROH cold spot region. This ROH hotspot region (21: 682199-2482464) included four genes, *SNRPN*, *SNORD115*, *UBE3A* and *bta-mir-11995*, which were described as mostly involved in embryo growth and development as well as pregnancy maintenance. *SNRPN* encodes small nuclear ribonucleoprotein polypeptide N and is a maternally imprinted gene that is highly expressed at the early stage of embryo development and plays a critical role in foetal development and placental function (Suzuki et al., 2009). In cattle, it has been demonstrated that in vitro embryo manipulation, including somatic cell nuclear transfer, can lead to abnormal reprogramming of *SNRPN* imprinting (Suzuki et al., 2009), which can affect embryo survival. The next gene, *SNORD115,* encodes a small nucleolar RNA (snoRNA) that modifies the transcriptional activity and function of numerous genes directly and indirectly by regulating *SNORD116* (Falaleeva et al., 2015). In turn, the lack of *SNORD116* expression during embryonic development affects feeding behaviour and energy homeostasis via modification of hypothalamic neuropeptide expression (Qi et al., 2016). Moreover, the detected miRNA *bta-mir-11995* was identified in a gene expression study as differentially expressed and secreted by bovine embryos under implantation conditions and can be utilized as a potential early biomarker of developmental competence (Melo-Baez et al., 2020). The last gene, *UBE3A*, also maternally imprinted, is activated during the most important developmental periods, and its expression is critical in the formation of normal embryonic development, especially in brain tissue (Wang et al., 2019; Sonzogni et al., 2020). Taking into account the recent studies of Zinovieva et al. (2020), the *UBE3A* gene was proposed as a selection signature in the two oldest Russian native cattle breeds, and this locus seems to be extremely important for embryo development and pregnancy maintained in cattle.

In general, the obtained results confirm that high-density SNP panels can provide additional information in regard to the identification of ROH hotspot regions, thus allowing for the detection of completely new candidate genes that are important for cattle production traits.

ROH cold spot regions

ROH distribution is not uniform across the genome, and unique ROH patterns develop for specific populations. In recent years, the topic of ROH hotspot regions has gained the attention of researchers. However, regions of the genome with very low or zero ROH frequency, called ROH coldspots or no-ROH regions, could be of interest. The existence of such regions might be associated with recombination hotspots in which a high variation at certain loci occurs and is beneficial (Mackiewicz et al., 2013) or possibly due to the selection pressure favouring heterozygotes (Eisen and Osthoff, 2014; Fijarczyk and Babik, 2015). The formation of ROH coldspots can also be associated with negative selection against homozygotes that could have a negative effect on fitness, promoting high variation that is in this case beneficial (Wang et al., 2022). In this study, we identified ROH cold spot regions using two SNP microarrays and identified and compared genes that spanned these regions. Only a fraction of ROH cold spot regions overlapped between microarray panels, resulting in only 23% (Figure 6) of common genes. Once again, the HD panel led to the identification of more ROH cold spot regions. This is presumably mostly due to the lower density of the 50k panel; however, both panels still led to the identification of ROH coldspots that overlapped with genes described as being related to cattle production traits and thus being of interest.

One interesting cold spot region identified with the use of the 50k SNP panel is localized at chromosome 12 (position 72285729-82791816). Interestingly, in the case

of the HD panel, this region was identified as an ROH hotspot. The SNP density for the 50k panel was very low for this region, spanning a total of only 12 SNPs over 6.7 Mb. This is especially evident when compared to the HD panel, in which 310 SNPs were observed. Within this region, four genes were detected: *PCCA, TMTC4, NALCN, ITGBL1* and a QTL that is associated with milk yield and composition, dry matter intake and metabolic body weight. The most interesting gene seems to be *PCCA*, which encodes one of the 2 subunits (alpha) of the biotin-dependent propionyl-CoA carboxylase (PCC). The PCC enzyme catalyses the first step of propionyl-CoA catabolism. In cattle, the *PCCA* gene has been associated with undercoat length and, as a result, with many aspects of thermoregulation adaptation. The regulation of fur length by the *PCC* gene is determined by its function in biotin transport and metabolism (Davila et al., 2020). Several reports have indicated that *PCCA* and *PCCB* gene activity is regulated by different diets and that the PCC enzyme is critical for propionate metabolism – a major substrate for gluconeogenesis in ruminants (Graber et al., 2010). Weber et al. (2013) showed that *PCCA* is associated with milk protein content and dry matter intake. In dairy cows, in given periods, for example, in the immediate postpartum period, propionate is the main source of energy via the regulation of glucose metabolism as well as supplying carbon converted to acetyl-CoA for oxidation in the TCA cycle (Kennedy et al., 2020). Interestingly, with the use of the HD panel, we also detected the *PCCB* gene located on the first chromosome. *PCCB*, as a metabolism regulator, was previously proposed as a candidate gene in cattle related to meat quality characteristics (Djari et al., 2013). Our results show that both *PCCA* and *PCCB* can be critical in cattle metabolism regulation due to being economically important in breeding. The identified region on BTA12 also contains the *NALCN* gene (sodium leak channel) responsible for nonselective cation transport, while in the vicinity of this region, a QTL associated with milk sodium content has been mapped.

Another interesting ROH cold spot region is localized at the initial part of chromosome 21 (21:802673–828905) and was identified only with the use of the 50k panel. Once again, the HD panel did show an ROH hotspot for this region, proving that the low SNP density provided by the 50k panel is sometimes not sufficient for ROH hotspot identification. Within this region, the current annotation does not contain any genes; however, there is a mapped QTL in close vicinity to this region associated with calving ease support. This potentially shows that this region is interesting with respect to Holsten cattle phenotypic traits.

In general, the obtained results confirm that high-density SNP panels can provide additional information in regard to the identification of ROH cold spot regions, and in some cases, ROH cold spot regions identified via medium-density panels can in fact be ROH hotspot regions when identified by high-density panels.

Conclusions

In summary, within the present study, we identified runs of homozygosity and heterozygosity in a vast population of Holstein cattle on the basis of two microarrays of medium (50k) and high SNP density (HD). Additionally, the genomic inbreeding coefficient (F_{ROH}) was calculated, and ROH hotspots and coldspots were identified and compared between the HD and 50k SNP panels, which were carefully examined with respect to production traits. The obtained results can shed new light on genomic regions under selection pressure in Holstein cattle and confirm that a high-density SNP panel compared to a medium-density SNP panel allows for more precise identification of ROH patterns, especially in the case of short ROH that could be associated with ancestral inbreeding. Moreover, in the case of the 50k SNP panel, when compared to the HD panel, some ROH hotspots and coldspots were not detected due to the lower density at certain genomic regions. This suggests that using medium-density panels might not be the best choice when the identification of ROH patterns is the main goal.

Conflict of interest statement

None of the authors has any other financial or personal relationships that could inappropriately influence or bias the content of the paper.

Data availability statement

The data used to support the findings are included within the article. The raw data can be shared by the corresponding author upon request.

References

Barbato O., Menchetti L., Brecchia G., Barile V.L. (2022). Using pregnancy-associated glycoproteins (PAGs) to improve reproductive management: From dairy cows to other dairy livestock. Animals, 12: 2033.

Biscarini F., Cozzi P., Gaspa G., Marras G. (2019). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. Retrieved from https://cran.rproject.org/web/packages/detectRUNS/index.html

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

Blott S., Kim J.J., Moisio S., Schmidt-Kuntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J., Georges M., Farnir F., Coppieters W. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics, 163: 253–266.

Buaban S., Lengnudum K., Boonkum W., Phakdeedindan P. (2022). Genome-wide association study on milk production and somatic cell score for Thai dairy cattle using weighted single-step approach with random regression test-day model. J. Dairy Sci., 105: 468–494.

Ceballos F.C., Joshi P.K., Clark D.W., Ramsay M., Wilson J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet., 19: 220–234.

Curik I., Ferencakovic M., Sölkner J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci., 166: 26–34.

Dai W., Zou Y., White R.R., Liu J., Liu H. (2018). Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. Funct. Integr. Genom., 18: 125–140.

Dash S., Singh A., Dixit S.P., Kumar A. (2022). Identification of selection signatures for milk performance traits among indigenous dairy cattle breeds using high density genomic information. Ind. J. Anim. Res., 10.18805/IJAR.B-4908.

Davila K.M.S., Howell A., Nunez A., Orelien A., Roe V., Rodriguez E., Dikmen S., Mateescu R.G. (2020). Genome-wide association study identifies variants associated with hair length in Brangus cattle. Anim Genet., 51: 811–814.

de Camargo G.M., Costa R.B., de Albuquerque L.G., Regitano L.C., Baldi F., Tonhati H. (2015). Polymorphisms in *TOX* and *NCOA2* genes and their associations with reproductive traits in cattle. Reprod. Fert. Develop., 27: 523–528.

Djari A., Esquerré D., Weiss B., Martins F., Meersseman C., Boussaha M., Klopp C., Rocha D. (2013). Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. BMC Genom., 14: 307.

Eisen D.P., Osthoff M. (2014). If there is an evolutionary selection pressure for the high frequency of MBL2 polymorphisms, what is it? Clin. Exp. Immunol., 176: 165–171.

El-Komy S.M., Saleh A.A., Abdel-Hamid T.M., El-Magd M.A. (2020) Association of GHR polymorphisms with milk production in buffaloes. Animals, 10: 1203.

Falaleeva M., Surface J., Shen M., de la Grange P., Stamm S. (2015). SNORD116 and SNORD115 change expression of multiple genes and modify each other's activity. Gene, 572: 266–273.

Ferencakovic M., Hamzic E., Gredler B., Curik I., Sölkner J. (2011). Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. Agric. Conspec. Sci., 76: 325–328.

Ferenčaković M., Sölkner J., Curik I. (2013). Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. Genet. Sel. Evol., 45: 42.

Fijarczyk A., Babik W. (2015). Detecting balancing selection in genomes: Limits and prospects. Mol. Ecol., 24: 3529–3545.

Fontanesi L., Scotti E., Dall'Olio S., Oulmouden A., Russo V. (2012). Identification and analysis of single nucleotide polymorphisms in the myosin VA (MYO5A) gene and its exclusion as the causative gene of the dilute coat colour locus in rabbit. World Rabbit Sci., 20: 35–41.

Forutan M., Ansari Mahyari S., Baes C., Melzer N., Schenkel F.S., Sargolzaei M. (2018). Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genom., 19: 98.

Gibson J., Morton N.E. Collins A. (2006). Extended tracts of homozygosity in outbred human populations. Hum. Mol. Genet., 15: 789–795.

Graber M., Kohler S., Kaufmann T., Doherr M.G., Bruckmaier R.M., van Dorland H.A. (2010). A field study on characteristics and diversity of gene expression in the liver of dairy cows during the transition period. J. Dairy. Sci., 93: 5200–5215.

Hadi Z., Atashi H., Dadpasand M., Derakhshandeh A., Ghahramani Seno M.M. (2015). The relationship between growth hormone polymorphism and growth hormone receptor genes with milk yield and reproductive performance in Holstein dairy cows. Iran J. Vet. Res., 16: 244–248.

Howrigan D.P., Simonson M.A., Keller M.C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. BMC Genom., 12: 460.

Jiang J., Ma L., Prakapenka D., Vanraden P.M., Cole J.B., Cole J.B. (2019). A large-scale genome-wide association study in US Holstein cattle. Front. Genet., 10: 412.

Kadri N.K., Guldbrandtsen B., Lund M.S., Sahana G. (2015). Genetic dissection of milk yield traits and mastitis resistance QTL on chromosome 20 in dairy cattle. J. Dairy. Sci., 98: 9015–9025.

Karim S., Saharti S., Alganmi N., Mirza Z., Alfares A., Turkistany S., Al-Attas M., Noureldin H., Al Sakkaf K., Abusamra H., Al-Qahtani M., Abuzenadah A. (2021). Two novel homozygous *HPS6* mutations (double mutant) identified by whole-exome sequencing in a Saudi consanguineous family suspected for oculocutaneous albinism. Life (Basel), 12: 14.

Keller M.C., Visscher P.M., Goddard M.E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics, 189: 237–249.

Kennedy K.M., Donkin S.S., Allen M.S. (2020). Effects of propionate concentration on short-term metabolism in liver explants from dairy cows in the postpartum period. J. Dairy. Sci., 103: 11449–11460.

Kim E.S., Sonstegard T.S., Van Tassell C.P., Wiggans G., Rothschild M.F. (2015). The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. PLoS One, 10(7):e0129967.

Kudo Y., Guardavaccaro D., Santamaria P.G., Koyama-Nasu R., Latres E., Bronson R., Yamasaki L., Pagano M. (2004). Role of F-box protein βTrcp1 in mammary gland development and tumorigenesis. Mol. Cell. Biol., 24: 8184–8194.

Li G., Tang J., Huang J., Jiang Y., Fan Y., Wang X., Ren J. (2022). Genome-wide estimates of runs of homozygosity, heterozygosity, and genetic load in two Chinese indigenous goat breeds. Front. Genet., 13: 774196.

Li Y., Zhou H., Cheng L., Zhao J., Hickford J. (2020). Variation in the stearoyl-CoA desaturase gene (*SCD*) and its influence on milk fatty acid composition in late-lactation dairy cattle grazed on pasture. Arch. Anim. Breed., 63: 355–366.

Ma L., Luo H., Brito L.F., Chang Y., Chen Z., Lou W., Zhang F., Wang L., Guo G., Wang Y. (2023). Estimation of genetic parameters and single-step genome-wide association studies for milk urea nitrogen in Holstein cattle. J. Dairy. Sci., 106: 352–363.

Mackiewicz D., Oliveira P.M.C., Oliveira S.M., Cebrat S. (2013). Distribution of recombination hotspots in the human genome – a comparison of computer simulations with real data. PLoS ONE., 8:e65272.

Makino T., Rubin C.J., Carneiro M., Axelsson E., Andersson L., Webster, M.T. (2018). Elevated proportions of deleterious genetic variation in domestic animals and plants. Genome Biol. Evol., 10: 276–290.

Marete A., Lund M.S., Boichard D., Ramayo-Caldas Y. (2018) A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. PLoS ONE 13(7): e0199931.

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J.L., Macciotta N.P.P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet., 46: 110–121.

Marras G., Wood B.J., Makanjuola B., Malchiodi F., Peeters K., van As P., Baes C.F., Biscarini F. (2018). Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (*Meleagris gallopavo*) population. Proc. 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand.

Mastrangelo S., Tolone M., Gerlando R.D., Fontanesi L., Sardina M.T., Portolano B. (2016). Genomic inbreeding estimation in small populations: Evaluation of runs of homozygosity in three local dairy cattle breeds. Animal, 10: 746–754.

McQuillan R., Leutenegger A.L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., MacLeod A.K., Farrington S.M., Rudan P., Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Campbell H., Wilson J.F. (2008). Runs of homozygosity in European populations. Am. J. Hum. Genet., 83: 359–372.

Melo-Baez B., Wong Y.S., Aguilera C.J., Cabezas J., Mançanares A.C.F., Riadi G., Castro F.O., Rodriguez-Alvarez L. (2020). MicroRNAs from extracellular vesicles secreted by bovine embryos as early biomarkers of developmental competence. Int. J. Mol. Sci., 21: 8888.

Mohammadi A., Alijani S., Rafat S.A., Abdollahi-Arpanashi R. (2020). Genome-wide association study and pathway analysis for female fertility traits in Iranian Holstein cattle. Ann. Anim. Sci., 20: 825–851.

Mulim H.A., Brito L.F., Pinto L.F.B., Ferraz J.B.S., Grigoletto L., Silva

M.R., Pedrosa V.B. (2022). Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. BMC Genom., 23: 209.
Nayeri S., Stothard P. (2016). Tissues, metabolic pathways and genes of key importance in lactating dairy cattle. Springer Sci. Rev., 4: 49–77.
Nothnagel M., Lu T.T., Kayser M., Krawczak M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum. Mol. Genet., 19: 2927–2935.
Pedrosa V.B., Schenkel F.S., Chen S.Y., Oliveira H.R., Casey T.M., Melka M.G., Brito L.F. (2021). Genomewide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. Genes (Basel), 12: 1830.
Pemberton T.J., Absher D., Feldmand M.W., Myers R.M., Rosenberg N.A., Li J.Z. (2012). Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet., 10: 275–292.
Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R., Baldi F. (2017). Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet., 48: 255–271.
Purfield D.C., Berry D.P., McParland S., Bradley D.G. (2012). Runs of homozygosity and population history in cattle. BMC Genet., 13: 70.
Qi Y., Purtell L., Fu M., Lee N.J., Aepler J., Zhang L., Loh K., Enriquez R.F., Baldock P.A., Zolotukhin S., Campbell L.V., Herzog H. (2016). Snord116 is critical in the regulation of food intake and body weight. Sci. Rep., 6: 18614.
Reese S.T., Geary T.W., Franco G.A., Moraes J.G.N., Spencer T.E., Pohler K.G. (2019). Pregnancy associated glycoproteins (PAGs) and pregnancy loss in high vs sub fertility heifers. Theriogenology, 135: 7–12.
Santos W., Schettini G., Fonseca M.G., Pereira G.L., Chardulo L.A., Neto O., Baldassini W.A., de Oliveira H.N., Curi R.A. (2021). Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. J. Anim. Breed. Genet., 138: 161–173.
Senczuk G., Guerra L., Mastrangelo S., Campobasso C., Zoubeyda K., Imane M., Marletta D., Kusza S., Karsli T., Gaouar S.B.S., Pilla F., Ciani E., The Bovita Consortium (2020). Fifteen shades of grey: combined analysis of genome-wide SNP data in steppe and Mediterranean grey cattle sheds new light on the molecular basis of coat color. Genes, 11: 932.
Sonzogni M., Zhai P., Mientjes E.J., van Woerden G.M., Elgersma Y. (2020). Assessing the requirements of prenatal UBE3A expression for rescue of behavioral phenotypes in a mouse model for Angelman syndrome. Mol. Autism., 11: 70.
Strucken E.M., Laurenson Y.C., Brockmann G.A. (2015). Go with the flow-biology and genetics of the lactation cycle. Front. Genet., 6: 118.
Suzuki J. Jr., Therrien J., Filion F., Lefebvre R., Goff A.K., Smith L.C. (2009). *In vitro* culture and somatic cell nuclear transfer affect imprinting of SNRPN gene in pre- and post-implantation stages of development in cattle. BMC Dev. Biol., 9: 9.
Szmatoła T., Gurgul A., Ropka-Molik K., Jasielczuk I., Zabek T., Bugno-Poniewierska M. (2016). Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. Livest. Sci., 188: 72–80.
Szmatoła T., Gurgul A., Jasielczuk I., Ząbek T., Ropka-Molik K., Litwińczuk Z., Bugno-Poniewierska M. (2019). A comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. Animals (Basel), 9: 1024.
Szmatoła T., Gurgul A., Jasielczuk I., Oclon E., Ropka-Molik K., Stefaniuk-Szmukier M., Polak G., Tomczyk-Wrona I., Bugno-Poniewierska M. (2022). Assessment and distribution of runs of homozygosity in horse breeds representing different utility types. Animals (Basel), 12: 3293.
Tian Z., Zhang Y., Zhang H., Sun Y., Mao Y., Yang Z., Li M. (2022). Transcriptional regulation of milk fat synthesis in dairy cattle. J. Funct. Foods, 96: 105208.
Topolski P., Jagusiak W. (2019). Inbreeding in a population of Polish Holstein-Friesian young bulls before and after genomic selection. Ann. Anim. Sci., 20: 71–83.
Tsartsianidou V., Sánchez-Molano E., Kapsona V.V., Basdagianni Z., Chatziplis D., Arsenos G., Triantafyllidis A., Banos G. (2021). A comprehensive genome-wide scan detects genomic regions related to local adaptation and climate resilience in Mediterranean domestic sheep. Genet. Sel. Evol., 53: 90.
van den Berg I., Fritz S., Rodriguez S., Rocha D., Boussaha M., Lund M.S., Boichard D. (2014). Concordance analysis for QTL detection in dairy cattle: a case study of leg morphology. Genet. Sel. Evol., 46: 31.
van den Berg I., Hayes B.J., Chamberlain A.J., Goddard M.E. (2019). Overlap between eQTL and QTL associated with production traits and fertility in dairy cattle. BMC Genom., 20: 291.
Viitala S., Szyda J., Blott S., Schulman N., Lidauer M., Mäki-Tanila A., Georges M., Vilkki J. (2006). The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. Genetics, 173: 2151–2164.
Vries L.D., Casey T., Dover H., VandeHaar M.J., Plaut K. (2011). Effects of transforming growth factor-β on mammary remodeling during the dry period of dairy cows. J. Dairy Sci., 94: 6036–6046.
Wang J., Lou S.S., Wang T., Wu R.J., Li G., Zhao M., Lu B., Li Y.Y., Zhang J., Cheng X., Shen Y., Wang X., Zhu Z.C., Li M.J., Takumi T., Yang H., Yu X., Liao L., Xiong Z.Q. (2019). UBE3A-mediated PTPA ubiquitination and degradation regulate PP2A activity and dendritic spine morphology. Proc. Natl. Acad. Sci. USA, 116: 12500–12505.
Wang X., Li G., Ruan D., Zhuang Z., Ding R., Quan J., Wang S., Jiang Y., Huang J., Gu T., Hong L., Zheng E., Li Z., Cai G., Wu Z., Yang J. (2022). Runs of homozygosity uncover potential functional-altering mutation associated with body weight and length in two Duroc pig lines. Front. Vet. Sci., 9: 832633.
Weber C., Hametner C., Tuchscherer A., Losand B., Kanitz E., Otten W., Sauerwein H., Bruckmaier R.M., Becker F., Kanitz W., Hammon H.M. (2013). Hepatic gene expression involved in glucose and lipid metabolism in transition cows: effects of fat mobilization during early lactation in relation to milk performance and metabolic changes. J. Dairy Sci., 96: 5670–5681.
Williams J.L., Hall S.J.G., Del Corvo M., Ballingall K.T., Colli L., Ajmone Marsan P., Biscarini F. (2016). Inbreeding and purging at the genomic level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. Anim. Genet., 47: 19–27.
Ye T., Shaukat A., Yang L., Chen C., Zhou Y., Yang L. (2022). Evolutionary and association analysis of buffalo FABP family genes reveal their potential role in milk performance. Genes (Basel), 13: 600.
Zhang H., Wu Z., Yang L., Zhang Z., Chen H., Ren J. (2021). Novel mutations in the *Myo5a* gene cause a dilute coat color phenotype in mice. FASEB J., 35: e21261.
Zhao G.F., Chen H., Lei C.Z., Zhang C.L., Zhang L., Lan X.Y., Fang X.T., Kang X.T. (2007). Study on SNPs of GHR gene and its associations with growth traits in Qinchuan cattle. Yi Chuan., 29: 319–323.
Zhou H., Cheng L., Azimu W., Hodge S., Edwards G.R., Hickford J.G.H. (2015). Variation in the bovine FABP4 gene affects milk yield and milk protein content in dairy cows. Sci. Rep., 5: 10023.
Zinovieva N.A., Dotsev A.V., Sermyagin A.A., Deniskova T.E., Abdelmanova A.S., Kharzinova V.R., Sölkner J., Reyer H., Wimmers K., Brem G. (2020). Selection signatures in two oldest Russian native cattle breeds revealed using high-density single nucleotide polymorphism analysis. PLoS One., 15(11):e0242200.

Received: 1 VIII 2023
Accepted: 4 X 2023