Genome-wide analysis of homozygosity regions in european simmental bulls

Original
Genome-wide analysis of homozygosity regions in european simmental bulls / Cesarani, A.; Gaspa, G.; Pauciullo, A.; Degano, L.; Vicario, D.; Macciotta, N. P. P.. - In: JOURNAL OF ANIMAL BREEDING AND GENETICS. - ISSN 0931-2668. - (2020). [10.1111/jbg.12502]

Published
DOI:10.1111/jbg.12502

Genome-wide analysis of homozygosity regions in european simmental bulls / Cesarani, A.; Gaspa, G.; Pauciullo, A.; Degano, L.; Vicario, D.; Macciotta, N. P. P.. - In: JOURNAL OF ANIMAL BREEDING AND GENETICS. - ISSN 0931-2668. - (2020).

Genome-wide analysis of Homozygosity Regions in European Simmental Bulls.

Running title: Genomic characterization of European dual-purpose cattle

Alberto Cesarani,[1,2] Giustino Gaspa[3], Alfredo Pauciullo[3], Lorenzo Degano[2], Daniele Vicario[2], Nicolò P.P. Macciotta[1]

[1] Department of Agricultural Sciences, University of Sassari, Sassari, Italy
[2] Associazione Nazionale Allevatori Pezzata Rossa Italiana (ANAPRI), Udine
[3] Department of Agricultural, Forestry and Alimentary Sciences, University of Torino, Grugliasco

Correspondence: Giustino Gaspa, giustino.gaspa@unito.it, Department of Agricultural, Forestry and Alimentary Sciences, University of Torino, Grugliasco

Abstract

Runs of homozygosity (ROH) has become a useful genetic tool to study several population phenomena. Due to their high relationship with autozygosity, ROH allows to infer population history, to estimate inbreeding level and tracking selective pressure in livestock genomes. Within-breed heterogeneity can be assessed by analyzing ROH pattern among individuals. Aim of this study was to give an overview about runs of homozygosity and the relationship between three production traits (milk yield, fat and protein contents) and autozygosity in bulls belonging to five European Simmental population. ROH count, distribution and ROH based inbreeding (F_{ROH}) were calculated for 3,845 Simmental bulls coming from five different European countries: Austria (AT), Switzerland (CH), Czech Republic (CZ), Germany (DE) and Italy (IT). Overall, average values of 77.8±20.7 and 205±74.4 Mb were found for number of ROH per animal and total genome length covered by ROH, respectively. Similar ROH features were highlighted for AT, DE and IT. Swiss bulls showed the highest F_{ROH} coefficient (12.6%), while CZ animals the lowest (4.6%). The relationship between ROH occurrence and milk production traits was investigated through a ROH-traits association at genome-wide level (GWRA). The GWRA approach allowed to identify 34 regions previously associated with milk (yield and/or composition). The ROH analyses presented in this study allowed to identify a mixed background in the Simmental breed, with the possible presence of three subgroups, and a strong relationship between autozygosity and production traits.

KEYWORDS: cattle, autozygosity, milk traits, single nucleotide polymorphism

1. INTRODUCTION

Current cattle breeds are the result of the evolutionary forces that left traces in their genome. Selective breeding of animals sharing similar performances and exterior characteristics transmittable to their progeny allowed to establish modern breeds and represented one of the pillars of animal breeding. In particular, artificial selection led to a dramatic increase of production levels but together with a reduction of the genetic variability within breed and increase of inbreeding level (Brotherstone and Goddard, 2005; Baes et al., 2019; Doekes et al., 2019). This had resulted in the generation of relatively few highly productive specialized breeds that spread worldwide during the last part of the 20th century to the detriment of local populations (Felius et al., 2011). In spite of such cattle genetics globalization, still genetic differences can be found between and within populations. Other factors such as geographical barriers, migratory events, crossbreeding, and specific different breeding goals among countries have contributed to create genetic differences between and within populations.

The Simmental cattle population is one example. Autochtonous of Switzerland, it is now spread worldwide counting about 40 Million heads (http://wsff.info). In spite of the common origin, the different Simmental populations have undergone to different selection histories that led to different production aptitudes. In the United States and in some other Countries (e.g. UK, Ireland or South Africa) the Simmental is farmed for meat production, whereas in Europe the Simmental is mainly a dual-purpose breed even with different emphasis on dairy/beef according to the breeding objectives. For example, the breeding program in Italy is oriented for 44% on milk production, 24% on meat production, 19.5% on morphology and 12.5% on fitness.

Cattle genomes sequencing has provided powerful tools for investigating genomic features (Bovine HapMap Consortium, 2009). Among these, the uninterrupted sequences of homozygous genotypes, known as Runs of Homozyogosity (ROH) have been widely investigated for many livestock species (Peripolli et al., 2017). Since the ROHs' occurrence could be attributed to their inheritance from a common ancestor (Keller et al., 2011; Purfield et al., 2012; Gurgul et al., 2014), these homozygous segments are useful for analyzing the degree of relatedness among individuals belonging to different livestock populations (Mastrangelo et al., 2016; Bertolini et al., 2018; Cesarani et al., 2019). Moreover, the ROH length allows for estimating the time distance of inbreeding occurrence: long ROH segments are likely to occur when the mating between relatives have taken place recently (Curik et al., 2014). The proportion of the genome covered by ROH is an estimate of the degree of autozygosity (Howrigan et al., 2011) and enable to infer the inbreeding in livestock (Purfield et al., 2012; Marras et al., 2015; Peripolli et al., 2018). ROH mapping has also been carried out to selection signature detection (Kim et al., 2013; Purfield et al., 2017; Kim et al., 2018). In this respect, the increased livestock performances due to selective breeding of bulls, often more related than average, can leave a traces in their genomes. Therefore, in a population undergoing selection the alleles involved in the determinism of a quantitative traits increase in frequency and tend to the fixation. Alike, the alleles in the neighboring loci does due to linkage disequilibrium resulting in an increase of regional autozygosity and, therefore, the genome proportion covered by ROH. This has been observed in a study lead by Kim et al. (2018) whose conclusions support the hypothesis of increased ROH occurrence in highly selected animals. For all these reasons, recently the attention of researchers' community is focusing on the relationship between ROH and economic traits in selected livestock population (Ferencakovic et al., 2017; Kim et al., 2018; Marchesi et al., 2018).

In the present study, the genomic structure of Simmental bulls of different European Countries has been investigated analysing ROH features in order to elucidate the variability among the different populations and to study relationships between ROHs and production traits.

2. MATERIALS AND METHODS

2.1. Data

Genotypes of 3,845 Simmental bulls have been provided by the Italian Simmental breeder association (ANAPRI). Animals from five different Countries were analyzed: Austria (AT 351 animals), Switzerland (CH, 215), Czech Republic (CZ, 248), Germany (DE, 550) and Italy (IT, 2481). A part of the genotypes came from international exchanges involving some of the European breeder associations. Since, the bulls were genotyped with chipset at different density (50K v1,v2-v3, 777K and custom chipset) BovineSNP50K Beadchip (http://www.Illumina.com) were taken as reference panel and only SNP in common among those panels were retained. Quality control of SNP data was the same used for national genomic evaluation and it was carried out with inhouse python script according the following settings: SNP were kept if minor allele frequency (MAF) >2%, call rate >97.5%, Hardy-Weinberg Equilibrium (HWE) p-value was >0.01. Moreover, SNPs mapping on sexual chromosomes or not mapped on UMD3.1 release were discarded. After data editing 40,099 SNP were retained for ROH detection and further analysis.

MACE proofs (MEBV) rescaled on Italian scale were made available for milk yield (MY, kg), milk fat (FP, %) and milk protein (PP, %). Bulls genotypes and MEBV are owned ANAPRI.

A multidimensional scaling analysis (MDS) based on raw Hamming distances was performed using the whole SNPs panel with PLINK software v1.9 (Chang et al., 2015). The first two principal dimensions were computed from the distance matrix.

2.2. Runs of homozygosity detections

Runs of homozygosity were computed using the algorithm implemented in “DetectRuns” package with function “ConsecutiveRUNS” (Biscarini et al., 2018) implemented in R software (R Core Team, 2019). This package also provides the “SlidingRUNS” function, which implements a sliding windows approach (similar to PLINK). More in general, different packages rely on different default ROH settings. However, in ROH computation common settings are not well established yet, (at least for livestock populations) and simulations or empirical studies formulate general advises only (e.g. Howrigan et al. 2011; Ferenkacovic et al. 2013). Since no “standard” methods have been established we reckon that the methodology developed in our previous paper best fit to our aims (Marras et al., (2015). Briefly,15 consecutive homozygous SNPs covering a segment of at least 1 Mb length were considered as the minimum threshold for defining a ROH. Neither heterozygous nor missing SNPs were admitted for ROH calling. Descriptive statistics were calculated on detected ROH: average and total number of ROH (n); average and total ROH length (Mb); average and total number of ROH count per SNP (n). When a ROH occurs only in an animal then it is labeled as non-repeated ROH (NR-ROH), otherwise (i.e. shared among 2 or more bulls) it is considered as repeated ROH (RP-ROH). Moreover, the same repeated ROH can appear in >2 animals. Thus, it is likely that the same ROH might be counted from 2 up to 3,845 times (total number of bulls). For these reasons we have counted those ROH once and labelled them as repeated-non-duplicated ROH (RPN-ROH). Finally, both the number of private ROHs (PR-ROH, occurring in bulls of one country but not in the others) and shared ROH between animals of different countries were computed (SH-ROH).

For each individual belonging to different countries, inbreeding coefficients based on ROH (F_{ROH}) was calculated as the percentage ratio between genome covered by ROH (L_{ROH}) and genome total length (L_{gen}), considering a total of about 2.57 Gbps.

2.3. Genome-wide ROH analysis (GWRA)

The effect of each unique ROH occurrence (presence or absence) was tested using the following linear model simulating a genome-wide association analysis using the R software (R Core Team, 2019):

$$y_{ijkl} = country_j + Yob(country)_{k(j)} + ROH_l + e_{ijkl} \quad (1)$$ where

- y_{ijkl} is the investigated trait (MEBV for milk, fat and protein) of i-th bulls;
- $Yob(country)_{k(j)}$ is the fixed effect of *k*-th year of birth (7 levels: 1=1972-1989; 2=1990-1993; 3=1994-1997; 4=1998-2001; 5=2002-2005; 6=2006-2009; 7=2010-2015) nested in the *j*-th country of birth (5 levels);
- ROH is the effect of the *l-th ROH* analyzed (2 levels, presence or absence);
- e_{ijkl} is the random residuals.

This model was performed for each ROH detected considering ROHs shared by at least 20 animals. Finally, in order to avoid false positive discovery, Bonferroni correction was applied dividing the significance level (type I error α=0.05) by the number of ROH tested (n=1,344). Using online databases (http:\\www.genomebrowser.org, http:\\www.Ensembl.org, http:\\www.GeneCards.org), annotated genes mapped in the identified genomic regions were highlighted as suitable candidate genes. Moreover, using milk, fat and protein as trait, mapped QTLs were retrieved in the online database (htttp:\\www.animalgenome.org).

3. RESULTS

The first two dimensions of the MDS plot (Figure 1) accounted for 2.61%, 1.52% of the variance, respectively. As far as the average values per country were considered, the first dimension allowed to separate AT, CH and DE (negative values) from CZ and IT (positive values), whereas the second dimension separated CH (positive values >0.05) from the other countries, with negative values (IT) or values very close to zero (AT, DE and CZ).

Figure 1

A total of 304,100 ROHs were detected with an average number per animal of 82.1±20.9. The average genome length covered by ROH per animal was 205.5±73.4 Mb (3,285±1,190 SNP long). Very similar number of ROH per animal was found for IT, AT and DE bulls (Table 1). Conversely, much lower and higher number ROH segments was observed for CZ and CH respectively. Similarly, CH bulls showed the largest ROH length (average and total) while CZ the smallest (Table 1).

Table 1

Total ROH length (Mb) and the number of ROH segment were positively correlated. The coordinate of each individual ROH feature clustered around the Country of origins (Figure 2) whereas bulls from Switzerland and Czech mapped in the opposite side presenting the highest and the lowest genome covered by ROH. A moderate linear association between the number of ROH per animal (nROH) and the mean ROH length (r= 0.30) was found, while all other ROH features were from highly (r=0.77) to strongly associated (r=0.97), all pairwise correlations were highly significant (p-value <0.001, Supplementary Table 1)

Figure 2

Regardless the ROH statistics considered, the number of detected ROHs was much higher for IT than for the other countries due to the sample size effect. Therefore, looking at ROH proportion, we have observed that despite about three-quarters of the ROH detected were RP-ROH (i.e. shared among at least two bulls regardless the country of origin) half these segments occurred less than four times (Table 2 and Supplemental Figure 1). Moreover, two countries (CH and CZ) were found to have lower than average proportion of RP-ROH. Since RP-ROH can be counted more than one animal, removing those duplicate, only 39% (119,005 of both repeated and non-duplicated) of the total detected ROH turned out in RN-ROH. The proportion of such segments ranged across different country from 43% of IT to 54% CH, the former having the highest number per bulls (61.6) (Table 2).

Supplementary Figure 1 and Table 2

Furthermore, we looked at those ROHs that were not shared between animals of different countries (i.e. typical of each populations and hereafter defined as “private”) and we found out that about 30% of the total ROH is classifiable as private ROHs for a total of 92,226 segments (Table 3). The occurrence of private ROH detected in Simmental bulls pointed out a difference among country of birth. Looking at the proportion of private ROH on the total number of segments detected within country, IT, DE and AT accounted for 25-30% of private ROH, whereas the remaining two had much higher figure (up to 42% of CH) (Figure 3). The average number of private ROH per bulls was 27.1±10.7 with the highest figure for CH bulls (48.5 ROH/bull)

Table 3 and Figure 3

Table 3 also shows the number of shared ROH among countries. ROH sharing among all possible pairs of countries partially reflected the pattern observed in figure 1, with the smallest and the largest values between CZ-CH and IT-DE, respectively. The raw number of the shared ROH is inflated by the sample sizes of bulls’ country of birth, for this reason we also reported the spearman rank correlation coefficients among all possible pairs of comparison. The correlations pattern indicated that DE-IT, DE-AT, IT-AT are the pair with the higher association. A complete list of ROH shared among all the possible combination of countries were provided in Supplementary Table 2. Among the 119,005 shared ROH, only 428 regions were identified in all the considered countries (Supplementary Table 3 and Figure 4). The chromosome with the highest number of shared regions was BTA6 (46 ROHs), while the chromosome with the lowest number of shared ROH was BTA25 (with just one ROH found in common among all countries). Thirty-four ROHs were shared at least by 100 different animals. The highest shared ROH was found in 432 animals, on BTA6 starting at 38.46 and ending at 39.5 Mb, followed by one repeated segment found in 245 different animals, on BTA9 at 0.12-1.59 Mb.

Figure 4

The mean ROH inbreeding coefficient ranged from 4.7% (CZ) to 12.6 % (CH), with an overall average value of 7.5% ± 2.9%. The distribution of F_{ROH} coefficients in four classes was calculated in the five considered countries (Table 4).

Table 4

In the GWRA analysis, a total of 1,344 ROHs shared at least by 20 animals were considered. The Manhattan plot reporting the $-\log_{10}$ of p-values for the three considered traits were shown in Figure 5. A total of 22, 16 and 27 regions were found as significant for MY, FP and PP, respectively.. The investigated traits showed almost the same significant regions (Supplementary Table 4). In particular, nine regions in common among all three traits were detected on chromosomes 2, 5, 11, 13, 20, 23 and 29. Moreover, the similarity of results for milk and protein was even stronger: other eight regions were found in common between these two traits, with BTA 1 and 11 showing the highest number of significant ROHs. As far as the non-duplicated significant regions were considered, a total of 34 ROHs were highlighted (Table 5).

Figure 5 and Table 5

4. DISCUSSION

In the 20th century, livestock populations underwent to strong artificial selection aimed at improving their production levels. The selection yielded traces in the animal genomes: the main consequence has been the increase of consanguinity and autozygosity levels and, therefore, the increase of ROH occurrence. The majority of studies on ROH in livestock have dealt with ruminants species (Peripolli et al., 2017; Kim et al., 2013; Mastrangelo et al., 2016; Purfield et al., 2017) and in particular cattle, but also other farm animals have been investigated (Bosse et al., 2012; Metzger et al., 2015).

Excluding CH and CZ, the number of ROH per animal and their length are in agreement with those reported in literature for some European Simmental population (Ferencakovic et al., 2013; Marras et al., 2015, Szmatola et al., 2016) but were much higher than those found by Signer-Haser et al. (2017) in Swiss cattle populations. In particular, Szmatoła et al., (2016) found 81.5±11.8 ROH/animals and 201.8±99.4 Mb of total ROH length. A slightly higher number of ROH per animal (94.3±12.2) but a similar total ROH length (210 Mb) were observed by Marras et al., (2015) for Italian Simmental. Ferenčaković et al., (2011) analyzing ROH in the Fleckvieh cattle (i.e. Austrian Simmental) found moderately larger ROH features: 96.79±13.37 ROH/animal in comparison to 80.62±15.23 from the current investigation (229.25 Mb vs. 195.29±55.82 Mb of total ROH length). Recently, Singer-Haser et al. (2017) analyzed ROH pattern in two related population reporting 7.1 and 10.9 ROH/animal for Swiss Fleckvieh and Simmental, respectively. Furthermore, they reported an average total length of 75.6 Mb (Swiss Fleckvieh) and 96.6 Mb (Simmental). The differences between their estimates and our findings might be explained by the more stringent settings used by Signer-Hasler et al. (2017) to call a ROH: 50 homozygous SNP (here 15 SNP) and 2 missing markers allowed inside a ROH (we did not allow any missing marker inside ROH). Moreover, they used a panel with 27,612 autosomal markers, i.e. nearly a half of the SNPs used in the present study.

The nROH/animals reported for Holstein and Brown Swiss were 81.7 (Marras et al., 2015) and 98.9 (Ferenčaković et al., 2013) respectively. In some cases, local cattle breeds exhibited even lower values: 72.3 ROH/animal for Tyrol Grey (Ferenčaković et al., 2013) and 46.4 for Polish Red (Szmatoła et al., 2016).

The Simmental groups involved in this study exhibited a variable ROH statistics: CH bulls showed higher values than cosmopolitan breeds (e.g. Holstein and Brown Swiss); AT, DE and IT revealed intermediate values and CZ exhibited lower values than local breeds.

Results about shared ROHs suggested that the most similar chromosome was the BTA6, especially the region between 71.11-72.03 Mb. The association between *KIT* gene and coat color, especially with degree of spotting, is well-known in cattle (Reinsch et al., 1999; Fontanesi et al., 2010; Hayes et al., 2010; Brening et al., 2013; Signer-Haser et al. 2017), but also in other species such as horse (Brooks et al., 2006; Haase et al., 2009), pigs (Moller et al., 1996) and rabbit (Fontanesi et al., 2014). Although coat color was not considered as phenotype in the current study, due the characteristic spotted coat of Simmental breed, the presence of *KIT* gene in the ROH region could suggest a possible explanation for the observed genetic similarity

The differences found in ROH features among the countries could be ascribed to their different genetic background, management and by the degree of sire exchanges among countries. The CH Simmental population is restrained with only 10,000 cows in pureness. The use of foreign bull semen inside the Switzerland is not so spread and Swiss male are mainly offspring of other national bulls. From this perspective, the likelihood for an animal to have a higher degree of consanguinity might partially explaining the F_{ROH} observed for this group. Conversely, a stronger trade and spreading of male semen can be observed among Austria, Germany and Italy. Bulls born in Italy are, for example, mostly sired by AT and DE imported bulls and their female offspring are mainly mated to different bulls coming from these countries. As far as IT Simmental is concerned lower F_{ROH} are the due to the combination of the adoption of a juvenile breeding scheme and limited number of semen straw stocked per bull (www.anapri.eu). A parallel exchange of bulls from Italy to AT and DE can be observed, even though of weaker intensity. Thus AT, IT and DE bulls can be considered as belonging to a large population, as confirmed by the results of the present study (Table 1, Figure 1-2). The lowest values showed by CZ bulls could be due to the larger population size of this group and to the large use of Red Holstein, Ayrshire and Montbeliarde bulls on CZ cows. It is worth remembering that until 80’s, CZ animals were considered as a crossbreed population (http://wsff.info). Results of the present study highlighted the presence of a common genetic base but also a large variability among the European Simmental bulls. CH population showed ROH characteristics similar to those that can be found in strongly selected breeds, while CZ populations seemed closer to local breeds. Bulls coming from AT, DE and IT showed halfway characteristics. These results suggest the presence of at least three different subgroups (AT-DE-IT, CH, CZ) in the five European Simmental populations that we analyzed. These subgroups could be identified based on similar characteristics (nROH, ROH length, F_{ROH} inbreeding coefficients) found in AT, DE and IT and different characteristics found in the other two countries. Moreover, the results of MDS approach reflected higher genetic similarity of AT-DE-IT bulls in comparison to CH and CZ.

The inbreeding class 5-10% grouped the highest number of animals for all countries, except CH that showed more animals in the 10-20% class. Only Switzerland and Italy presented animals in the upper class of inbreeding (>20%) with 3 and 5 animals, respectively. As far as the overall F_{ROH} coefficient (7.5±2.9%) is concerned, close values were found for Simmental breed by Marras et al., (2015) and Szmatoła et al., (2016), while slightly higher (9±2%) by Ferenčaković et al., (2011). As expected, F_{ROH} average coefficients highlighted in this study were lower than those reported for other cosmopolitan breeds that were subjected to stronger selection: e.g. 12.7%, 14.5% and 15.6% for Brown Swiss (Cesarani et al., 2018; Marras et al., 2015; Ferenčaković et al., 2013); and 11.6% for Holstein (Marras et al., 2015; Szmatoła et al., 2016). ROH-based inbreeding in Swiss cattle populations was recently investigated by Signer-Hasler et al. (2017), who reported lower F_{ROH} values for Swiss Fleckvieh (2.9±2.1%) and Simmental (3.9±2.3%), whereas we found an FROH value of (12.6±2.9%). Despite these values are hardly comparable to those reported in the present study (F_{ROH} derived from different ROH settings) Signer-Hasler recorded an higher degree of inbreeding for the original Simmental population in comparison to Swiss Fleckvieh.

The analysis of F_{ROH} performed separately by country still highlighted a heterogeneous pattern. In particular, CH bulls showed F_{ROH} values (12.6%) closer to those reported for cosmopolitan breeds, while inbreeding value found for CZ (4.7%) was of the same magnitude of the values reported in literature for indigenous and not selected breeds; for example, Szmatoła et al. (2016) reported a F_{ROH}

value of 5.7% for Polish Red. In addition, the different selective pressure on different countries might explain the F_{ROH} coefficient results. In this respect, AT, DE and IT showed very close values, with CH and CZ showing higher and lower values, respectively.

Gene discovery carried out in the significant regions identified by GWRA approach highlighted some interesting genes and a certain degree of overlapping with known QTL regions. Half of the significant regions (17) overlapped, completely or partially, with known QTLs: a total of 43 QTLs were found (Supplementary Table 4) to be associated with milk production traits (yield and composition) in cattle. Among the genes found in the genomic regions, some of them were already associated with milk traits in literature. *UDP-Glucose Pyrophosphorylase 2* (*UGP2*) gene (BTA11, 62,244,354-62,299,942 bp), significant for all the three traits, was reported to be associated with lactose synthesis (Connor et al., 2008; Littlejohn et al., 2009; Osorio et al., 2016). In particular, the *UGP2* gene expression was found to be downregulated (Connor et al., 2008) and upregulated (Littlejohn et al., 2009) when daily milking frequency was reduced or increased, respectively. *ADP Ribosylation Factor Related Protein 1* (*ARFRP1*), mapped on BTA13 (54,519,582-54,525,755 bp), was related to lactation performance by Doelman (2011). In the same study, also the *Enolase 2* (*ENO2*), mapped on BTA5 at 103,899,877-103,906,801 bp, and *Nudix Hydrolase 16* (*NUDT16*), mapped on BTA1 at 140,044,682-140,046,219 bp, genes were associated with lactation performance. The *Carbonic Anhydrase 6* (*CA6*), mapped on BTA16 (45,289,116-45,313,929 bp) and found as significant for protein, gene was reported to be associated with lactogenesis in human (Sharp et al., 2016). The *CD27 Molecule* (*CD27*) gene was mapped on BTA5 (104,302,582-104,308,155 bp) in a region found as significant for all three traits; this gene was reported to be associated with milk production traits and, in particular, with protein (Raven et al., 2014). In the same significant shared ROH, also the *CD4 Molecule* (*CD4*) gene was mapped (BTA5, 103,994,655-104,010,513 bp); He et al., (2011) associated this gene with somatic cell score and milk/protein yields. The *Galectin Like* (*LGALSL*) gene was mapped on BTA11 (62,812,719-62,820,736 bp) in a ROH found as significant for both milk and protein. Bai et al., (2016), counterpoising high yielders and low yielders, found a different expression level of *LGALSL* gene. The *Prostaglandin F Receptor* (*PTGFR*) gene, mapped on BTA3 (66,383,734-66,445,668 bp) in a ROH significant for fat and protein, was already reported to be associated with fat yield in Canadian Holstein Cattle (Li et al., 2010).

5. CONCLUSION

Results of the present study confirmed the usefulness of Runs of Homozygosity as a tool for investigating genomic features, because they are able to disentangle peculiar features of the genome allowing a more complete characterization of considered populations. The variability existing among different groups of the Simmental population farmed in different European countries has been investigated and differences related to gene flow and different genetic management have been highlighted. A possible existence of three main sub-groups of European Simmental could be suggested.

ACKNOWLEDGMENTS

Financial support for this research was provided by the Italian ministry of Agriculture (project *DUAL BREEDING* grant no. J21J18000010005). Corresponding author was financially supported by the University of Turin (*grant. GASG_RILO_19_01 –* “*Ricerca Locale” 2019*). Parts of the genotype data were provided by the international consortium responsible for genetic and genomic evaluation of Brown Swiss and Simmental cattle in Germany, Austria and the Czech Republic, which is gratefully acknowledged. The authors are also thankful to Swiss herd book cooperative Zollikofen, to provide Swiss bull genotypes.

CONFLICT OF INTEREST

The authors declare that they have no competing interest.

AVAILABILITY OF DATA

Data subject to third party restrictions. The data that support the findings of this study are available from ANAPRI. Restrictions apply to the availability of these data, which were used under license for this study.

REFERENCES

Baes, C. F., Makanjuola, B. O., Miglior, F., Marras, G., Howard, J. T., Fleming, A., & Maltecca, C. (2019). Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. *Journal of Dairy Science, 102*(3), 2807-2817. doi:10.3168/jds.2018-15520

Bai, X., Zheng, Z. Q., Liu, B., Ji, X. Y., Bai, Y. S., & Zhang, W. G. (2016). Whole blood transcriptional profiling comparison between different milk yield of Chinese Holstein cows using RNA-seq data. *Bmc Genomics, 17*. doi:10.1186/S12864-016-2901-1

Bertolini, F., Cardoso, T. F., Marras, G., Nicolazzi, E. L., Rothschild, M. F., Amills, M., & consortium, A. (2018). Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution, 50*. doi:10.1186/s12711-018-0424-8

Biscarini, F., Cozzi, P., Gaspa, G., Marras, G., 2018. Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes. CRAN.Packages

Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A. F., Schook, L. B., . . . Groenen, M. A. M. (2012). Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. *Plos Genetics, 8*(11). doi: 10.1371/journal.pgen.1003100Brooks, S. A., Lear, T. L., Adelson, D. L., & Bailey, E. (2007). A chromosome inversion near the KIT gene and the Tobiano spotting pattern in horses. *Cytogenetic and Genome Research, 119*(3-4), 225-230. doi:10.1159/000112065

Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science, 324(5926), 528-532.

Brotherstone, S., & Goddard, M. (2005). Artificial selection and maintenance of genetic variance in the global dairy cow population. *Philosophical Transactions of the Royal Society B-Biological Sciences, 360*(1459), 1479-1488. doi:10.1098/rstb.2005.1668

Brenig, B., Beck, J., Floren, C., Bornemann-Kolatzki, K., Wiedemann, I., Hennecke, S., . . . Schutz, E. (2013). Molecular genetics of coat colour variations in White Galloway and White Park cattle. Animal Genetics, 44(4), 450-453. doi:10.1111/age.12029

Cesarani, A., Sechi, T., Gaspa, G., Usai, M. G., Sorbolini, S., Macciotta, N. P. P., & Carta, A. (2019). Investigation of genetic diversity and selection signatures between Sarda and Sardinian Ancestral black, two related sheep breeds with evident morphological differences. *Small Ruminant Research, 177*, 68-75. doi:10.1016/j.smallrumres.2019.06.014

Connor, E. E., Siferd, S., Elsasser, T. H., Evock-Clover, C. M., Van Tassell, C. P., Sonstegard, T. S., . . . Capuco, A. V. (2008). Effects of increased milking frequency on gene expression in the bovine mammary gland. *Bmc Genomics, 9*. doi:10.1186/1471-2164-9-362

Curik, I., Ferencakovic, M., & Solkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science, 166*, 26-34. doi:10.1016/j.livsci.2014.05.034

Doekes, H. P., Veerkamp, R. F., Bijma, P., de Jong, G., Hiemstra, S. J., & Windig, J. J. (2019). Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. *Genetics Selection Evolution, 51*(1). doi:10.1186/s12711-019-0497-z

Doelman, J. (2011). Hepatic Gene Expression Profiling to Predict Future Lactation Performance in Dairy Cattle (Doctoral dissertation).

Ferencakovic, M., Hamzic, E., Gredler, B., Curik, I., & Solkner, J. (2011). Runs of homozygosity reveal genome-wideautozygosity in the Austrian fleckvieh cattle. *Agric Conspec Sci, 76*, 325 - 328.

Ferencakovic, M., Hamzic, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I., & Soelkner, J. (2013). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics, 130*(4), 286-293. doi:10.1111/jbg.12012

Ferencakovic, M., Solkner, J., Kaps, M., & Curik, I. (2017). Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. Journal of Dairy Science, 100(6), 4721-4730. doi:10.3168/jds.2016-12164

Fontanesi, L., Tazzoli, M., Russo, V., & Beever, J. (2010). Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus. *Animal Genetics, 41*(3), 295-303. doi:10.1111/j.1365-2052.2009.02007.x

Fontanesi, L., Vargiolu, M., Scotti, E., Latorre, R., Pellegrini, M. S. F., Mazzoni, M., . . . De Giorgio, R. (2014). The KIT Gene Is Associated with the English Spotting Coat Color Locus and Congenital Megacolon in Checkered Giant Rabbits (Oryctolagus cuniculus). *Plos One, 9*(4). doi:ARTN e93750

Fontanesi, L., Vargiolu, M., Scotti, E., Latorre, R., Pellegrini, M. S. F., Mazzoni, M., ... & De Giorgio, R. (2014). The KIT gene is associated with the English spotting coat color locus and congenital megacolon in Checkered Giant rabbits (Oryctolagus cuniculus). PLoS One, 9(4), e93750.
10.1371/journal.pone.0093750

Gibson, J., Morton, N., & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics, 15*, 789 - 795.

Gurgul, A., Semik, E., Pawlina, K., Szmatoa, T., Jasielczuk, I., & Bugno-Poniewierska, M. (2014). The application of genome-wide SNP genotyping methods in studies on livestock genomes. *Journal of Applied Genetics, 55*(2), 197-208. doi:10.1007/s13353-014-0202-4

Haase, B., Brooks, S. A., Tozaki, T., Burger, D., Poncet, P. A., Rieder, S., . . . Leeb, T. (2009). Seven novel KIT mutations in horses with white coat colour phenotypes. *Animal Genetics, 40*(5), 623-629. doi:10.1111/j.1365-2052.2009.01893.x

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., & Goddard, M. E. (2010). Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. *Plos Genetics, 6*(9). doi:10.1371/journal.pgen.1001139

He, Y. H., Chu, Q., Ma, P. P., Wang, Y. C., Zhang, Q., Sun, D. X., . . . Zhang, Y. (2011). Association of bovine CD4 and STAT5b single nucleotide polymorphisms with somatic cell scores and milk production traits in Chinese Holsteins. *Journal of Dairy Research, 78*(2), 242-249. doi:10.1017/S0022029911000148

Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *Bmc Genomics, 12*. doi:Doi 10.1186/1471-2164-12-460

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*(1), 237-U920. doi:DOI 10.1534/genetics.111.130922

Kim, E.-S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., . . . Sonstegard, T. S. (2013). Effect of Artificial Selection on Runs of Homozygosity in US Holstein Cattle. *Plos One, 8*(11). doi:10.1371/journal.pone.0080813

Kim, K., Jung, J., Caetano-Anolles, K., Sung, S., Yoo, D., Choi, B.-H., . . . Kim, H. (2018). Artificial selection increased body weight but induced increase of runs of homozygosity in Hanwoo cattle. *Plos One, 13*(3). doi:10.1371/journal.pone.0193701

Littlejohn, M. D., Tiplady, K., Fink, T. A., Lehnert, K., Lopdell, T., Johnson, T., . . . Spelman, R. J. (2016). Sequence-based Association Analysis Reveals an MGST1 eQTL with Pleiotropic Effects on Bovine Milk Composition. *Scientific Reports, 6*. doi:10.1038/Srep25376

Marchesi, J. A. P., Buzanskas, M. E., Cantao, M. E., Ibelli, A. M. G., Peixoto, J. O., Joaquim, L. B., . . . Ledur, M. C. (2018). Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. *Animal, 12*(6), 1126-1134. doi:10.1017/S1751731117002671

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., . . . Macciotta, N. P. P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal Genetics, 46*(2), 110-121. doi:10.1111/age.12259

Mastrangelo, S., Tolone, M., Di Gerlando, R., 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*(5), 746-754. doi:10.1017/S1751731115002943

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Agueda, L., Gut, M., . . . Distl, O. (2015). Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *Bmc Genomics, 16*. doi:10.1186/S12864-015-1977-3

Moller, M. J., Chaudhary, R., Hellmen, E., Hoyheim, B., Chowdhary, B., & Andersson, L. (1996). Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome, 7*(11), 822-830. doi:DOI 10.1007/s003359900244

Osorio, J. S., Lohakare, J., & Bionaz, M. (2016). Biosynthesis of milk fat, protein, and lactose: roles of transcriptional and posttranscriptional regulation. *Physiological Genomics, 48*(4), 231-256. doi:10.1152/physiolgenomics.00016.2015

Peripolli, E., Metzger, J., Antunes de Lemos, M. V., Stafuzza, N. B., Kluska, S., Olivieri, B. F., . . . Baldi, F. (2018). Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits. *Bmc Genomics, 19*. doi:10.1186/s12864-018-5060-8

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. *Animal Genetics, 48*(3), 255-271. doi:10.1111/age.12526

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *Bmc Genetics, 13*. doi:10.1186/1471-2156-13-70

Purfield, D. C., McParland, S., Wall, E., & Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *Plos One, 12*(5). doi:ARTN e017678010.1371/journal.pone.0176780

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raven, L. A., Cocks, B. G., & Hayes, B. J. (2014). Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *Bmc Genomics, 15*. doi: Doi 10.1186/1471-2164-15-62

Reinsch, N., Thomsen, H., Xu, N., Brink, M., Looft, C., Kalm, E., . . . Averdunk, G. (1999). A QTL for the degree of spotting in cattle shows synteny with the KIT locus on chromosome 6. *Journal of Heredity, 90*(6), 629-634.

Szmatola, 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. *Livestock Science, 188*, 72-80. doi:10.1016/j.livsci.2016.04.006

http://wsff.info/cm-2007-czech-republic.htm#H1 latest access 21 February 2020.

TABLE 1 Runs of homozygosity (ROH) features according to the bulls' country of birth.

	Country[†]				
ROH feature	**AT**	**CH**	**CZ**	**DE**	**IT**
nROH[‡] (n)					
average	80.62	114.42	56.02	81.97	77.47
SD	15.53	11.7	13.11	15.53	15.23
maximum	116	147	92	131	139
Mean length (Mb)					
Average	2.42	2.83	2.16	2.40	2.45
SD	3.01	3.58	2.93	2.98	3.24
Maximum	65.67	97.8	64.34	72.62	92.02
Mean length, SNP (n)					
Average	38.65	45.51	34.3	38.46	39.12
SD	47.88	56.93	47.24	47.19	51.55
maximum	1089	1596	1019	1167	1242
Total length (Mb)					
average	195.29	323.87	121.25	197.09	189.76
SD	55.82	73.59	60.57	55.67	66.37
Maximum	458.43	579.54	433.89	432.24	720.2
Total length, SNP (n)					
average	3,116.43	5,206.85	1,921.39	3,152.53	3,030.87
SD	903.25	1,193	979.05	898.2	1073.03
maximum	7,488	9,250	6,913	7,104	11,599

[†] AT=Austria (n=351), CH=Switzerland (n=215) CZ=Czech Republic (n=248) DE=Germany (n=550), IT=Italy (n=2481). [‡] Number of ROH segments per animals.

TABLE 2 Total number (nROH), non-repeated (NR-ROH), repeated (NR-ROH) and repeated-non-duplicated (RN-ROH) ROH by bulls' country of birth (count and %).

ROH statistic[†]	**AT**	**CH**	**CZ**	**DE**	**IT**
nROH[‡] (n)	28,299	24,600	13,922	45,067	192,212
NR-ROH (n)	7,546	8,551	4,885	10,665	43,992
RP-ROH (n)	20,753	16,049	9,037	34,402	148,220
RN-ROH (n)	13,775	13,243	7,079	20,339	64,509
RP-ROH[§] (%)	73	65	65	76	77
RN-ROH[¶] (%)	49	54	51	45	43
RN-ROH[‖] (n/bulls)	39.2	61.6	28.5	37.0	26.0

[†] n= number of Segments for the i-th country: AT=Austria (n=351), CH=Switzerland (n=215) CZ=Czech Republic (n=248) DE=Germany (n=550), IT=Italy (n=2481). [‡] $\text{NR-ROH}_i+\text{RP-ROH}_i$, [§] $\text{RP-ROH}_i/\text{nROH}_i\times100$, [¶] $\text{RN-ROH}_i/\text{nROH}_i\times100$, [‖] $\text{RN-ROH}_i/\text{no. bulls}_i$

TABLE 3 Number of private ROH per country of birth are presented on the diagonal (sum=92,226). Number of ROHs found in common between all pairs of Country (above the diagonal) and spearman rank correlation between shared ROH between all pairs of country (below the diagonal).

	Country[†]				
	AT	**CH**	**CZ**	**DE**	**IT**
AT	**7,915**	254	205	885	3,980
CH	0.18	**10,421**	191	336	1,896
CZ	0.33	0.12	**5,163**	250	2,420
DE	0.65	0.22	0.38	**11,421**	7,466
IT	0.67	0.24	0.44	0.74	**57,306**

[†] AT=Austria (n=351), CH=Switzerland (n=215), CZ=Czech Republic (n=248), DE=Germany, (n=550) and IT=Italy (n=2481).

TABLE 4 Inbreeding coefficients calculated using ROH (F_{ROH}) and their distribution in four classes, number of animals and (proportion), across country of birth and average value per country.

F_{ROH} class	**Country**[†]				
	AT	**CH**	**CZ**	**DE**	**IT**
< 5%	31 (0.09)	1 (0.01)	182 (0.73)	33 (0.06)	326 (0.13)
5-10%	288 (0.82)	33 (0.15)	56 (0.23)	454 (0.82)	1,858 (0.75)
10-20%	32 (0.09)	178 (0.83)	10 (0.04)	63 (0.12)	292 (0.12)
> 20%	0 (0)	3 (0.01)	0 (0)	0 (0)	5 (0)
Mean ± sd, %	7.6±2.2	12.6±2.9	4.7±2.4	7.7±2.2	7.4±2.6
Max, %	17.8	22.5	16.9	16.8	28.0

[†] AT=Austria (n=351), CH=Switzerland (n=215), CZ=Czech Republic (n=248), DE=Germany, (n=550) and IT=Italy (n=2,481).

TABLE 5 Runs of homozygosity exceeding the significance threshold in the genome-wide ROH analysis (GWRA).

ROH	BTA	Number of SNP	Start (Mb)	End (Mb)	Trait[†]
1	1	19	31.55	32.60	MY
2		17	103.62	104.84	MY, PP
3		22	106.60	107.91	PP
4		22	116.93	118.19	MY, PP
5		22	139.09	140.36	MY, PP
6		147	151.00	158.23	MY, PP
7		51	152.51	155.10	MY
8	2	24	22.09	23.50	PP
9		21	72.61	73.90	MY, FP, PP
10		19	89.71	91.06	MY, FP, PP
11	3	33	66.11	67.59	FP, PP
12	5	31	103.66	105.42	MY, FP, PP
13	10	22	100.54	101.65	MY
14	11	222	50.43	67.56	MY, FP, PP
15		76	51.08	58.42	MY, FP, PP
16		21	56.11	57.88	MY, PP
17		24	61.14	62.85	MY, PP
18		17	69.67	70.75	PP
19		18	93.24	94.71	MY, FP
20	13	16	14.62	15.78	MY, PP
21		19	54.20	55.29	MY, FP, PP
22	16	20	23.33	25.00	PP
23		25	42.63	44.99	FP
24		21	43.81	46.09	MY, PP
25		18	44.99	46.74	PP
26	20	24	49.81	51.17	MY, FP, PP
27	21	23	0.04	3.50	FP, PP
28		21	14.13	15.27	FP
29	22	29	0.16	2.01	FP
30	23	28	9.00	10.67	MY, FP, PP
31	28	23	18.15	19.89	MY, PP
32		16	31.64	33.00	PP
33	29	22	41.99	43.30	FP
34		16	50.41	51.50	MY, FP, PP

[†] MY = milk yield; FP = fat percentage; PP = protein percentage.

FIGURE CAPTIONS

FIGURE 1 Results of multidimensional scaling performed on all considered markers. The first two principal coordinates extracted from the inter-sample distance matrix were plotted, whereas the third was depicted by color gradient (AT = Austria; CH = Switzerland; CZ = Czech Republic; DE = Germany; IT = Italy).

FIGURE 2 Relationship between number of ROH segments and total ROH length (Mb) per animal by country of birth (AT = Austria; CH = Switzerland; CZ = Czech Republic; DE = Germany; IT = Italy). Country averages were depicted with larger symbols.

FIGURE 3 Distribution of ROH in common among bulls of the five country (AT = Austria; CH = Switzerland; CZ = Czech Republic; DE = Germany; IT = Italy).

FIGURE 4 Common Shared ROH segments along the 29 autosomal Chromosome (n=428). Thicker lines indicates higher ROH occurrence.

FIGURE 5 Manhattan plot of genome-wide ROH analysis (GWRA): –log(p-values) against the position along the chromosome (midpoint between starting and ending point of each ROH). The red line represents the significance threshold.

SUPPORTING INFORMATION

SUPPLEMENTARY TABLE 1 Distribution of ROH within class of length and bulls' country of birth (AT = Austria; CH = Switzerland; CZ = Czech Republic; DE = Germany; IT = Italy)

SUPPLEMENTARY TABLE 2. ROH found exclusively in one country (private ROH) and ROH shared exclusively among all possible combination of countries.

SUPPLEMENTARY TABLE 3. ROHs identified in all five groups across the 29 autosomal chromosomes.

SUPPLEMENTARY TABLE 4. Quantitative trait loci associated with milk traits (milk yield, fat percentage and protein percentage) that overlapped the runs of homozygosity found as significant with the GWRA approach.

SUPPLEMENTARY FIGURE 1 Distribution of Repeated ROH regardless the bulls' country of origin. The dashed red line represent the median of the distribution