# Genome-Wide Homozygosity in Italian Holstein Cattle using HD SNP Panel

*G. Gaspa*[*], G. Marras[*], S. Sorbolini[*], P. Ajmone Marsan[†], J. L. Williams[‡], A. Valentini[§], C. Dimauro[*], N. P. P. Macciotta[*]

[*] Università di Sassari, Italy, [†]Università Cattolica del Sacro Cuore, Piacenza, Italy, [‡]Parco Tecnologico Padano, Lodi, Italy [§]Università della Tuscia, Viterbo, Italy.

**ABSTRACT:** High throughput genotyping techniques allow to identify long and uninterrupted stretches of homozygous genotypes named runs of homozygosity (ROH). The objective of this work was to calculate ROH in Italian Holstein Friesian bulls in order to identify genomic regions potentially under selection. A total of 2,993 bulls were genotyped using medium and high density SNP panels. A total of 161,566 and 67,915 ROH were detected using 54K and 777K panels respectively. The average number of ROH per animal was 77.2±9.5 (54K) and 74.2±15.2 (777K). Regions with a high occurrence of ROH were identified in several chromosomes. Of particular interest was a ROH hotspot identified on BTA26, where several genes involved in the metabolism of the mammary gland map. ROH could be used to detect genomic regions involved in traits of economic importance.

**Key words:** runs of homozygosity; dairy cattle; selection signatures; autozygosity.

## INTRODUCTION

Runs of homozygosity (ROH) are defined as uninterrupted stretches of DNA harboring homozygous genotypes. They represent an estimate of the degree of autozygosity at genome-wide level (Gibson et al., (2006)). The occurrence of ROH in an individual may be the result of inbreeding events but they may also be present in outbreed populations as result of other phenomena. An increased frequency of common extended haplotypes can also be a consequence of selection pressure on genomic regions involved in functional roles. In Humans, ROH have been related to the prevalence of some complex diseases (Ku et al., (2011)) and they have been used to map the recessive variants of many other disorders with high density SNP panel (>500,000 SNP). In cattle, ROH have been used to track the history of their recent selection (Purfield et al., (2012)) and to estimate molecular inbreeding coefficients (Pryce et al., (2012), Ferencakovic et al., (2013)). The present work investigates ROH distribution in Italian Holstein Friesian Bulls using both 54K and 777K SNP panels. Moreover, annotated genes that map in these IBD regions were retrieved as presumably exposed to artificial selection.

## MATERIALS AND METHODS

**Data.** We used genotypes of 3,009 Italian Holstein bulls. Two platforms were used: 2,093 bulls were assayed using medium density (MD) SNP chip (Illumina BovineSNP50), 916 bulls were genotyped using high density (HD) platform (Illumina BovineHD). All of them were genotyped in the framework of three Italian research projects. Only 25 bulls were in common between two datasets. Data quality control was performed both on animals and SNP for each dataset separately. SNP that did not map to any chromosome or that were in the X chromosome were eliminated from the dataset as well as SNP with more than 2.5% of missing data. No pruning based on LD was performed. After data editing 2,093 (44,395) and 900 (718,557) bulls (SNP) were used for MD and HD respectively.

**Runs of homozygosity detection.** A python script was designed to find uninterrupted stretches of homozygous genotypes in the analyzed bulls. The criteria used for ROH detections were: i) minimum ROH size of 15 SNP; ii) minimum length of a ROH = 1 Mb; iii) two adjacent SNP are considered in the same ROH if their relative distance <1 Mb; iv) neither heterozygous nor missing were allowed; v) no sliding windows were applied to assess the presence of a ROH (Ferencakovic et al., (2013)). The total number of ROH, the average number of ROH per animal and the sum of all ROH (Mb) per animal were calculated. The average number of SNP falling into a ROH was also calculated. With the aim of identifying extended regions of homozygosity, the number of times (%) that a SNP falls into a ROH was plotted against the position of the SNP along the chromosome.

**Gene search.** Annotated genes in genomic regions corresponding to the detected extended regions of homozygosity were derived from the UCSC Genome Browser Gateway (http://genome.ucsc.edu/). Functional annotations were derived consulting genecards database (http://www.genecards.org/).

## RESULTS AND DISCUSSION

**ROH Analysis.** The absolute number of detected ROH was higher in the MD (161,566) in comparison to HD (67,915) SNP panel. However, observed differences in the number of ROH are mainly due to the smaller sample size of HD dataset. In fact, the average number of ROH per animal was only slightly lower in HD (74.2±15.2) than in MD (77.2±9.5). The total length of ROH per animal followed the same pattern of number of ROH per animal (236.5±66.5 and 285.7±73.2 Mb for MD and HD respectively). The average number of SNP into a ROH raised from 63.5±13.5 in MD up to 920.8±171 in HD dataset, respectively. The number of SNP into a ROH largely varied across chromosomes, between 1,734 (BTA 4) and 10,849 (BTA 8) in MD and HD datasets respectively. The number of ROH per animal and the number of SNP per ROH confirmed what found by other authors in Holstein (Purfield et al. (2012)). A reduction in the number of ROH detected by the HD SNP panel compared to MD was also

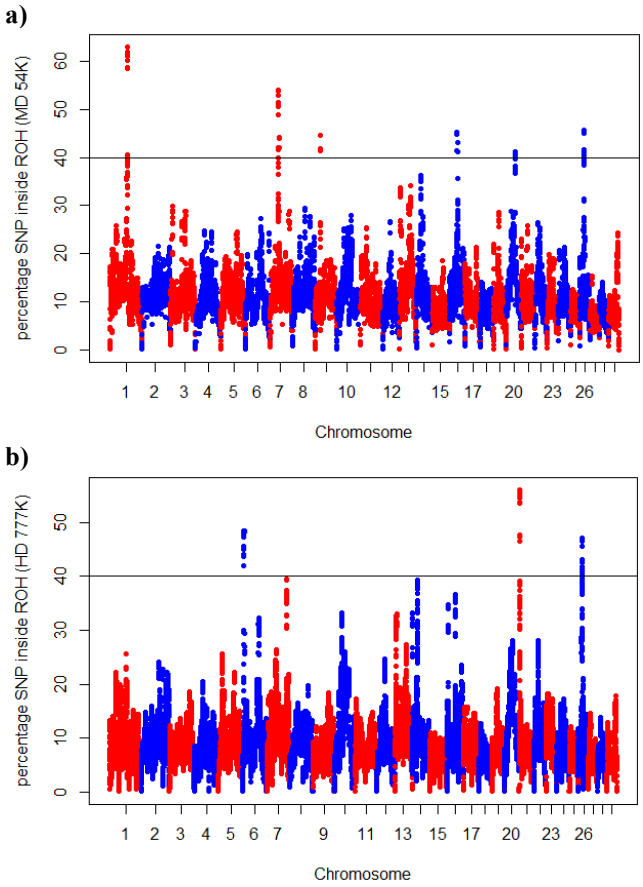


**Figure 1.** Occurrence (%) of a SNP into a ROH across animals using (a) 54K or (b) 777K SNP panels in Italian Holstein Bulls.

found by Ferencakovic et al., (2013) using other breeds. This may be due to the fact that reliable ROH (confirmed using both panels) are just those > 4Mb of length.

The distribution of ROH segments along the genome was assessed looking at the frequency of a SNP occurring into a ROH across different individuals (%). A percentage higher than the threshold of 40% was chosen as an indication of a possible ROH hotspot in the genome. Hotspot regions partially overlapped between MD and HD panels (Figure 1). In particular, 6 and 5 ROH hotspots were suggested by MD (BTA1-7-9-16-20-26) and HD (BTA6-7-9-14-21-26) panels respectively. The highest proportion of SNP falling into a ROH was found on BTA1 (>60%) and BTA21 (>50%) in MD and HD respectively. Results confirmed the presence of ROH-enriched genomic regions in cattle, as suggested by other authors (e.g. BTA 6 between 5.19-6.75 Mb) (Ferencakovic et al. (2013)). Differences between MD and HD were particularly evident on BTA1 and BTA6, whereas consensus regions were mostly located on BTA7 and BTA26. These differences probably derived from the non-uniform distribution of the ROH both in length and position along the genome. Hence, ROH detected in MD are likely not to be confirmed in HD along the genome, and this happens especially for the shortest ones (Purfield et al. (2012)).

**Gene search.** In our study an intriguing result was the peak identified on BTA26 at 21-23 Mb (Figure 2). In

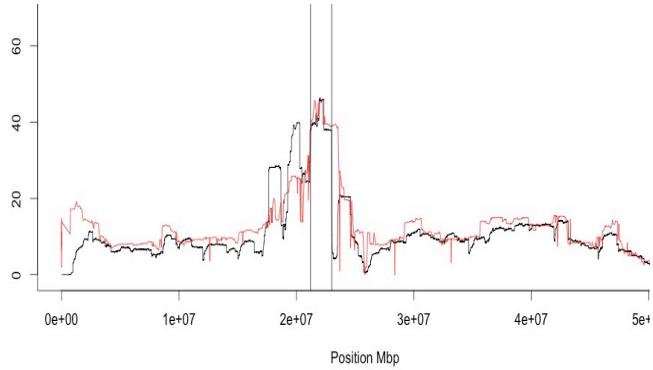


**Figure 2.** Occurrence (%) of a SNP into a ROH across Italian Holstein Bulls using 54K (red line) or 777K SNP panels (black line) on BTA 26.

this chromosomal region we observed a good agreement between MD and HD panels and a corresponding drop in the heterozygosity (-10%). A relevant number of protein coding genes (n=42) are annotated in this chromosomal region on the Btau3.1 assembly. Several of them play a relevant role in the mammary gland biology (*SCD, ELOVL3, LZTS2, FGF8, KCNIP2*). At 21,137,945-21,148,318 bp is located the *SCD* locus (Table 1). This gene encode for the *stearoyl-CoA desaturase (delta-9-desaturase)* a key enzyme in the cellular biosynthesis of unsaturated fatty acids (Macciotta et al. (2008), Alim et al., (2012)) expressed in the bovine mammary gland during lactation (Bionaz and Loor, (2008)). The *leucine zipper, putative tumor suppressor 2* (*LZTS2)* has been mentioned by Lemay et al., (2009) who carried out an extensive in silico analysis using publicly available milk proteomic data and mammary expressed sequence tags in the bovine mammary gland. Finally, also *FGF8* (*fibroblast growth factor 8*) and *KCNIP2* (*Kv channel-interacting protein 2*) are considered two genes controlling the functionality of mammary gland as they seem to be involved in human

**Table 1.** List of Genomic regions of extended homozygosity detected using HD panel and list of candidate genes.

| BTA[§] | Start, bp | End, bp | Length, bp | SNP[‡] | Genes |
|---|---|---|---|---|---|
| 6 | 5,234,762 | 6,638,045 | 1,508,258 | 42 | *MAD2L1 MGC134 093* |
| 7 | 96,949,619 | 98,336,404 | 1,386,785 | 419 | *FAM81B ARSK SPATA9 GLRX ELL2* |
| 21 | 898,385 | 1,829,761 | 1,180,466 | 77 | |
| 26 | 17,638,682 | 18,680,250 | 1,041,568 | 323 | |
| | 21,146,794 | 23,000,155 | 1,853,361 | 412 | *SCD ELO VL3 LZTS2 FGF8 KCNIP2* |

§ Bos Taurus Autosome (BTA) that presented a percentage of SNP into a ROH ≥ 40% across animals.

‡ Number of Single Nucleotide Polymorphisms (SNP) included in a ROH.

breast cancer (Hens and Wysolmerski (2005), Saito-Hisaminato et al., (2002)) (Table 1).

## CONCLUSIONS

Results obtained in the present work highlight differences in the detection and in the distribution of ROH between MD and HD panels, respectively. However, for specific genomic regions partially overlapping results were found. ROH hotspots were identified in different chromosomes. In particular, an interesting region of ~2 Mb was identified on BTA26. This region harbors some genes involved in the metabolism of the mammary gland. These results suggest that in cattle ROH may contribute to detect genomic regions involved in the determinism of traits of economic importance.

## ACKNOWLEDGMENTS

Research funded by the Italian Ministry of Agriculture (grants INNOVAGEN and SELMOL) and Fondazione CARIPLO (grant PROZOO).

## LITERATURE CITED

Alim, A. M., Fan ,Y. P., Wu X. P. et al. (2012). Mol. Biol. Rep., 39:8733-40.

Bionaz, M., and Loor, J. (2008). BMC Genomics, 9:366-387.

Ferencakovic, M., Solkner, J. and Curik, I. (2013). Genet. Sel. Evol., 45, 42.

Gibson, J., Morton, N. and Collins, A. (2006). Hum. Mol. Genet., 15, 789 - 795.

Hens, J. and Wyslomerski, J. J. (2005). Breast Cancer Res., 7:220-224.

Ku, C. S., Naidoo N., Teo, S. M. et al. (2011). Hum. Mol. Genet. 129, 1-15.

Lemay, D. G., Lynn D. J., Martin, W. F. et al. (2009). Genome Biol., 10:R42.

Macciotta, N. P. P., Mele, M., Conte, G. et al. (2008). J. Dairy Sci. 91, 3184-3189.

Pryce, J. E., Hayes, B. J. and Goddard, M. E. (2012). J. Dairy Sci. 95, 377-388.

Purfield, D., Berry, D., McParland, S. et al. (2012). BMC Genet. 13, 70.

Saito-Hisaminato, A., Katagiri, T., Kakiuchi, S. et al. (2002). DNA Res., 9:35-45.