Genome-wide variability and selection signatures in Italian island cattle breeds

Cesarani, A; Sorbolini, S; Criscione, A; Bordonaro, S; Pulina, G; Battacone, G; Marletta, D; Gaspa, G; Macciotta Nicolò Pietro Paolo,

Genome-wide variability and selection signatures in Italian island cattle breeds,

ANIMAL GENETICS, 49: 371:383, 2018,
doi: 10.1111/age.12697

Genome-wide variability and selection signatures in Italian island cattle breeds

A. Cesarani*, S. Sorbolini*, A. Criscione†, S. Bordonaro†, G. Pulina*,§, G. Battacone*, D. Marletta†, G. Gaspa*, N.P.P. Macciotta*

*** Dipartimento di Agraria, Università degli Studi di Sassari, 07100 Sassari, Italy**

† Dipartimento di Agricoltura, Alimentazione e Ambiente, Università degli Studi di Catania, 95131 Catania, Italy.

§Agenzia FORESTAS, Regione Autonoma della Sardegna, 09123 Cagliari, Italy

Summary

In the present study, a sample of 88 animals belonging to four local (Modicana, Sarda, Sardo-Bruna and Sardo-Modicana) and one cosmopolitan (Italian Brown Swiss) cattle breeds were genotyped with a medium density SNP beadchip and compared in order to investigate their genetic diversity and the existence of selection signatures. A total of 43,012 SNPs scattered across all twenty-nine autosomal chromosomes were retained after the data quality control. Basic population statistics, Wright Fixation Index and Runs of Homozygosity (ROH) analyses confirmed that Italian Brown genome was mainly shaped by selection, as underlined by the low values of heterozygosity and minor allele frequency. As expected, local cattle exhibited a large within breed genetic heterogeneity. The F_{ST} comparison with the largest number of significant SNPs was Sardo-Bruna *vs* Sardo-Modicana, whereas the smallest was observed for Italian Brown Swiss *vs* Sardo-Modicana, respectively. Modicana exhibited the largest number of detected ROH, whereas the smallest was observed for Sardo-Modicana. Signatures of selection were detected in genomic regions that harbor genes involved in milk production traits for the Italian Brown Swiss and fitness traits for local breeds. According to the results of Multi-Dimensional scaling and admixture analysis the Sardo-Bruna is more similar to the Sarda rather than to the Italian Brown Swiss. Moreover, the Sardo-Modicana is genetically closer to

the Modicana rather than to the Sarda breed. Results of the present work confirm the usefulness of Single Nucleotide Polymorphisms in deciphering the genetic architecture of livestock breeds.

Keywords: indigenous breeds, selection signatures, inbreeding, admixture, biodiversity

Introduction

The bovine domestication occurred presumably about 8-10 thousand years ago in southwest Asia (Zeder 2017). This process led to the zebuine and taurine breeds (Loftus *et al*. 1994; Upadhyay *et al.* 2016) derived both from the extinct wild aurochs (*Bos primigenius*) that spread in Europe and Africa in successive waves of migration. With domestication, cattle acquired a large variety of distinctive traits compared to their wild ancestors: for example, they became smaller in size and developed the capacity to adapt to various environments. During the Neolithic revolution, cattle accompanied human migrations and crosses between individuals of different ethnic groups generated a gene flow that changed the genetic makeup of their populations (Ajmone-Marsan *et al.* 2010).

The continuously increasing demand for work, milk and meat has enhanced between population differences over the centuries. In particular, changes in the farming systems, intense implementation of artificial selection, crossbreeding, and widespread use of artificial insemination that occurred in the last decades resulted in a huge genetic improvement of few highly specialized cattle breeds. However, as a consequence the within breed genetic variability has been seriously constrained in these populations (Brotherstone and Goddard, 2005). Biodiversity has been drastically endangered, a relevant reduction in the number of farmed cattle breeds has been observed leading to the extinction of many local breeds. Indigenous populations, better suited to extensive farming but not very productive, have been often abandoned in favor of highly productive breeds (Scherf 2000; Medugorac *et al.* 2009).

Concerns about climate changes, ethical issues, and evolution of consumer needs, including ecosystem services and landscape protection, are bringing towards sustainable livestock farming

systems. Such an evolving situation seems to offer new opportunities to indigenous breeds, because of their strong linkage to the production area, large genetic variability, and great fitness. Local breeds, are now considered as important reservoirs of resilience and biodiversity (Giovambattista *et al*. 2001). Their genomes represent an ideal model for studying and understanding the evolutionary history of livestock species, essential goal for evolutionary biology and population genetics. Moreover, local breeds represent a source of income in marginal areas (Ruto *et al*. 2008) and a chance to answer to the environmental changes (Medugorac *et al.* 2009). Their typical products support a sustainable development of the rural environment and respond to new consumer demands for healthy foods.

In Italy there is a particular attention for biodiversity, due to the high number of native animal and plant populations distributed throughout the whole country (Maiorano *et al*. 2007). Seventeen indigenous cattle breeds have been officially recognized by the Italian Ministry of Agriculture. Of particular interest is the situation of four cattle breeds farmed in extensive traditional systems in the two main Italian Islands, Sicily and Sardinia. The Sarda (SAR) breed is present in the Island of Sardinia since about 3,000 years BC. It originates from west Mediterranean cattle populations (mainly from the Iberic peninsula) with influences from North African and Middle East breeds (Della Maria 1936; Brandano *et al*. 1983a). At the end of the XIX century, crossbreeding with Brown Swiss (BSW) bulls imported from Switzerland and Modicana (MOD) bulls imported from Sicily were carried out in order to improve the aptitude of SAR to draught, milk and meat production respectively. These crosses have led to the current Sardo-Bruna (SB) and Sardo-Modicana (SM) breeds, respectively. The three Sardinian breeds have been officially recognized in 1985 with the establishment of the Herd book. The current population size, based on the number of animals recorded in the Herd book, is 25,315 and 923 herds for the SAR, 2,822 and 150 herds for the SM, and 33,662 and 1,426 herds for the SB respectively (www.aia.it).

The Modicana herdbook was established since 1952. Currently there are 5,209 animals recorded in the herd book, farmed in 235 herds (www.aia.it). An early genetic characterization of these breeds

was carried out using morphologic measurements (Brandano *et al*.1983b), milk and blood protein polymorphisms (Brandano *et al.* 1983c). Recently SM and MOD were compared in a study on coat color genetic determinism using the *Melanocortin 1 receptor* gene (Guastella *et al.* 2011) and the distribution of Runs of Homozygosity (ROH) was studied in MOD by Mastrangelo *et al*. (2016). The SAR, MOD, and BSW can be considered as founder breeds and SB and SM are the derived ones. In this work, a comparison between the five breeds is carried out using a medium density (50K) SNP panel in order to investigate the genetic diversity and in particular to assess the extent of diversity between pure-breeds and derived crosses. Moreover, gene discovery was performed in the genomic regions that exhibited difference between breeds.

Materials and methods

Animals and genotypic data

A total of 88 animals of five different breeds were genotyped in outsourcing with the Illumina BovineSNP50 beadchip: 22 BSW, 27 MOD, 19 SAR, 10 SB, and 12 SM, respectively. Genomic DNA was obtained from blood samples for SB, MOD, SM, and from nasal swab for SAR, using the NucleoSpin DNA rapidLyse Kit (Macherey-Nagel) according to manufacturer's instructions. For BSW animals, genotype data were generated within the SELMOL research project using the Genomix kit (Talent, Trieste, Italy). Animals of local breeds were randomly sampled from different herds located in various areas of Sardinia and Sicily. Given the difficulty in gathering large samples in local breeds, criteria used in the present work to include animals in the analysis were absence of relatedness, distribution in the territory, morphological appearance and information based on farmer interviews.

Since BSW animals were genotyped using Illumina BovineSNP50 v1 BeadChip in contrast to the other genotypic data (Illumina BovineSNP50 v2), common markers were retained and remapped on the UMD 3.1 release of the Bovine genome assembly. Only autosomal SNPs were considered.

Quality control was performed with Plink 1.9 (Purcell *et al*. 2007). Animals with a call rate > 95% were retained. SNP selection was based on call rate (>97.5%), minor allele frequency (MAF>0.05), and significant deviation for Hardy -Weimberg equilibrium (P<0.00001). After quality control, 43,012 common SNPs between the two Beadchip versions were retained. Missing genotypes were imputed using Beagle 4 (Browning & Browning, 2016).

Heterozygosity, Minor allele frequency and Linkage Disequilibrium

Heterozygote count (HET) and the minor allele frequency (MAF) were calculated for each SNP separately by breed using Plink 1.9. Linkage disequilibrium (LD) between markers was calculated within 1000 kb distance (McKay *et al.* 2007) using Haploview (Barrett *et al.* 2005).

Multi-dimensional scaling and admixture analysis

The Multi-Dimensional scaling plot (MDS) and admixture analysis were performed using the Zanardi pipeline (Marras *et al*. 2016) and "ggplot2" R package (Wickham, 2009). In MDS analysis, a principal component (PC) analysis is performed on the genomic correlation matrix **G** and PC scores are calculated for each individual. In order to confirm the animal classification in five different breeds, the K parameter of admixture was fixed at 5.

Wright Fixation Index and LOWESS

Ten pair-wise comparisons were performed using the Wright fixation index (F_{ST}) calculated using the equation proposed by Nei (1977):

$F_{ST} = (H_T – H_S) / H_T$

where H_T is the observed total heterozygosity and H_S is the observed heterozygosity in each population, respectively. For the F_{ST} calculation, an in house Python script was used. In order to simplify the graphic interpretation of raw F_{ST} data, a Locally Weighted Scatterplot Smoothing

(LOWESS) procedure was used (Pintus *et al*. 2014). The LOWESS is a local smoothing regression in which the space of the independent variable (in this case the progressive order of adjacent SNPs along the chromosome) is fragmented into different intervals for which separate regressions are fitted. The method is aimed at removing noise from raw data and at improving graphical representation. A smoothing parameter corresponding to an interval of 20 SNPs for each local regression was used.

A common problem when interpreting genetic difference metrics is the lack of proper statistical tests. Some authors have proposed to fix a threshold based on the F_{st} distribution (Kijas *et* al. 2012; Pintus et al., 2014). Although the distribution of raw F_{st} values tends to be skewed, LOWESS smoothed values could be considered approximately normally distributed. Thus, the significance threshold in the present work was set to three standard deviation from the mean. Such a stringent threshold was adopted considering the limited sample size.

Runs of homozygosity

Runs of Homozygosity (ROH) were detected using the Zanardi pipeline. Some constraints were fixed in order to limit the number of spurious ROH segments (Marras *et al.* 2015): the minimum length of ROH was set at 1 Mb, homozygous segments of minimum fifteen SNPs were considered and neither heterozygous or missing genotypes were allowed. The following ROH statistics were calculated by animal and by breed: number of ROH, the average ROH length (in Mb) and the sum of all ROH segments by animal (S_{ROH}, in Mb). ROH were grouped into five classes of length (1 < Mb ≤ 2, 2 < Mb ≤ 4, 4 < Mb ≤ 8, 8 < Mb ≤ 16 and Mb > 16).

The ROH-based inbreeding coefficient (F_{ROH}) for each animal was calculated as

$$F_{ROH} = \frac{\sum S_{ROH(>8Mb)}}{Lgen}$$

where L_{gen} is the total length of genome. The minimum length of ROH to be included in the calculation was fixed to 8 Mb based on previous reports in cattle (Marras *et al*. 2015). Moreover, the ROH count per SNP (SNP_{ROH}), i.e. the number of animals having a given SNP included in a ROH (Nothnagel *et al*. 2010) was calculated. A threshold of 50% was fixed to consider a SNP_{ROH} value as significant.

Gene discovery

Gene discovery was performed in regions flagged by F_{ST} values exceeding the control chart upper limit. Intervals spanning 0.25 Mb upstream and downstream the significant marker were considered. Moreover, regions identified by ROH distribution were studied. In particular, markers having SNP_{ROH} > 50% within a breed were considered as significant and the region spanning 0.25 Mb upstream and downstream surrounding them was investigated. Annotated genes were retrieved from UCSC Genome Browser Gateway (http://genome.ucsc.edu./) and National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov) databases.

Results

HET and MAF showed a little variation between the five considered breeds (Table 1). BSW showed smallest values of both HET and MAF, whereas MOD and SAR exhibited the largest values for these parameters, respectively.

A clear distinction between the breeds could be observed along the first axis (PC1) of the MDS plot (Fig. 1). In particular the PC1, that explains about 5.4% of the total variance, depicts a geographic cline: starting from the bottom of the graph there are individuals from BSW (origin from the Switzerland, North of Italy), then SAR and SB (centre of Italy), and at the top SM and MOD (native of Sicily, Southern Italy). Furthermore, it could be seen that along this dimension, SM breed is more similar to MOD than SAR. The second axis (PC2), explaining about 3% of the total variance,

highlights a separation within the SAR breed. The PC2 seemed to be able to discriminate animals according to the percentage of SAR genetic contribution: an increase in PC2 scores indicates the passage from SAR purebred to crosses, and then to MOD and BSW breeds. Population structure analysed by admixture (Fig. 2) revealed a clear definition of BSW animals (95% assigned to a single cluster, the one of red colour), and less precise for MOD and SAR (90% and 93% assigned to two different clusters, respectively). Finally, also the derived breeds were grouped into two distinct clusters (70% of both SB and SM cattle). The LD pattern (Fig. 3) shows the lowest value for MOD, the highest for BSW and SB, respectively.

The F_{ST} comparison with the largest number of significant SNPs was SB *vs* SM, whereas the smallest was observed for BSW *vs* SM (Table S1). Figure 4 reports Manhattan plots of F_{ST} predicted by LOWESS for the comparisons between pure breeds and crosses. It can be observed that the highest F_{ST} values between BSW and SB were found for BTA6 (Fig. 4a), with the top significant markers (Table S2) located between 38.20 and 38.83 Mb. In this region map some known genes controlling milk production traits (*ABCG2*, *PKD2*, *SPP1*, *LAP3*), and body size (*NCAPG* and *LCORL*) in cattle. BTA8 and BTA13 showed the highest F_{ST} peaks in the SAR *vs* SB comparison (Fig. 4b) with seven and three significant markers respectively (Table S2). In the region highlighted on BTA8 is located the *microRNA2471 (MIR2471*), whereas in the highlighted segment of BTA13 is annotated the *Eukaryotic translation initiation factor 6* (*EIF6*) gene.

SAR and SM were different mainly on BTAs 7, 14, and 21 (Fig. 4c and Table S2). An interesting gene retrieved from the database was the *Ubiquitin Protein Ligase E3A* (*UBE3A*) that maps in the region between 2.1 and 2.3 Mb of BTA21.

As far as the comparison between SM and MOD is concerned (Fig. 4d), the highest values of F_{ST} have been found on BTAs 5, 16 and 20 (Table S2). On BTA20 the region from 70.9 to 71.7 Mb presents a QTL associated with milk somatic cell score. Moreover, this segment contains

several annotated genes, among which of interest is the *Solute Carrier Family 9 Member A3* (*SLC9A3*).

Finally, for the SM vs SB comparison the highest values of F_{ST} have been detected on chromosomes 7 and 24 (Fig. 4e and Table S2). On BTA7, five significant markers define a region (47.2-47.3 Mbp) were the *Transcription Factor 7* (T-Cell Specific, HMG-Box) (*TCF7*) gene maps.

The total number of detected ROH (Table 2) exhibited a large variation between breeds, with MOD and SM having the largest and the smallest value, respectively. The BSW had the largest average ROH length, even if together with a huge variability as evidenced by the value of the standard deviation (Table 2). This breed had also the highest average number of SNP per ROH (Table 2). On the contrary MOD showed the smallest values of both statistics. As expected, most represented ROH classes in all breeds were those of length <4Mb (relative frequency ranging from 0.736 in BSW to 0.868 in MOD and SM, respectively). The largest number of ROH in the class of highest length (>16 Mb) was observed in BSW, and it was markedly larger than in all the other considered breeds (Table 2).

ROH count per SNP showed some interesting peaks along the genome. The highest peak was observed on BTA6 for BSW at approximately 30-40Mb (Fig. 5a). In this region map several known genes as *ABCG2*, *SPP1*, *LCORL*, *NCAPG*. BSW exhibited another signal between 10-30 Mb on BTA20 (Fig. 5b). Moreover, BTA1, BTA10 and BTA11 showed interesting signals of SNPs in homozygosity for over 50% of the animals. In particular, BSW showed a SNP_{ROH} peak on BTA1 (Fig. 6a) between 103.5 and 105.5 Mb. On the same chromosome, a peak was detected for MOD at 139.0 Mb. On BTA10 an interesting homozygous region was observed in the SAR breed between 72.2 and 72.8 Mb (Fig. 6b). Among the genes that map in this region the *Dehydrogenase/Reductase 7* (*DHRS7*) can be mentioned. Finally, the SB showed a relevant value of SNP_{ROH} on BTA11 (Fig. 6c) between 65.0 and 67.0 Mb where the *Ewing Tumor Associated Antigen 1* (*ETAA1*) was annotated.

223 BSW exhibited also the largest average F_{ROH} (Table 3) whereas the smallest value was observed by
224 SM.
225
226 **Discussion**
227 The practice of crossbreeding has represented a major cause of gene flow across cattle populations,
228 providing a relevant contribution to the shaping of worldwide current breeds. The history of the
229 Sarda breed and its crosses with Modicana and Brown Swiss represents a typical example. Results
230 of the present study confirm the genetic relationships between the considered breeds. The admixture
231 analysis (Fig. 2) clearly detected the five different genetic groups, highlighting the genetic
232 background of the crossbred derived population in comparison of the original purebreds.
233 Furthermore, the analyses of the genome features with different approaches gave useful insights on
234 effects of selection and environmental adaptation on the cattle genome.
235 A first indication was provided by basic population statistics. The lower genetic variability
236 exhibited by the BSW in comparison with the other two pure-breeds, SAR and MOD, was expected
237 due to the intense artificial selection this breed has been subjected to in the last decades
238 (www.anarb.it). A low allelic diversity for BSW cattle in comparison with other cattle breeds has
239 been already reported (Schmid *et al.* 1999; Melka & Schenkel 2012).
240 The genome feature analysis carried out using the MDS decomposition, and the ROH detection
241 highlighted an interesting structure of the considered sample of animals. The North-South
242 geographical gradient highlighted by the first axis of the MDS is in agreement with several studies
243 where a dimension reduction method is applied to molecular data on populations from different
244 geographical origin (Price 2006; Chessa *et al.* 2009; Ciani *et al.* 2014). Also, the variation of the
245 ROH statistics and of the inbreeding coefficient F_{ROH} exhibited the same cline. In particular the
246 average ROH length, the average number of SNP per ROH, and the F_{ROH} showed an increase moving
247 from South to North. This gradient was also confirmed by the LD analysis (Fig. 3). Purfield *et al.*

(2012) found a higher number of ROH in cattle breeds of British Isles compared to other European breeds and ascribed such a diversity to the closed population histories of these cattle. Results obtained in the present study can be probably due to a low effective population size of BSW and to the population history of the SAR, MOD, and their crosses. A geographical South-North gradient in ROH feature distribution has been observed also in human populations (Nothnagel *et al.* 2010), and it has been explained with the most pronounced genetic isolation of Northern populations compared to Mediterraneans. The second axis of the MDS analysis highlights two clusters in the sample of Sarda cattle (Fig. 1). Previous studies on this population highlighted a large morphological heterogeneity (Brandano *et al.* 1984). Moreover, in the traditional extensive cattle farming system of Sardinia it is not very common to exchange bulls between herds, resulting in a high average relatedness of individuals within farm and a low degree of kinship among farms.

Different degree of genetic relationships between original and derived breeds have been observed. The similarity between SM and MOD was quite expected (Fig. 1). Although the first importation of MOD bulls from Sicily started at the end of the nineteenth century in the Montiferru area (Center-North Sardinia), it probably occurred again in more recent times and therefore the genetic component of Modicana purebred is still preserved into current SM. On the other hand, the separation between SB and the two founder breeds, i.e. BSW and SAR (Fig. 1), seems to indicate an absence of recent genetic exchange.

The genetic history of the breeds is also depicted by other structural elements of their genome, as their linkage disequilibrium (Fig. 3) and the extent of regions of autozygosity (Fig. 5 and 6). The intensive genetic selection of BSW in comparison with the other investigated breeds resulted in the highest level of LD and in the largest values of all ROH statistics. These results agree with previous reports on this breed (Ferenčaković *et al.* 2013; Marras *et al.* 2015). A previous study on MOD breed reported a smaller value of F_{ROH} (Mastrangelo *et al.* 2016) but using different ROH settings (i.e. minimum number of SNP in a ROH equal to 40, minimum ROH length 4Mb, two missing SNP

allowed in a ROH etc.). An interesting result is the distribution across individuals of specific ROHs, i.e. a segment that starts and ends exactly in the same position. The largest ROH frequency was about 0.06 (Table 4) and it can be seen that in general local breeds tend to share ROH whereas the autozygous segment detected on BTA6 can be found only within the BSW breed. In particular, the latter ROH flagged a region where several known genes affecting milk traits are located. These results confirm the role of ROH as indicators not only of inbreeding but also of signatures of selection (Marras *et al.* 2014; Kim *et al.* 2015).

Signatures of selection were highlighted in the present study. Some of them flagged genome regions already detected in many studies on cattle. An example is represented by the markers exhibiting the largest F_{ST} values in the BSW *vs* SB comparison, all located in the region of BTA6 spanning between 36-39Mb that harbors some known genes controlling milk production traits (*ABCG2*, *PKD2*, *SPP1*, *LAP3*) (Olsen *et al*. 2005; Cohen-Zinder *et al.* 2005) and body size (*NCAPG* and *LCORL*) (Takasuga 2016) (Table S2). This region was also flagged by a significant value of ROH count per SNP in BSW.

Other two well known selection signatures were detected in BSW on BTA6 (Fig. 5a) by SNP_{ROH} significant values (>50%). The first was located at around 70 Mb, where the *V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog* (*KIT*) locus maps. This gene is involved in mammalian coat colour determinism (Fontanesi *et al.* 2010; Stella *et al.* 2010). The second signature of selection, at around 85 Mb, identified the caseins cluster (Blott *et al.* 2003). Another interesting peak value of SNP_{ROH} was found on BTA20 (14-25 Mb) (Fig. 5b), in a region where a large QTL associated with milk protein percentage was reported (Ashwell *et al*. 2004) . Among the several genes that map in this region, of interest is the *Importin 11* (*IPO11*) locus. This gene has been found to be associated with the displacement of the abomasum in German Holstein cattle breed (Mömke *et al*. 2013).

297 Interestingly, the F_{ST} pairwise comparison between the SAR and the SB did not detect SNPs located
298 in genomic regions known to contain genes associated with milk production traits. These results,
299 together with the pattern highlighted by the MDS, confirm that current SB is closer to SAR than to
300 BSW, probably due to backcrossing.
301 Of interest are the signatures of selection found in the comparisons between local breeds. Some of
302 them include interesting genes that were found to be associated with fitness traits. In the comparison
303 between the SAR and its derived SB, the seven highly significant SNPs found on BTA8 between
304 40.4 and 40.6 Mbp (Fig. 4b) identified a region where maps the *microRNA2471* (*MIR2471*). In
305 animals, microRNAs are molecules involved in diverse biological processes such as development,
306 cell differentiation, proliferation and metabolism. They are among major post-transcriptional
307 regulators of gene expression through promoting mRNA degradation or translational repression
308 (Glazov *et al.* 2009; Guo *et al.* 2010; Meunier *et al.* 2013). Recently they have been found to be
309 essential for the regulation of the immune response (Xiao & Rajewsky 2009). The highest peak of
310 F_{ST} comparison between the SAR and the other derived breed, the SM, was detected on BTA14 (Fig.
311 4c and Table S2) in a region where maps the *gasdermin C* (*GSDMC*) locus. This gene was
312 associated to UV-protective eye pigmentation in Fleckvieh cattle (Pusch *et al.* 2012). Another
313 peak was located on BTA21, between 2.1 and 2.3 Mb, where the *Ubiquitin Protein Ligase E3A*
314 (*UBE3A*) gene is annotated. This locus has been associated with the calving ease (Pausch *et al.*
315 2011; Meszaros *et al.* 2016) in cattle. This trait represents very often a distinguishing feature in
316 indigenous breed that are mainly reared in extensive and semi-extensive systems (Boggio *et al.*
317 1988).
318 Other genes detected in the local breeds are related to milk production traits and fatty acid
319 metabolism. Among genetic differences found between SM and MOD, of interest is the region
320 located on BTA20, from 70.9 to 71.7 Mb. Among the annotated genes, is worth of mention the
321 *Solute Carrier Family 9 Member A3* (*SLC9A3*), involved in the rumen sodium transport (Rabbani *et*

322 *al.* 2011). A high Na^{2+} tissue concentration improves milk production in warm/humid conditions
323 (Granzin & Gaughan 2002). Moreover, a QTL associated with milk somatic cell score was reported
324 in this region (Durán Aguilar *et al*. 2016). The comparison between the two derived breeds, i.e., SB
325 *vs* SM. found a selection signature defined by five significant markers (47.2-47.3 Mbp) on BTA7,
326 where maps the *Transcription Factor 7* (*T-Cell Specific, HMG-Box*) (*TCF7*) gene. Recently, this
327 locus was associated with milk production in Chinese Holstein (Mao *et al.* 2015).
328 An interesting candidate gene highlighted by SNP_{ROH} in the SAR breed on BTA10 is the
329 *Dehydrogenase/Reductase 7* (*DHRS7*) locus. It catalyses the oxidation/reduction of a wide range of
330 substrates, including retinoid and steroids (Haeseleer & Palczewski 2000) and it has high expression
331 levels in adipocytes and skeletal muscles (Wu *et al.* 2009). In addition, this gene is responsible for
332 the final step in the cholesterol production (Porter 2000). This gene was already associated in
333 Nellore cattle with the intramuscular fat deposition and composition (Cesar *et al.* 2014). Finally,
334 another signature of selection that included a gene involved in fatty acid metabolism was found in
335 the SAR *vs* SB comparison (three significant markers on BTA13 between 65.1 and 65.2 Mb) (Fig.
336 4b). This region harbours the *Eukaryotic translation initiation factor 6* (*EIF6*) locus. This gene
337 controls fatty acid synthesis and glycolysis in tissues responsive to insulin such as adipose and
338 muscular.
339
340 **Conclusion**
341 Results of the present work confirm the usefulness of genome structural features in deciphering the
342 genetic architecture of livestock breeds. The different approaches used to explore medium density
343 SNP genotypes gave a comprehensive picture of genetic relationships between the three original and
344 the two derived breeds, reflecting their recent genetic history. As expected, a larger heterogeneity was
345 highlighted for the local breeds. Signatures of selection located in genomic regions harboring
346 candidate genes for milk production traits have been detected in the comparisons involving the

specialized BSW breed, whereas for local breeds the flagged genes involved in fitness and fatty acid metabolism. The study confirmed the importance of these populations as resevoir of biodiversity and as models for studying the genetic basis of adaptability.

Acknowledgments

This Research was funded by the Banco di Sardegna Foundation and by the Regional Government of Sardinia (project grant 07/G1-20, POR-FSE 2007-13). The author wishes to thank Dr. Salvatore Mele of Agenzia FORESTAS for his contribution to the animal sampling.

References

Ajmone-Marsan P., Garcia J.F. & Lenstra J.A. (2010) On the origin of cattle: How aurochs became cattle and colonized the world. Evolutionary Anthropology: Issues, News, and Reviews **19**, 148-157.

Ashwell M.S., Heyen D.W., Sonstegard T.S., Van Tassell C.P., Da Y., VanRaden P.M., Ron M., Weller J.I. & Lewin H.A. (2004) Detection of Quantitative Trait Loci Affecting Milk Production, Health, and Reproductive Traits in Holstein Cattle. Journal of Dairy Science **87**, 468–475.

Barrett J.C., Fry B., Maller J. & Daly M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21**, 263–265.

Blott S., Kim J.J., Moisio S., Schmidt-Küntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vikki 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.

Boggio F., Pracchi R. & Asole A. (1998) Atlante economico della Sardegna 1. Jaca Book, Edizioni Universitarie Jaca, Italia.

Borg I. & Groenen P. (2003) Modern multidimensional scaling: theory and applications. Journal of Educational Measurement **40**, 277-280.

Brandano P., Asara P., Pulina G., Bolla P. & Crimella C. (1983a). The Sardinian cattle. 1. Morphological and biological characters. Annals of the Faculty of Agriculture of the University of Sassari **30**, 161-177.

Brandano P., Asara P., Pulina G., Bolla P. & Crimella C. (1983b) The Sardo- Modicana cattle Breed. 1. Annals of the Faculty of Agriculture of the University of Sassari **30**, 197-214.

Brandano P., Pulina G. & Asara P. (1983c) The indigenous cattle of Sardinia. Breeds and herds characteristics. Annals of the Faculty of Agriculture of the University of Sassari **30**, 1-23.

Brotherstone S., & Goddard M. (2005). Artificial selection and maintenance of genetic variance in the global dairy cow population. Philos Trans R Soc Lond B Biol Sci. 360: 1479–1488.

Browning B.L. & Browning S.R. (2016) Genotype imputation with millions of reference samples. The American Journal of Human Genetics **98**, 116-126.

Cesar A.S., Regitano L.C., Mourão G.B., Tullio R.R., Lanna D.P., Nassu R.T., Mudado M.A., Oliveira P.S.N., do Nascimento M.L., Chaves A.S., Alencar M.M., Sonstegard T.S., Garrick D.J., Reecy J.M. & Coutinho L.L. (2014) Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. BMC genetics **15**, 39.

Ciani E., Crepaldi P., Nicoloso L. *et al*. (2014) Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. Animal Genetics, **45**, 356-366.

Chessa, B., F. Pereira, F. Arnaud, A. Amorim, F. Goyache *et al.* 2009. Revealing the history of sheep domestication using retrovirus. Science **324**, 532

Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., der Wind A.E., Lee J.H., Drackley J.K., Band M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I. & Ron M. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Research **15**, 936-944.

Della Maria G. (1936) Ancient stories of Sardinian bovine breed. Rivista di Zootecnia **13**, 47-57.

Durán Aguilar M., Román Ponce S.I., Ruiz López F.J., González Padilla E., Vásquez Peláez C.G., Bagnato A. & Strillacci M.G. (2016) Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. Journal of Animal Breeding and Genetics 134, 49-59.

421 Ferenčaković M., Hamzić E., Gredler B., Solberg T.R., Klemetsdal G., Curik I., & Sölkner J.
422 (2013) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from
423 selected cattle populations. Journal of Animal Breeding and Genetics **130**, 286–293.
424
425 Fontanesi L., Scotti E. & Russo V. (2010) Analysis of SNPs in the KIT Gene of Cattle with
426 Different Coat Colour Patterns and Perspectives to Use These Markers for Breed Traceability and
427 Authentication of Beef and Dairy Products. Italian Journal of Animal Science **9**, e42.
428 Giovambattista G., Ripoli M. V., Peral-Garcia P. & Bouzat J. L. (2001) Indigenous domestic breeds
429 as reservoirs of genetic diversity: the Argentinean Creole cattle. Animal Genetics **32**, 240–247.
430
431 Glazov E.A., Kongsuwan K., Assavalapsakul W., Horwood P.F., Mitter N. & Mahony, T.J. (2009)
432 Repertoire of Bovine miRNA and miRNA-Like Small Regulatory RNAs Expressed upon Viral
433 Infection. PLOS ONE **4**, e6349.
434
435 Granzin B.C. & Gaughan J.B. (2002) The effect of sodium chloride supplementation on the milk
436 production of grazing Holstein Friesian cows during summer and autumn in a humid sub-tropical
437 environment. Animal Feed Science and Technology **96**, 147-160.
438
439 Guastella A.M., Sorbolini S., Zuccaro A., Pintus E., Bordonaro S., Marletta D. & Macciotta N.P.P.
440 (2011) Melanocortin 1 receptor (MC1R) gene polymorphisms in three Italian cattle breeds. Animal
441 Production Science **51**, 1039–1043.
442
443 Guo L & Lu Z. (2010) Global expression analysis of miRNA gene cluster and family based on
444 isomiRs from deep sequencing data. Computational Biology and Chemistry **34,** 165-171.
445

446 Haeseleer F. & Palczewski K. (2000) Short-chain dehydrogenases/reductases in retina. Methods in
447 enzymology **316**, 372-383.
448
449 Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B.,
450 McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan L.,
451 Dalrymple B., and International Sheep Genomics Consortium (2012) Genome-Wide Analysis of the
452 World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PloS*
453 *Biology* **10,** e 10001258.
454 Loftus R. T., MacHugh D. E., Bradley D. G., Sharp P. M. & Cunningham P. (1994) Evidence for
455 two independent domestications of cattle. Proceedings of the National Academy of Sciences **91**,
456 2757–2761.
457
458 Maiorano L., Falcucci A., Garton E.O. & Boitani L (2007) Contribution of the Natura 2000 network
459 to biodiversity conservation in Italy. Conservation Biology **21**, 1433-1444.
460
461 Mao Y., Zhu X., Xin S., Zhang M., Wang X., Cheng D., Zhang H., Konig S., Yang Z. & Yang L.
462 (2015) Polymorphisms in the promoter of interleukin-12β2 and interleukin-23 receptor genes
463 influence milk production traits in Chinese Holstein cows. Livestock Science **178**, 1–8.
464
465 Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J.L. &
466 Macciotta, N.P.P. (2015) Analysis of runs of homozygosity and their relationship with inbreeding in
467 five cattle breeds farmed in Italy. Animal Genetics **46**, 110–121.
468

469 Marras G., Rossoni A., Schwarzenbacher, H., Biffani S., Biscarini F. & Nicolazzi E.L. (2016)
470 Zanardi: an open-source pipeline for multiple-species genomic analysis of SNP array data. Animal
471 Genetics **48**, 121-128.
472
473 Mastrangelo S., Di Gerlando R., Tolone M., Tortorici L., Sardina M.T. & Portolano B. (2014)
474 Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. BMC
475 Genetics **15**, 108.
476
477 Mastrangelo S., Tolone M., Gerlando R.D., Fontanesi L., Sardina M.T. & Portolano B. (2016)
478 Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three
479 local dairy cattle breeds. Animal **10**, 746–754.
480
481 McKay S.D., Schnabel R.D., Murdoch B.M., Matukumalli L.K., Aerts J., Coppieters W., Crews D.,
482 Dias Neto E., Gill C.A., Gao C., Mannen H., Stothard P., Wang Z., Van Tassell C.P., Williams J.L.,
483 Taylor J.F. & Moore S.S. (2007) Whole genome linkage disequilibrium maps in cattle. BMC
484 Genetics **8**, 74.
485
486 Medugorac I., Medugorac A., Russ I., Veit-Kensch C.E., Taberlet P., Luntz, B., Mix H.M. &
487 Förster M. (2009) Genetic diversity of European cattle breeds highlights the conservation value of
488 traditional unselected breeds with high effective population size. Molecular Ecology **18**, 3394-3410.
489 Melka M.G. & Schenkel F.S. (2012) Analysis of genetic diversity in Brown Swiss, Jersey and
490 Holstein populations using genome-wide single nucleotide polymorphism markers. BMC Research
491 Notes **5**, 161.
492

493 Mészáros G., Taferner R. & Sölkner J. (2016) Pleiotropic and epistatic interactions between
494 stillbirth and calving ease in cattle. Acta Agriculture Slovenica **5**, 56.
495
496 Meunier J., Lemoine F., Soumillon M., Liechti A., Weier M., Guschanski K., Hu H., Khaitovich P.
497 & Kaessmann, H. (2013) Birth and expression evolution of mammalian microRNA genes. Genome
498 Research **23,** 34-45.
499
500 Mömke S., Sickinger M., Lichtner P., Doll K., Rehage J. & Distl O. (2013) Genome-wide
501 association analysis identifies loci for left-sided displacement of the abomasum in German Holstein
502 cattle. Journal of Dairy Science **96**, 3959–3964.
503
504 Nei M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of
505 Human Genetics **41**, 225-233.
506
507 Nothnagel M., Lu T.T., Kayser M. & Krawczak M. (2010) Genomic and geographic distribution of
508 SNP-defined runs of homozygosity in Europeans. Human Molecular. Genetics **19**, 2927–2935.
509
510 Olsen H.G., Lien S., Gautier M., Nilsen H., Roseth A., Berg P.R., Sundaasen K.K., Svendsen M. &
511 Meuwissen T.H.E. (2005) Mapping of a Milk Production Quantitative Trait Locus to a 420-kb
512 Region on Bovine Chromosome 6. Genetics **169**, 275–283.
513
514 Pausch H, Wang X, Jung S, Krogmeier D, Edel C, et al. (2012) Identification of QTL for UV-
515 Protective Eye Area Pigmentation in Cattle by Progeny Phenotyping and Genome-Wide
516 Association Analysis. PLoS ONE **7**: e36346.
517

518 Pausch H., Flisikowski K., Jung S., Emmerling R., Edel C., Götz K.U. & Fries R. (2011) Genome-
519 Wide Association Study Identifies Two Major Loci Affecting Calving Ease and Growth-Related
520 Traits in Cattle. Genetics **187**, 289–297.
521
522 Pintus E., Sorbolini S., Albera A., Gaspa G., Dimauro C., Steri R., Marras G. & Macciotta N.P.P.
523 (2014) Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection
524 signatures in Piedmontese and Italian Brown cattle breeds. Animal Genetics **45**, 1–11.
525
526 Porter F.D. (2000) RSH/Smith–Lemli–Opitz syndrome: a multiple congenital anomaly/mental
527 retardation syndrome due to an inborn error of cholesterol biosynthesis. Molecular Genetics and
528 Metabolism **71**, 163-174.
529
530 Price A.L., Patterson N.J, Plenge R.M., Weimblatt M.E., Shadick N.A. & Reich D. (2006) Principal
531 components analysis corrects for stratification in genome-wide association studies. Nature Genetics
532 **38**, 904-909.
533
534 Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P.,
535 de Bakker P.I.W., Daly M.J. & Sham P.C. (2007) PLINK: a toolset for whole-genome association
536 and population-based linkage analyses. American Journal of Human Genetics **81**, 559–575.
537
538 Purfield D.C., Berry D.P., McParland S. & Bradley D.G. (2012) Runs of homozygosity and
539 population history in cattle. BMC Genetics **13**, 70.
540
541 R Core Team (2015) R: A language and environment for statistical computing. R Foundation for
542 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ruto E., Garrod G. & Scarpa R. (2008) Valuing animal genetic resources: a choice modeling application to indigenous cattle in Kenya. Agricultural Economics **38**, 89–98.

Scherf B.D. (2000) World watch list for domestic animal diversity (No. Ed. 3) Food and Agriculture Organization (FAO).

Schmid B.M., Saitbekova N., Gaillard C. & Dolf G. (1999) Genetic diversity in Swiss cattle breeds. Journal of Animal Breeding and Genetics **116**, 1-8.

Stella A., Ajmone-Marsan P., Lazzari B. & Boettcher P. (2010) Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. Genetics **185**, 1451.

Takasuga A. (2016) PLAG1 and NCAPG-LCORL in livestock. Animal Science Journal **87**, 159–167.

Upadhyay M.R., Chen W., Lenstra J.A., Goderie C.R.J., MacHugh D.E., Park S.D.E., Magee D.A., Matassino D., Ciani F., Megens H.J., van Arendonk J.A.M, Groenen M.A.M, & European Cattle Genetic Diversity Consortium, Crooijmans R.P.M.A. (2016) Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. Heredity **118,** 169-176.

Wickham H. (2009) Gplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

567 Xiao C. & Rajewsky K. (2009) MicroRNA control in the immune system: basic principles. Cell
568 **136**, 26-36.
569
570 Zeder M. (2017) Domestication and early agriculture in the Mediterranean basin: Origins, diffusion,
571 and impact. Proceedings of the National Academy of Sciences **105**, 11592-11604.
572
573
574

575 **Table 1**. Mean value of heterozygosity (HET) and

576 Minor allele frequency (MAF) in the five breeds.

	HET		MAF	
	Mean	s.d.	Mean	s.d.
BSW	0.318	0.011	0.232	0.010
MOD	0.348	0.008	0.249	0.006
SAR	0.335	0.011	0.252	0.005
SB	0.343	0.012	0.251	0.007
SM	0.347	0.013	0.251	0.006

577 BSW = Italian Brown Swiss; MOD = Modicana;

578 SAR =Sarda; SB =Sardo Bruna; SM = Sardo Modicana.

579

580 **Table 2**. Statistics of ROH size and frequency in the five investigated cattle breeds.

	BSW	MOD	SAR	SB	SM
Average length (Mb)	3.9 ± 5.0	2.3 ± 1.8	2.9 ± 2.4	2.6 ± 2.3	2.4 ± 2.0
Average number of SNP per ROH	67.2 ± 85.8	40.2 ± 30.3	49.1 ±40.8	44.7 ± 39.1	41.2 ± 33.6
Number of ROH					
1-2 Mb	780	1270	834	423	447
2-4 Mb	404	571	420	220	195
4-8 Mb	231	242	251	87	83
8-16 Mb	138	34	74	21	13
>16 Mb	56	2	2	4	2
Total	1609	2119	1581	755	740

581 BSW = Italian Brown Swiss; MOD =Modicana; SAR =Sarda; SB =Sardo Bruna; SM = Sardo

582 Modicana.

583 **Table 3**. ROH-based inbreeding coefficient (F_{ROH})

584 calculated using ROH>8Mb.

	F_{ROH}			
	Mean	s.d.	Max	Min
BSW	0.127	0.043	0.210	0.043
MOD	0.073	0.056	0.290	0.031
SAR	0.095	0.086	0.360	0.015
SB	0.080	0.078	0.282	0.019
SM	0.060	0.058	0.227	0.023

585 BSW = Italian Brown Swiss; MOD =Modicana; SAR =Sarda;

586 SB =Sardo Bruna; SM = Sardo Modicana.

587 **Table 4.** Most frequent ROH detected in the five breeds

Chromosome	Start	End	Length (Mb)	Frequency[1]	Breed
1	73924347	75505402	1.58	5	SB, MOD,SAR
29	23762023	25780595	2.02	5	SB, SM, MOD,SAR
6	32241952	34661866	2.41	5	BSW
9	27516531	28538817	1.02	5	SB, SM, SAR
9	821062	2677236	1.86	5	SM, MOD, SAR, BSW

588 BSW = Italian Brown Swiss; MOD =Modicana; SAR =Sarda; SB =Sardo Bruna; SM = Sardo

589 Modicana.

590 [1]Number of individuals that possess the specific ROH across breeds

591

Figures legend

Figure 1 Multi-Dimensional Scaling plot of the five investigated breeds: Italian Brown Swiss (BSW), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

Figure 2 Genetic structure and admixture plot obtained through coefficients of individual membership to clusters (K=5) assumed to be present in the sample of investigated breeds. Red columns = cluster 1; Light green columns = cluster 2; Blue columns = cluster 3; Green columns = cluster 4; Purple columns = cluster 5.

Figure 3 Average LD (r^2) between markers within an interval of 1000 kb in the five Italian cattle breeds: Italian Brown Swiss (BSW), Modicana (MOD), Sarda (SAR), Sardo-Bruna (SB) and Sardo-Modicana (SM).

Figure 4 Manhattan plot of F_{ST} values predicted by the LOWESS. **a**) Comparison between Italian Brown and Sardo-Bruna. **b**) Comparison between Sarda and Sardo-Bruna. **c**) Comparison between Sarda and Sardo-Modicana. **d**) Comparison between Sardo-Modicana and Modicana. **e**) Comparison between Sardo-Modicana and Sardo-Bruna. Red color dots indicate significant F_{ST} values (i.e. greater than 3 standard deviations from the mean).

Figure 5 Occurrence of SNP counted in a ROH measured by the percentage of animals belonging to the five investigated breeds for which a particular SNP falls into a ROH versus the position along the chromosome. **a**) Comparison of BTA6. **b**) Comparison of BTA20.

Figure 6 Occurrence of SNP counted in a ROH measured by the percentage of animals belonging to the five investigated breeds for which a particular SNP falls into a ROH versus the position along the chromosome. **a**) Comparison of BTA1. **b**) Comparison of BTA10. **c**) Comparison of BTA11.

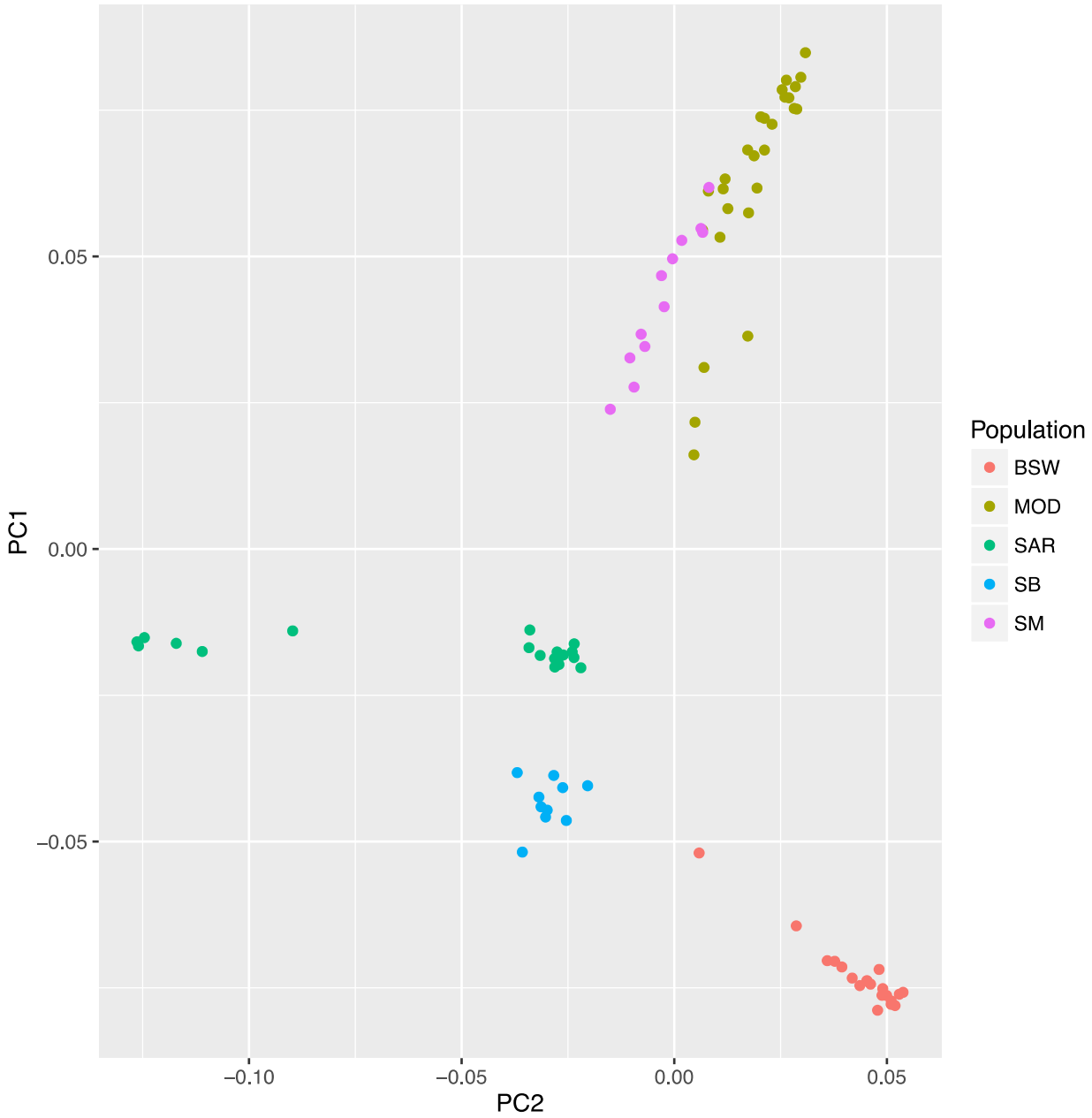


Figure 1

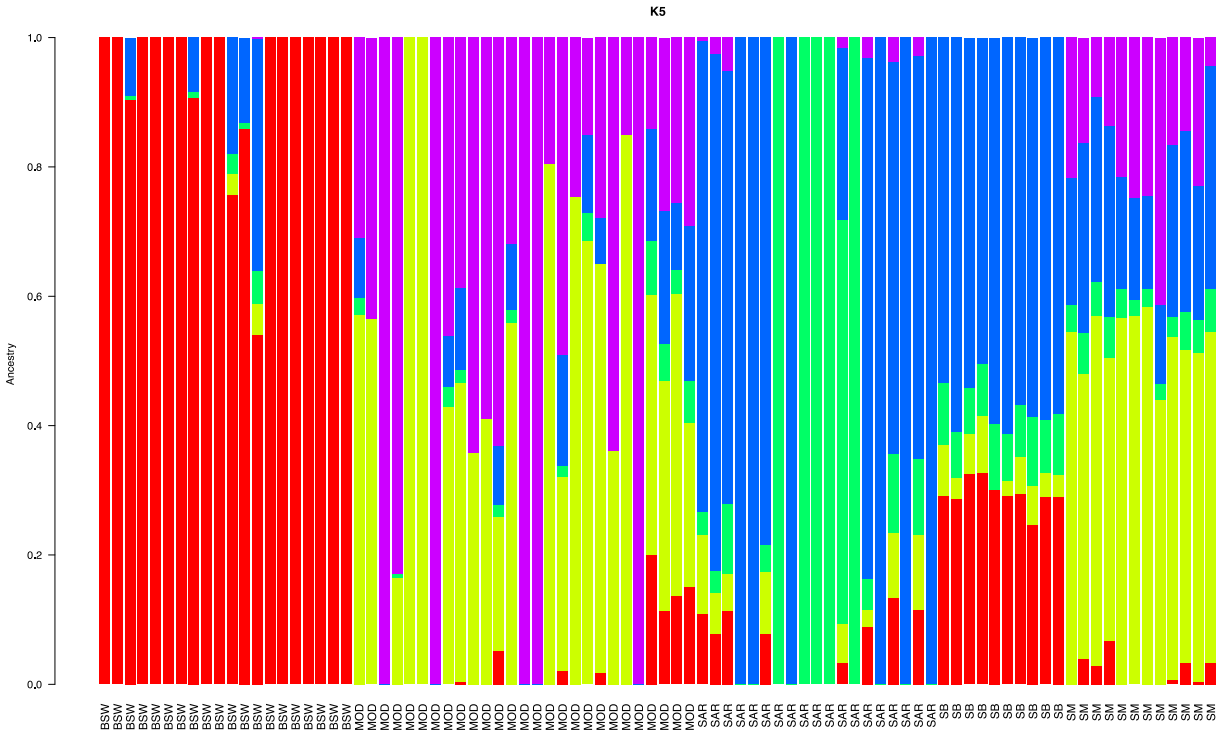


Figure 2

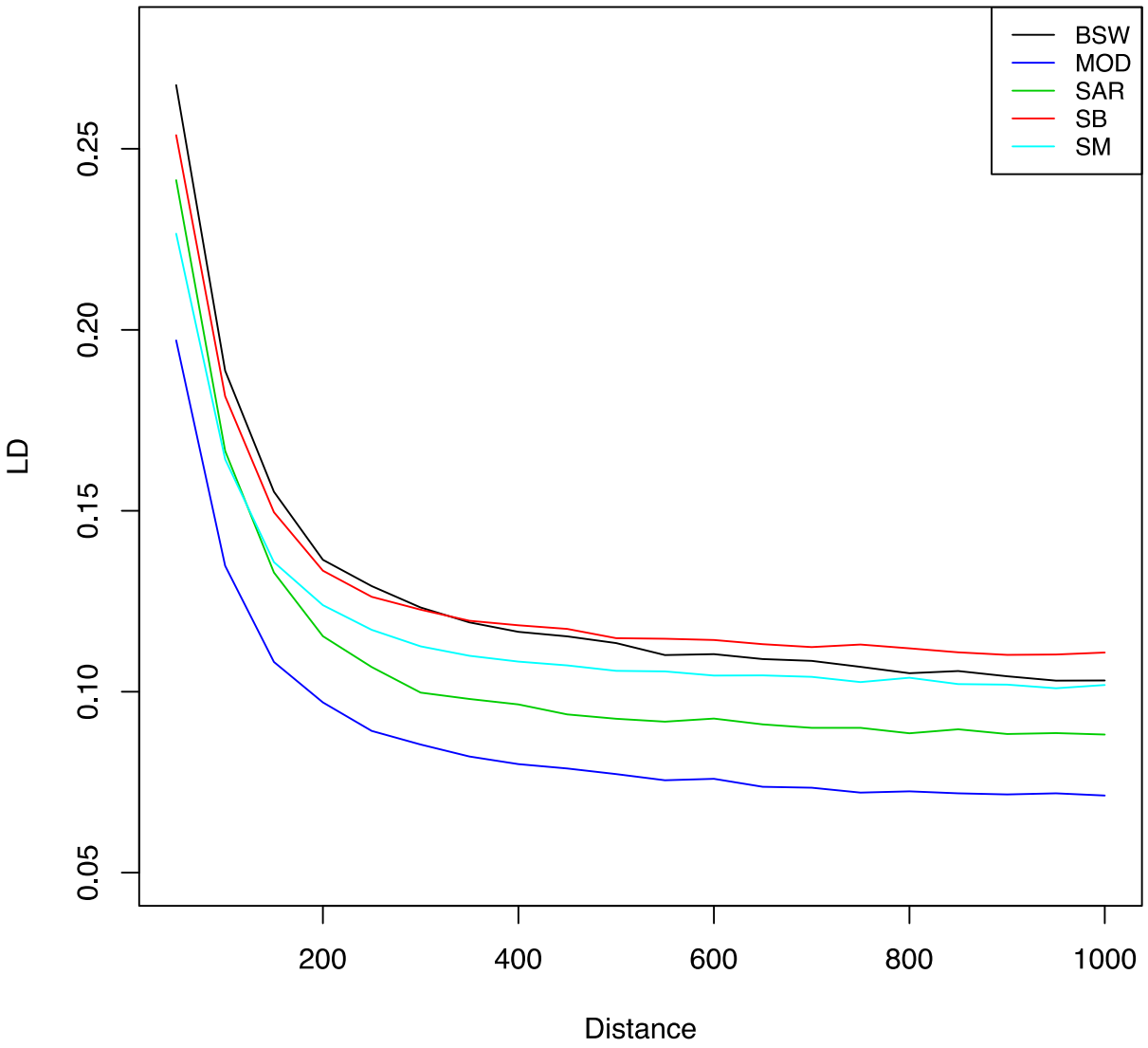
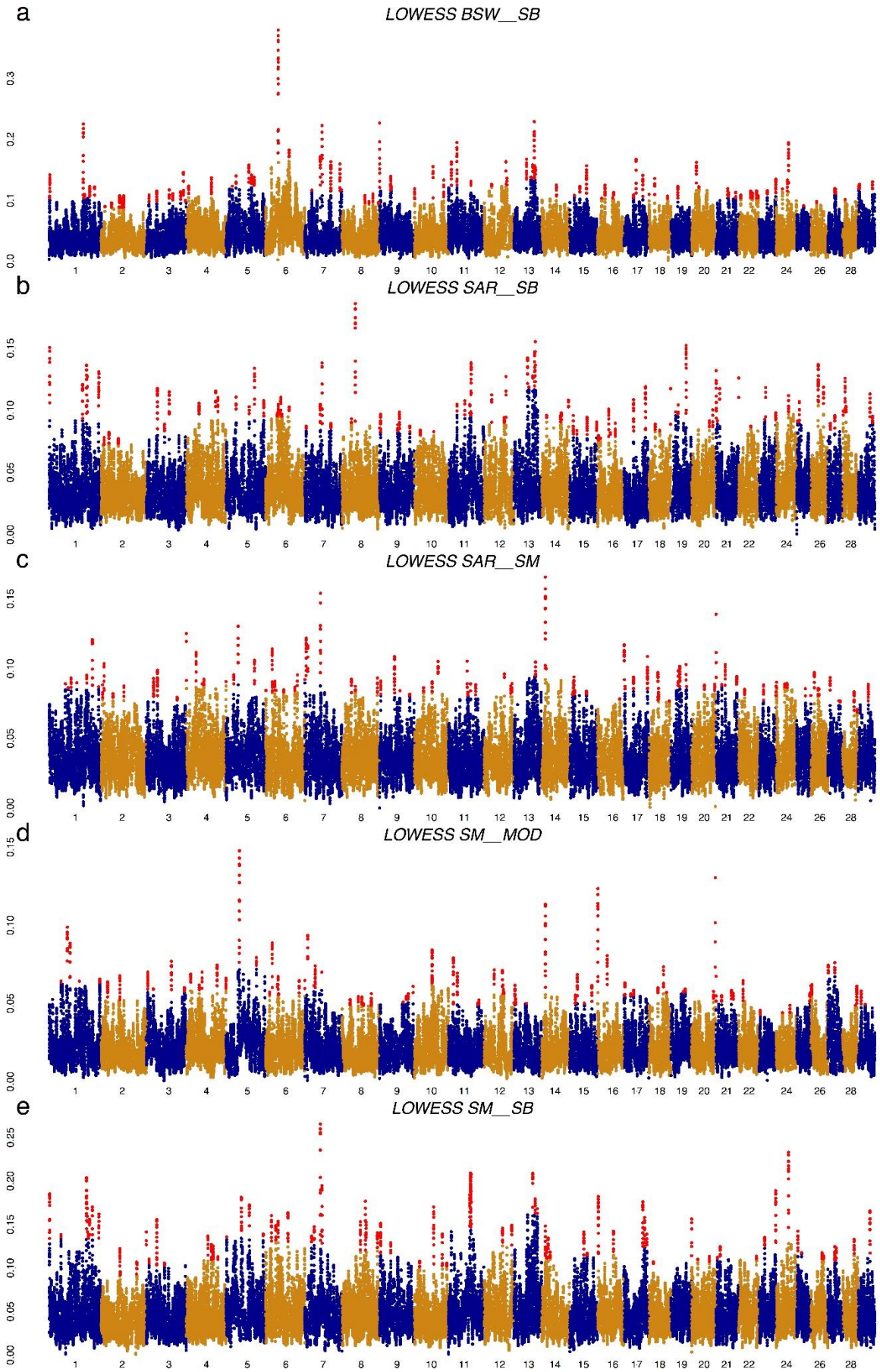
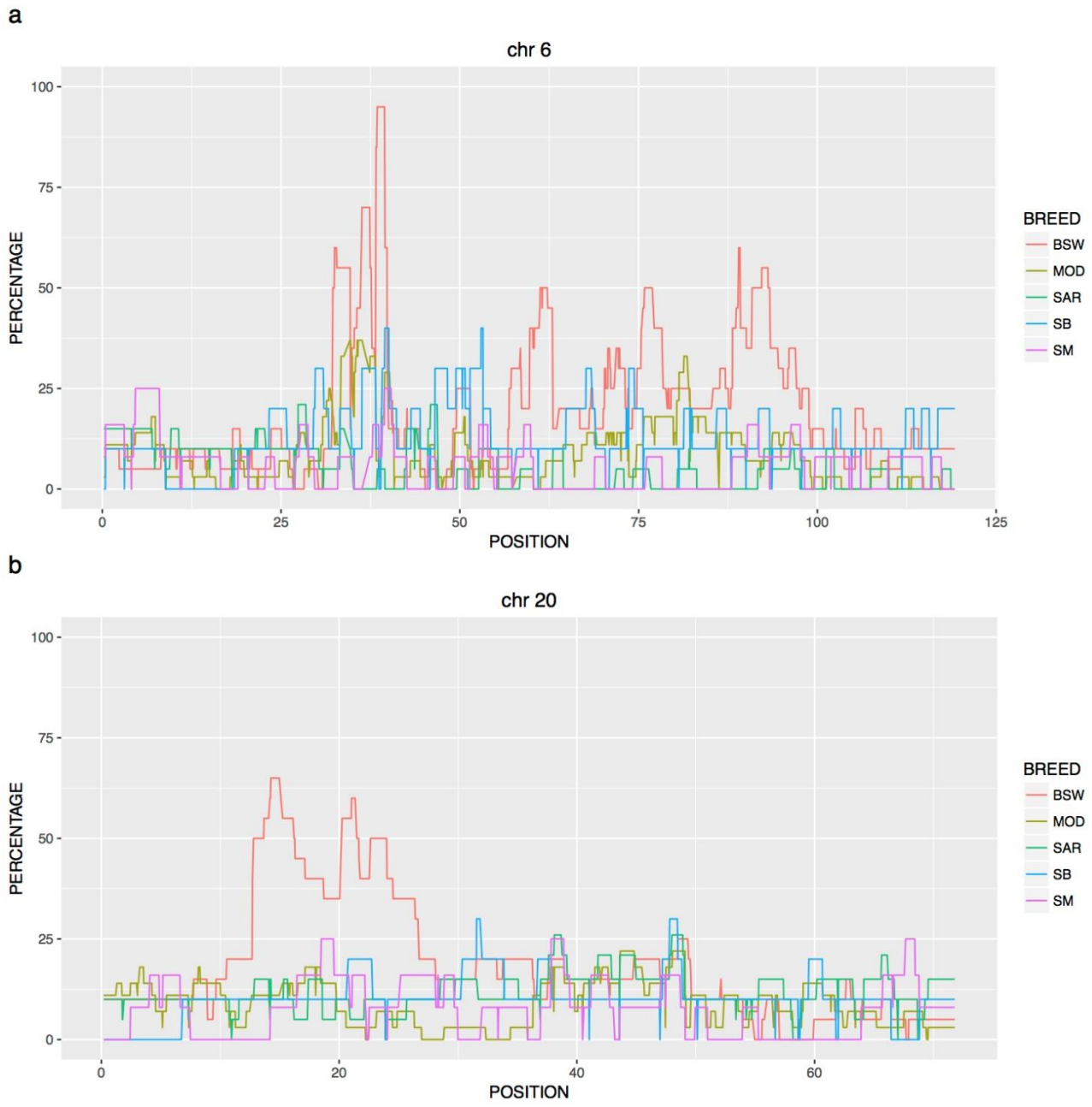


Figure 3



635

636 **Figure 4**



637

638 **Figure 5**

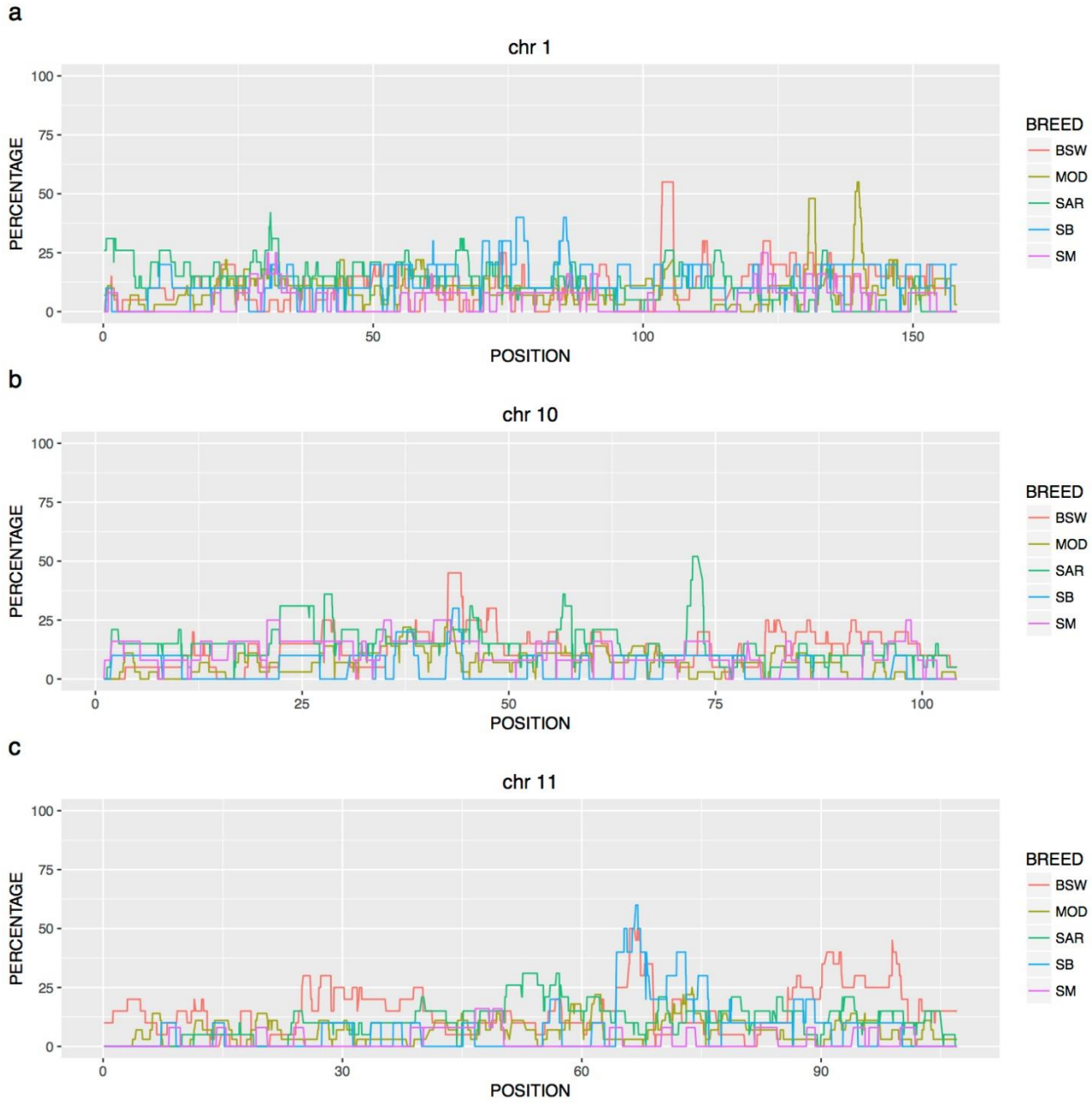


Figure 6