UNIVERSITÀ DEGLI STUDI FIRENZE

DOTTORATO DI RICERCA IN

Scienze Agrarie e Ambientali

CICLO XXXIV

CONSERVATION STATUS, GENETIC DIVERSITY AND SELECTION SIGNATURES OF ITALIAN AUTOCHTHONOUS BEEF BREEDS UTILISING PEDIGREE AND GENOMIC INFORMATION

Settore Scientifico Disciplinare AGR/17

Dottoranda

Dott.ssa Maria Chiara Fabbri

Tutore

Prof. Riccardo Bozzi

Coordinatore

Prof. Giacomo Pietramellara

Anni 2018/2021

Abstract

The recent alarming reports on global climate change and the challenges facing the agricultural sector to meet the increase in meat consumption, impose research in biodiversity. An important genetic pool of local breeds might play a crucial role in the near future to address these challenges. Italy is considered as one of the richest countries in biodiversity, but several Italian autochthonous cattle breeds are at risk of extinction. To safeguard biodiversity and increase genetic diversity within breeds, appropriate management tools must be developed. To achieve this, precise knowledge of the population structure and genetic diversity per breed are required. The thesis focuses on these needs, fixing on the local beef breeds hold by ANACLI (Italian national breeders association of Limousine and Charolaise breeds - http://www.anacli.it/), i.e., six different breeds, three from Tuscany (Calvana, Mucca Pisana and Pontremolese) and three from Sardinia (Sarda, Sardo Bruna and Sardo Modicana). All the six breeds have been recognized from the Italian breeders Association (AIA; Associazione Italiana Allevatori, Rome) to be at risk of extinction and at present are enrolled to the register of cattle breeds at limited diffusion. In each study, Limousine and Charolaise breeds are included in order to compare results between local beef breeds and two of the most important beef breeds reared in Italy.

The first study performed was based on pedigree analysis using the software ENDOG. Findings immediately describe the urgent situation especially for the three Tuscan breeds in terms of inbreeding and effective population sizes, which ranges between 14.62 (Pontremolese) to 39.79 (Sardo Modicana) in local breeds, values extremely lower than the two cosmopolitan breeds Charolaise (90.29) and Limousine (135.65). The average inbreeding coefficients were higher in Tuscan breeds (7.25%, 5.10%, and 3.64% for Mucca Pisana, Calvana, and Pontremolese, respectively) compared to the Sardinian breeds (1.23%, 1.66%, and 1.90% in Sardo Bruna, Sardo Modicana, and Sarda, respectively), while for Charolaise and Limousine they were <1%. Another informative parameter was the rates of matings between relatives. The highest rates of mating between half-siblings were observed for Calvana and Mucca Pisana (~9% and 6.5%, respectively), while the highest rate of parent–offspring mating was ~8% for Mucca Pisana.

A medium - low completeness of pedigree data was found in all the six local breeds, consequently the necessity to use genomic information to study population structure was strongly raised. Single nucleotide polymorphisms (SNPs) were used as markers in the next

analyses and animals were genotyped with GeneSeek GGP-LDv4 33k, a low-density SNP chip containing 30,111 SNPs. The first step in genomic analysis, was linkage disequilibrium investigation (LD) detection, because it is related to various evolutionary forces, such as inbreeding, nonrandom mating, population bottleneck, drift, recombination, and mutations, and hence is an essential parameter to examine population history. Average squared correlation between pairs of loci was similar in Calvana and Mucca Pisana (~0.14) and higher in Pontremolese (0.17); Limousine presented the lowest LD extent (0.07), and a more rapid LD decay. This analysis confirmed the genetic diversity loss, as dictated by the genomic effective population size estimates, of local breeds here analyzed.

To better investigate the genomic regions which describe the low genetic diversity detected, a runs of homozygosity (ROH) analysis was then performed. Runs of homozygosity consist of contiguous regions of the genome where an individual is homozygous in all sites, and so describe the breed autozygosity and define more accurate inbreeding estimates than pedigree data. High frequency of ROH might reflect selection signatures. The Charolaise, Limousine, Sarda, and Sardo Bruna breeds were found to have a high frequency of short ROH (~ 15.000); Calvana and Mucca Pisana presented also runs longer than 16 Mbp. Longer ROH are due to recent inbreeding, as recombination has not had the possibility of breaking up the homozygous segment, on the other hand, short ROH demonstrate an older origin because several meiosis have been occurred. This suggests that mating between relatives in the smallest population (especially in Tuscan breeds) occur until now and obviously it reflects on inbreeding. Indeed, the highest level of average genomic inbreeding was observed in Tuscan breeds, around 0.3, while Sardinian and cosmopolitan breeds showed values around 0.2. The frequency of ROH occurrence revealed eight breed-specific genomic regions where genes of potential selective and conservative interest are located (e.g. MYOG, CHI3L1, CHIT1 (BTA16), TIMELESS, APOF, OR10P1, OR6C4, OR2AP1, OR6C2, OR6C68, CACNG2 (BTA5), COL5A2 and COL3A1 (BTA2)).

All the obtained results in this thesis represent the first exhaustive genomic background description of Calvana, Mucca Pisana, Pontremolese, Sarda, Sardo Bruna, Sardo Modicana, Italian Limousine and Italian Charolaise breeds. Consequently, they may be used as a tool for preserving biodiversity of the aforementioned populations, and may provide the guidelines for correct management and conservation schemes.

Dedication

To my family,

Who believed,

believes

and will believe in me.

And I in them.

Acknowledgements

A special gratitude I give to my supervisor Prof. Riccardo Bozzi for giving me this opportunity and the possibility to understand what I want to do in my life.

Many thanks to Dr. Christos Dadousis who teaches me that limits exist to be overcome.

Thank you to my colleagues who made every working day simple and enjoyable.

Contents

List of tables

Table 4. Characterization of within-breed common runs of homozygosity at a threshold of 80% and with at least 15 SNPs and the number of genes included

List of figures

Chapter 1

Chapter 4

Chapter 1

1.1 Genetic diversity definition

Genetic diversity has been defined as the variety of alleles and genotypes present in a population and it reflects the differences between individuals in morphological, physiological and behavioral aspects [1]. Consequently, it can provide important information regarding the history and the evolution of the species as well as the population structure, becoming an aid to conserve potential genetic resources for future use.

In addition, the genetic diversity of the species depends on the diversity of the gene pool of each breed and also on the overlap between the gene pools of different breeds. Indeed, the genetic diversity of the target breed and the estimate of the part of its gene pool that does not overlap with other breeds are of paramount interest for population conservation and management. This non-overlapping part support the importance of breeds as reservoir for advantageous mutations that could be detected and used in breeding methods or for disadvantageous mutations that can be removed from population reared [2]. The loss of genetic diversity leads to increased homozygosity, increasing the probability to fix recessive deleterious alleles, and in this case an effect on the phenotype could happen. Furthermore, the homozygosity causes the decrease of fitness and fertility of the breed, which may also cause a decrease in performance [3,4]. The decrease in fitness, fertility, and performance that is associated with an increased level of inbreeding is called inbreeding depression [5].

1.2 Conservation of local breeds

In the last few decades growing attention has been placed to maintain biodiversity and to apply local breeds conservation strategies; several European projects have been developed such as TREASURE (https://treasure.kis.si/), EuReCa (https://www.regionalcattlebreeds.eu/) NEXTGEN (https://www.epfl.ch/labs/nextgen/) and many others.

Breed conservation is fundamental especially for those species whose wild ancestors are now extinct, such as cattle and horses [6] because there are not any other sources that can be used to protect the biodiversity. The causes of the loss of biodiversity are numerous: if on one hand the natural evolutionary forces such as mutation, adaptation, population bottleneck and genetic drift are natural agents that occurred on genetic diversity, on the other hand, the increasing human demand of animal products led to increase intensive selection programs, accelerating the formation of specialized breeds. As a result, a few highly productive breeds replaced local ones across the world [7], causing a strong erosion of genetic resources [8].

Local breeds are essential for the exploitation of heterosis which could avoid the selection plateau. Moreover, genetic diversity offers the only weapon against climate change, disease, changing availability of foodstuffs, social change etc. [9], which are the problems that more affect the 21st century. Thus, the conservation of genetic diversity is becoming the basis to respond to possible unexpected environmental changes because it guarantees adaptation abilities [1].

According to Food and Agriculture Organization (FAO) [10], the 12.6% of European and Caucasian local breeds are extinct, 23.8% are endangered and endangered maintained, and 19.7% are classified at critical and critical maintained risk. From these percentages, cattle represent a great part: 169 breeds are extinct, 117 at critical (and critical maintained) risk and 135 are endangered (endangered maintained). Breeds at risk are mostly threatened following some criteria which classify them into extinct, at critical, critical-maintained risk, endangered and endangered-maintained, i.e. the current population individuals (number of females and number of males), alterations in population size and the degree of crossbreeding with other breeds. Below, the risk status classification as reported by FAO [11]:

- *extinct*: a breed is categorized as extinct when there are no breeding males or breeding females remaining. Nevertheless, genetic material might have been

cryoconserved which would allow recreation of the breed. In reality, extinction may be realized well before the loss of the last animal or genetic material.

- *critical*: a breed is categorized as critical if the total number of breeding females is less than or equal to 100 or the total number of breeding males is less than or equal to five; or the overall population size is less than or equal to 120 and decreasing and the percentage of females being bred to males of the same breed is below 80 percent, and it is not classified as extinct.
- *critical-maintained*: are those critical populations for which active conservation programs are in place or populations are maintained by commercial companies or research institutions.
- *endangered*: a breed is categorized as endangered if the total number of breeding females is greater than 100 and less than or equal to 1 000 or the total number of breeding males is less than or equal to 20 and greater than five; or the overall population size is greater than 80 and less than 100 and increasing and the percentage of females being bred to males of the same breed is above 80 percent; or the overall population size is greater than 1 000 and less than or equal to 1 200 and decreasing and the percentage of females being bred to males of the same breed is below 80 percent, and it is not assigned to any of above categories.
- *endangered-maintained*: are those endangered populations for which active conservation programs are in place or populations are maintained by commercial companies or research institutions.

Several strategies are applied to the animal genetic resources conservation. One of them is *in situ* conservation which means that the breeds are reared in their natural surrounding where have developed their distinctive properties. For the aforementioned characteristics this approach is complex, because should include performance recordings, development of breeding and management schemes, and it is also expensive, especially if the breed is not used for production. The most used alternative to the latter, is *ex situ* conservation, namely the conservation away from the habitat and production system where the resources are developed. This approach includes both the maintenance of alive animals (*in vivo*) and the cryopreservation of reproductive organs (*in vitro*), which could consist of the cryoconservation of sperm, oocytes, embryos, but also the entire gonads [12]. These methods may have the aim to increase the virtual effective population size and minimizing inbreeding in local breeds. In addition, the long term storage of germplasm may assure the

re-establishment of a lost breed, the restoration of genetic diversity within a breed, or the use of specific genotypes to address new breeding goals [13].

1.3 Population under study

At the end of 20th century, the register of cattle breeds at limited diffusion (Registro Anagrafico delle razze bovine autoctone a limitata diffusione) has been instituted in Italy. It represents the instrument to safeguard the cattle breeds which are not included in a national selection program. The breeds enrolled in it are 16: Agerolese, Burlina, Cabannina, Calvana, Cinisara, Garfagnina, Modenese, Modicana, Mucca Pisana, Pezzata Rossa Oropa, Pontremolese, Pustertaler Sprinzen, Sarda, Sardo Bruna, Sardo Modicana and Varzese-Ottonese-Tortonese.

In Italy, starting from 2014 to 2020, a national project approved by MIPAAF - Management Authority of the PSRN Biodiversity sub-measure 10.2 - is funded, in order to foster knowledge, to enhance and safeguard the genetic diversity and conservation of farm animals (https://www.psrn.it/insights/tutelare-aumenta-le-potenzialita-produttive-attuali-e-future/). Indeed, the main purpose of this project is to characterize the Italian animal genetic resources and to maintain biodiversity, aimed at improving selection programs and at guaranteeing the conservation of breeds at risk of extinction. Nine different livestock sectors are included in PSRN 10.2 sub measure, namely milk, beef and dual-purpose cattle, sheep - goats, pigs, horses, asses, poultry and buffalo. Each branch included the autochthonous breeds present in Italy, based on species and production purpose.

The thesis takes into account and focuses on the local beef breeds hold by ANACLI (Italian national breeders association of Limousine and Charolaise breeds - http://www.anacli.it/), i.e., six different breeds, three from Tuscany (Calvana, Mucca Pisana and Pontremolese) and three from Sardinia (Sarda, Sardo Bruna and Sardo Modicana).

All the six breeds have been recognized from the Italian breeders Association (AIA; Associazione Italiana Allevatori, Rome) to be at risk of extinction and at present are enrolled to the register of cattle breeds at limited diffusion as previously reported. All the population sizes reported in the following paragraphs are updated in March 2021 and refer only to alive animals legally enrolled to the aforementioned register.

- **Pontremolese**

Pontremolese faces a critical risk of extinction, consisting a limited number of animals (n = 55). Historically, Pontremolese originated from the provinces of Massa Carrara, La Spezia, and Parma and, in the past, was used for the transport of Carrara marble. Nowadays, this breed is reared as beef breed, and is famous for its adaptability and rusticity capability (Figure 1).

Figure 1: Pontremolese sample and distribution of the farms in Tuscany, Italy.

- **Calvana**

Calvana (n = 391) originates from the Calvana mountain, in Prato's province. It is considered to be within the bio-type of Chianina breed, from which it differs in the smaller size due to a more difficult breeding environment. It is particularly suitable to live in marginal areas and in difficult grazing; Calvana breed has ever been considered as a beef breed, with a carcass yield of 65% (Figure 2).

Figure 2: Calvana sample and distribution of the farms in Tuscany, Italy.

- **Mucca Pisana**

Nowadays Mucca Pisana (n = 486) is mainly reared in the province of Pisa. The first documental evidence of the Mucca Pisana dates back to the early 1800s, in the lower valley of the Serchio river. The breed derives from the crossing of a local Podolica breed, which is genetically intermediate between the Maremmana and the Pontremolese breeds, with Alpine Brown Swiss (Schwyz) cattle, imported by the Lorena family in the second half of the 1700s. Since 1850 further breeding has been carried out with other cattle breeds, including Chianina.

The Mucca Pisana has a good aptitude to work, good production of milk which enables it to feed more than one calf and, above all, the good quality of the meat. It has a notable maternal behavior and can adapt to difficult environments and diets low in energy and rich in fodder (Figure 3).

Figure 3: Mucca Pisana sample and distribution of the farms in Tuscany, Italy.

- **Sarda**

The alive animals number of Sardinian breeds are definitely greater than Tuscan, but they are anyway classified at risk of extinction. Sarda breed consists in 21972 alive animals.

Until the last century, Sarda breed was characterized by high phenotypic variability because of its geographical distribution in Sardinia Island (mountain or plain). The quality of the meat is considered of high quality (Figure 4).

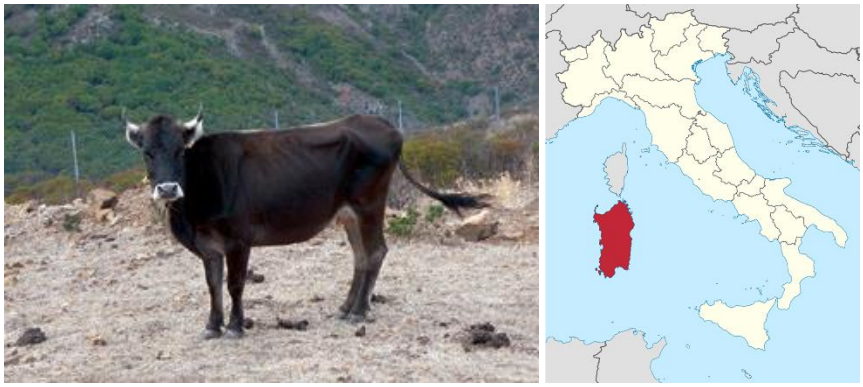

Figure 4: Sarda sample and the region in Italy where it is reared.

- **Sardo Bruna**

This breed shows 26923 alive animals reared in Sardinia, consequently, it is the more numerous studied group. It origins from several crossbreeding of autochthonous breeds with Bruna Alpina breed starting from XIX century (Figure 5). Its rusticity and adaptability make Sardo Bruna perfect to be rear in extremely marginal areas, becoming an alternative to sheep breeding.

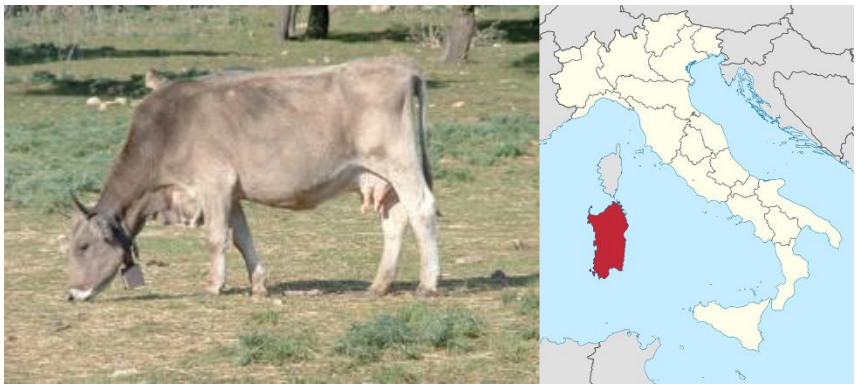

Figure 5: Sardo Bruna sample and the region in Italy where it is reared.

- **Sardo Modicana**

Sardo Modicana originates from a crossbreeding between Sarda breed (aforementioned) and Modicana Breed (reared in Sicily, the second main island of Italy), at the beginning of 19th century; the aim was to improve the Sarda breed work aptitude. With the industrialization, this breed suffered a drastic demographic reduction, with an increasing conversion to beef production. Nowadays, 2121 alive animals are reared (Figure 6).

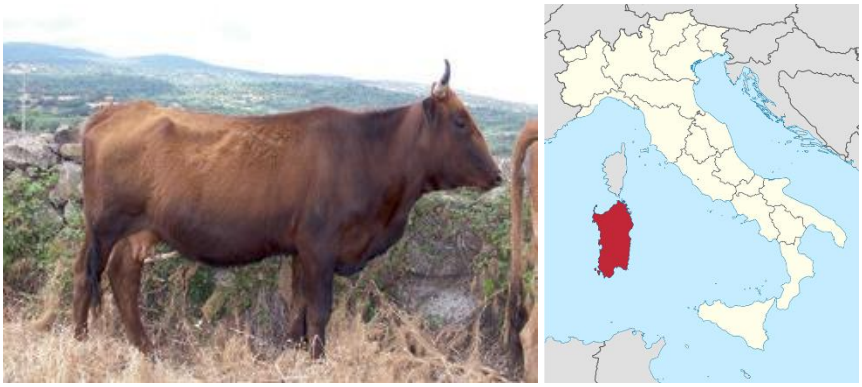

Figure 6: Sardo Modicana and the region in Italy where it is reared.

Hence, the historical, cultural, and ecological values of these six breeds are enormous and undisputed. Furthermore, the landscape of Tuscany and Sardinia placed barriers on widespread intensive breeding. As a result, the conservation and increase in sample sizes of local breeds could be of economic importance for these regions.

Chapter 2

2.1 Estimates of genetic diversity

Genetic diversity investigation and characterization lead to more effective conservation strategies.

Traditionally, the genetic diversity is evaluated with some parameters obtained by pedigree information, e.g. the probability that an individual is homozygous at a locus (inbreeding coefficient), the probability of sampling two identical alleles in a pair of individuals (coancestry coefficient) and the proportion of genetic information in the population derived from a specific ancestor [14]. More recently, with the advent of DNA related techniques, molecular data are more used to estimate genetic diversity. DNA markers are easier to analyze and observe, more reliable because the genetic information is unique for each species (and individual) and are independent of various aspects, such as age, physiological conditions, and environmental factors [15]. Two different classes of molecular markers have been mostly used to investigate genetic diversity: microsatellites and single nucleotide polymorphisms (SNPs). Microsatellites were largely used in the first part of 21st century; this kind of analysis have provided useful genetic information on cattle populations widespread around the world, from Europe [16] to South America [17] and Asia [18,19]. Recently, the availability of genotyping has made possible to provide a detailed evaluation of cattle genetic diversity globally [20]. Several methods could be applied to analyze genetic diversity with SNPs markers, such as Runs Of Homozygosity analysis [21], admixture analysis [22] and Linkage Disequilibrium detection [23]. In addition, the genomic inbreeding coefficient can be estimated with different formula and software, becoming an additional parameter to take into account when genetic diversity and population structure are investigated. In the studies here reported, two approaches have been chosen to calculate genomic inbreeding coefficients: the first examining identical by state (IBS) information SNP by SNP using a genomic relationship matrix (F_{GRM}; [24]) and the second using runs of homozygosity (F_{ROH}), which was proposed for the first time in humans by McQuillan et al. [25]. Several studies on cattle populations investigated the correlations between inbreeding calculated from pedigree (F_{PED}), F_{GRM} and F_{ROH}. The correlations between F_{GRM} and F_{ROH} were generally lower than those between F_{ROH} and F_{PED} [26–28]. Indeed, estimates based on ROH directly reflect homozygosity on the genome and have the advantage of not being affected by estimates of

allele frequency or incompleteness of the pedigree. For these reasons, F_{ROH} has become a good indicator of inbreeding levels, but, unfortunately, was shown that it is affected by the marker chip density used [29].

In the following sections, different approaches and parameters to estimate genetic diversity are described, using pedigree and molecular data.

2.1.1 Pedigree analysis

Pedigree analysis is an important tool to describe genetic variability and its evolution across generation.

This method can be traced back to Wright and McPhee (1925) [30], who examined the genetic structure in British Shorthorn cattle breed. Lacy (1989) [31] improved the method of pedigree analysis and described both the effective number of founders and the effective number of founder genomes. Later, Boichard et al. (1997) [32] developed the idea to use the probabilities of gene origin to measure genetic variability in a population. From those years, several studies have been performed in livestock populations.

One of the most common parameters calculated through pedigree is the Effective population size (Ne). The Ne of a population was elaborated by Wright [33] and is defined as the size of an idealized population undergoing the same rate of genetic drift as the population under study [34].

The maintenance of genetic diversity within a population is reached by maximizing the Ne of a population, or by minimizing the increase in inbreeding across generations [35]. This is because quantitative genetic theory argued that Ne describes the capacity of a population to respond to natural selection and the ability to evolve and adapt to the changes in its environment [34].

Ne is not usually equal to the census size (N) of a population, many discrepancies exist in cosmopolitan cattle breeds, indeed, in highly selected dairy cows Ne was found lower than 100 [3,36,37], as well as in some beef breeds, such as Simmental and Hereford, where Ne is critically low [38,39]. Ne has been also estimated in local breeds, but the pedigree-based method requires an adequate completeness of data over several generations and for local breeds, this is not an easy task for practical reasons related to breeding management. Boichard et al. [32] argued that the low pedigree completeness could result in overestimation of the Ne. However, it has been suggested that the threshold required to avoid the effect of inbreeding depression in short-term must be equal to 50 as well as larger than 500 to maintain the evolutionary potential of the population over long-term [9]. In conclusion, Ne is the most preferred parameter for determining the endangerment status of a breed [40] and it also allows to calculate the rate of inbreeding (ΔF), which describes the dynamics of variation in a population, and it is independent of either a reference population or pedigree depth [41]. The ΔF is inversely related to the effective population size. Pedigree analysis also allows to lead the estimation of another parameter which is considered

informative, i.e. generation intervals. It is the average age of parents when the next generation is born. If older animals are used as parents, the generation interval is longer and genetic change is slower. But, if we use younger animals and replace older generations with younger generations, the generation interval is shorter and genetic progress is more rapid. Obviously, decreasing the generation interval is the purpose of selection programs, but in local breeds generation interval describes the farm and matings management, indicating the demographic status.

2.1.2 Linkage disequilibrium

Linkage disequilibrium (LD) is the non-random association of alleles at different loci, it can lead to recombination events in the genome at low rates and, thus, to the conservation of segments from one generation to the other [42]. The size of conserved genomic segments depends on the time that recombination has occurred since the development of LD in the population. However, larger is the number of generations, shorter is the segments in LD. Furthermore, LD is influenced by other events such as selection, genetic drift and mutation [43], for all these reasons LD is also used to describe the population history. The first whole-genome LD study in cattle, to quantify the extent and pattern of LD, was performed using 284 microsatellite markers sampled from 581 maternally inherited gametes in Dutch black and white dairy cattle [44]. With the advent of high throughput genotyping several studies have been performed on the investigation of LD [45,46]. During the past decades, LD analysis has been successful in identifying genes for Mendelian diseases and/or traits in human [47] and livestock populations [48].

The power of genome-wide association study (GWAS) and accuracy of genomic selection (GS) largely depend on LD between quantitative trait loci (QTL) and markers [49], and for these reasons, LD is analyzed in highly selected breeds, such as Holstein, Angus, Limousine [23,50] but also in local breeds [42], to investigate the genetic diversity. Furthermore, since LD decays were normally found across generations, the LD level was widely utilized to estimate genomic Ne at any particular time in the past generations [51]. A small N_e means that alleles in the current population derive from a common ancestor belonging to few generations ago, because few recombination events occurred: the chromosome segments that are IBD are large, and so LD affects long distance [48]. This explanation assumes a constant N_e but, in practice, the N_e of a population can change over time. For instance, in *Bos taurus* cattle N_e was large before domestication (~ 90,000) and declined to approximately 100 after breed formation [52]. Thus, Ne estimated from LD is a very informative parameter when population structure and demographic history are evaluated.

2.1.3 Selection signatures

Selection signatures detection is the most recent approach used in population structure investigation.

Artificial selection in cattle resulted in breeds differentiation that are specialized in milk or meat production. These selection strategies tend to cause changes in specific genomic regions that control the breed traits such as morphology, production performance, reproduction, adaptation to different environments, and resistance to diseases [53].

The unique genetic patterns or footprints in the genomic regions subjected to selection are called “selection signatures”. Various statistical approaches have been proposed for the detection of selection signatures, such as extended haplotype homozygosity (EHH) and integrated haplotype score (IHS), both based on linkage disequilibrium because firstly, the core haplotypes is identified, and then the age of each core is assessed by the decay of LD according to distance [54]. An alternative approach to the detection of selection signatures is based on the measure of population differentiation due to locus-specific allele frequencies between populations, which is quantified using the F_{ST} statistic [33]. F_{ST} value ranges from 0 (no differentiation) to 1 (high differentiation). Highly differentiated allele frequency between the populations at any given locus indicates positive selection, whereas low F_{ST} values suggest negative selection. Nevertheless, this index could be overestimated when the sample size of the population under study is small [54].

Methods which focus on identifying genomic regions with reduced variation respect to the genome average exist, and the most known and recently used is runs of homozygosity (ROH) approach. ROH are contiguous lengths of homozygous genotypes that occur within an individual when two haplotypes share a recent common ancestor, consequently they are identical by descent (IBD) segments. Broman and Weber (1999) [55] were the first to recognize that long stretches of homozygous segments in human populations, most likely reflect autozygosity and may have implications for human health. Gibson et al. (2006) [56] developed this concept by analyzing the distribution of ROH in outbred human populations, giving importance to the number and the length of segments. At the moment, ROH are widely used by researchers in livestock populations [28], especially to assess genomic inbreeding levels, to define population structure and demography history. To the best of our knowledge, were Sölkner et al. (2010) [57] and Ferenčaković et al.(2011) [58] the first to apply ROH concept to cattle, followed by Purfield et al. (2012) [59]. Several studies came in quick succession in the last ten years, focusing on local cattle breeds, discovering genomic

regions linked to environmental adaptation [60], resistance and/or susceptibility to infection and diseases [61] and resilience capacity [62].

Chapter 3

3. Aim

The aim of this PhD thesis was to analyze in depth the genetic architecture of six Italian local cattle beef breeds: Calvana (CAL), Mucca Pisana (MUP), Pontremolese (PON), Sarda (SAR), Sardo Bruna (SAB) and Sardo Modicana (SAM). The aforementioned breeds are at risk of extinction thus, genetic and genomic analysis are primary step to study populations in order to elaborate conservation measures and handle inbreeding. Pedigree and genomic analysis were performed with the aim to individuate parentage, to explore the population structure, genetic diversity within and between breeds, to identify signatures of selection and to provide instruments to plan assortative matings to control the inbreeding. Two cosmopolitan beef breeds (Limousine, LIM and Charolaise, CHA) were added to the analysis in order to better compare results.

References

1. Toro MA, Caballero A. Characterization and conservation of genetic diversity in subdivided populations. Philos Trans R Soc Lond B Biol Sci. 2005;360:1367–78.

2. Wellmann R, Bennewitz J. Key Genetic Parameters for Population Management. Front Genet [Internet]. Frontiers; 2019 [cited 2021 May 20];10. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2019.00667/full

3. Doekes HP, Veerkamp RF, Bijma P, de Jong G, Hiemstra SJ, Windig JJ. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. Genet Sel Evol. 2019;51:54.

4. Carolino N, Gama LT. Inbreeding depression on beef cattle traits: estimates, linearity of effects and heterogeneity among sire-families. Genet Sel Evol GSE. 2008;40:511–27.

5. Charlesworth D, Willis JH. The genetics of inbreeding depression. Nat Rev Genet. Nature Publishing Group; 2009;10:783–96.

6. Sponenberg DP, Martin A, Couch C, Beranger J. Conservation Strategies for Local Breed Biodiversity. Diversity. Multidisciplinary Digital Publishing Institute; 2019;11:177.

7. Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, et al. Genetic diversity in farm animals--a review. Anim Genet. 2010;41 Suppl 1:6–31.

8. Scherf B, Food and Agriculture Organization of the United Nations, editors. World watch list for domestic animal diversity. 3rd ed. Rome: Food and Agriculture Organization of the United Nations; 2000.

9. Bolaji U-FO, Ajasa AA, Ahmed RO, Bello SF, Ositanwosu OE. Cattle Conservation in the 21st Century: A Mini Review. Open J Anim Sci. Scientific Research Publishing; 2021;11:304–32.

10. Risk Status of Animal Genetic Resources | Domestic Animal Diversity Information System (DAD-IS) | Food and Agriculture Organization of the United Nations [Internet]. [cited 2019 Sep 16]. Available from: http://www.fao.org/dad-is/risk-status-of-animal-genetic-resources/en/

11. Rischkowsky B, Pilling D. The state of the world's animal genetic resources for food and agriculture. Rome: Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations; 2007.

12. Wu JY, Sun YX, Wang AB, Che GY, Hu TJ, Zhang XM. Effect of newborn bovine serum on cryopreservation of adult bovine testicular tissue. Andrologia. 2014;46:308–12.

13. Woelders H, Joost Hiemstra S. The potential of cryopreservation and reproductive technologies for animal genetic resources conservation strategies. Cryobiology. 2011;63:316–7.

14. Eusebi PG, Martinez A, Cortes O. Genomic Tools for Effective Conservation of Livestock Breed Diversity. Diversity. Multidisciplinary Digital Publishing Institute; 2020;12:8.

15. Hamouda M. Molecular analysis of genetic diversity in population of Silybum marianum (L.) Gaertn in Egypt. J Genet Eng Biotechnol [Internet]. 2019 [cited 2021 May 20];17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6890919/

16. Maudet C, Luikart G, Taberlet P. Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. J Anim Sci. 2002;80:942–50.

17. Machado MA, Schuster I, Martinez ML, Campos AL. Genetic diversity of four cattle breeds using microsatellite markers. Rev Bras Zootec. Sociedade Brasileira de Zootecnia; 2003;32:93–8.

18. Suh S, Kim Y-S, Cho C-Y, Byun M-J, Choi S-B, Ko Y-G, et al. Assessment of Genetic Diversity, Relationships and Structure among Korean Native Cattle Breeds Using Microsatellite Markers. Asian-Australas J Anim Sci. 2014;27:1548–53.

19. Zhou GL, Jin HG, Zhu Q, Guo SL, Wu YH. Genetic diversity analysis of five cattle breeds native to China using microsatellites. J Genet. 2005;84:77–80.

20. Mastrangelo S, Tolone M, Ben Jemaa S, Sottile G, Di Gerlando R, Cortés O, et al. Refining the genetic structure and relationships of European cattle breeds through meta-analysis of worldwide genomic SNP data, focusing on Italian cattle. Sci Rep. Nature Publishing Group; 2020;10:14522.

21. Dixit SP, Singh S, Ganguly I, Bhatia AK, Sharma A, Kumar NA, et al. Genome-Wide Runs of Homozygosity Revealed Selection Signatures in Bos indicus. Front Genet [Internet]. Frontiers; 2020 [cited 2020 Jul 15];11. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2020.00092/full

22. Gebrehiwot NZ, Strucken EM, Aliloo H, Marshall K, Gibson JP. The patterns of admixture, divergence, and ancestry of African cattle populations determined from genome-wide SNP data. BMC Genomics. 2020;21:869.

23. Porto-Neto LR, Kijas JW, Reverter A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genet Sel Evol GSE. 2014;46:22.

24. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.

25. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83:359–72.

26. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet. 2015;46:110–21.

27. Zhang Q, Calus MP, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet [Internet]. 2015 [cited 2020 Jul 15];16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4509611/

28. Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, et al. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. J Anim Breed Genet. 2013;130:286–93.

29. Szmatoła T, Gurgul A, Jasielczuk I, Ząbek T, Ropka-Molik K, Litwińczuk Z, et al. A Comprehensive Analysis of Runs of Homozygosity of Eleven Cattle Breeds Representing Different Production Types. Animals. Multidisciplinary Digital Publishing Institute; 2019;9:1024.

30. Wright S, McPhee HC. An Approximate Method of Calculating Coefficients of Inbreeding and Relationship from Livestock Pedigrees. 1925 [cited 2021 Jul 7]; Available from: https://naldc.nal.usda.gov/catalog/IND43966972

31. Lacy RC. Analysis of founder representation in pedigrees: Founder equivalents and founder genome equivalents. Zoo Biol. 1989;8:111–23.

32. Boichard D, Maignel L, Verrier É. The value of using probabilities of gene origin to measure genetic variability in a population. Genet Sel Evol. 1997;29:5.

33. Wright S. Evolution in Mendelian Populations. Genetics. 1931;16:97–159.

34. Falconer DS, Mackay FCT. Introdutction to Quantitative Genetics. Introdutction Quant Genet. 1996;464–464.

35. Howard JT, Pryce JE, Baes C, Maltecca C. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J Dairy Sci. 2017;100:6009–24.

36. Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D, Villanueva B. Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in the Spanish Holstein Population. PLOS ONE. Public Library of Science; 2015;10:e0124157.

37. Sørensen AC, Sørensen MK, Berg P. Inbreeding in Danish dairy cattle breeds. J Dairy Sci. 2005;88:1865–72.

38. de Araujo Neto FR, Vieira DA, Santos DJ de A, Pessoa MC, Borquis RRA, de Oliveira HN, et al. Population structure of Simmental beef cattle using pedigree analysis. Trop Anim Health Prod. 2020;52:1513–7.

39. Mc Parland S, Kearney JF, Rath M, Berry DP. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. J Anim Sci. 2007;85:322–31.

40. Gandini GC, Ollivier L, Danell B, Distl O, Georgoudis A, Groeneveld E, et al. Criteria to assess the degree of endangerment of livestock breeds in Europe. Livest Prod Sci. 2004;91:173–82.

41. Hiemstra SJ, de Haas Y, Mäki-Tanila A, Gandini G. Local cattle breeds in Europe [Internet]. Wageningen Academic Publishers; 2010 [cited 2021 Sep 2]. Available from: https://www.wageningenacademic.com/doi/book/10.3920/978-90-8686-697-7

42. Toro Ospina AM, Maiorano AM, Curi RA, Pereira GL, Zerlotti-Mercadante ME, Dos Santos Gonçalves Cyrillo JN, et al. Linkage disequilibrium and effective population size in Gir cattle selected for yearling weight. Reprod Domest Anim Zuchthyg. 2019;54:1524–31.

43. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9:477–85.

44. Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N, Grisart B, et al. Extensive Genome-wide Linkage Disequilibrium in Cattle. Genome Res. 2000;10:220–7.

45. Prasad A, Schnabel RD, McKay SD, Murdoch B, Stothard P, Kolbehdari D, et al. Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. Anim Genet. 2008;39:597–605.

46. Yurchenko AA, Daetwyler HD, Yudin N, Schnabel RD, Vander Jagt CJ, Soloshenko V, et al. Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. Sci Rep [Internet]. 2018 [cited 2019 Mar 7];8. Available from: http://www.nature.com/articles/s41598-018-31304-w

47. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet. Nature Publishing Group; 2002;3:299–309.

48. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. Nature Publishing Group; 2009;10:381–91.

49. Xu L, Zhu B, Wang Z, Xu L, Liu Y, Chen Y, et al. Evaluation of Linkage Disequilibrium, Effective Population Size and Haplotype Block Structure in Chinese Cattle. Anim Open Access J MDPI. 2019;9.

50. de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage Disequilibrium and Persistence of Phase in Holstein–Friesian, Jersey and Angus Cattle. Genetics. 2008;179:1503–12.

51. Flury C, Tapio M, Sonstegard T, Drögemüller C, Leeb T, Simianer H, et al. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. J Anim Breed Genet. 2010;127:339–47.

52. MacEachern S, Hayes B, McEwan J, Goddard M. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (Bos taurus) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. BMC Genomics. 2009;10:181.

53. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet Sel Evol [Internet]. 2015 [cited 2019 Mar 7];47. Available from: http://www.gsejournal.org/content/47/1/49

54. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. Selection signatures in livestock genome: A review of concepts, approaches and applications. Livest Sci. 2020;241:104257.

55. Broman KW, Weber JL. Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain. Am J Hum Genet. 1999;65:1493–500.

56. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet. 2006;15:789–95.

57. Sölkner J, Ferencakovic M, Gredler-Grandl B, Curik I. Genomic metrics of individual autozygosity, applied to a cattle population. Book Abstr 61st Annu Meet Eur Assoc Anim Prod. 2010;

58. Ferencakovic M, Hamzic E, Gredler B, Curik I, Sölkner J. Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. Agric Conspec Sci. 2011;76:325–9.

59. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC Genet. 2012;13:70.

60. Xu L, Zhao G, Yang L, Zhu B, Chen Y, Zhang L, et al. Genomic Patterns of Homozygosity in Chinese Local Cattle. Sci Rep. Nature Publishing Group; 2019;9:16977.

61. Mastrangelo S, Sardina MT, Tolone M, Di Gerlando R, Sutera AM, Fontanesi L, et al. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. Animal. 2018;12:2480–8.

62. Strillacci MG, Vevey M, Blanchet V, Mantovani R, Sartori C, Bagnato A. The Genomic Variation in the Aosta Cattle Breeds Raised in an Extensive Alpine Farming System. Anim Open Access J MDPI. 2020;10:E2385.

Chapter 4

animals

MDPI

Article

Population Structure and Genetic Diversity of Italian Beef Breeds as a Tool for Planning Conservation and Selection Strategies

Maria Chiara Fabbri [1,*], Marcos Paulo Gonçalves de Rezende [2], Christos Dadousis [1], Stefano Biffani [3], Riccardo Negrini [4,5], Paulo Luiz Souza Carneiro [6] and Riccardo Bozzi [1]

[1] Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università di Firenze, 50144 Firenze, Italy; christos.dadousis@unifi.it (C.D.); riccardo.bozzi@unifi.it (R.B.)
[2] Associazione Nazionale Allevatori Bovini di Razza Piemontese, 12061 Carrù, Italy; mpgrezende@gmail.com
[3] Consiglio Nazionale delle Ricerche, Istituto di Biologia e Biotecnologia Agraria, 20133 Milano, Italy; biffani@ibba.cnr.it
[4] Associazione Italiana Allevatori, 00161 Roma, Italy; riccardo.negrini@unicatt.it
[5] Istituto di Zootecnica, Facoltà di Scienze Agrarie, Alimentari e Ambientali, Università Cattolica del S.Cuore,29100 Piacenza, Italy
[6] Universidade Estadual Sudoeste da Bahia, Jequié, Bahia 45205-490, Brazil; carneiropls@gmail.com
* Correspondence: mariachiara.fabbri@unifi.it

Received: 20 August 2019; Accepted: 22 October 2019; Published: 29 October 2019

Simple Summary: The recent alarming reports on global climate change and the challenges facing the agricultural sector to meet the increase in meat consumption, impose research in biodiversity. An important genetic pool of local breeds might play a crucial role in the near future to address these challenges. Although Italy is considered as one of the richest countries in biodiversity, there are autochthonous cattle breeds under extinction. To safeguard biodiversity and increase genetic diversity within breeds, appropriate management tools must be developed. To achieve this, precise knowledge of the population structure and genetic diversity per breed are required. This study analyzed pedigree data of six local beef breeds: Calvana, Mucca Pisana, and Pontremolese (from the region of Tuscany), all under extinction, and Sarda, Sardo Bruna, and Sardo Modicana, from the island of Sardinia, that are larger in number but of lower productivity. In addition, the study investigated the population structure of the

cosmopolitan beef breeds, Charolaise and Limousine, reared in the same regions and undergoing selection. The high mating percentage between relatives for Mucca Pisana and Calvana is an alarming situation for these breeds. The population structure of the Sardinian breeds suggests the application of breeding programs.

Abstract: The aim was to investigate the population structure of eight beef breeds: three local Tuscan breeds under extinction, Calvana (CAL), Mucca Pisana (MUP), and Pontremolese (PON); three local unselected breeds reared in Sardinia, Sarda (SAR), Sardo Bruna (SAB), and Sardo Modicana (SAM); and two cosmopolitan breeds, Charolaise (CHA) and Limousine (LIM), reared in the same regions. An effective population size ranges between 14.62 (PON) to 39.79 (SAM) in local breeds, 90.29 for CHA, and 135.65 for LIM. The average inbreeding coefficients were higher in Tuscan breeds (7.25%, 5.10%, and 3.64% for MUP, CAL, and PON, respectively) compared to the Sardinian breeds (1.23%, 1.66%, and 1.90% in SAB, SAM, and SAR, respectively), while for CHA and LIM they were <1%. The highest rates of mating between half-siblings were observed for CAL and MUP (~9% and 6.5%, respectively), while the highest rate of parent–offspring mating was ~8% for MUP. Our findings describe the urgent situation of the three Tuscan breeds and support the application of conservation measures and/or the development of breeding programs. Development of breeding strategies is suggested for the Sardinian breeds.

Animals **2019**, *9*, 880; doi:10.3390/ani9110880www.mdpi.com/journal/animals

Keywords: genetic diversity; beef cattle; pedigree analysis; autochthonous breeds; conservation

---

1. Introduction

Cattle domestication started in Southwest Asia, in the 9th millennium BC [1,2], while in Europe it began between 8800 to 8000 BC [3], due to migration from the Near East. The effective size of female cattle founders has been estimated to be ~80 animals [2,3]. Over time, natural and artificial selection resulted in the development of various breeds in the world. Artificial selection became more intense due to the Industrial Revolution

and urbanization which begun in the 19th century and drastically changed global food consumption as well as increased the request for meat production [4]. This further led to the development of modern breeding both in plants and animals at the beginning of the 20th century [5] to meet the levels of increased food consumption. For several decades breeding programs were mainly focused on the development of high-performance cattle breeds, specialized for dairy, beef, or dual-purpose (milk and meat). As a result, today's global food market is heavily based on cosmopolitan breeds. Nevertheless, the high success of cosmopolitan breeds worldwide resulted in the loss of interest in local breeds, which represented valuable genetic resources [6–8]. This situation isvery likely to continue in the future based on predictions for the upcoming decades of a continued increase in the global population and food consumption [5,9–11].

The Food and Agricultural Organization (FAO) has reported 1224 local cattle breeds worldwide [12]. From those, 181 are extinct, 105 at critical risk, and 140 are considered endangered. The majority of those breeds are of European and Caucasian origin (119 extinct, 91 at critical risk, and 108 endangered). In Italy, 61 cattle breeds are registered, 51 of which represent local breeds. From those, 18 are extinct, 7 at critical risk, 8 endangered, and 3 in a vulnerable situation [13]. The Italian Breeders Association (AIA; Associazione Italiana Allevatori, Rome), has officially recognized 16 local cattle breeds at risk of extinction. From those, six are considered as beef (Calvana, Mucca Pisana, Pontremolese, Sarda, Sardo Bruna, and Sardo Modicana) and the remaining 10 as dual-purpose breeds. All 16 breeds are enrolled in the register of autochthonous cattle populations at limited diffusion (Registro Anagrafico delle razze bovine autoctone a limitata diffusione), with the aim to safeguard and adopt conservation measures for them. There is no breeding program running for any of these breeds and some populations are under extinction or at critical numbers, with a high risk of extinction. Regarding the Italian beef sector, the preference of Italian farmers for the Piedmontese, firstly, and later on for the imported Limousine and Charolaise, over other beef breeds has resulted in a loss of interest for the rest of the Italian local beef breeds in the past several decades.

Nevertheless, there exist advantages related to the presence and maintenance of local breeds [14]. Firstly, local breeds represent a significant genetic and economic resource, being able to adapt to various landscapes where cosmopolitan breeds cannot benefit. They are more rustic and resistant to their local environment than their cosmopolitan

counterparts. Moreover, they represent an important gene bank that could be essential to address future climate changes, or potential disease outbreaks [15], and hence to preserve the global food production chain [16]. In addition, they play an important role in the preservation of human cultural inheritance. For example, local breeds have been used by farmers for organic farming and manufacturing of niche products in mountainous regions. Nevertheless, low production remains the limiting factor for the farming of local breeds that endangers their existence.

Genetic diversity is a primary step for the establishment of a breeding program or to take conservation measures. It is defined as the measure of genetic differences between and within groups of animals and is highly related to the selection and adaptation of a breed to the local environment. There are several causes that influence genetic diversity, such as migration, mutation, selection, drift, bottleneck, and inbreeding [16]. The first two processes may also bring an increase in genetic diversity, while selection via assortative mating and high levels of inbreeding (as a result of unsupervised mating among relatives) can lead to allele fixation. The inbreeding rate (ΔF; per year or generation) has been used to evaluate how genetic diversity evolves during breed history [17]. In particular, ΔF summarizes the increase in inbreeding values per generation (or year) at a population level, providing an overview of genetic diversity and the risk of inbreeding depression. This last phenomenon is essential to be controlled in livestock species, because it can cause a decrease in performance (such as growth, meat quality, and quantity) [18,19] and reduced fitness [17]. The rate of inbreeding is also related to the effective population size (Ne) by the equation $\Delta F = \frac{1}{2}Ne$. Ne is defined as the number of individuals that effectively participate in producing the next generation and is often lower than the census of the total number of individuals [20]. Consequently, it is a factor that contributes to the genetic diversity ofa breed and its conservation status.

Pedigree analysis is a primary step that enables the characterization of the genetic diversity of populations: it identifies genetic variability and changes in the population structure in consecutive generations. It describes and quantifies the increase of homozygosity and the level of inbreeding in the population, important factors to be considered either for breeding or conservation schemes. Indeed, demographic analyses can also help to understand factors regarding genetic history, conservation status of a population, and relationships within and between breeds [21]. Although several studies have been carried out on the population structure and genetic diversity of

cattle [22–24], only a fewwere focused on small populations.

Our aim was to investigate the population structure and genetic diversity of six Italian local beefbreeds, three reared in Tuscany (Calvana (CAL), Mucca Pisana (MUP), Pontremolese (PON)) and three in Sardinia (Sarda (SAR), Sardo Bruna (SAB), and Sardo Modicana (SAM)), utilizing pedigree information to support the development of strategies for conservation or breeding. Following FAO legislation, CAL, MUP, and SAM were classified as endangered breeds, while PON was set to a critical situation. No risk of extinction exists for SAR and SAB breeds, with each of these two breeds having a few thousand animals. In addition, to compare with cosmopolitan breeds, we analyzed pedigree datafrom the Italian populations of Charolaise (CHA) and Limousine (LIM).

2. Materials and Methods

2.1.Data

Pedigree information of the local breeds was supplied by the Italian Breeders Association (AIA; Associazione Italiana Allevatori, Rome) and by the breeding association (Associazione Nazionale Allevatori delle razze bovine Charolaise e Limousine–ANACLI, Rome) for the two cosmopolitan breeds. Our full dataset included animals born between 1980 and 2018, and consisted of 2798 CAL, 3399 MUP, 328 PON, 97,163 SAR, 74,981 SAB, 25,355 SAM, 99,464 CHA, and 322,321 LIM cattle. Table 1 summarizes the pedigree data per breed.

Table 1. Number of pedigree records (*N*), number of males, females, and generations for each breed.

Breed [1]	*N*	Males	Females	Generations
Tuscan				
CAL	2798	1201	1597	10
MUP	3399	1447	1952	14
PON	328	147	181	13
Sardinian				
SAR	97,163	28,869	68,294	11
SAB	74,981	13,697	61,284	10
SAM	25,355	10,398	14,957	12
Cosmopolitan				
CHA	99,464	39,171	60,293	18
LIM	322,321	133,445	188,876	15

[1] CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = SardoModicana; CHA = Charolaise; LIM = Limousine.

2.2. Pedigree Analysis

Genetic analysis was carried out with the ENDOG v4.8 software [25]. Pedigree completeness was evaluated with the following parameters: (i) equivalent complete generations (equiGen; defined as the sum for $(1/2)^n$, where n is the number of generations separating the individual from each of its known ancestors [26]); (ii) maximum complete generations (maxGen; number of generations separating an animal from its furthest ancestor [27]); and (iii) full complete generations (fullGen; number of generations separating the offspring from the furthest generation, where both parental lines of the individual are known. Ancestors with both parents unknown were considered as founders). To identify individuals with insufficient pedigree information to estimate inbreeding, the pedigree completeness index (PCI) was also calculated [28].

$$PCI = \frac{4 \times C_{sire} \times C_{dam}}{C_{sire} + C_{dam}}$$

where C_{sire} and C_{dam} were contributions from the paternal and maternal lines calculated (separately foreach line) as

$$C = \frac{1}{d} \sum_{i=1}^{d} g_i$$

where g_i is the proportion of known ancestors in generation *I* and *d* is the total number of generations. Generation interval (GI) was defined with two measures: (i) the average age of parents at the birth of all their progenies and (ii) the average age of parents at the birth of the progenies that were used for reproduction. Both parameters were calculated for the classical four pathways (father–son, father–daughter, mother–son, and mother–daughter).

2.3. Genetic Diversity

Genetic diversity was described with three parameters: (i) inbreeding coefficient (*F*; the probabilitythat an individual has two identical alleles by descent) calculated according to Meuwissen and Luo [29]; average relatedness coefficient (AR; the probability that an allele randomly chosen from the whole population belongs to a given animal) that defines the mean relationship of each individual with the rest of the population, and was computed following Gutiérrez et al. [25]; and (iii) rate of inbreeding ($\Delta F =$

$\frac{1}{2*Ne}$) where *Ne* is the effective population size. *Ne* was computed via regression as following:

$$Ne = \frac{1}{2 * b}$$

where *b* is the regression coefficient of the individual *F* over equiGen. To overcome pedigree incompleteness, equiGen was used.

In addition, for a better description of the population structure within each breed, the frequency of mating between close relatives—full-siblings (sibs), half-sibs, and parent with offspring—were calculated.

Also, the effective number of founders (*fe*) [22], and the effective number of ancestors (*fa*) [25] were considered. The *fe*/*fa* ratio and the number of ancestors explaining 50% [22,24] of the genetic contribution (ANC_50), expressed as a percentage on reference population [25], were also calculated. Ratio *fe*/*fa* = 1 shows the absence of bottleneck in the population under study, and low ANC_50 is an indicator of the founder effect [30].

Pedigree content (i.e., the proportion of known parents in each generation) was analyzed for each breed to estimate the contribution of each ancestor (for male and female lines) up to the fifth parental generation [25].

2.4. Principal Component Analysis

Principal component analysis (PCA) summarizes correlated variables into a reduced set of mutually uncorrelated variables (PCs, principal components), allowing a dimensionally reduced visualization while keeping a certain amount of the original variance. Each of the PCs contains all the original variables. The PCs are constructed by maximum variability explained in the data and with the constrain to be orthogonal to each other. This helps to summarize information and to better study relationships among the samples [31]. PCA was performed on a set of population parameters to identify potential differences among the breeds under study. The parameters considered were average inbreeding coefficient (AVG_F), true mean inbreeding (TMI; including only the animals with at least three complete generations traced), GI, ANC_50, average relatedness (AR), average pedigree content (P_CONT), and *Ne* calculated based on equiGen. Moreover, the relative change of population size, expressed both as average population size ratio ($\mathrm{APSR} = mean\left[\left(\frac{Ny}{Ny+1} \times 100\right) - 100\right]$, where *N* is the total number of animals per year and *y* = {2000, . . . , 2016}), and average standard deviation of population

size ratio (APSSD), were included. Past software was used for the PCA [32].

3. Results

3.1. Pedigree Analysis

The pattern of male to female ratio per generation was similar for CAL, MUP, SAM, SAR, CHA, and LIM, with more stable numbers between males and female during the generations, while SAB and PON had fluctuating trends (Supplementary Material, Figure S1). Table 2 summarizes the average of the four pedigree completeness parameters for each breed.

Table 2. Pedigree completeness parameters.

Breed [1]	EquiGen [2]	FullGen [3]	MaxGen [4]	PCI (%) [5]
Tuscan				
CAL	2.87	2.04	4.44	66
MUP	3.91	2.44	7.55	74
PON	2.10	1.06	4.24	38
Sardinian				
SAR	1.10	0.64	1.89	22
SAB	0.75	0.45	1.20	15
SAM	1.85	1.08	3.18	39
Cosmopolitan				
CHA	2.79	1.51	6.39	50
LIM	3.07	1.79	5.83	59

[1] CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine. [2] equiGen = average values of equivalent complete generations. [3] fullGen = average values of full complete generations. [4] maxGen = average values of maximum completegenerations. [5] PCI = pedigree completeness index expressed as a percentage.

MUP had the highest PCI values (74%), followed by CAL and LIM (66% and 59%, respectively). Intermediate PCI values were observed for CHA, PON, and SAM, while SAR and SAB had the lowest PCI values. As expected, pedigree quality increased over time with a similar pattern for both PCI and equiGen indices (Figures S2 and S3). GI in years was rather high for all local breeds, with the lowest values observed in SAM (7.8) and the highest in SAB (13.3). SAR and CAL had similar GI (~10). As shown in Table 3, LIM and CHA had a lower average GI (7.0 and 6.7, respectively).

It should be noted, however, that there was variation within each breed, with standard deviation (SD) estimates of GI being equal to the mean or slightly higher. Moreover, the highest GI for the father–son and father–daughter paths was found for PON. The lowest values for the father–son path were observed for SAB and CHA,

while CHA also had the lowest GI for the father–daughter path.

The maternal intervals were shorter than the paternal for MUP and PON, while for SAB, SAM, CHA, and LIM, maternal pathways were greater. Equal father/mother to daughter GI was observed for SAR (10.7). CAL showed similar GI in the four pathways which varied between 9.55 (father–son) and 10.54 (mother–son) (Table 3). The largest differences between maternal and paternal pathways were observed for PON (paternal pathway ~15 years, maternal pathways ~9 years) as shown in Figure S4.

Table 3. Generation interval [1] in years for each breed [2], calculated in the classical four pathways (standard deviations in parenthesis).

Pathway	CAL	MUP	PON	SAR	SAB	SAM	CHA	LIM
Father to sons	9.55(13.83)	9.26(11.33)	13.20(14.01)	11.22(14.79)	6.17(7.69)	7.26(11.61)	6.35(7.46)	7.25(10.11)
Father to daughters	10.37(14.17)	9.86(11.04)	17.12(17.59)	10.66(14.24)	7.24(8.70)	7.35(11.45)	5.46(4.72)	6.22(6.89)
Mother to sons	10.54(13.30)	6.82(7.32)	8.10(10.25)	8.57(8.47)	10.86(12.68)	7.88(6.47)	7.62(8.80)	9.13(12.55)
Mother to daughters	10.27(12.77)	8.20(9.78)	9.85(9.95)	10.69(12.68)	16.49(18.96)	8.19(8.28)	7.78(8.11)	7.52(8.04)
Total Interval	10.29(13.49)	8.94(10.38)	12.51(13.86)	10.60(13.19)	13.30(16.72)	7.80(9.54)	6.69(7.00)	7.05(8.13)

[1] Generation interval was measured as the average age of parents at the birth of all their progenies; [2] CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine.

3.2. Genetic Diversity

Estimated genetic parameters per breed are summarized in Table 4. The average *F* for the Tuscan breeds were 7.25%, 5.10%, and 3.64% for MUP, CAL, and PON, respectively. Sardinian breeds showed lower values (1.23%, 1.66%, and 1.90% for SAB, SAM, and SAR, respectively), while CHA and LIM had inbreeding coefficients less than 1%. True mean inbreeding (TMI) was higher than *F* values in all the breeds studied.

Table 4. Genetic variability parameters for each breed.

Breed [1]	*Ne* [2]	Δ*F* (%) [3]	*F* (%) [4]	TMI (%) [5]	AR (%) [6]
Tuscan					
CAL	19.68	2.54	5.10	6.00	6.39
MUP	18.52	2.70	7.25	8.00	10.54
PON	14.62	3.42	3.64	5.60	7.15
Sardinian					
SAR	16.64	3.00	1.90	5.10	0.04
SAB	18.91	2.64	1.23	5.10	0.05

SAM	39.79	1.26	1.66	2.80	0.37
Cosmopolitan					
CHA	90.29	0.55	0.96	1.30	0.20
LIM	132.65	0.37	0.71	0.90	0.20

[1] CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine. [2] *Ne* = effective population size based on equivalent generations. [3] Δ*F* = rate of inbreeding. [4] *F* = inbreeding coefficient. [5] TMI = true mean inbreeding. [6] AR = average relatedness.

In general, inbreeding increased by generation. For CHA and LIM, the increase was relatively small (Figure 1).

Estimates of *Ne* also varied among breeds. For the local breeds, the values ranged from 14.62 (PON) to 39.79 (SAM). *Ne* was similar for MUP, SAB, SAR, and CAL. Concerning the two cosmopolitan breeds, *Ne* estimates were much higher (Table 4). The average Δ*F* was very low in LIM, CHA, and SAM, while the Tuscan breeds, SAR and SAB, had inbreeding rates ranging between 2.54% (CAL) and 3.42% (PON). The AR values, expressed as a percentage, were generally higher in breeds with high *F*. The *fe*/*fa* ratio was practically around 1 for all the local breeds indicating the absence of narrow bottlenecks, whereas CHA and LIM had a higher ratio (3 and 2.1, respectively) (Table 5).

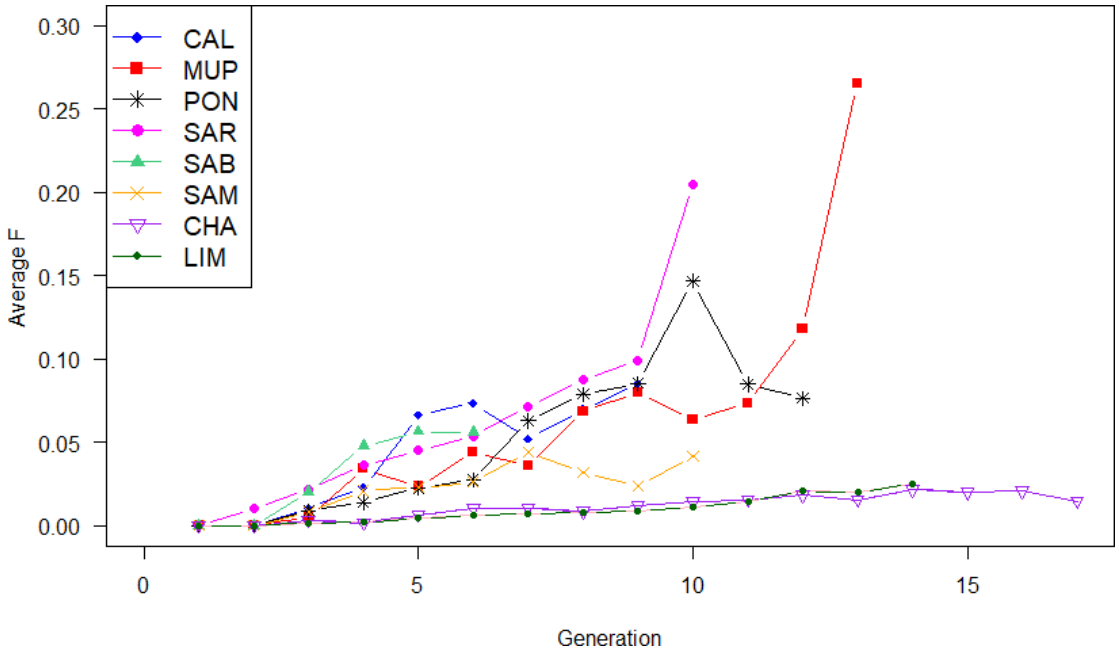

Figure 1. Rate of inbreeding (Δ*F*) per generation for each breed, where CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine.

Table 5. Population parameters for each breed.

Breed [1]	*fe/fa* [2]	ANC_50 [3]	P_CONT (%) [4]	APSR [5]	APSSD [6]
Tuscan					
CAL	1.1	8	92	1.40	14.34
MUP	1.1	5	94	−1.42	27.57
PON	1.1	5	73	−103.23	454.11
Sardinian					
SAR	1.2	542	61	−4.03	21.85
SAB	1.2	294	48	2.26	20.08
SAM	1.2	96	78	−4.38	9.50
Cosmopolitan					
CHA	3.0	219	77	3.60	7.05
LIM	2.1	330	83	6.04	6.25

[1] CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine. [2] *fe/fa* = ratio of effective number of founders to effective number of ancestors. [3] ANC_50 = ancestors explaining 50% of the genetic contribution. [4] P_CONT = pedigree content. [5] APSR = population size expressed as average ratio throughout the years. [6] APSSD = population size expressed as average standard deviation throughout the years.

The estimated ANC_50 reflects the size of the different populations, with extremely low values found for the Tuscan breeds, intermediate for SAM (~100), and high values (>200) for SAB, SAM, CHA, and LIM (Table 5). APSR and APSSD were much higher (>|100|) in PON, compared to the rest, indicating large fluctuations in the population size throughout the years. For the rest of the breeds, APSR varied between −4.38 (SAM) and 6.04 (LIM), while APSSD ranged between 6.25 and 27.57 (for LIM and MUP, respectively).

Regarding P_CONT, CAL and MUP breeds presented the highest average values of the first generation (92% and 94%, respectively), followed by LIM, whereas the lowest value was found for SAB (48%) (Table 5). The percentages of known parents, grandparents, great grandparents, and so on, for both sire and dam lines are expressed in Figure S5. CAL, CHA, and LIM had more complete paternal than maternal lines, while the opposite was found in PON. Incomplete pedigree was found for all the Sardinian breeds, even in recent generations. In contrast MUP had almost complete information up to the fifth generation.

All breeds had a very low percentage of mating between full-sibs (<1%) with the highest value being observed for PON (0.6%) (Figure 2).

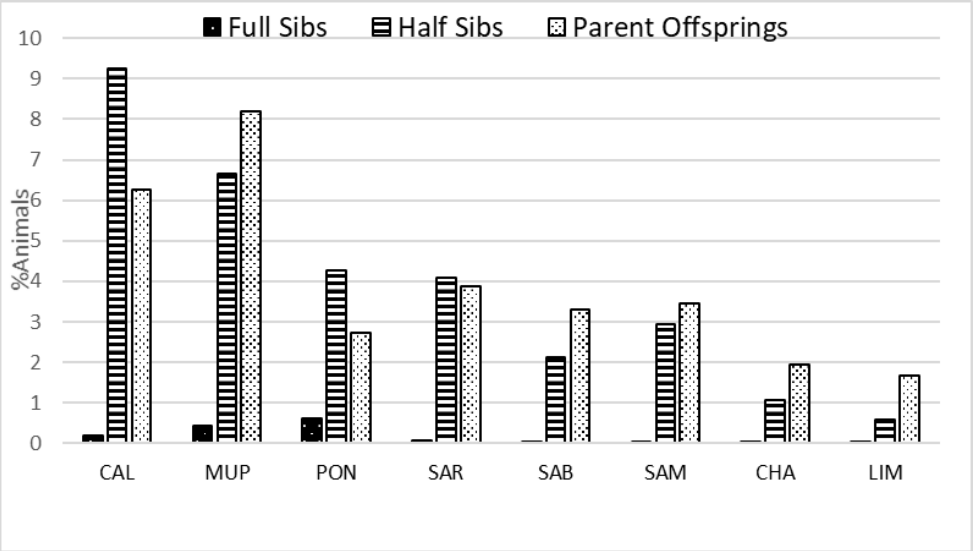

Figure 2. Percentage of animals involved in matings between close relatives (between full-siblings (sibs), half-sibs, and parent–offspring) in each breed, where CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine.

However, clear differences were observed for the half-sibs and the parent–offspring matings. In these cases, CAL and MUP had the highest percentages (9.25% and 6.65%, respectively) for the matings between half-sibs; for parent–offspring percentages were 8.2% and 6.25% for MUP and CAL, respectively. The cosmopolitan breeds had the lowest values in all cases.

J.3. PCA of the Population Structure Parameters

The biplot of the first two PCs, explaining together 78.31% of the total variation among breeds, is shown in Figure 3. In general, PC1 (capturing 50.27% of the total variability) separated the local from the cosmopolitan breeds, with the exception of SAM which was placed closer to the cosmopolitan breeds, while PC2 (explaining 28.04% of the variability) further separated the Tuscan from the Sardinian breeds. More precisely, three groups were formed: (i) CAL and MUP; (ii) SAR and SAB; and (iii) LIM and CHA. SAM clustered between the Sardinian and the cosmopolitan breeds, while PON was on the sideline. CAL and MUP were located near the parameters linked to inbreeding and relatedness (i.e., AVG_F, AR, and TMI), while CHA, LIM, and SAM were connected to the effective population size (*Ne*) and the *fe*/*fa* ratio. In PC1 all parameters were

loaded except P_CONT, while PC2 was mainly described by the P_CONT and AVG_F, with *Ne* and *fe/fa* ratio having loadings close to zero.

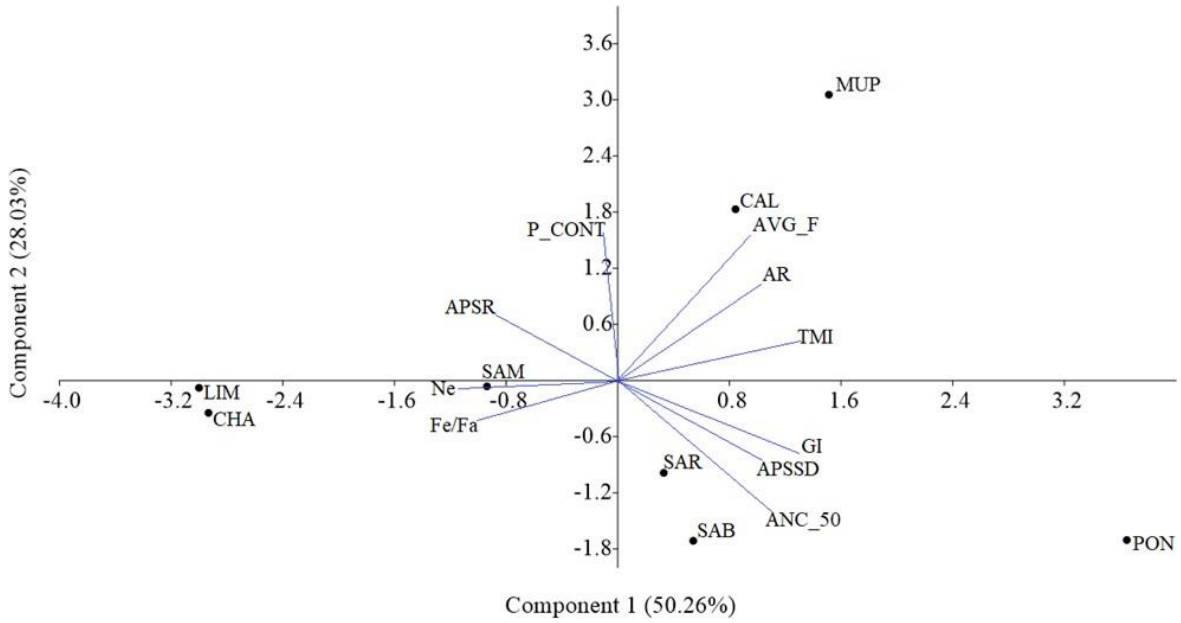

Figure 3. Biplot of the first two principal components. Principal component analysis (PCA) performed on the following population parameters: true mean inbreeding (TMI); average coefficient inbreeding (AVG_F); average relatedness (AR); effective population size (*Ne*); effective number of founders/effective number of ancestors (*fe/fa*); ancestors explaining 50% (ANC_50); pedigree content (P_CONT); population size expressed as average ratio throughout the years (APSR); population size expressed as average standard deviation throughout the years (APSSD); generation interval (GI). The vectors represent the variables and the points represent the breeds.

4. Discussion

To the best of our knowledge, this is the first study utilizing full pedigree records for the CAL, MUP, PON, SAR, SAB, and SAM local beef breeds together with two Italian beef populations of the CHA and LIM. Merging those breeds, our dataset consisted of three subgroups: three Tuscan breeds under extinction (CAL, MUP, and PON); three local breeds from the island of Sardinia (SAR, SAB, and SAM, each consisting of a large population and without undergoing a breeding program); and two cosmopolitan beef breeds (CHA, LIM), that are mainly reared in Italy in the regions of Tuscany and Sardinia and have recently set up a national breeding scheme.

The three Tuscan breeds are under extinction, hence drastic measures need to be taken for their conservation. At present, only 263 CAL (37 males and 226 females), 346 MUP (52 males and 294 females), and 52 PON (8 males and 44 females) cattle are alive. A pedigree

analysis to investigate the relationships among individuals and the levels of inbreeding within each breed is a primary step.

4.1. Pedigree Analysis

The level of inbreeding within a breed is closely related and dependent on the pedigree completeness [33], because incomplete pedigree data can underestimate the level of inbreeding in a population [34]. In general, the degree of pedigree completeness was lower in Sardinian than in Tuscan and cosmopolitan breeds (Table 2). Cappelloni [35] reported pedigree completeness of 2.44, 3.18, and 1.72 equiGen for CAL, MUP, and PON, respectively, which is lower than those found in our analysis. This was somehow expected, since the quality of the pedigree data has increased over time and in more recent years. Pedigree completeness of all the local breeds in our study was also higher compared to Spanish local beef breeds investigated by Gutiérrez et al. [36], but similar to the more recent study by Cañas-Álvarez et al. [37], who focused on the same Spanish populations analyzing demographical changes until 2009. Torrecillas et al. [38] also analyzed MUP pedigree data, but equiGen values were lower than in the current study (2.26). This could be attributed to the smaller number of animals in the pedigree (*n* = 1231) as well as to higher pedigree incompleteness.

Regarding the cosmopolitan breeds, the number of equiGen found in LIM was similar to Slovenian Limousine (3.38) reported by Kadleč ík et al. [27]. In general, LIM and CHA had lower values of pedigree completeness compared to previously reported data on European Charolaise (ranging from 8.3 in Swedish populations to 9.3 in French populations) and European Limousine (6.5 in Irish population to 7.5 in Swedish and British populations) [22]. However, values of equiGen reported in these studies were averaged over a specific time period (e.g., between 2004–2008 in European Charolaise).

Gutiérrez et al. [36] and Cañas-Álvarez et al. [37] analyzed Spanish local beef breeds; their average GI in years was smaller (from 3.75 in Sayaguesa to 7.83 in Morucha) than those found in our study. The GI values of LIM and CHA were comparable to other commercial breeds like Angus and Nellore [39]. For PON, MUP, and SAR our analysis showed the longest GI of the sire–offspring pathways compared with the dam–offspring pathways (Table 3). Similar results have been reported by Mc Parland et al. [40] in Charolaise, Limousine, Hereford, Angus, Simmental, and Holstein Friesian breeds. For the two cosmopolitan breeds and SAB results were opposite. Similar findings, however,

have been previously reported in Illawarra Shorthorn, Hereford, and two Asturiana breeds [41–43]. This could be partly attributed to the use of artificial insemination in cosmopolitan breeds nowadays, which is almost entirely missing in the local breeds, and farming practices (e.g., the longevity of the dams within each breed).

4.2. Genetic Diversity

The highest average inbreeding values were observed for Tuscan breeds (from 3.64% for PON to 7.25% for MUP) as shown in Table 4. The inbreeding level of CHA (0.96%) was similar to Swedish, Irish, and Danish Charolaise, but higher than French Charolaise (0.67%) [22]. The LIM had lower inbreeding than American (1%) [44] and Irish (1.08%) Limousine [40].

The TMI values were higher in breeds with a PCI lower than 40% (PON, SAR, SAB, and SAM) (Tables 2 and 4), confirming the underestimation of the inbreeding coefficient when pedigree is incomplete [30]. The level of inbreeding had, in general, a linear increase in local breeds among generations (Figure 1), except for PON where the changes in the number of animals throughout generations produced an erratic trend.

Another measure commonly used to assess the genetic variability within a breed is *Ne*. Meuwissen [45] proposed a threshold value of 50 animals to prevent the loss of genetic variability. The local breeds in our study had lower values of *Ne* than the threshold proposed, ranging between 14 and 40. However, a common problem related to the analysis of *Ne* is the amount of missing data in the pedigree. Boichard et al. [30] argued that the low pedigree completeness could result in overestimation of the *Ne*. In our study, the Sardinian breeds had very low pedigree quality, as depicted by the estimation of equiGen and the PCI (Table 4), and although the population size of those breeds is large enough, *Ne* was low (SAR = 16.64; SAB = 18.91; SAM = 39.79), suggesting that the situation could be more alarming in terms of loss of genetic variability.

The situation is different for the Tuscan breeds. The higher pedigree quality (greater values of equiGen and PCI) allowed for more accurate estimation of *Ne*. The low values found for the three populations, ranging between 14.6 (PON) and 19.7 (CAL), together with the small populations and the number of farms, report these breeds to be in an alarming situation. Notably, nowadays ~48% of the alive CAL cattle and ~70% of PON belong to four farms, while for MUP, five farms keep ~79% of the total population. This is a worrisome fact. In the case of an outbreak disease in the area there will be a thread on the existence of the Tuscan breeds.

Compared to other Italian local beef breeds (Chianina, Marchigiana, and Romagnola) analyzed by Bozzi et al. [46], the three Tuscan breeds had lower *Ne* estimates. In a more recent study, Mastrangelo et al. [21] analyzed genomic data for the same local and cosmopolitan breeds presented in this study. The reported *Ne* estimates differed from our estimates (Calvana = 33.5; Mucca Pisana = 8.7; Pontremolese = 7.2; Sarda = 62.2; Sardo Bruna = 1021.3; Sardo Modicana = 54.8; Charolaise = 67.8; Limousine = 468.9); however, this discrepancy could be attributed to a different approach used from the authors, who estimated *Ne* from the relationship between linkage disequilibrium (LD), *Ne,* and recombination rate. Moreover, their analysis was based on a small sample of the total population per breed (24, 23, 24, 30, 10, 28, 25, and 20, for CAL, MUP, PON, SAR, SAB, SAM, CHA, and LIM,respectively), which might not be representative of the population. Future genomic analysis utilizing a larger number of animals could help in reducing this discrepancy. Nevertheless, this is another indicator of the sensitivity of the *Ne* estimates upon the methodology applied.

Regarding the level of inbreeding, FAO suggests a threshold of 1% per generation to maintain reproductive fitness [47]. Several studies had analyzed Δ*F* with different approaches, estimating annual Δ*F* [22,36] or per generation [48], while others emphasize Δ*F* during the last generation [46]. The six local breeds had Δ*F* greater than 1% per generation with the highest value observed for PON (3.42%).The AR between individuals of the Tuscan populations ranged from 6.4 to 10.5, indicating that animals shared a high percentage of alleles in relation to the population. The AR values observed in Sardinian breeds were low (<0.4%), but this could also be an artefact due to the lack of complete pedigree data.

The *fe*/*fa* ratio close to 1 that was found in all local breeds indicates a high balance between the founders' contributions and consequently, an absence of bottleneck effect [30]. As expected, LIM and CHA presented greater values, which could be mainly attributed to selection. Estimates of ANC_50 (Table 5) suggested the presence of founder effect for the Tuscan breeds. Regarding the pedigree content, in general, the completeness of sire pathways was higher in more distant generations. Similar findings have been reported in Spanish local beef breeds [36]. The maternal line information was more complete only in the last generations.

The proportions of mating between close relatives were also examined. The most alarming situations were observed for CAL, PON, MUP, and SAR, with rates of half-sibs matings >4% and up to 9.25% for CAL (Figure 2). Although matings between full-sibs are

avoided, high rates of mating between half-sibs and parent–offspring are worrisome. In contrast, the cosmopolitan breeds had a very low percentage of matings between close relatives. This was to some extent expected, since both populations in Italy were based so far on imported animals and semen from abroad, while the national breeding program was recently initialized. Finally, a PCA performed on a set of estimated population parameters revealed the similarity and a common structure between CHA–LIM, SAR–SAB, and CAL–MUP (Figure 3).

4.3. Measures of Conservation

Present results suggest the necessity of safeguarding measures that will guarantee the physical, economic, and logistic viability of the three Tuscan breeds and thereby their existence. High rates of parent–offspring and half-sibs matings in MUP outline the necessity for the development of an appropriate mating scheme. Several ways exist to preserve animal populations from extinction and maintain genetic diversity within a population: (i) genetic/genomic tools, (ii) biotechnology, management, (iv) scientific support, (v) cultural relatedness, and (vi) political engagement. In the first two categories, cryoconservation of semen, embryos and oocytes could be reported [4,49]. In addition, control of mating targeting either sustaining biodiversity or maintaining favorable characteristics of the animals, or both, could be applied [50,51]. Optimal contribution offers one possible way to achieve this [52]. Moreover, animal genotyping will improve not only the correct parental assignment but also will provide with a clearer description of relationships among individuals as well as similarities among different breeds at the genomic level [21]. The use of multiple ovulation embryo transfer (MOET) is another strategy that could be utilized to keep favorable genetic material in the population. Scientific support could be further enhanced via the development of research nucleus per breed to investigate the variability in a set of phenotypes and the potential of selective breeding. For local breeds, historical bonds with the culture can be found and should not be overlooked. To complement this, the development of appropriate marketing of the final product (mainly meat in our case) that will guarantee reliance from the consumers as well as a satisfying income to the farmers should be considered. A techno-economic analysis could be a further step to assess the applicability of different scenarios and to quantify cost–benefit.

5. Conclusions

Our analysis outlined the critical situation, in terms of population size and genetic diversity, for the Pontremolese, Calvana, and Mucca Pisana breeds. Concerning the breeds of Sardinia (Sarda, Sardo Bruna, and Sardo Modicana) low pedigree quality poses restrictions for an accurate assessment of the genetic diversity. However, trends of pedigree completeness in recent years are encouraging and towards the desired direction. Genomic data, favorably of a large and representative sample per breed, are expected to bridge this gap and to shed more light on the genomic background of the aforementioned breeds and the level of their genomic diversity.

As expected, based on the history of the CHA and LIM breeds, there is space for intense selection. However, this should not be thoughtlessly applied, but under an optimized scheme taking into account both an increased genetic gain of the traits under selection and maintenance of genetic diversity.

Supplementary Materials: The following are available online at . Figure S1. Number of males and females per generation for each breed, where CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine. Figure S2. Pedigree completeness investigated for each breed as the average pedigree completeness index (PCI) for each generation, where CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine. Figure S3. Pedigree completeness investigated as the average equivalent generations (equiGen) for each breed for each generation, where CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine. Figure S4. Generation intervals for each breed, for the four pathways: father–son, father–daughter, mother–son, and mother–daughter, where CAL = Calvana; MUP= Mucca Pisana; PON = Pontremolese; SAR = Sarda; SAB = Sardo Bruna; SAM = Sardo Modicana; CHA = Charolaise; LIM = Limousine. Figure S5. Pedigree content for each breed, for (**a**) Tuscan breeds, (**b**) Sardinian breeds, (**c**) cosmopolitan breeds. Each line is structured as follows: the top is the paternal line and the bottom isthe maternal line.

Author Contributions: R.B. and S.B. conceived the idea, formulated the objectives of this study, and supervised the project. M.C.F. and M.P.G.d.R. conducted the analysis and wrote the first draft of the paper. C.D. helped in data preparation and in supervising the project. R.N. and P.L.S.C. contributed to discussions and critically contributed to the final version of the manuscript. All authors read and approved the final manuscript.

Funding: This research was funded by ANACLI, grant number 2015.99.2264.1127.

Acknowledgments: The research was funded by ANACLI through the I-BEEF project PSRN 2014-2020. Sottomisura 10.2: Biodiversità animale. We acknowledge the Italian Breeders Association (AIA) and Associazione Nazionale Allevatori delle razze bovine Charolaisee e Limousine (ANACLI) for providing the data.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Larson, G.; Bürger, J. A population genetics view of animal domestication. *Trends Genet.* **2013**, *29*, 197–205. [CrossRef] [PubMed]
2. Scheu, A.; Powell, A.; Bollongino, R.; Vigne, J.-D.; Tresset, A.; Çakırlar, C.; Benecke, N.; Bürger, J. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet.* **2015**, *16*, 54. [CrossRef] [PubMed]
3. Bollongino, R.; Burger, J.; Powell, A.; Mashkour, M.; Vigne, J.-D.; Thomas, M.G. Modern Taurine Cattle Descended from Small Number of Near-Eastern Founders. *Mol. Biol. Evol.* **2012**, *29*, 2101–2104. [CrossRef] [PubMed]
4. Hiemstra, S.J. *Local Cattle Breeds in Europe: Development of Policies and Strategies for Self-Sustaining Breeds*; Academic Publishing: Wageningen, The Netherlands, 2010; ISBN 978-90-8686-144-6.
5. Hickey, J.M.; Chiurugwi, T.; Mackay, I.; Powell, W. Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **2017**, *49*, 1297–1303. [CrossRef]
6. Anderson, S. Animal genetic resources and sustainable livelihoods. *Ecol. Econ.* **2003**, *45*, 331–339. [CrossRef]
7. Wollny, C.B. The need to conserve farm animal genetic resources in Africa: Should policy makers be concerned? *Ecol. Econ.* **2003**, *45*, 341–351. [CrossRef]
8. Ovaska, U.; Soini, K. Local Breeds—Rural Heritage or New Market Opportunities? Colliding Views on the Conservation and Sustainable Use of Landraces. *Sociol. Rural.* **2017**, *57*, 709–729. [CrossRef]
9. Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food Security: The Challenge of Feeding 9 Billion People. *Science* **2010**, *327*, 812–818. [CrossRef]
10. Nelson, G.C.; Rosegrant, M.W.; Palazzo, A.; Gray, I.; Ingersoll, C.; Robertson, R.; Tokgoz, S.; Zhu, T.; Sulser, T.B.; Ringler, C.; et al. *Food Security, Farming, and Climate Change to 2050: Scenarios, Results, Policy Options*; International Food Policy Research Institute: Washington, DC, USA, 2010; ISBN 978-0-89629-186-7.
11. The Economist Intelligence Unit. *Global Food Security Index 2018*; Economist Intelligence Unit: London, UK, 2018.
12. Risk Status of Animal Genetic Resources|Domestic Animal Diversity Information System (DAD-IS)|Food and Agriculture Organization of the United Nations. Available online: http://www.fao.org/dad-is/risk-status-of-animal-genetic-resources/en/ (accessed on 16 September 2019).

13. Domestic Animal Diversity Information System (DAD-IS) Food and Agriculture Organization of the United Nations. Available online: http://www.fao.org/dad-is/en/ (accessed on 16 September 2019).
14. Mendelsohn, R. The challenge of conserving indigenous domesticated animals. *Ecol. Econ.* **2003**, *45*, 501–510.
[CrossRef]
15. Curone, G.; Filipe, J.; Cremonesi, P.; Trevisi, E.; Amadori, M.; Pollera, C.; Castiglioni, B.; Turin, L.; Tedde, V.; Vigo, D.; et al. What we have lost: Mastitis resistance in Holstein Friesians and in a local cattle breed. *Res. Veter Sci.* **2018**, *116*, 88–98. [CrossRef]
16. Oldenbroek, K. *Utilisation and Conservation of Farm Animal Genetic Resources*; Wageningen Academic Publisher:
Wageningen, The Netherlands, 2007; ISBN 978-90-8686-032-6.
17. Hill, W.G.; Mackay, T.F.C. Falconer and Introduction to quantitative genetics. *Genetics* **2004**, *167*, 1529–1536. [PubMed]
18. Sumreddee, P.; Toghiani, S.; Hay, E.H.; Roberts, A.; Agrrey, S.E.; Rekaya, R. Inbreeding depression in line 1 Hereford cattle population using pedigree and genomic information1. *J. Anim. Sci.* **2019**, *97*, 1–18. [CrossRef] [PubMed]
19. Pereira, R.; Santana, M.; Ayres, D.; Bignardi, A.; Menezes, G.; Silva, L.; Machado, C.; Josahkian, L.; Albuquerque, L. Inbreeding depression in Zebu cattle traits. *J. Anim. Breed. Genet.* **2016**, *133*, 523–533. [CrossRef]
20. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **2009**, *10*, 195–205. [CrossRef] [PubMed]
21. Mastrangelo, S.; Ciani, E.; Marsan, P.A.; Bagnato, A.; Battaglini, L.; Bozzi, R.; Carta, A.; Catillo, G.; Cassandro, M.; Casu, S.; et al. Conservation status and historical relatedness of Italian cattle breeds. *Genet. Sel. Evol.* **2018**, *50*, 35. [CrossRef] [PubMed]
22. Bouquet, A.; Venot, E.; Laloë, D.; Forabosco, F.; Fogh, A.; Pabiou, T.; Moore, K.; Eriksson, J.-A.; Renand, G.; Phocas, F. Genetic structure of the European Charolais and Limousin cattle metapopulations using pedigree analyses. *J. Anim. Sci.* **2011**, *89*, 1719–1730. [CrossRef] [PubMed]
23. Piccoli, M.L.; Neto, J.B.; Brito, F.V.; Campos, L.T.; Bértoli, C.D.; Campos, G.S.; Cobuci, J.A.; McManus, C.M.;
Barcellos, J.O.J.; Gama, L.T. Origins and genetic diversity of British cattle breeds in Brazil assessed by pedigree analyses1. *J. Anim. Sci.* **2014**, *92*, 1920–1930. [CrossRef]

24. Cavani, L.; Silva, R.M.D.O.; Carreño, L.O.D.; Ono, R.K.; Bertipaglia, T.S.; Farah, M.M.; Millen, D.D.; Fonseca, R.D. Genetic diversity of Brazilian Brahman cattle by pedigree analysis. *Pesqui. Agropecuária Bras.* **2018**, *53*, 74–79. [CrossRef]
25. Gutierrez, J.; Goyache, F. A note on ENDOG: A computer program for analysing pedigree information.
J. Anim. Breed. Genet. **2005**, *122*, 172–176. [CrossRef]
26. Maignel, L.; Boichard, D.; Verrier, E. Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bull.* **1996**, *14*, 49–56.
27. Kadleč ík, O.; Pavlík, I.; Moravč íková, N.; Kasarda, R. Inbreeding and genetic diversity loss of four cattle beef
breeds in Slovakia. *Acta Fytotech. Zootech.* **2016**, *19*, 59–63. [CrossRef]
28. Maccluer, J.W.; Boyce, A.J.; Dyke, B.; Pfenning, D.W.; Parsons, C.J.; Weitkamp, L.R. Inbreeding and pedigree structure in Standardbred horses. *J. Hered.* **1983**, *74*, 394–399. [CrossRef]
29. Meuwissen, T.; Luo, Z. Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* **1992**, *24*,
305–313. [CrossRef]
30. Boichard, D.; Maignel, L.; Verrier, E. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* **1997**, *29*, 5–23. [CrossRef]
31. Abegaz, F.; Chaichoompu, K.; Génin, E.; Fardo, D.W.; König, I.R.; Mahachie John, J.M.; Van Steen, K.
Principals about principal components in statistical genetics. *Brief. Bioinform.* **2018**, bby081. [CrossRef]
32. Hammer, O.; Harper, D.; Ryan, P. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **2001**, *4*, 1–9.
33. Cassell, B.; Adamec, V.; Pearson, R. Effect of Incomplete Pedigrees on Estimates of Inbreeding and Inbreeding Depression for Days to First Service and Summit Milk Yield in Holsteins and Jerseys. *J. Dairy Sci.* **2003**, *86*,2967–2976. [CrossRef]
34. Lutaaya, E.; Misztal, I.; Bertrand, J.; Mabry, J. Inbreeding in populations with incomplete pedigrees. *J. Anim. Breed. Genet.* **2001**, *116*, 475–480. [CrossRef]
35. Cappelloni, M. Sviluppo di Metodi per la Gestione di Razze Autoctone Italiane. Ph.D. Thesis, Università degli Studi di Padova, Padova, Italy, 2013.
36. Gutiérrez, J.P.; Altarriba, J.; Díaz, C.; Quintanilla, R.; Cañón, J.; Piedrafita, J. Pedigree analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.* **2003**, *35*, 43–63. [CrossRef]

37. Cañas-Álvarez, J.J.; Gónzalez-Rodríguez, A.; Martín-Collado, D.; Avilés, C.; Altarriba, J.; Baro, J.A.; De La Fuente, L.F.; Díaz, C.; Molina, A.; Varona, L.; et al. Monitoring changes in the demographic and genealogical structure of the main Spanish local beef breeds1. *J. Anim. Sci.* **2014**, *92*, 4364–4374. [CrossRef]

38. Torrecillas, C.P.; Bozzi, R.; Negrini, R.; Filippini, F.; Giorgetti, A. Genetic variability of three Italian cattle breeds determined by parameters based on probabilities of gene origin. *J. Anim. Breed. Genet.* **2002**, *119*, 274–279. [CrossRef]

39. Falleiro, V.B.; Malhado, C.H.M.; Malhado, A.C.M.; Carneiro, P.L.S.; Carrillo, J.A.; Song, J. Population Structure and Genetic Variability of Angus and Nellore Herds. *J. Agric. Sci.* **2014**, *6*, 276. [CrossRef]

40. Mc Parland, S.; Kearney, J.F.; Rath, M.; Berry, D. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *J. Anim. Sci.* **2007**, *85*, 322–331. [CrossRef] [PubMed]

41. Cañón, J.; Gutiérrez, J.; Dunner, S.; Goyache, F.; Vallejo, M. Herdbook analyses of the Asturiana beef cattle
breeds. *Genet. Sel. Evol.* **1994**, *26*, 65–75. [CrossRef]

42. Herron, N.; Pattie, W. Studies of the Australian Illawarra Shorthorn breed of dairy cattle. II.* Genetic analysis.
Aust. J. Agric. Res. **1977**, *28*, 1119–1132. [CrossRef]

43. Ozkütük, K.; Bichard, M. Studies of pedigree Hereford cattle breeding. 1. Herdbook analyses. *Anim. Sci.*
1977, *24*, 1–13. [CrossRef]

44. Gengler, N.; Misztal, I.; Bertrand, J.K.; Culbertson, M.S. Estimation of the dominance variance for postweaninggain in the U.S. Limousin population. *J. Anim. Sci.* **1998**, *76*, 2515. [CrossRef]

45. Meuwissen, T.H.E. Operation of Conservation Schemes. In *Genebanks and the Management of Farm Animal*
Genetic Resources; DLO Institute for Animal Science and Health: Lelystad, The Netherlands, 1999.

46. Bozzi, R.; Franci, O.; Forabosco, F.; Pugliese, C.; Crovetti, A.; Filippini, F. Genetic variability in three Italian beef cattle breeds derived from pedigree information. *Ital. J. Anim. Sci.* **2010**, *5*, 129–137. [CrossRef]

47. FAO. *Secondary Guidelines for Development of National Farm Animal Genetic Resources Management*
Plans—Management of Small Populations at Risk; FAO: Rome, Italy, 1998.

48. Melka, M.; Stachowicz, K.; Miglior, F.; Schenkel, F. Analyses of genetic diversity in five Canadian dairy breeds using pedigree data. *J. Anim. Breed. Genet.* **2013**, *130*, 476–486. [CrossRef]

49. Howard, J.T.; Pryce, J.E.; Baes, C.; Maltecca, C. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *J. Dairy Sci.* **2017**, *100*, 6009–6024. [CrossRef]

50. Biscarini, F.; Nicolazzi, E.L.; Stella, A.; Boettcher, P.J.; Gandini, G. Challenges and opportunities in genetic improvement of local livestock breeds. *Front. Genet.* **2015**, *6*, 6. [CrossRef] [PubMed]

51. Caballero, A.; Santiago, E.; Toro, M.A. Systems of mating to reduce inbreeding in selected populations.

Anim. Sci. **1996**, *62*, 431–442. [CrossRef]

52. Woolliams, J.; Berg, P.; Dagnachew, B.; Meuwissen, T. Genetic contributions and their optimization. *J. Anim.Breed. Genet.* **2015**, *132*, 89–99. [CrossRef] [PubMed]

Chapter 5

animals

MDPI

Article

Estimation of Linkage Disequilibrium and EffectivePopulation Size in Three Italian Autochthonous Beef Breeds

Maria Chiara Fabbri *, Christos Dadousis and Riccardo Bozzi

Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università di Firenze,50144 Firenze, Italy; christos.dadousis@unifi.it (C.D.); riccardo.bozzi@unifi.it (R.B.)

* Correspondence: mariachiara.fabbri@unifi.it

Received: 15 May 2020; Accepted: 11 June 2020; Published: 14 June 2020

Simple Summary: Linkage disequilibrium (LD) of genomic markers is related to various evolutionary forces, such as inbreeding, nonrandom mating, population bottleneck, drift, recombination, and mutations, and hence is an essential parameter to examine population history. In this analysis, we examined the LD pattern of three Italian local beef breeds (Calvana, Mucca Pisana, and Pontremolese) facing the risk of extinction, using the commercial Limousine beef breed as a control. Our results provide important information on the population history and the current status of the breeds and they can be further used for conservation and breeding purposes.

Abstract: The objective was to investigate the pattern of linkage disequilibrium (LD) in three local beef breeds, namely, Calvana (n = 174), Mucca Pisana (n = 270), and Pontremolese (n = 44). As a control group, samples of the Italian Limousine breed (n = 100) were used. All cattle were genotyped with the GeneSeek GGP-LDv4 33k SNP chip containing 30,111 SNPs. The genotype quality control for each breed was conducted separately, and SNPs with call rate < 0.95 and minor allele frequency (MAF) > 1% were used for the analysis. LD extent was estimated in PLINK v1.9 using the squared correlation between pairs of loci (r^2) across autosomes. Moreover, r^2 values were used to calculate historical and contemporary effective population size (N_e) in each breed.

Average r^2 was similar in Calvana and Mucca Pisana (~0.14) and higher in Pontremolese (0.17); Limousine presented the lowest LD extent (0.07). LD up to 0.11–0.15 was persistent in the local breeds up to 0.75 Mbp, while in Limousine, it showed a more rapid decay. Variation of different LD levels across autosomes was observed in all the breeds. The results demonstrated a rapid decrease in N_e across generations for local breeds, and the contemporary population size observed in the local breeds, ranging from 41.7 in Calvana to 17 in Pontremolese, underlined the demographic alarming situation.

Animals **2020**, *10*, 1034; doi:10.3390/ani10061034 www.mdpi.com/journal/animals

Keywords: linkage disequilibrium; conservation; effective population size; local breeds

---

1. Introduction

In recent years, there has been a greater interest in recovering and preserving local breeds, especially for their adaptation's capacity in marginal areas and for their importance as reservoir of genetic diversity.

In this context, Tuscany (central region of Italy) represents an important pool of genetic diversity with six different cattle breeds native from this area [1]. Three of these six breeds (Calvana (CAL), Mucca Pisana (MUP), and Pontremolese (PON)) have been recognized from the Italian breeders Association (AIA; Associazione Italiana Allevatori, Rome) to be at risk of extinction and at present are enrolled to the register of cattle breeds at limited diffusion (Registro Anagrafico delle razze bovine autoctone a limitata diffusione). From those, PON faces a critical risk of extinction, consisting a limited number of animals (n = 49). Historically, PON originated from the provinces of Massa Carrara, La Spezia, and Parma and, in the past, was used for the transport of Carrara marble, being a robust and rustic breed. CAL (n = 366) originates from the Calvana mountain, in Prato's province. It is particularly suitable to live in marginal areas and has ever been considered as a beef breed. MUP (n = 413) is mainly reared nowadays in the province of Pisa. The breed is a crossbreed, mainly between Schwyz and Chianina breeds [2,3]. Hence, the historical, cultural, and ecological values of these three breeds are enormous and undisputed. Furthermore, the landscape of Tuscany placed barriers on widespread intensive breeding. As a result, the conservation and increase in sample sizes of local breeds could be of economic importance for this region.

Exploring genetic diversity is essential for developing conservation programs in autochthonous breeds, and one of the most commonly used parameters to assess genetic diversity is the effective population size (N_e) [4]. N_e is defined as the size of an ideal population that explains the same rate of random genetic changes as the current population [5,6], and has been traditionally estimated from the pedigree. However, with the advent of genomic technology, N_e can also be inferred from genomic data. The pedigree-based method requires an adequate completeness of data over several generations [7], and for local breeds, this is not an easy task for practical reasons related to breeding management. In a previous study [8], N_e from pedigrees was calculated for the same Tuscan breeds, but generally, the low quality of pedigrees and the alarming demographic situation of these breeds have raised the necessity to estimate the effective population size with genomic data; the linkage disequilibrium (LD) method was chosen for this purpose.

LD of genomic markers is related to various evolutionary forces, such as inbreeding, nonrandom mating, population bottleneck, drift, recombination, and mutation, and hence is an essential parameter to examine population history and genetic diversity. LD is defined as the nonrandom association between alleles at two (or more) loci [9]. In general, it is expected that the strength of LD decreases with an increased distance between markers located on the same chromosome. The term linkage disequilibrium (also known as gametic disequilibrium) goes back to Lewontin and Kojima, in 1960 [9,10]. The interest in studying LD patterns grew together with the research in genes associated with diseases [11,12]. Moreover, LD evaluation is an important prerequisite in genome wide association studies (GWAS), useful to detect the number of markers that will be sufficient for quantitative trait locus mapping [13,14]. Information of LD has been utilized in genomic breeding programs, such as marker-assisted selection and whole genome predictions [15]. The cost reduction and the efficient implementation of genotyping in animal breeding have made LD a common analysis in various species, such as pigs [16,17], poultry [18], sheep [19], goats [20], and cattle [21–23]. The most common LD measures are the squared genetic correlation coefficient (r^2) described by Hill and Robertson [24] and D' reported by Lewontin [10]. LD patterns in livestock populations have been analyzed to investigate (i) the structure and the history of populations [25], (ii) gene mapping [26], and (iii) the effective population size (N_e) with molecular data [7,27,28].

The aim of this study was to investigate the LD and the LD-based N_e patterns of three Italian beef cattle breeds (CAL, MUP, and PON) facing risk of extinction. The commercial

Limousine beef breed (LIM) was included in the analysis to allow for comparisons between local unselected breeds and a cosmopolitan counterpart.

2. Materials and Methods

2.1. Animals and Sampling

A total of 588 beef cattle from four breeds (CAL = 174, MUP = 270, PON = 44, and LIM = 100) were genotyped. The percentage of the sampled cattle relative to the entire pool per breed for the local breeds was 47.5%, 65.4%, and 89.8% for CAL, MUP, and PON, respectively. Regarding the LIM, a sample of 100 cattle was extracted at random from a pool of 533 genotyped cattle, belonging to the last three generations and balanced by sex (52 males and 48 females). Genotypic data from LIM were provided by ANACLI (Associazione nazionale allevatori delle razze bovine Charolaise e Limousine, Roma) [29].

2.2. Genotyping and Quality Control

All cattle were genotyped with GeneSeek GGP-LDv4 33k (Illumina Inc., San Diego, CA, USA). Genotype quality control (QC) and data filtering were performed with PLINK v1.9 [30], and was conducted separately for each breed. Only SNPs located on the 29 autosome chromosomes were included (n = 28,289). SNPs with minor allele frequency (MAF) lower than 1%, and with call rate <0.95 were removed. Further, SNPs with more than 10% missingness values and deviated from Hardy Weinberg Equilibrium ($p < 0.000001$), as well as animals with more than 10% missingness, were also removed. After filtering, 164, 263, 41, and 100 cattle and 23,646, 23,436, 22,791, and 23,279 SNPs remained for CAL, MUP, PON, and LIM, respectively (Table 1).

2.3. Genomic Relationship Matrix

The genomic relationship matrix (GRM) was created per breed to (i) investigate the among breed identical by state relationships and (ii) compare the status of the genotypic samples among the populations under study. The GRM was calculated with the following formula by VanRaden [31]:

$$GRM = \frac{\boldsymbol{ZZ'}}{2\sum p_i(1-p_i)} \quad (1)$$

where $\boldsymbol{Z}$ is a centered matrix of marker genotypes of all individuals and p_i is the frequency of the second allele at locus *i*. $\boldsymbol{Z}$ was calculated from genotypes of reference population

subtracting $2p_i$ from the matrix X that defines the genotypes for each individual as 0, 1 or 2. Heatmap graphs for the GRMof each breed were produced in R software [32].

For each GRM, the following parameters were taken into account: (i) the mean of the diagonal values; (ii) the mean of all the off-diagonal; (iii) the minimum and maximum of the diagonal values; and (iv) the minimum and maximum of the off-diagonal values. For all the above-mentioned parameters, the absolute value and squared root were also estimated.

2.4. Linkage Disequilibrium

Linkage disequilibrium was measured using r^2, which is the squared correlation of the alleles at two loci [24]. The r^2 is considered to be a better measure of LD than *D'* because it is more robust and less sensitive to changes in effective population size and gene frequency [33,34]. The r^2 ranges between 0 and 1 and was calculated as follows:

$$r^2 = \frac{(freq\ (AB) \times\ freq\ (ab) - freq\ (Ab) \times\ freq\ (aB))^2}{(freq\ (A)\ \times\ freq\ (a)\ \times\ freq\ (B)\ \times\ freq\ (b))} \quad (2)$$

where *freq* (*A*), *freq* (*a*), *freq* (*B*), and *freq* (*b*) are the allele frequencies and *freq* (*AB*), *freq* (*ab*), *freq* (*Ab*), and *freq* (*aB*) are the genotype frequencies. The LD extent was calculated for all SNPs pairs of each chromosome using PLINK v1.9 [30] under the command: *–r2 –ld-window 99999 –ld-window-r2 0*, in order to take an interval less than 99,999 SNPs and to save in the output all SNPs pairs.

The LD decay was analyzed in order to compare differences between and within breeds: (i) LDs were binned into four intervals of 0.25 Mbp (0–0.25, 0.25–0.5, 0.5–0.75, and 0.75–1 Mbp), the mean and the standard deviation (SD) of r^2 values were computed for each interval; (ii) LD was also investigated for each autosome, considering windows of 1 Kbp. The LD-decay plots were visualized using the *ggplot2* R package [35].

2.5. Estimation of Historical and Contemporary Effective Population Size

The historical and recent N_e for all breeds were estimated with the SNeP software [36], which is based on the relationships between LD, N_e, and the recombination rate. The default options were used. N_e was analyzed starting 13 generations ago because the default maximum distance in SNeP was 4000 Kbp. High LD in closely linked SNPs reflects ancient population history (50 Kbp ≈ 1500 generations ago), while high LD between distant

SNPs describes more recent history (4000 Kbp ≈ 12.5 generations ago) [37]. The following formula was used to estimate N_e from LD [28]:

$$N_{e(t)} = \frac{1}{(4f(c_t))}\left(\frac{1}{E[r^2_{adj}|c_t|} - \alpha\right) \quad (3)$$

where $N_{e(t)}$ is the effective population size estimated for t generation ago, which is calculated as $t = 1/(2\,f\,(c_t))$ [38]; c_t is the recombination rate at t generations ago, defined for a specific physical distance between markers; r^2_{adj} is the LD estimate adjusted for sample size (164, 263, 41, and 100 for CAL, MUP, PON, and LIM, respectively); and α is a constant and set to 1, as suggested by Ohta and Kimura [39]. The contemporary effective population size (*cNe*) was calculated with NEESTIMATOR v.2 [40] with the mating model set to random; *cNe* means that the results referred to the time period of the sample size included in the analysis.

3. Results

3.1. Quality Control and Genomic Relationship Matrix

Table 1 summarizes the number of SNPs and individuals from each breed, before and after the quality control.

Table 1. Number of autosomal SNPs and individuals before (pre-) and after (post-)quality control (QC) per breed.

Breed [1]	N SNPs Pre-QC	N SNPs Post-QC	N Individuals Pre-QC	N Individuals Post-QC
CAL	28,289	23,646	174	164
MUP	28,289	23,436	270	263
PON	28,289	22,791	44	41
LIM	28,289	23,279	100	100

[1] CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese; LIM = Limousine.

The GRM heatmaps per breed are presented in Figure 1. Individuals were ordered by farm to consider the farm management. Diagonal blocks, indicating highly related individuals, were mainly found for CAL and MUP. Moreover, this grouping was mainly attributed to the farm level. The off-diagonal values among the different farms indicate that there is reduced gene-flow among the farms, with the exception of CAL. In CAL, higher levels of relationship between individuals both between and within farms were observed. The lowest relatedness average between animals was found for LIM.

GRM summary statistics for each breed are reported in Table S1. The mean of the

diagonal value was <1 in all breeds. The average of the diagonal values was 0.99 for CAL, MUP, and LIM, and 0.97 for PON. The highest diagonal values were 1.75, 1.72, 1.54, and 1.22 for MUP, PON, CAL, and LIM, respectively. The minimum diagonal values ranged from 0.67 (MUP) to 0.78 (LIM). The average of off-diagonal was negative for all breeds. The highest off-diagonal maximum was found in MUP (1.09), followed by PON (0.84) and CAL (0.79), while the lowest was found in LIM (0.50).

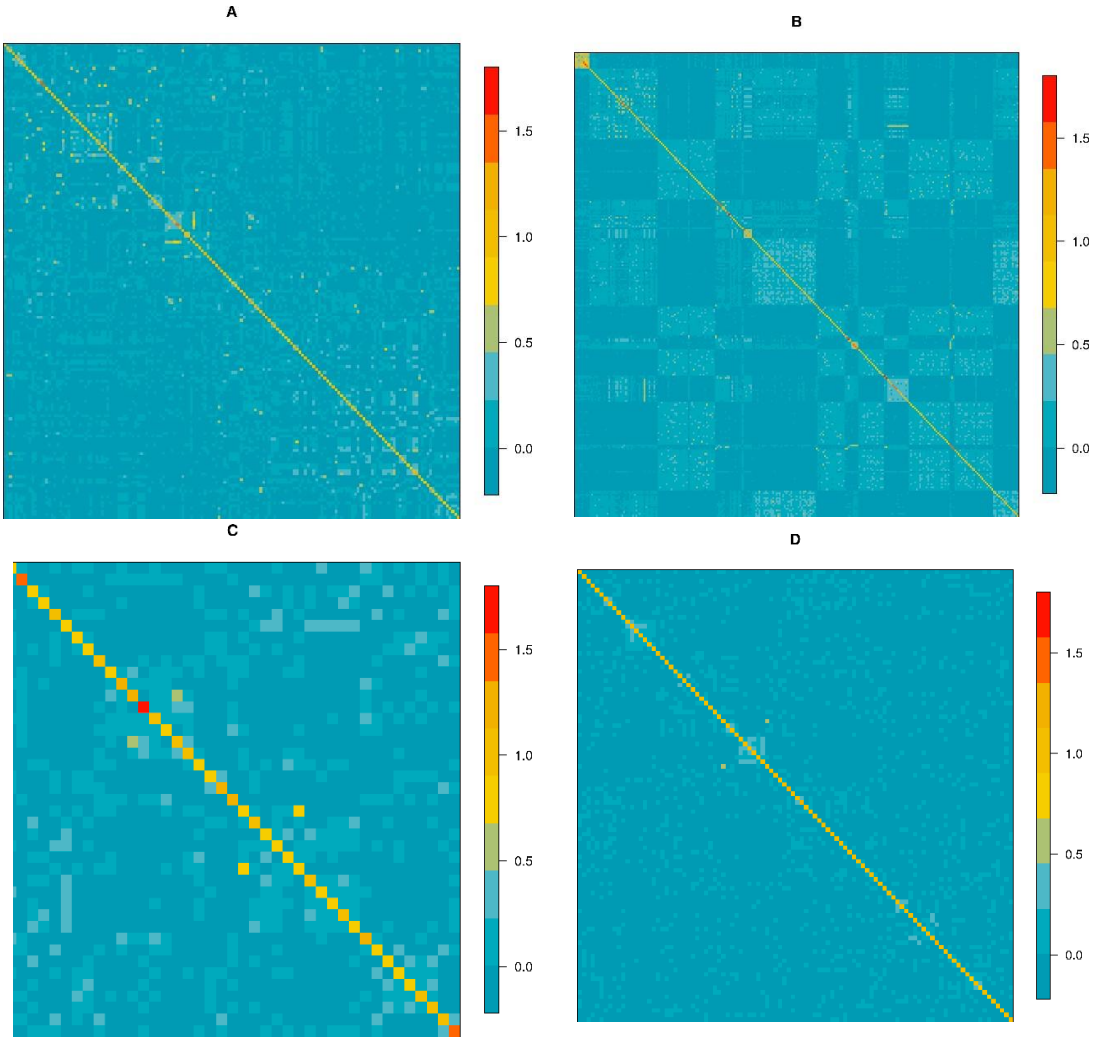

Figure 1. Heatmaps of genomic relationship matrix of (**A**) Calvana, (**B**) Mucca Pisana, (**C**) Pontremolese, and (**D**) Limousine breeds.

3.2. Linkage Disequilibrium

The total autosomal length in the analyzed SNP chip was 2512 Mbp, with the shortest *Bos taurus* autosome (BTA) being the BTA25 (~42.8 Mbp) and the longest being the BTA1 (~158.5 Mbp). The dimensions of each chromosome, the number of SNPs, the mean distance between SNPs, and the longest interval between pairwise SNPs for each chromosome are shown in Table S2. In all breeds analyzed, the highest number of SNPs

was found on BTA1 (ranging from 1328 to 1282, in CAL and PON, respectively), and the smallest was found in BTA27 (from 451 to 435 SNPs in CAL and PON, respectively). The average distances between adjacent SNPs were 0.105, 0.106, and 108 ± 0.08 Mbp in CAL and MUP, LIM, and PON, respectively. The largest distances between two adjacent SNPs were found on BTA12, that is, 4.44, 2.89, 2.87, and 2.09 Mbp in LIM, CAL, MUP, and PON, respectively.

The average of the highest r^2 values was found in PON (0.17), and the lowest in LIM (0.07), with CAL and MUP (0.14) at intermediate values. The mean and SD of r^2 per chromosome and breed are summarized in Table 2. For CAL, the highest values were found in BTA20 (0.21), followed by BTA6 (0.18). BTA21 (0.19), BTA4, and BTA9 (0.17) were the three autosomes that showed the highest extent of LD in MUP. For PON, BTA16 had the highest r^2 (0.23), followed by BTA21 (0.22), BTA20, and BTA1 (0.20). BTA20 and BTA16 were also the chromosomes with higher LD extent for LIM, albeit at a much lower degree (0.13 and 0.12, respectively). The lowest LD was found in BTA22 (for PON and LIM) and in BTA27 (for CAL, MUP, and LIM). For LIM, the lowest LD values (0.05) were found in more BTA apart from BTA22 and BTA27 (Table 2).

Table 2. The average and standard deviation (SD) of linkage disequilibrium (r^2) for *Bos taurus* autosomes(BTAs) per breed.

Breed [1]	**CAL**		**MUP**		**PON**		**LIM**	
Autosome	**Average r^2**	**SD**	**Average r^2**	**SD**	**Average r^2**	**SD**	**Average r^2**	**SD**
BTA1	0.14	0.20	0.13	0.20	0.20	0.25	0.08	0.16
BTA2	0.16	0.23	0.13	0.19	0.17	0.22	0.10	0.18
BTA3	0.15	0.21	0.12	0.17	0.16	0.22	0.08	0.15
BTA4	0.14	0.19	0.17	0.26	0.20	0.26	0.10	0.22
BTA5	0.15	0.20	0.15	0.20	0.18	0.23	0.07	0.12
BTA6	0.18	0.26	0.14	0.20	0.19	0.23	0.08	0.14
BTA7	0.12	0.16	0.13	0.17	0.17	0.21	0.07	0.12
BTA8	0.14	0.19	0.11	0.16	0.16	0.21	0.06	0.11
BTA9	0.15	0.22	0.17	0.23	0.15	0.21	0.08	0.17
BTA10	0.13	0.18	0.13	0.17	0.15	0.18	0.06	0.10
BTA11	0.12	0.18	0.12	0.16	0.15	0.18	0.06	0.11
BTA12	0.12	0.17	0.14	0.19	0.17	0.21	0.06	0.10
BTA13	0.15	0.21	0.12	0.17	0.16	0.20	0.05	0.09
BTA14	0.13	0.18	0.13	0.18	0.19	0.23	0.06	0.11
BTA15	0.14	0.18	0.15	0.21	0.20	0.24	0.09	0.17
BTA16	0.15	0.22	0.15	0.22	0.23	0.28	0.12	0.20
BTA17	0.13	0.18	0.15	0.20	0.15	0.18	0.06	0.10
BTA18	0.13	0.17	0.15	0.20	0.17	0.20	0.06	0.12

BTA19	0.15	0.20	0.14	0.19	0.16	0.21	0.07	0.13
BTA20	0.21	0.27	0.15	0.21	0.20	0.25	0.13	0.21
BTA21	0.14	0.21	0.19	0.24	0.22	0.25	0.10	0.18
BTA22	0.12	0.16	0.11	0.16	0.14	0.17	0.05	0.08
BTA23	0.14	0.20	0.14	0.19	0.15	0.19	0.08	0.15
BTA24	0.14	0.19	0.12	0.16	0.17	0.22	0.06	0.11
BTA25	0.13	0.17	0.12	0.16	0.15	0.19	0.06	0.10
BTA26	0.13	0.17	0.13	0.18	0.18	0.22	0.05	0.09
BTA27	0.11	0.15	0.10	0.14	0.15	0.18	0.05	0.09
BTA28	0.12	0.16	0.13	0.16	0.15	0.19	0.06	0.10
BTA29	0.15	0.22	0.12	0.17	0.20	0.25	0.05	0.09

[1] CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese, LIM = Limousine.

The distribution of r^2 across BTA is shown in Figure 2. The highest variation was observed for PON. On the contrary, LIM had both the lowest variation and the lowest r^2 values. Median values were similar across BTA within breed except for PON, where more fluctuations were present. The reduced area of the first quartiles indicated that the observations had similar values and were close to 0. The third quartiles were wider than the first in all breeds. Moreover, the whiskers, representing maximum (and minimum) values, suggested the high differentiation of LD extent within the local breed and between the local and commercial breeds.

In order to analyze the LD decay within a specific distance between SNPs, independently of differences within chromosomes, LD was investigated considering intervals of 0.25 Mbp up to 1 Mbp (Figure 3). Differences among breeds were observed, especially for LIM, which was clearly separated from the local breeds with a steeper decay with an increasing distance.

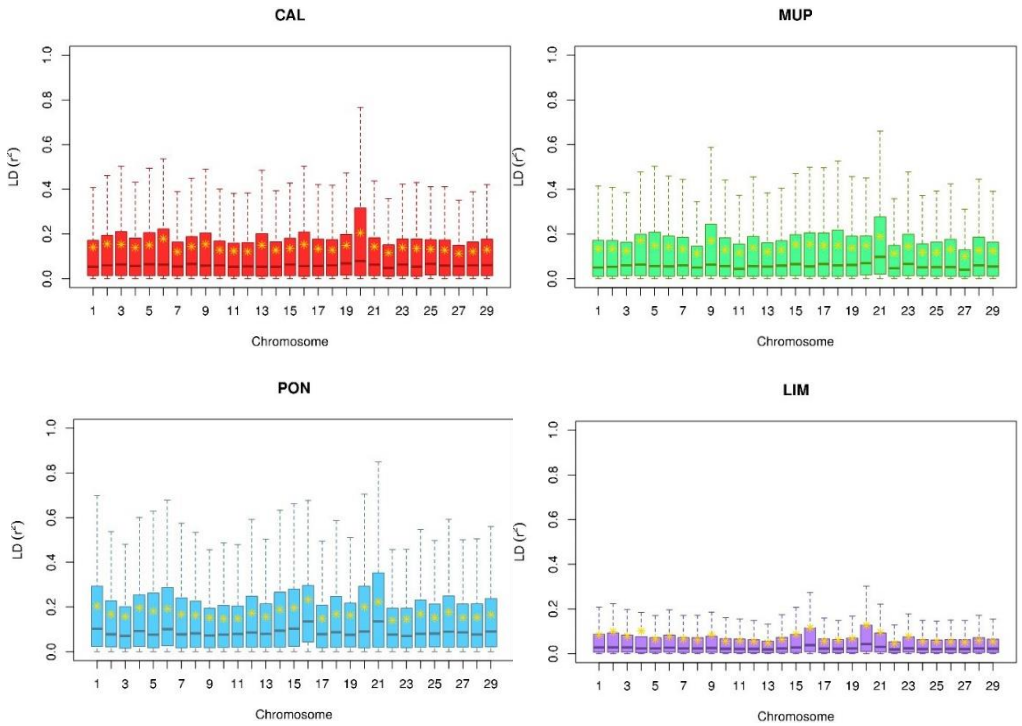

Figure 2. Boxplots of linkage disequilibrium (LD) (r^2) of *Bos taurus* autosomes (BTAs) per breed. The yellow asterisks represent the mean and horizontal lines within each boxplot are the medians (outliers have been removed). CAL = Calvana; MUP = Mucca Pisana; PON = Pontremolese, LIM = Limousine.

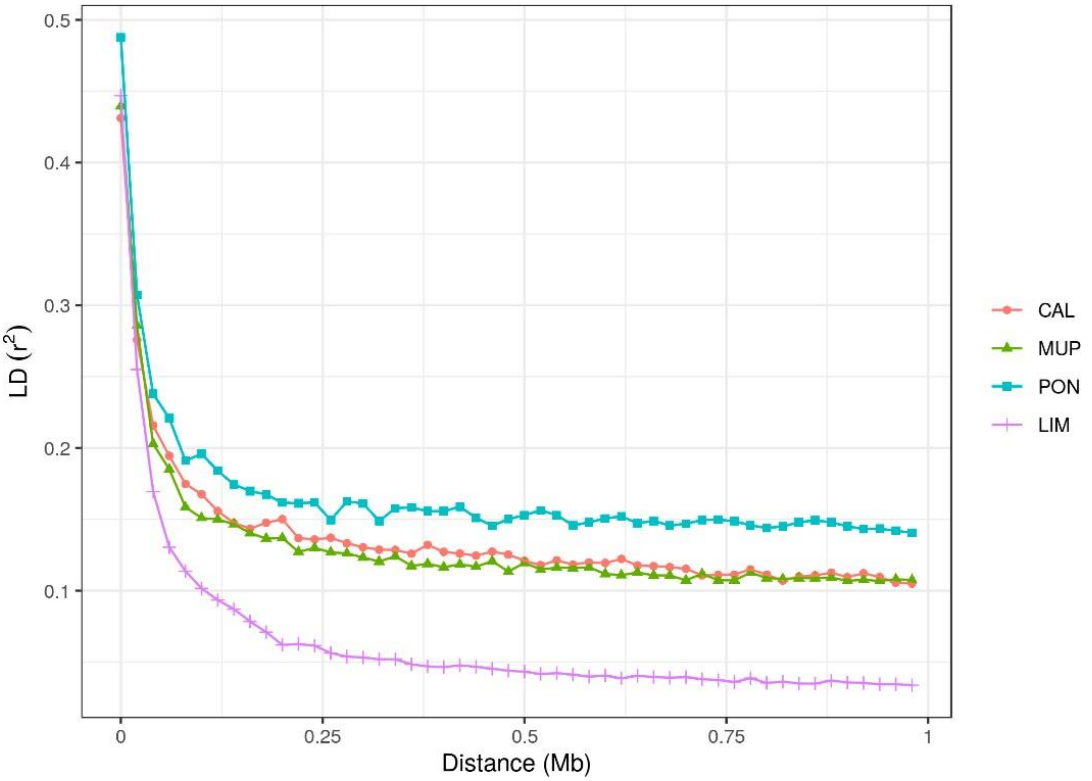

Figure 3. Linkage disequilibrium (LD), calculated as r^2, in different distances of the genome and up to 1 Mbp for Calvana (CAL), Mucca Pisana (MUP), Pontremolese (PON), and Limousine (LIM).

PON and LIM presented the two opposite extremes (top and bottom on Figure 3, respectively) while CAL and MUP had similar and intermediate trends. For PON, an r^2 close to 0.15 was maintained even at 1 Mbp distance, while for CAL and MUP, it was maintained at ~0.1. For LIM, r^2 < 0.1 was found for distances >0.12 Mbp (Figure 3). In general, for the local breeds, a sharp decay was observed till ~0.12 Mbp, while for LIM, the r^2 was stabilized after ~0.25 Mbp. Average r^2 for each bin was reported in Table S3. Values of r^2 > 0.2 in the first bin were found only for PON, while from 0.25 to 1 Mbp, values remained close to 0.15. MUP and CAL had similar trends, with r^2 ~0.19 0.25 in the first interval, whereas from 0.25 to 0.5 Mbp, CAL showed higher mean r^2 value than MUP;

while with the longest distance (>0.75 and up to 1 Mbp), r^2 was close to 0.10 for both breeds. LIM showed the more rapid LD decay, with average r^2 ranging from 0.141 (<0.25 Mbp) to 0.035 (>0.75 Mbp).

The r^2 was also calculated per BTA in windows of 1 Kbp. The LD decay was found to be different among chromosomes, but with a similar trend in the local breeds (Figure S1). CAL and PON presented a higher variation of r^2 within the analyzed interval of 1 Kbp; indeed, the decay did not linearly decrease, while LIM showed a faster decay.

3.3. Estimation of Historical and Contemporary Effective Population Size

On the basis of LD estimates, the trend of N_e over time (per generation) for each breed was investigated (Figure 4). Similar to LD, the N_e pattern of the local breeds was different from that of LIM. The decrease in N_e was clear in all breeds, with a sharper decay for LIM. N_e decreased from~275 (80th generation ago) to 79 and 65 (13th generation ago) in CAL and MUP, respectively, while for PON, N_e estimates decreased from 204 to 45 (in the most distant and recent generation, respectively). The historical trend of LIM was very different than those for the local breeds, decreasing from 920 to 310 (80th to 13th generations ago).

Regarding the contemporary N_e (*cNe*), CAL, MUP, PON, and LIM showed values of 41.7, 18.7, 17.0, and 327.9, respectively.

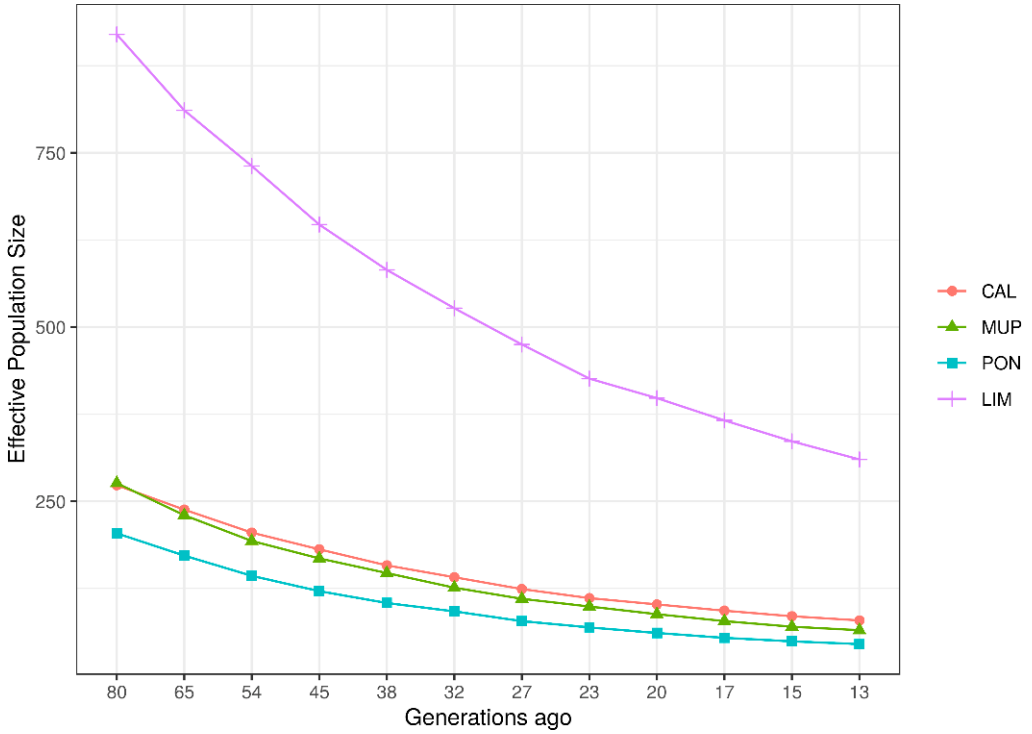

Figure 4. Average estimated historical effective population size (*Ne*) in the Italian breeds: Calvana (CAL), Mucca Pisana (MUP), Pontremolese (PON), and Limousine (LIM).

4. Discussion

In this study, LD was investigated to assess the genomic architecture and the evolutionary history of autochthonous populations. LD provides various information that is used in several applications. For example, LD between linked markers determines the power and the precision of association mapping studies, because it influences the ability to localize genes, loci affecting economic traits, and diseases. Indeed, understanding the extent of LD improves the planning and the performance of genomic breeding programs [41]. However, direct comparison between the studies should be done with caution, owing to some factors that influence LD estimates, that is, sample size, population history and structure, LD measure (r^2 or D'), marker type (microsatellites or SNPs), marker filtering, density, and distribution [34]. In the present study, a genome-wide LD extent and N_e parameters were calculated for three Italian local beef breeds (CAL, MUP, and PON). The results were contrasted to the cosmopolitan Limousine, a commercial beef breed undergoing selection. This is the first study that performed an in-depth analysis of LD extent in these three Italian local breeds. In the last years, great importance has been given to genetic diversity and has underlined local breeds as important genetic resources as they harbor unique gene pools as a result of adaptation to the local environment [42]. In the absence of high-quality pedigree information, a GRM was constructed for each breed to define the degree of relatedness within a sampled breed. This information was used to investigate potential differences among the breeds that might be responsible for differences in LD and N_e patterns. It is known that, in a closed inbred population, the recombination decreases and the LD increases [43]. The heatmaps showed that this situation has been avoided; farm groups were present, especially in CAL, but the relatedness within them was not found to be worrisome. Differences in LD and N_e estimates of LIM compared with local breeds could be a result of different levels of relationship among the animals sampled, but, most of all, differences could be caused by the greater sample size and the breeding program used for LIM, which is absent for local breeds.

Regarding LD extent, our results revealed an r^2 variation among BTA within the breed. This could be partly attributed to different lengths of the chromosomes [22]. However, BTA1, which was the longest chromosome, presented a high LD level only in PON. Interestingly, BTA16, BTA20, and BTA21 were characterized by high r^2 values shared in more than one breed: BTA16 showed high r^2 values in PON and LIM; BTA20 in CAL, PON, and LIM; and

BTA21 in PON and MUP. These autosomes should be investigated in further works because the markers with high r^2 values could reveal potential Quantitative Trait Loci (QTLs).

Average LD was different between local breeds and LIM. LIM is one of the most reared beef breeds in Italy, and was analyzed in this study to compare local breeds LD patterns with that belonging to a breed undergoing directional selection. In general, in commercial breeds, artificial selection and insemination allow the control of matings, and thus the inbreeding, which is further linked to LD, as inbreeding augments the covariance between alleles at different loci [9]. Furthermore, if the breed had an expansion in population size or consists of a large population, genetic drift is weaker and, as a consequence, LD decreases, converging to an equilibrium [43]. The higher LD observed in the local breeds is likely related to a higher ancestral relatedness and to a historically smaller N_e [44], as shown in Figure 4. Moreover, the LD decay of the local breeds had the characteristic trend of populations that have suffered a collapse in population size and/or a bottleneck, as described by Rogers [43]. A previous study on Italian local cattle breeds carried out by Mastrangelo et al. [45] confirmed the slow LD decay found in this study. Mastrangelo et al. [44] analyzed two local breeds reared in Sicily, an island in the South of Italy. The r^2 values were 0.16 for Cinisara and 0.20 for Modicana. No other studies on the level of LD in Italian local cattle were found in the literature. Hence, comparisons were done with foreign breeds. Mustafa et al. [46] investigated the Sahiwal dairy breed, which is under threat of extinction, showing an average r^2 value equal to 0.18, which is similar to our Tuscan local breeds (ranged from

0.14 to 0.17). Nevertheless, differences were found for the LD decay, which was significantly more rapid than in Tuscan populations. Tunisian local breeds, consisting of big populations, studied by Jemaa et al. [22], showed more similar LD patterns to LIM than the Tuscan breeds, with r^2 values lower than 0.05 in larger distances (>0.5 Mbp). Another study that corroborated the hypothesis that the decay of Tuscan breeds was slower than other local breeds was performed by Makina et al. [47], who studied four local African cattle compared with Angus and Holstein breeds. In all six breeds, at 1 Mbp, r^2 values were lower than 0.1 and in the four local breeds (Afrikaner, Nguni, Drakensberger, and Bonsmara), and r^2 did not exceed the value of 0.05 (in our study, all local populations had $r^2 > 0.1$ at 1 Mbp distance).

More studies on LD on commercial beef cattle breeds can be found in the literature. For

instance, in Chinese Simmental breed [48] in the window of 0.5–1 Mbp, the r^2 was 0.05. This is similar to our estimates in LIM (0.03) and significantly lower than our findings in the local breeds (0.11 for CAL and MUP and 0.15 for PON). Biegelmeyer et al. [49] investigated the LD patterns in Hereford and Braford, and found average r^2 values equal to 0.07 and 0.06, respectively, for the same window of 1 Mbp. Both breeds showed a great level of LD in the short term, suggesting a faster decay than in Tuscan breeds and more similar to LIM decay.

Other studies always based on selected breeds reported smaller average r^2 by chromosomes for Limousine breed [15,50] than that found in the present research, likely owing to the different history of populations and to a more ancient breeding selection system compared with Italian Limousine.

Sample size is another factor that influences the estimate of LD. Khatkar et al. [51] declared that, if r^2 parameter is used to calculate LD, a minimum sample size of 75 animals is required both for an accurate estimation and to avoid bias. LIM, as well as CAL and MUP, consisting of larger populations than PON, were sampled considering the threshold of the aforementioned study. In our analysis, only PON did not reach this threshold. However, this was because, at present, only 49 PON alive animals are officially enrolled to the register of cattle at limited diffusion. Hence, our samples referred to ~81% of the entire PON population. Regarding the two other local breeds, the genotypic data analyzed represented 47.5% and 65.4% of Calvana and Mucca Pisana, respectively. Overall, our study considered a much higher number of genotyped animals per breed compared with previous studies on autochthonous cattle breeds [7,16,20,22,23,45,52].

In fact, sample size could be the cause of the different results found in the present research when compared with previous studies. For example, Kukučková et al. [37] investigated the trend of historical N_e of 15 European cattle breeds, Limousine included, and reported lower values (N_e was equal to ~300 in the 60th generation ago) than the Italian LIM, but the sample size of that study was limited to 20 animals. Mastrangelo et al. [3] investigated the trends of N_e of the same Tuscan breeds, among a plethora of cattle breeds, and related different trends of historical effective population size. The recent N_e (13th generation) was less than 100, as in our study, but the trends for PON and MUP contrast with this latter with the lowest values and no overlapping with Calvana as instead related in the present study. Regarding LIM, both studies showed different trends between local and

cosmopolitan breeds, but the LIM historical N_e was higher in our study (80th generation N_e was 920, whereas it was <500 in Mastrangelo et al. [3]). Obviously, the contemporary N_e (*cNe*) also presented some differences, even maintaining the same ranking: *cNe* was greater in the present study for the Tuscan breeds, namely, 41.7, 18.7, and 17 (for CAL, MUP, and PON, respectively) and 33.5, 8.7, and 7.2, respectively, for the same breeds in the cited paper. Only LIM had lower *cNe* value (327.9 instead of 468.9), probably because of the sampling strategy (only animals of the last three generations were taken). However, CAL, and definitely MUP and PON, were at risk of loss of genetic diversity, presenting N_e lower than 50 animals for generation, which is the threshold suggested by Food and Agriculture Organization (FAO) [53].

5. Conclusions

Our results, utilizing a medium-density SNP chip, demonstrated differences in LD and *Ne* patterns in four beef breeds, three locals under extinction (Calvana, Mucca Pisana, and Pontremolese) and one cosmopolitan (Limousine), reared in Italy. The greater genetic diversity loss, as dictated by the N_e estimates, was found for Mucca Pisana. This work complements previous analysis carried out on pedigree information, both describing the population structure of the three local breeds. It is necessary to carefully monitor these populations and to strengthen the use of accurate mating plans in order to both increase the census size and control relatedness and inbreeding. LD results could be utilized in association studies as well as in the development of low-density SNP chip panels, for example, for parentage testing. Future studies could focus on the genomic regions with high r^2 values, as well as investigate runs of homozygosity in the three local breeds.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2615/10/6/1034/s1, Figure S1: Linkage Disequilibrium decay described for each chromosome considering intervals of 1 Kbp; Table S1: Genomic relationship matrix (GRM) summary statistics for each breed; Table S2: Summary of analyzed SNPs for each breed, split by autosomal chromosomes; Table S3: Average and standard deviation (SD) of linkage disequilibrium (r^2), estimated in four intervals of 0.25 Mbp, per breed.

Author Contributions: R.B. and C.D. conceived the idea, formulated the objectives of this study, and supervised the project. M.C.F. conducted the analysis and wrote the first draft

of the paper. C.D. helped in data preparation and in supervising the project. All authors read and approved the final manuscript.

Funding: This research was funded by ANACLI, grant number 2015.99.2264.1127.

Acknowledgments: The research was funded by ANACLI through the I-BEEF project PSRN 2014–2020. Sottomisura 10.2: Biodiversità animale. We acknowledge Associazione Nazionale Allevatori delle razze bovine Charolaise e Limousine (ANACLI) for providing the data.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Bozzi, R.; Alvarez, I.; Crovetti, A.; Fernández, I.; De Petris, D.; Goyache, F. Assessing priorities for conservation in Tuscan cattle breeds using microsatellites. *Anim. Int. J. Anim. Biosci.* **2012**, *6*, 203–211. [CrossRef]
2. Negrini, R.; Milanesi, E.; Bozzi, R.; Pellecchia, M.; Ajmone-Marsan, P. Tuscany autochthonous cattle breeds: An original genetic resource investigated by AFLP markers. *J. Anim. Breed. Genet.* **2006**, *123*, 10–16. [CrossRef]
3. Mastrangelo, S.; Ciani, E.; Ajmone Marsan, P.; Bagnato, A.; Battaglini, L.; Bozzi, R.; Carta, A.; Catillo, G.; Cassandro, M.; Casu, S.; et al. Conservation status and historical relatedness of Italian cattle breeds. *Genet. Sel. Evol.* **2018**, *50*, 35. [CrossRef] [PubMed]
4. Leroy, G.; Mary-Huard, T.; Verrier, E.; Danvy, S.; Charvolin, E.; Danchin-Burge, C. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genet. Sel. Evol.* **2013**, *45*, 1. [CrossRef]
5. Palstra, F.P.; Ruzzante, D.E. Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol. Ecol.* **2008**, *17*, 3428–3447. [CrossRef] [PubMed]
6. Crow, J.F.; Kimura, M. An introduction to population genetics theory. *Introd. Popul. Genet. Theory* **1970**, *2*, 174–208.
7. Flury, C.; Tapio, M.; Sonstegard, T.; Drögemüller, C.; Leeb, T.; Simianer, H.; Hanotte, O.; Rieder, S. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *J. Anim. Breed. Genet.* **2010**, *127*, 339–347. [CrossRef]
8. Fabbri, M.C.; Gonçalves de Rezende, M.P.; Dadousis, C.; Biffani, S.; Negrini, R.; Souza Carneiro, P.L.; Bozzi, R. Population Structure and Genetic Diversity of Italian Beef Breeds as a Tool for Planning Conservation and Selection Strategies. *Animals* **2019**, *9*, 880. [CrossRef] [PubMed]
9. Slatkin, M. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* **2008**, *9*, 477–485. [CrossRef]
10. Lewontin, R.C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **1964**, *49*, 49–67.
11. Park, L. Population-specific long-range linkage disequilibrium in the human genome and its influence on identifying common disease variants. *Sci. Rep.* **2019**, *9*, 1–13. [CrossRef] [PubMed]
12. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **1999**, *22*, 139–144. [CrossRef] [PubMed]
13. Lee, S.H.; Cho, Y.M.; Lim, D.; Kim, H.C.; Choi, B.H.; Park, H.S.; Kim, O.H.; Kim, S.; Kim, T.H.; Yoon, D.; et al. Linkage Disequilibrium and Effective Population Size in Hanwoo Korean Cattle. *Asian-Australas. J. Anim. Sci.* **2011**, *24*, 1660–1665. [CrossRef]
14. Meuwissen, T.H.; Goddard, M.E. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* **2000**, *155*, 421–430. [PubMed]
15. Lu, D.; Sargolzaei, M.; Kelly, M.; Li, C.; Vander Voort, G.; Wang, Z.; Plastow, G.; Moore, S.; Miller, S.P. Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Front. Genet.* **2012**, *3*. [CrossRef] [PubMed]
16. Muñoz, M.; Bozzi, R.; García-Casco, J.; Núñez, Y.; Ribani, A.; Franci, O.; García, F.; Škrlep, M.; Schiavo, G.; Bovo, S.; et al. Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. *Sci. Rep.* **2019**, *9*, 1–14. [CrossRef]

17. Du, F.-X.; Clutter, A.C.; Lohuis, M.M. Characterizing Linkage Disequilibrium in Pig Populations. *Int. J. Biol. Sci.* **2007**, 166–178. [CrossRef] [PubMed]
18. Seo, D.; Lee, D.H.; Choi, N.; Sudrajad, P.; Lee, S.-H.; Lee, J.-H. Estimation of linkage disequilibrium and analysis of genetic diversity in Korean chicken lines. *PLoS ONE* **2018**, *13*, e0192063. [CrossRef]
19. Ghoreishifar, S.M.; Moradi-Shahrbabak, H.; Parna, N.; Davoudi, P.; Khansefid, M. Linkage disequilibrium and within-breed genetic diversity in Iranian Zandi sheep. *Arch. Tierz.* **2019**, *62*, 143–151. [CrossRef]
20. Monau, P.I.; Visser, C.; Muchadeyi, F.C.; Okpeku, M.; Nsoso, S.J.; Van Marle-Köster, E. Population structure of indigenous southern African goats based on the Illumina Goat50K SNP panel. *Trop. Anim. Health Prod.* **2020**. [CrossRef]
21. Edea, Z.; Dadi, H.; Dessie, T.; Lee, S.-H.; Kim, K.-S. Genome-wide linkage disequilibrium analysis of indigenous cattle breeds of Ethiopia and Korea using different SNP genotyping BeadChips. *Genes Genom.* **2015**, *37*, 759–765. [CrossRef]
22. Jemaa, S.B.; Thamri, N.; Mnara, S.; Rebours, E.; Rocha, D.; Boussaha, M.; Jemaa, S.B.; Thamri, N.; Mnara, S.; Rebours, E.; et al. Linkage disequilibrium and past effective population size in native Tunisian cattle. *Genet. Mol. Biol.* **2019**, *42*, 52–61. [CrossRef] [PubMed]
23. Cañas-Álvarez, J.J.; Mouresan, E.F.; Varona, L.; Díaz, C.; Molina, A.; Baro, J.A.; Altarriba, J.; Carabaño, M.J.; Casellas, J.; Piedrafita, J. Linkage disequilibrium, persistence of phase, and effective population size in Spanish local beef cattle breeds assessed through a high-density single nucleotide polymorphism chip. *J. Anim. Sci.* **2016**, *94*, 2779–2788. [CrossRef] [PubMed]
24. Hill, W.G.; Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **1968**, *38*, 226–231. [CrossRef]
25. Gautier, M.; Faraut, T.; Moazami-Goudarzi, K.; Navratil, V.; Foglio, M.; Grohs, C.; Boland, A.; Garnier, J.-G.; Boichard, D.; Lathrop, G.M.; et al. Genetic and Haplotypic Structure in 14 European and African Cattle Breeds. *Genetics* **2007**, *177*, 1059–1070. [CrossRef]
26. Arias, J.A.; Keehan, M.; Fisher, P.; Coppieters, W.; Spelman, R. A high density linkage map of the bovine genome. *BMC Genet.* **2009**, *10*, 18. [CrossRef] [PubMed]
27. Wang, J. Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. B Biol. Sci.* **2005**, *360*, 1395–1409. [CrossRef]
28. Corbin, L.J.; Liu, A.Y.H.; Bishop, S.C.; Woolliams, J.A. Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet.* **2012**, *129*, 257–270. [CrossRef]
29. Statistics of the Italian Limousine and Charolaise Breeds, ANACLI Site. Available online: http://www.anacli.it/WEBSITE/index.php?&pagid=2499&sessione= (accessed on 12 February 2020).
30. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
31. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
32. The R Project for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 31 January 2020).

33. Zhao, H.; Nettleton, D.; Dekkers, J.C.M. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genet. Res.* **2007**, *89*, 1–6. [CrossRef]
34. Jasielczuk, I.; Gurgul, A.; Szmatoła, T.; Semik-Gurgul, E.; Pawlina-Tyszko, K.; Szyndler-Nędza, M.; Blicharski, T.; Szulc, K.; Skrzypczak, E.; Bugno-Poniewierska, M. Comparison of linkage disequilibrium, effective population size and haplotype blocks in Polish Landrace and Polish native pig populations. *Livest. Sci.* **2020**, *231*, 103887. [CrossRef]
35. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.
36. Barbato, M.; Orozco-terWengel, P.; Tapio, M.; Bruford, M.W. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* **2015**, *6*. [CrossRef]
37. Kukučková, V.; Moravčíková, N.; Ferenčaković, M.; Simčič, M.; Mészáros, G.; Sölkner, J.; Trakovická, A.; Kadlečík, O.; Curik, I.; Kasarda, R. Genomic characterization of Pinzgau cattle: Genetic conservation and breeding perspectives. *Conserv. Genet.* **2017**, *18*, 893–910. [CrossRef]
38. Hayes, B.J. Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. *Genome Res.* **2003**, *13*, 635–643. [CrossRef]
39. Ohta, T.; Kimura, M. Linkage Disequilibrium between Two Segregating Nucleotide Sites under the Steady Flux of Mutations in a Finite Population. *Genetics* **1971**, *68*, 571–580.
40. Do, C.; Waples, R.S.; Peel, D.; Macbeth, G.M.; Tillett, B.J.; Ovenden, J.R. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol. Ecol. Resour.* **2014**, *14*, 209–214. [CrossRef]
41. Qanbari, S. On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. *Front. Genet.* **2020**, *10*. [CrossRef] [PubMed]
42. Upadhyay, M.; Eriksson, S.; Mikko, S.; Strandberg, E.; Stålhammar, H.; Groenen, M.A.M.; Crooijmans, R.P.M.A.; Andersson, G.; Johansson, A.M. Genomic relatedness and diversity of Swedish native cattle breeds. *Genet. Sel. Evol.* **2019**, *51*, 56. [CrossRef] [PubMed]
43. Rogers, A.R. How Population Growth Affects Linkage Disequilibrium. *Genetics* **2014**, *197*, 1329–1341. [CrossRef]
44. Pérez O'Brien, A.M.; Mészáros, G.; Utsunomiya, Y.T.; Sonstegard, T.S.; Garcia, J.F.; Van Tassell, C.P.; Carvalheiro, R.; da Silva, M.V.B.; Sölkner, J. Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livest. Sci.* **2014**, *166*, 121–132. [CrossRef]
45. Mastrangelo, S.; Saura, M.; Tolone, M.; Salces-Ortiz, J.; Di Gerlando, R.; Bertolini, F.; Fontanesi, L.; Sardina, M.T.; Serrano, M.; Portolano, B. The genome-wide structure of two economically important indigenous Sicilian cattle breeds1. *J. Anim. Sci.* **2014**, *92*, 4833–4842. [CrossRef] [PubMed]
46. Mustafa, H.; Ahmad, N.; Heather, H.J.; Eui-soo, K.; Khan, W.A.; Ajmal, A.; Javed, K.; Pasha, T.N.; Ali, A.; Kim, J.J.; et al. Whole genome study of linkage disequilibrium in Sahiwal cattle. *S. Afr. J. Anim. Sci.* **2018**, *48*, 353–360. [CrossRef]
47. Makina, S.O.; Taylor, J.F.; van Marle-Köster, E.; Muchadeyi, F.C.; Makgahlela, M.L.; MacNeil, M.D.; Maiwashe, A. Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. *Front. Genet.* **2015**, *6*. [CrossRef] [PubMed]
48. Zhu, M.; Zhu, B.; Wang, Y.H.; Wu, Y.; Xu, L.; Guo, L.P.; Yuan, Z.R.; Zhang, L.P.; Gao, X.; Gao, H.J.; et al. Linkage Disequilibrium Estimation of Chinese Beef Simmental Cattle Using High-density SNP Panels. *Asian-Australas. J. Anim. Sci.* **2013**, *26*, 772–779. [CrossRef] [PubMed]

49. Biegelmeyer, P.; Gulias-Gomes, C.C.; Caetano, A.R.; Steibel, J.P.; Cardoso, F.F. Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. *BMC Genet.* **2016**, *17*, 32. [CrossRef]
50. Rodríguez-Ramilo, S.T.; García-Cortés, L.A.; de Cara, M.Á.R. Artificial selection with traditional or genomic relationships: Consequences in coancestry and genetic diversity. *Front. Genet.* **2015**, *6*. [CrossRef]
51. Khatkar, M.S.; Nicholas, F.W.; Collins, A.R.; Zenger, K.R.; Cavanagh, J.A.L.; Barris, W.; Schnabel, R.D.; Taylor, J.F.; Raadsma, H.W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genom.* **2008**, *9*, 187. [CrossRef]
52. Sharma, A.; Lim, D.; Chai, H.-H.; Choi, B.-H.; Cho, Y. Demographic Trends in Korean Native Cattle Explained Using Bovine SNP50 Beadchip. *Genom. Inform.* **2016**, *14*, 230–233. [CrossRef]
53. Woolliams, J.A.; Gwaze, D.P.; Meuwissen, T.H.E.; Planchenault, D.; Renard, J.-P.; Thibier, M.; Wagner, H. *Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans–Measurement of Domestic Animal Diversity (MoDAD)*; FAO: Rome, Italy, 1998; p. 58.

Chapter 6

PLOS ONE

RESEARCH ARTICLE

Genetic diversity and population history of eight Italian beef cattle breeds using measures of autozygosity

Maria Chiara Fabbri [1*], Christos Dadousis [2], Francesco Tiezzi [3], Christian Maltecca [3], Emmanuel Lozada-Soto [3], Stefano Biffani [4], Riccardo Bozzi [1]

[1] Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università di Firenze, Firenze, Italy.

[2] Dipartimento di Scienze Medico-Veterinarie, Università di Parma, Parma, Italy.

[3] Department of Animal Science, North Carolina State University, Raleigh, NC 27695, United States.

[4] Institute of Agricultural Biology and Biotechnology (CNR), Milano, Italy

* Corresponding author: mariachiara.fabbri@unifi.it

Received: February 17, 2021, Accepted: October 6, 2021, Published: October 25, 2021

PLOS ONE | https://doi.org/10.1371/journal.pone.0248087 October 25, 2021

Abstract

In the present study, GeneSeek GGP-LDv4 33k single nucleotide polymorphism chip was used to detect runs of homozygosity (ROH) in eight Italian beef cattle breeds, six breeds with distribution limited to Tuscany (Calvana, Mucca Pisana, Pontremolese) or Sardinia (Sarda, Sardo Bruna and Sardo Modicana) and two cosmopolitan breeds (Charolaise and Limousine). ROH detection analyses were used to estimate autozygosity and inbreeding and to identify

genomic regions with high frequency of ROH, which might reflect selection signatures. Comparative analysis among breeds revealed differences in length and distribution of ROH and inbreeding levels. The Charolaise, Limousine, Sarda, and Sardo Bruna breeds were found to have a high frequency of short ROH (~ 15.000); Calvana and Mucca Pisana presented also runs longer than 16 Mbp. The highest level of average genomic inbreeding was observed in Tuscan breeds, around 0.3, while Sardinian and cosmopolitan breeds showed values around 0.2. The population structure and genetic distances were analyzed through principal component and multidimensional scaling analyses, and resulted in a clear separation among the breeds, with clusters related to productive purposes. The frequency of ROH occurrence revealed eight breed-specific genomic regions where genes of potential selective and conservative interest are located (e.g. MYOG, CHI3L1, CHIT1 (BTA16), TIMELESS, APOF, OR10P1, OR6C4, OR2AP1, OR6C2, OR6C68, CACNG2 (BTA5), COL5A2 and COL3A1 (BTA2)). In all breeds, we found the largest proportion of homozygous by descent segments to be those that represent inbreeding events that occurred around 32 generations ago, with Tuscan breeds also having a significant proportion of segments relating to more recent inbreeding.

Introduction

Runs of homozygosity (ROH) consist of contiguous regions of the genome where an individual is homozygous in all sites [1]. This occurs when the haplotypes transmitted from both parents are identical due to being inherited from a common ancestor. The length of a ROH is an imprint of the history of a population linked to its effective population size and provides evidence for phenomena such as inbreeding, mating system, and population bottlenecks. In theory, longer ROH are due to recent inbreeding, as recombination has not had the possibility of breaking up the homozygous segment, on the other hand, short ROH demonstrate an older origin because several meiosis have been occurred [2]. Information on inbreeding is crucial in the design of breeding and conservation programs to control the increase in inbreeding levels and to avoid the unfavorable effect of inbreeding depression in progeny [3].

The inbreeding coefficient of an individual (F) is defined as the probability that two randomly chosen alleles at a specific locus within an individual are identical by descent (IBD) [4]. Homozygosity caused by two IBD genomic segments is defined “autozygosity”, F is therefore an estimate of genome-wide autozygosity [5] and ROH are highly likely to be autozygous [6].

The estimation of the inbreeding coefficient from the proportion of the genome covered by ROH (F_{ROH}) has been considered a powerful and accurate method of detecting inbreeding effects [5] and a valid alternative to pedigree inbreeding coefficient [7,8], which doesn't take into account the stochastic nature of recombination. Pedigree information could be incomplete and/or incorrect especially for local breeds, where the extensive breeding system and the natural mating system could allow a limited control of relatedness.

The high occurrence of ROH in chromosomes could potentially represent a selection signature, i.e. a genomic footprint that could provide an overview for understanding the mechanism of selection and adaptation, and could help to uncover regions related to important physiological, economical and adaptive traits [9]. A selection signature is characterized by a reduced haplotype variability, defined as ROH island [10]. Two different methods have been applied to detect ROH islands: the first one based on an arbitrarily defined frequency of common ROH within the population (for e.g., 20 % [11]; 45% [12]; 70% [13]), while the second approach on a percentile threshold (99th percentile) based on the top 1% of SNPs observed in a ROH [14,15].

However, the use of ROH as markers for the identification of genomic regions potentially subjected to non-recent evolutionary events is not straightforward. It requires that homozygous segments have been inherited from old ancestors and were not caused by recent demographic events [16]. A further approach to estimate global inbreeding (F_G) for each population, which links the genomic homozygous segments to the time of living of the most recent common ancestor, is the Homozygous-Identical by Descent (HBD) state probabilities. Druet and Gautier [17] presented an approach to investigate local and global inbreeding, based on the hidden Markov model (HMM). This approach assumes that the genome is formed by HBD and non-HBD segments, where each segment has a HBD state probability. Solé et al. [18] implemented a new HMM with multiple age based HBD-classes in which the length of HBD segments have distinct distributions: longer segments for more recent common ancestors, and shorter for more ancient ancestors. The expected HBD segment lengths are inversely related to the number of generations to the common ancestor and their frequency to past effective population size and individual inbreeding coefficients [17].

Assessment of genetic diversity and population structure is an important task to understand the evolutionary history of the breeds, but also to provide important information for the conservation and management of biodiversity [19]. Italy has a biodiversity reservoir for local

breeds, but generally, local populations have a small sample size and one of the most important obstacle is the increase in inbreeding, leading to negative effect on production and reproduction traits [20]. The maintenance of genetic diversity should be the priority for countries such as Italy, where local breeds guarantee the economical survivor of marginal areas. Selection programs are not easy to apply to local populations for the reduced sample size which also implies a higher level of inbreeding than in selected breeds [21]. For the former populations, it is even more necessary to organize conservation programs aimed at maintaining genetic diversity and controlling inbreeding. Within this context, improving the knowledge about the genomic background of local breeds is crucial.

The aim of this study was to assess genome-wide autozygosity in eight Italian beef breeds, six at critical risk of extinction, namely Calvana (CAL), Mucca Pisana (MUP), Pontremolese (PON), mainly reared in Tuscany, Sardo Bruna (SAB), Sardo Modicana (SAM), Sarda (SAR) reared in Sardinia. The cosmopolitan breeds, i.e. Charolaise (CHA) and Limousine (LIM), were included in the analysis to compare results and to highlight the differences between local breeds and two of the most widespread breeds reared in Italy. ROH distribution and characterization have been investigated across the genome, and consequently, the inbreeding coefficients (F_{ROH}) within breeds were calculated; the HDB state probabilities have been used to estimate global inbreeding (F_G) and to investigate its change across generations, in order to describe the demographic history of the populations.

Materials and methods

Ethics statement

DNA sampling for all the eight breeds was conducted using nasal swabs and no invasive procedures were applied. Thus, in accordance to the 2010/63/EU guide and the adoption of the Law D.L. 04/03/2014, n.26 by the Italian Government, an ethical approval was not required for our study.

Animal sampling, quality control and SNPs characterization

A total of 1,308 animals, belonging to eight breeds, have been genotyped with GeneSeek GGP-LDv4 33k (Illumina Inc., San Diego, California, USA) single nucleotide polymorphism (SNP) DNA chip. Sampled animals for the three Tuscan breeds were 179, 190 and 45 for CAL, MUP and PON, respectively. The limited number of alive animals of these breeds restricted

the total samples. Also, for SAM, being at risk of extinction, only 101 genotypes have been recovered. Samples of SAR (n=199) and SAB (n=194) were animals born from 2005 and 2000, respectively. CHA and LIM samples consisted of cattle born from 2015 to 2019 (200 samples for each breed). Genotype quality control and data filtering were performed with PLINK v1.9 [22] and were conducted separately for each breed: only SNPs located on the 29 autosomes were included (n = 28,289). Linkage Disequilibrium (LD) pruning was not performed as suggested by Dixit et al. [23], as LD is related to various evolutionary forces which are the phenomena ROH analysis investigates (e.g. inbreeding, nonrandom mating, population bottleneck, artificial and natural selection). Editing for SNP MAF was not applied to the dataset because it does not improve ROH detection, on the contrary homozygous regions could be ignored [24]. A SNP characterization was performed based on MAF categories. SNPs were classified into 5 classes: monomorphic SNPs, SNPs with MAF ranged from 0 to 0.005, from 0.005 to 0.01, from 0.01 to 0.05 and >0.05. The aim was to evaluate the number of monomorphic SNPs within and between breeds, given that numerous common monomorphic SNPs could influence ROH investigation.

Multidimensional scaling plot analysis

A multidimensional scaling plot analysis (MDS) was performed to investigate the population structure between the eight breeds based on genetic distances. The first three dimensions were obtained with PLINK v1.9 [22] using the *--mds-plot* flag, which were estimated on the matrix of genome wide pairwise Identical by State (IBS) distances [25]. Results were plotted using Scatterplot3d R package [26]

Runs of homozygosity detection

Analysis of runs of homozygosity was conducted with the R package *detectRUNS* v. 0.9.5 [27]. The following parameters were applied in order to detect a ROH: i) the minimum number of consecutive SNPs was set to 15; ii) the minimum ROH length required was 1 Mbp; iii) the maximum gap between consecutive homozygous SNPs was 1 Mbp; iv) the maximum number of opposite genotypes in the run was set to 2; v) the maximum number of missing genotypes allowed was 2. The consecutive method was preferred than the sliding windows one in order to avoid the detection of artificial ROH shorter than the window described above (15 SNPs) [28].

A principal component analysis (PCA) was conducted on the number of ROH per chromosome for each breed, to infer the similarities between populations based on ROH

chromosomic distribution. All ROH were classified into five classes of length as suggested by Kirin et al. [2], and Ferenčaković et al. [29]: 0-2, 2-4, 4-8, 8-16, >16 Mbp. For each of the eight breeds the total number of ROH, the ROH average number per individuals, the average length of ROH, the number of ROH per breed per chromosome, and the number of ROH per class of length were estimated.

Genomic inbreeding based on ROH

The genomic inbreeding (F_{ROH}) was calculated as suggested by McQuillan et al. [30]:

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{genome}}$$

Where $\sum L_{ROH}$ was the sum of the length of all ROH found in an individual and L_{genome} was the total autosomal genome length. The F_{ROH} per class of ROH length was calculated.

Selection signatures and Gene enrichment

In order to investigate the selection signatures in the eight cattle breeds, the occurrences of ROH across genome was explored. The SNPs frequencies (%) in detected ROH were evaluated for each breed and plotted against the position of the SNP across autosomes. The threshold considered was the 80% of ROH occurrence for each breed, which were filtered taking only the genomic regions containing a minimum number of 15 SNPs. These genomic regions were analyzed and overlapped to Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/) of NCBI (National Center for Biotechnology information) to identify genes. The UMD 3.1 assembly was used for mapping.

Homozygosity by descent (HBD) segments and global inbreeding

The hidden Markov model (HMM)-based approach was used to scan the individual genome for the HBD segments as described in Solé et al. [18]. The analysis was computed with the *RZooROH* R package [31]. The HBD state probability values for each marker were averaged across individual in each population. Averaging HBD probabilities of all loci across the genome led to global (genome-wide) inbreeding (F_G) calculation. Each class (K) has its own rate parameter, R_K, which indicates the length of the segments for its respective class. The length of HBD classes is exponentially distributed with rate R_K, which is double the number of generations to the common ancestor of the respective class. The length of the

HBD segment is $1/R_K$ Morgans, indicating high rates associated with shorter segments. The study focused on <16 Mbp ROH length, so the model applied was six HBD classes with respective rates ($R_K = 2^1, 2^2, 2^3, \ldots, 2^6$) and one non-HBD with an R_K rate of 2^6, so that 32 generations (generation = R/2) and short HBD segments from 1.5 Mbp ($1/2^6$) of length were included in the analysis. The rate of the non-HBD class was fixed as the most ancient class.
A MixKR [18] model with K = 7 was performed. To estimate the inbreeding coefficient, we considered the ancestors with an R_K rate higher than a threshold T as unrelated. The corresponding genomic inbreeding coefficient (F_{G-T}) was then estimated, with $R_K \leq T$ averaged over the whole genome (as reported by Druet et al. [32]).

Results

Animal sampling, quality control and SNPs characterization

In total, 28,178 SNPs were divided into 5 classes of MAF (Table 1) while 111 SNPs remained unclassified. PON presented the highest number of monomorphic SNPs (n=4,151; i.e. the 15.6% of the total number of SNPs), followed by LIM and CHA, namely 3,915 (14.4%) and 3,456 (12.8 %) respectively. CAL had an intermediate value (2,968 - 11.40%) while MUP, SAB, SAM and SAR had less than 2,000 monomorphic SNPs, which maintained lower than the 8% of the total amount of them. The first two classes of MAF (0-0.005 and 0.005-0.01) contained few SNPs, exceeding 1,000 markers only in MUP.

Table 1. Number of autosomal SNPs per breed[1] classified into 5 classes of minor allele frequency. Sample size for each breed was reported.

	CAL (n = 179)	CHA (n = 200)	LIM (n = 200)	MUP (n = 190)	PON (n = 45)	SAB (n = 194)	SAM (n = 101)	SAR (n = 199)
Monomorphic	2,968	3,456	3,915	1,999	4,151	1,903	1,663	1,447
0-0.005	589	544	234	912	0	588	313	396
0.005-0.01	323	123	109	477	0	290	242	253
0.01-0.05	1,359	714	759	1,653	1,628	893	1,162	1,308
>0.05	23,032	23,45	23,271	23,23	22,473	24,609	24,887	24,878

[1] CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

In order to investigate the presence of common SNPs between breeds, pairs comparisons have been performed. Fig 1 explains the total number of monomorphic SNPs on the diagonal, and the off-diagonal represents the common SNPs deriving from pairs comparisons among breeds. PON had the highest number of monomorphic SNPs (n = 4,151), while they amounted to 3,456 for CHA and 3,915 for LIM. MUP and SAB presented numbers close to 2,000 while SAM and SAR had the lowest values of monomorphic SNPs (1,663 and 1,447, respectively). As expected, the two breeds under selection shared the greatest number of monomorphic SNPs and were followed by LIM vs. PON and PON vs. CAL. However, no monomorphic SNP was found in common in all eight breeds.

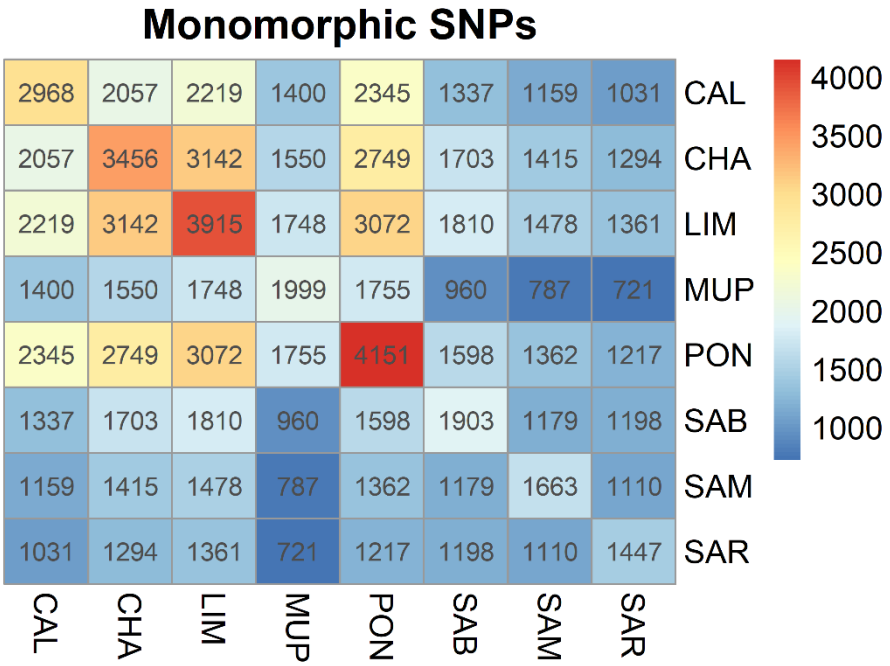

Fig 1. Heatmap of monomorphic SNPs pairs comparison among breeds; CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

The monomorphic SNPs distribution across autosomes has been investigated and it is showed in Fig 2. As expected, the amounts of monomorphic SNPs decreased relative to chromosomic length. However, the highest total number was reported in BTA5, except for SAR and SAM (BTA6); the lowest number of monomorphic SNPs was found in BTA25 for Tuscan breeds and LIM, in BTA26 for Sardinian breeds and lastly, in BTA28 for CHA.

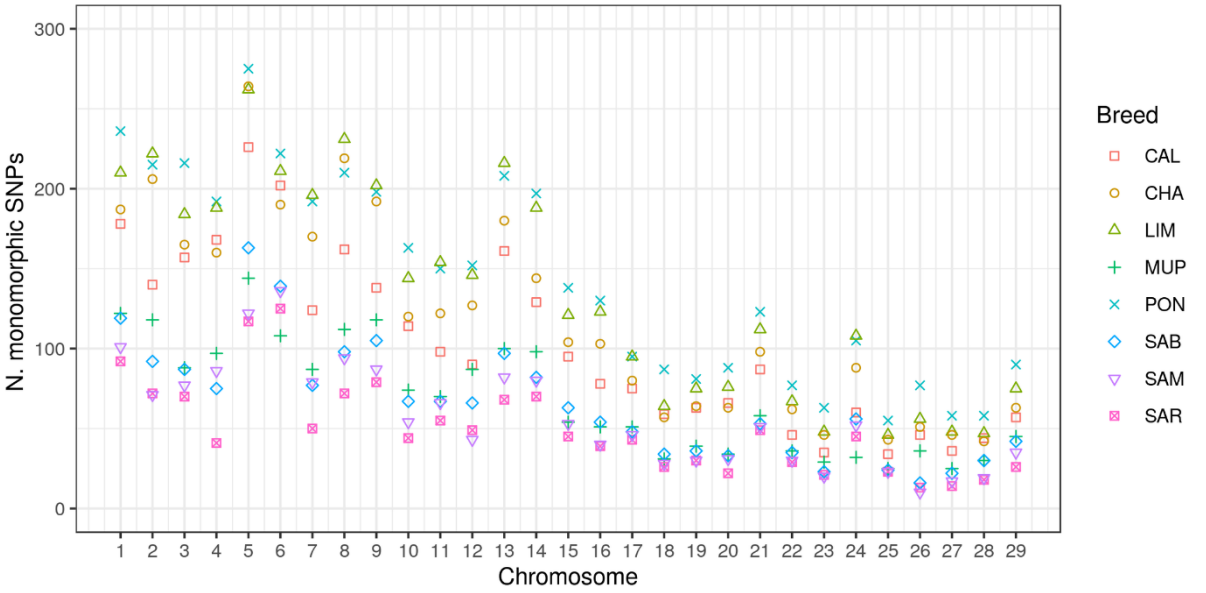

Fig 2. Monomorphic Single Nucleotide Polymorphisms distribution across chromosomes in each breed[1], where [1] CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

Multidimensional scaling plot analysis

The MDS plot evidenced clustering of breeds (Figure 3).

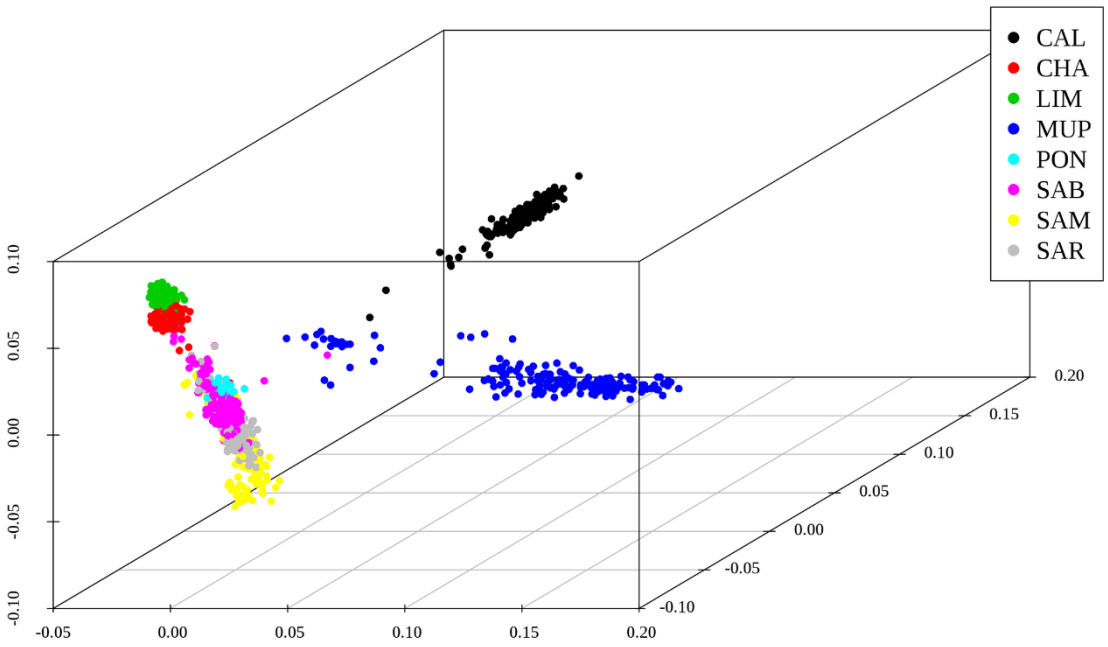

Fig 3. Multidimensional scaling plot for 8 cattle beef breeds, where CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

MUP and CAL were extremely distant from each other and from the other six breeds, which grouped in a unique large cluster. Only a small group of MUP samples were nearer to the third cluster (Sardinian and Cosmopolitan breeds). Both LIM and CHA showed extremely compact clusters, suggesting as expected, a low genetic variability within each breed, and also close to each other, underlining their different genetic background compared to local breeds. SAM and SAB individuals were more scattered than SAR and PON which, however, overlapped with these former.

Runs of Homozygosity detection

Table 2 shows the total number of ROH detected per breed, the average number per individual and the ROH total length.

Table 2. Descriptive statistics of ROH for the eight breeds.

Breed[1]	N. animals	N. total[2]	Average N. per indiv[3]	Average Length (Mbp)[4]
CAL	179	42,873	244	3.11
CHA	200	49,812	254	2.14
LIM	200	53,01	266	2.09
MUP	190	44,748	224	3.68
PON	45	10,044	234	3.06
SAB	194	47,776	246	2.21
SAM	101	22,758	225	2.20
SAR	199	48,339	243	2.24

[1] CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda. [2] N. = the sum of ROH events per breed. [3] Average N. indiv = the average number of ROH per individual; [4] Average Length = the average length of ROH across individuals for each breed.

The two cosmopolitan breeds displayed the highest number of identified ROH (LIM=53,010; CHA= 49,812), whereas SAM and PON showed the lowest values, 22,758 and 10,044, respectively. However, the average number per individuals was almost the same in each breed, ranging from 224 in SAB to 266 in SAR. For each chromosome the number of ROH per breed (S1 Table) has been calculated. *Bos taurus* autosome (BTA) 1 had consistently the highest ROH number in all breeds, except for MUP, where BTA2 had the highest number of ROH. The average length was found to be higher in Tuscan breeds (ranging from 3.06 to 3.68 Mbp). The Sardinian breeds had intermediate values (2.20 – 2.24 Mbp) while the average ROH length for the cosmopolitan breeds was 2.14 Mbp for CHA and 2.09 Mbp for LIM. PCA analysis on the number of ROH per chromosome revealed groups among breeds (Fig 4). PC1 clearly separated Tuscan breeds from Sardinian and cosmopolitan, while PC2 placed SAR, SAB, CHA and LIM close to 0, while MUP and CAL located to the opposite site to PON. The first two PCs explained together the 95.26% of the total variation among samples.

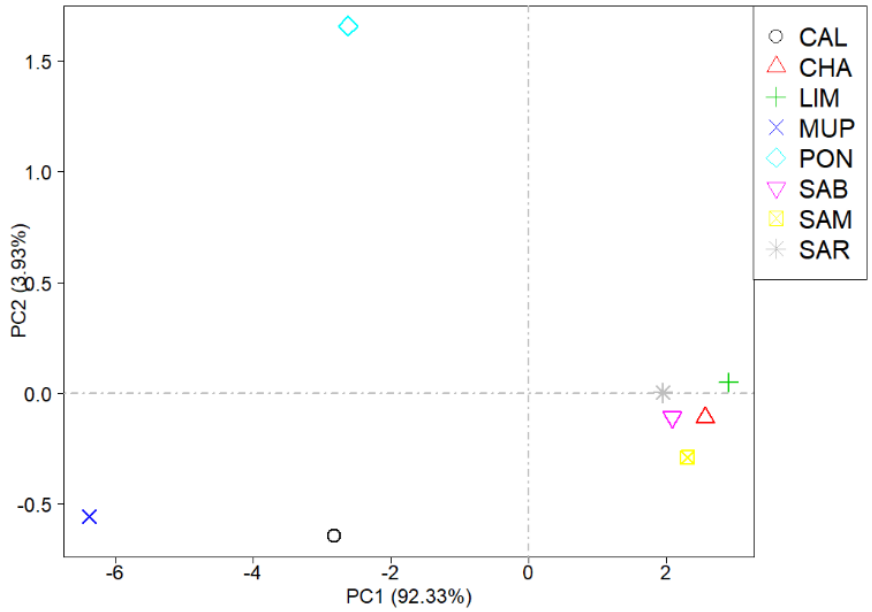

Fig 4 Scatterplot of the first two principal components (PCs). Principal component analysis (PCA) was performed on the number of identified ROH per chromosome in each breed. CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

Five classes of length were considered in order to investigate the ROH pattern. The ROH distribution by length and number for each breed were reported in Fig 5.

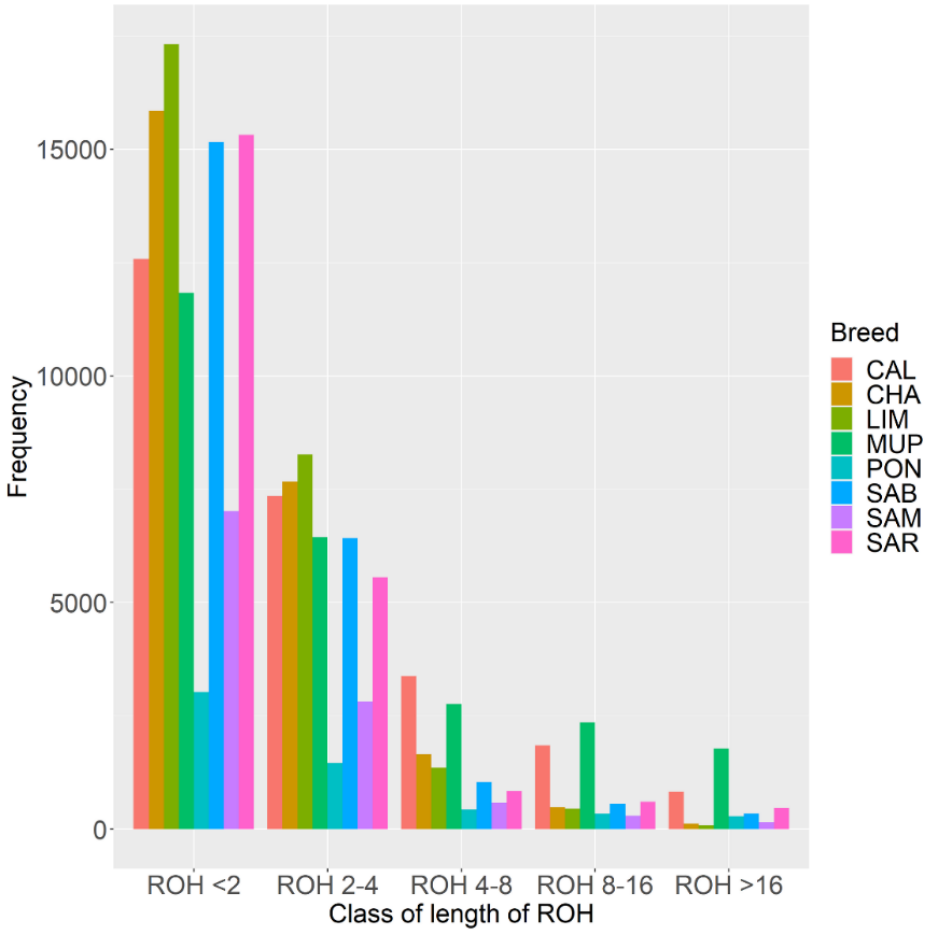

Fig 5. ROH classified into 5 classes of length. CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

The majority of ROH detected belonged to the first two classes (<2 and 2-4 Mbp) for all breeds. LIM and SAR had ~34,000 ROH with length less than 2 Mbp, followed by CHA and SAB with ~32,000. PON and SAM were the two breeds with a lower number of short ROH detected (6,200 and 15,602, respectively). The second class of ROH length (2-4 Mbp) maintained similar pattern, with LIM, CHA, SAB and SAR having the higher number of ROH. Regarding the classes of longer length (>4Mbp), CAL and MUP had the highest number of ROH, even in the class of >16 Mbp (1,943 for MUP and 1,018 for CAL). The other six breeds in this last class showed 510, 392, 306, 169, 146 and 108 long ROH, for SAR, SAB, PON, SAM, CHA and LIM, respectively.

Genomic inbreeding (F_{ROH})

Descriptive statistics for F_{ROH} were reported in Fig 6.

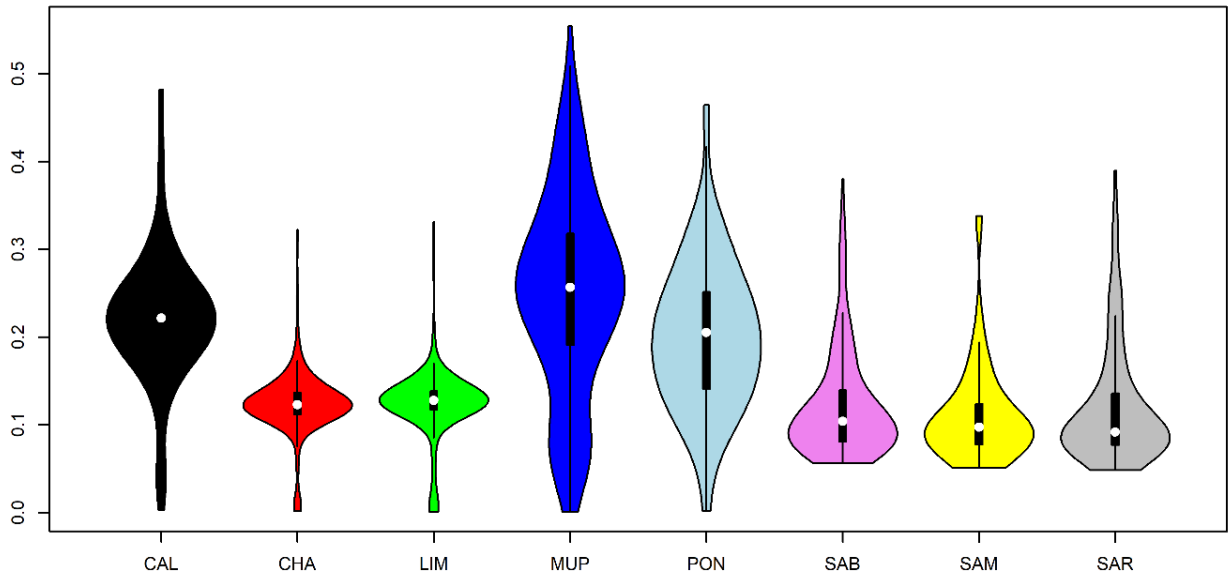

Fig 6. Descriptive statistics of FROH distribution for the eight breeds. CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

Tuscan breeds presented the highest level of average genomic inbreeding coefficients (0.33, 0.30, 0.28 for MUP, CAL and PON, respectively) with maximum values that exceeded 0.5. F_{ROH} for LIM, CHA, SAB and SAR was identical (~0.22), while the lowest average inbreeding coefficient was found in SAM (0.20). Tuscan and cosmopolitan breeds presented minimum values close to 0, instead Sardinian breeds had a minimum F_{ROH} near to 0.14.

To investigate the recent and past inbreeding in each breed, F_{ROH} was calculated for the previous classes of length (Table 3).

Table 3. Mean and standard deviation (SD) of F_{ROH} per class of length for each breed.

	<2 Mbp		2 – 4 Mbp		4 – 8 Mbp		8 – 16 Mbp		>16 Mbp	
Breed[1]	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CAL	0.30	0.07	0.22	0.08	0.15	0.07	0.11	0.07	0.06	0.06
CHA	0.22	0.05	0.12	0.03	0.04	0.03	0.02	0.02	0.02	0.02

LIM	0.22	0.06	0.12	0.04	0.04	0.03	0.02	0.02	0.02	0.03
MUP	0.33	0.11	0.25	0.12	0.21	0.11	0.18	0.10	0.13	0.08
PON	0.29	0.08	0.20	0.08	0.14	0.09	0.12	0.08	0.09	0.07
SAB	0.22	0.06	0.12	0.06	0.05	0.06	0.05	0.06	0.05	0.05
SAM	0.20	0.05	0.10	0.06	0.05	0.06	0.04	0.05	0.04	0.05
SAR	0.22	0.06	0.12	0.07	0.05	0.07	0.06	0.07	0.06	0.06

[1] CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda

F_{ROH} in Tuscan breeds was higher for all the classes analyzed when compared to Sardinian and cosmopolitan breeds. The mean for the first class corresponded to the general average inbreeding for each breed, but values decreased with an increasing length. Sardinian breeds maintained values near to 0.05 from 4-8 Mbp up to >16 Mbp classes while in the latter length class, CHA and LIM inbreeding coefficients were close to 0. MUP showed the highest F_{ROH} in all the 5 categories, having an average genomic inbreeding coefficient of 0.13 in long ROH.

Selection signatures and Gene enrichment

To identify genomic regions potentially important for selection and/or conservation, the SNPs' frequency contained in the runs were plotted across autosomes for each breed (Fig 7). As presented in Fig 7, the autosomes generally more interested by ROH with high occurrence was BTA21, except for CAL, which were BTA13 and BTA6. LIM and CHA presented ROH peaks also on BTA2 and BTA5, respectively.

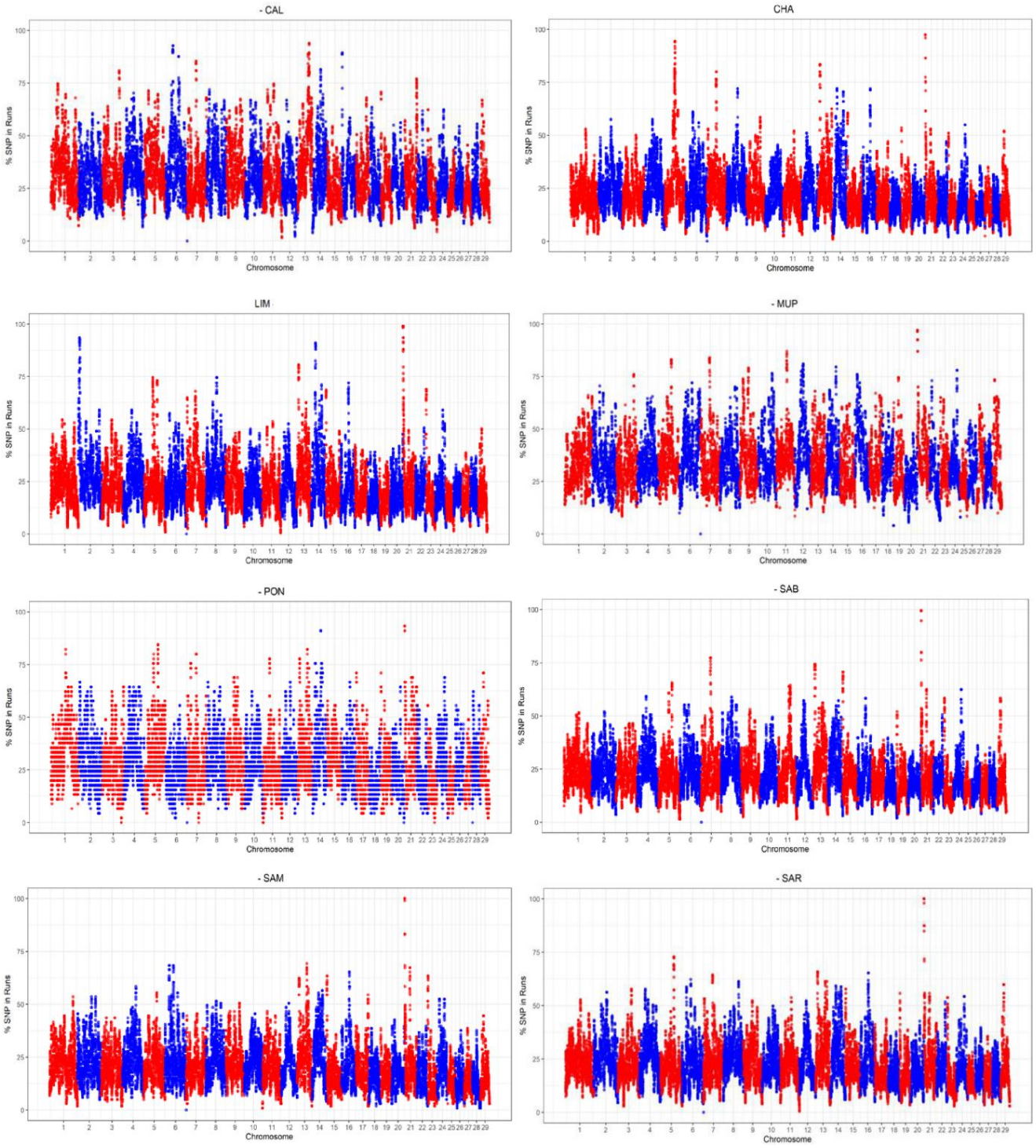

Fig 7. Manhattan plots of the distribution of ROH in the eight cattle breeds. The x-axis is the SNP position and the y-axis shows the frequency (%) at which each SNP was observed in ROH across individuals (CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda).

Applying the abovementioned threshold of 80%, 35 ROH have been detected in total across all breeds (S2 Table); the highest number of genomic regions identified was found in Tuscan breeds (n=11, CAL; n= 9, MUP; n= 6, PON). Five ROH with high occurrence were found both for CHA and LIM, while each Sardinian breed presented only one run. The longest runs were found in LIM (2.65 and 2.38 Mbp), followed by MUP (2.17 Mbp), SAM and SAR (2 Mbp).

One genomic region on BTA21 was found in common in several breeds starting at 83,766 Mbp with BovineHD2100000012. The region was almost identical for CHA, PON, SAB with a length of ~1.70 Mbp; this run contained 18 SNPs for CHA and PON, 17 for SAB. This ROH was present also in LIM, MUP, SAM, and SAR starting at the same SNP but finishing with different markers (BovineHD2100000320 for LIM located at 2,467,774 bp, BovineHD2100000283 for MUP at 2,256,102 bp and BovineHD2100000258 for SAM and SAR positioned at 2,085,345 bp). For the aforementioned breeds, the number of SNPs within this region ranged from 25 (LIM) to 20 (Sardinian). Within this shared run, four genes were detected: *IGHM* (Immunoglobulin heavy constant Mu), *MKRN3* (Makorin finger protein 3), *MAGEL2* (MAGE family member L2) and *NDN* (Necdin), located upstream of the Prader-Willi syndrome (PWS) region.

From the 35 ROH detected, within-breed specific regions containing a minimum number of 15 SNPs were selected to investigate the genes (Table 4). The list of genes in each run is reported in S3 Table.

Table 4. Characterization of within-breed common runs of homozygosity at a threshold of 80% and with at least 15 SNPs and the number of genes included.

Breed[1]	CHR	Start_SNP	End_SNP	Length_Mbp	N. genes
CAL	6	BovineHD0600010715	Hapmap26233-BTA-75846	1.49	2
	16	BovineHD1600000011	BovineHD1600000286	1.06	11
CHA	5	BovineHD0500016070	BovineHD0500016088	0.09	3
	5	BovineHD0500016090	BovineHD0500016469	1.74	57
LIM	2	ARS-BFGL-NGS-21306	BTB-01111412	2.65	22
	14	BovineHD1400007190	Hapmap40958-BTA-34312	1.70	7
MUP	5	5_74951342	5_75130860	0.18	1
PON	5	BovineHD0500021258	5_75130860	0.37	6

[1] CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese.

Among genes presented in S3 Table, we found some of special interest. For CAL, we found the *MYOG* (Myogenin) (a muscle-specific transcription factor), *CHI3L1* (Chitinase 3 Like 1) and *CHIT1* (Chitinase 1) genes on BTA16, both of these are involved in inflammatory processes. On BTA5, for CHA we found the largest number of genes (n = 57). These were linked to cell survival after damage or stress (*TIMELESS*; Timeless Circadian Regulator), transport and/or esterification of cholesterol (*APOF*; Apolipoprotein F), growth regulation, development and differentiation (*SLC39A5* - Solute Carrier Family 39 Member 5, *PA2G4* - Proliferation-Associated 2G4, *CD63* - CD63 Molecule) and olfactory receptors (*OR10P1*, *OR6C4*, *OR2AP1*, *OR6C2*, *OR6C68*). For LIM, the most interesting detected genes were located on BTA2: *CYP27C1* (encoding a member of the cytochrome P450 superfamily), *MSTN* (Myostatin) and genes encoding collagen chains (*COL5A2* - Collagen Type V Alpha 2 Chain, *COL3A1* - Collagen Type III Alpha 1 Chain). MUP and PON shared *CACNG2* (Calcium Voltage-Gated Channel Auxiliary Subunit Gamma 2), which is on BTA5; It is a gene involved in synaptic plasticity, learning and memory.

Homozygosity by descent (HBD) segments and global inbreeding

The percentage of non-HBD segments was higher in SAB (42.7%), followed by SAM, SAR and MUP (~ 39%), while CHA, LIM and PON showed values around 32.5%. The HBD segments identified (Fig 8) belonged mainly to HBD class with R_K equal to 64. Tuscan breeds showed a greater proportion of HBD genome also for R_K ranged from 4 to 8, while CAL is the breed with higher proportion of HBD segments when R_K was 16.

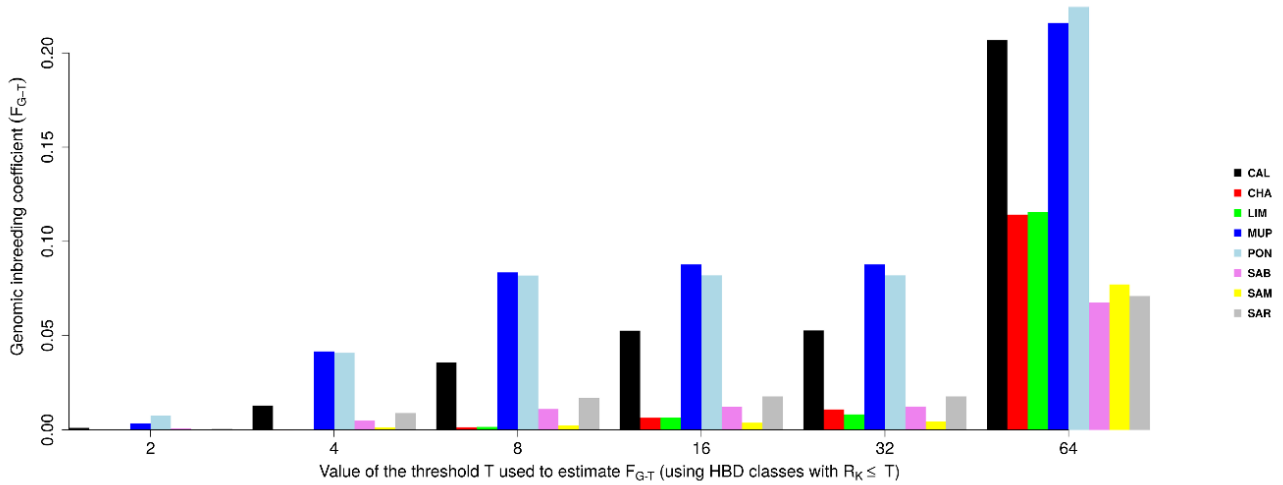

Fig 8. Proportion of the genome consisted of HBD classes in different RK. CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

The results suggested that all the breeds suffered an increase in inbreeding during ancient generations (around 32 generations ago), and only Tuscan breeds have been involved in new consistent inbreeding events, approximately 2-4 generations ago. Results were similar when inbreeding (F_{G-T}) was calculated respect to different base populations (Fig 9). The F_{G-T} was estimated as the probability of belonging to any of the HBD classes with a $R_K \leq$ a threshold T, averaged over the whole genome.

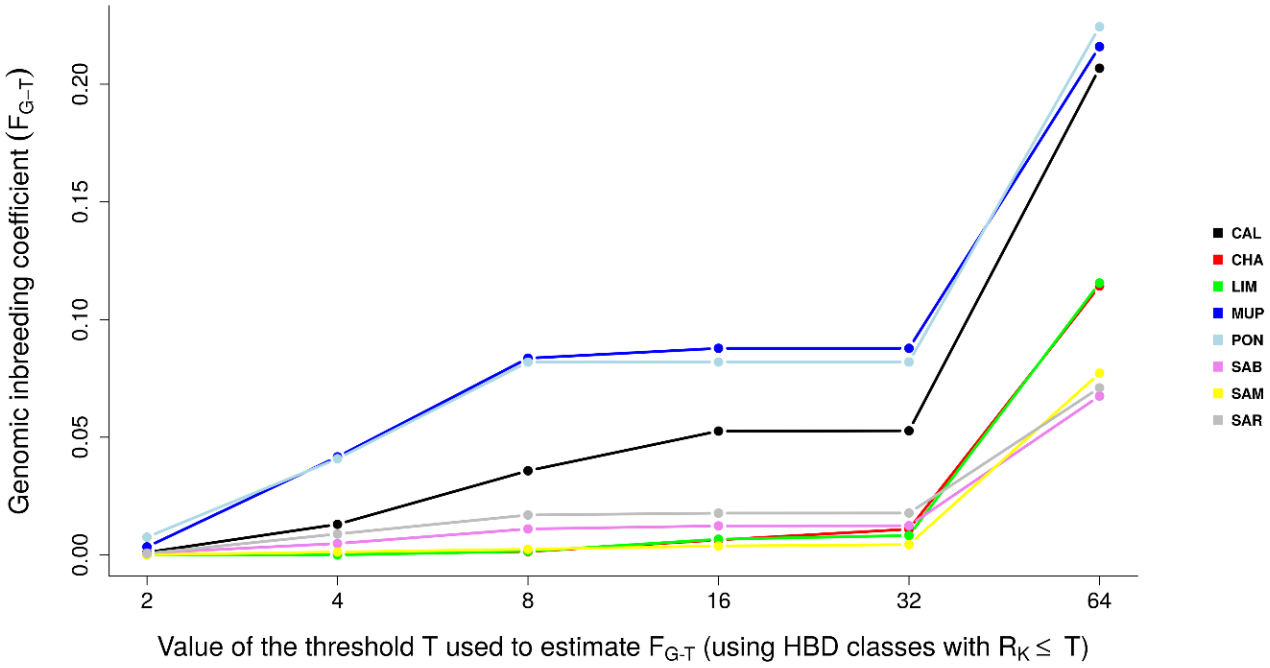

Fig 9. Genomic inbreeding coefficients estimated respect to different base populations (F_{G-T}) selecting different thresholds T, setting the base population approximately 0.5 * T generations ago.

The F_G estimated with the most remote base population showed values exceeding 0.2 for Tuscan breeds, while LIM and CHA had F_G around 0.1. Sardinian maintained F_G close to 0.06. Classes associated with smaller R_K rates (i.e., with longer HBD segments) explained a smaller HBD proportion: the average inbreeding coefficient was close to 0 in cosmopolitan and

Sardinian breeds, while for PON and MUP F_G was slightly lower than 0.1 from classes with $8 \leq R_K \leq 32$, while CAL didn't exceed 0.05. The inbreeding coefficient associated with common ancestors tracing back up to approximately two generations ago (corresponding to HBD-classes with $R_K \leq 4$) tended to 0 in all Tuscan breeds.

Partitioning of individual genomes in different HBD classes was also performed and Fig 10 reports the plot of 40 randomly chosen individuals per breed. Each bar represents an individual, the white spaces are individuals with no HBD segments belonging to HBD classes analyzed, the height of the different stacks is the proportion of the genome associated with the HBD class of the corresponding color and the total height showed the overall level of inbreeding.

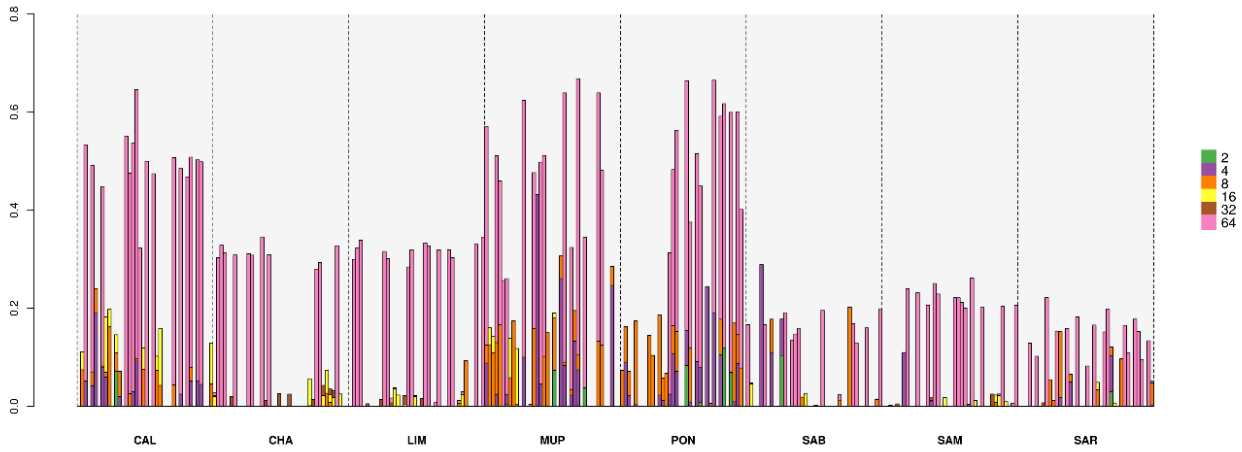

Fig 10. Proportion of the genome interested by HBD classes in different R_K. CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda.

The results confirmed that all breeds acquired the majority of their inbreeding load derived from ancient ancestors (32 generations ago), but once again, the Tuscan breeds appeared with a different demographic historical structure compared to Sardinian and cosmopolitan breeds; the level of inbreeding was higher in Tuscan and deriving both from past and recent phenomena. Indeed, PON and MUP were affected also by inbreeding acquired in recent generations (2-4, i.e. R_K=4 and 8, respectively). CAL also showed traces of ancestors from 8

generation ago in several individuals. Sardinian and cosmopolitan breeds showed lower levels of inbreeding and HBD segments derived from ancient ancestors.

Discussion

The advent of high-throughput genotyping arrays facilitated the study of genetic diversity and population structure in cattle [20], but local breeds remained understudied, even if in the last years greater importance has been given to the maintenance of biodiversity and the adoption of conservation measures for breeds at risk of extinction. Several advantages are brought with the conservation of local breeds, such as economical and genetic benefits [33]. This study comprehensively describes the genome-wide autozygosity and the consequent population structure of six local breed reared in Italy, namely Calvana, Mucca Pisana, Pontremolese, Sardo Bruna, Sardo Modicana and Sarda, by exploring the distribution of ROH, the level of genomic inbreeding (F_{ROH}) and the partitioning of homozygous identical by descent (HBD) segments across generations.

Multidimensional scaling plot analysis

MDS approach has been preferred to Principal component analysis (PCA) because it detects meaningful dimensions that explain observed genetic distance, i.e. pairwise IBS distance, while PCA method calculates the population structure based on genetic correlations among individuals [28]. The genome feature analysis carried out using MDS decomposition (Fig 3) was in accordance with the ROH-based PCA (Fig 4), highlighting a grouping among breeds. In particular, in both analyses Sardinian and cosmopolitan breeds were more similar to each other than Tuscan populations, except for PON, which clustered together with Sardinian and cosmopolitan breeds in MDS but not in PCA. This might be because sampled animals of PON were few due to the little size of population (only 49 living animals are reported [34]), and this could affect PCA results but it does not influence MDS analysis, solely based on genetic distances clustering. Plotting of the eight breeds was therefore a description of the breeds sample size. The SAB and SAR breeds counted approximately ~ 25,000 alive animals [35], Charolaise ~ 18,000 and Limousine ~ 50,000 (http://www.anacli.it/), while Calvana and Mucca Pisana samples amounted to a few hundred [33]. CHA and LIM formed the most compact cluster indicating that the breeding management in these breeds has a narrower genetic basis.

Runs of Homozygosity and Genomic inbreeding (F_{ROH})

The total number of ROH detected in each breed, was higher than what found in other studies focusing on cattle [14,35,36]. Differences might be due to the low-density panel used for genotyping, the quality control of the genotypes, the parameters used to define a ROH and the sample size. The small number of detected ROH in PON and SAM might be a result of limited sample size. The relatively high average length of ROH, ranged from 2.09 to 3.68 (Table 2), suggested that ancient inbreeding is present in all breeds. Short ROH represented the vast majority of ROH detected in all breeds (Fig 5), being more profound for LIM and CHA. This is in line with the history of cosmopolitan breeds, which have seen a crucial increase in sample size in the last 15 years (http://www.anacli.it). The growing interest in selection programs have probably led to a slight increase in inbreeding compared to other European Limousine and Charolaise populations; indeed, Szmatola et al. [14] described F_{ROH} in Polish Limousine ranging from 0.059 (>1Mbp) to 0.011 (>16 Mbp), while, Polish Charolaise showed values from 0.065 to 0.009. In this study F_{ROH} decreased in both selected breeds from 0.22 to 0.02 for the aforementioned classes of length.

Nowadays, SAB, SAR and SAM are distributed across 1,432, 950 and 146 farms, and it is known that Sardinian farmers exchange bulls between herds, causing a high average relatedness of individuals within farm but allowing a low degree of kinship among farms [35]. This could explain why F_{ROH} in SAR, SAB and SAM has been maintained near 0.05 (Fig 6 and Table 3) in medium (4-8 Mbp) and long ROH (>16 Mbp), suggesting that ancient and recent inbreeding has created a plateau of consanguinity across Sardinian populations. Results reported by Cesarani et al. [35], where the average length per individual ranged from 2.9 for SAR to 2.4 Mbp for SAM, showed a trend of decreased inbreeding in these populations. Tuscan breeds are in a more worrisome situation with their population sizes and inbreeding levels being at critical levels. The MUP breed presented average F_{ROH} values equal to 0.33 and 0.13 for short and long runs, respectively. The PON and CAL breeds had genomic inbreeding equal to 0.30 in the first mentioned class and 0.9-0.6 in the last, respectively. To the best of our knowledge, studies on Tuscan breeds here investigated were not present in literature, but several studies on local cattle breeds reported lower F_{ROH} values than our results. Addo et al. [37] analyzed two German local cattle populations (Angler and Red-and-White dual-purpose breeds) compared to Red Holstein and, on the contrary of this study, is the cosmopolitan population to have the greater values of genomic inbreeding. However, F_{ROH}

decreased quickly in the two local breeds, arriving to 0.009 and 0.02 in >16 Mbp length class, while in the Tuscan breeds F_{ROH} ranged from 0.06 to 0.13.

Quantification of the genome wide autozygosity for genetic conservation aims is fundamental because several studies correlated F_{ROH} with inbreeding depression in production and fertility traits [38–40]. In addition, recent inbreeding could fix recessive deleterious variants because there was a strong positive correlation between the number of deleterious homozygotes and the genomic ROH proportion [41].

Selection signatures and Gene enrichment

The higher threshold used in this study (80% of frequency) for the investigation of selection signatures, led to the identification of a common run between breeds. It is located on BTA21, starting from 83.766 kbp to 1,786.020 kbp.

An interesting genomic region with an occurrence of more than 80%, has been found in CAL on BTA16 (from 99,900 to 1,163,809 bp), where *MYOG* and *FMOD* genes were located. *MYOG* is related with *MSTN* gene which regulates muscle mass. The different myogenin genotypes are related to the variation in the number of muscle fibers and the growth rate, which lead to a variation in the muscle mass [42]. Indeed, has been suggested to use *MYOG* in marker-assisted selection for improving the growth trait in chicken [43].

FMOD plays an important role in the maintenance of mature tissues and has been discovered that reduces scar formation without diminishing the tensile strength in adult wound models (i.e. mice, rats and pig) [44]. It might be related to the higher rusticity and adaptability to harshly areas of local breeds. CHA reported a genomic region dense in genes (Table 4) on BTA5 (from 56,722,571 to 58,464,570 bp). Here two groups of genes are located: the first one included *TIMELESS*, *APON*, *APOF*, STAT2, IL23A and PAN2; the second one contained several genes of Olfactory Receptor Families.

APOF and *APON* are apolipoproteins which are component of lipoproteins and it has been showed that overexpression of Apolipoprotein F in mice reduced HDL cholesterol levels by 20-25% by accelerating clearance from the circulation [45].

Olfactory receptors (*ORs*) are essential for mammals to avoid dangers and search food [46]. Nowadays, a few genome-wide association studies reported associations between *ORs* and intake-related traits of livestock. Magalhães et al. [47] argued that olfactory receptors play a role in transferring energy within the cell, participating in the change of *GDP* (guanosine

diphosphate) to *GTP* (guanosine triphosphate). Other explanations for the effect of *ORs* on meat traits are their action by promoting the absorption of fatty acids and by differentiating adipocyte, that leads to an increase in accumulation of fat; in addition, their known role to increase the search of food improve the weight gain.

This cluster of genes found in CHA, is in line with artificial selection purposes, as for LIM, which included in the first significant run (BTA2; 5,305,197 - 7,958,492 bp), the presence of *MSTN* gene. It is known that *MSTN* inhibits the proliferation of muscle fibers, regulating muscle mass by negatively influencing cell differentiation via the myogenic regulatory factors (such as *MYOG*) [48]; three traits are associated with this gene: meat color L* (QTL:11644), meat percentage (QTL:11883; QTL:18424) and meat weight (QTL:11694) (https://www.animalgenome.org). Previous studies identified *MSTN* within ROH island in Limousine cattle, highlighting that *MSTN* is a gene under selective pressure for the phenotypic features in Limousine breed, indeed, *MSTN* has a strong positive effect on muscling and it is negative correlated with fat deposition [14,49].

MUP and PON presented consecutive genomic regions on BTA5, sharing the *CACNG2* gene. Interestingly, this gene was found to be associated with milk protein percentage QTL (https://www.animalgenome.org), which is probably because these two breeds originate from several past crossbreeding events including with Holstein and Schwyz (MUP) [34] and Reggiana (PON) (http://www.anacli.it) .

Homozygosity by descent (HBD) segments and global inbreeding

The parameters used in this analysis were chosen according to ROH results. The length of ROH ranged from >0 and <16 Mbp, , consequently, we are interested to investigate HBD classes with R_K equal to 2, 4, 8, 16, 32, 64, which correspond to 1, 2, 4, 8, 16 and 32 generations ago. F_G values are higher than F_{ROH} observing the equivalent length of segments, and this could be because the algorithm has difficulties to detect very short ROH when a low and/or medium density chips have used [10], even if the HBD probability of the SNP in these regions can be estimated, leading to F_G values higher than F_{ROH} values. However, within 32 generations, no pronounced differences in F_G levels have been found by Solè et al. [18] when low, medium or high density SNP chips were compared, defining GGP-LDv4 33k SNP chip adequate and cheaper for HBD segments identification. The greater proportion of HBD genome originated from ancient ancestors dates back to 32 generations ago and this is in line

with the numerous short ROH detected. An unexpected result was that Sardinian breeds showed almost halved values in HBD classes (when R_K is 64) compared to cosmopolitan, suggesting that for these breeds, R_K should be increased in order to detect shorter HBD segments which have been found during ROH analysis (Fig 5). It would be interesting to investigate historic events during the 16^{th} and 32^{th} generation observing the inbreeding increase from each to the other generation. No pedigree data on these generations was available when pedigree inbreeding has been calculated in a previous study for these Italian breeds, except for CHA whose results are comparable [33].

However, in the last generations the inbreeding coefficients decreased. This suggests that the increased attention to the maintenance of biodiversity have led to a greater mating control by farmers. Unexpectedly, the investigation of individual proportion of HBD genome, identified some individuals that are not HBD. Further analyses are needed but these individuals could be identified and selected for their use in mating programs to decrease inbreeding. Furthermore, animals with a small proportion of HBD genome compared to population could be also useful in conservation plans of local endangered cattle. Nevertheless, the worrisome situation for Tuscan breeds in terms of inbreeding has been underlined. Also, issues in mating management have been arising since the global inbreeding depends on past ancestors but also to recent generations. Given the limited diffusion of CAL, MUP and PON, the number of potential matings is extremely reduced.

Conclusion

The genomic results using a low-density SNP chip panel showed critical inbreeding levels in smaller local populations. Cosmopolitan breeds showed lower genetic variability but also negligible inbreeding levels, demonstrating the soundness of the ongoing breeding scheme. The population structure and genetic distances highlighted a clear separation among the breeds, with clusters related to productive purposes and sample sizes. The results obtained in this study represent a useful tool for preserving biodiversity, proving background information for the correct genetic management and conservation for the described populations.

Acknowledgments

We acknowledge Associazione Nazionale Allevatori delle razze bovine Charolaise e Limousine (ANACLI) for providing the data.

References

[1] Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet 2018;19:220–34. https://doi.org/10.1038/nrg.2017.109.

[2] Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic Runs of Homozygosity Record Population History and Consanguinity. PLOS ONE 2010;5:e13996. https://doi.org/10.1371/journal.pone.0013996.

[3] Zhang Q, Calus MP, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet 2015;16. https://doi.org/10.1186/s12863-015-0227-7.

[4] Wright S. Coefficients of Inbreeding and Relationship. Am Nat 1922;56:330–8.

[5] Keller MC, Visscher PM, Goddard ME. Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. Genetics 2011;189:237–49. https://doi.org/10.1534/genetics.111.130922.

[6] Broman KW, Weber JL. Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain. Am J Hum Genet 1999;65:1493–500.

[7] Saura M, Fernández A, Varona L, Fernández AI, de Cara MÁR, Barragán C, et al. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genet Sel Evol GSE 2015;47:1. https://doi.org/10.1186/s12711-014-0081-5.

[8] Yengo L, Zhu Z, Wray NR, Weir BS, Yang J, Robinson MR, et al. Detection and quantification of inbreeding depression for complex traits from SNP data. Proc Natl Acad Sci 2017;114:8602–7. https://doi.org/10.1073/pnas.1621096114.

[9] Mastrangelo S, Jemaa SB, Ciani E, Sottile G, Moscarelli A, Boussaha M, et al. Genome-wide detection of signatures of selection in three Valdostana cattle populations. J Anim Breed Genet 2020;137:609–21. https://doi.org/10.1111/jbg.12476.

[10] Nandolo W, Mészáros G, Banda LJ, Gondwe TN, Lamuno D, Mulindwa HA, et al. Timing and Extent of Inbreeding in African Goats. Front Genet 2019;10. https://doi.org/10.3389/fgene.2019.00537.

[11] Mastrangelo S, Tolone M, Sardina MT, Sottile G, Sutera AM, Di Gerlando R, et al. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. Genet Sel Evol 2017;49:84. https://doi.org/10.1186/s12711-017-0360-z.

[12] Dadousis C, Cecchi F, Ablondi M, Fabbri MC, Stella A, Bozzi R. Keep Garfagnina alive. An integrated study on patterns of homozygosity, genomic inbreeding, admixture and breed traceability of the Italian Garfagnina goat breed. PLOS ONE 2021;16:e0232436. https://doi.org/10.1371/journal.pone.0232436.

[13] Strillacci MG, Marelli SP, Martinez-Velazquez G. Hybrid Versus Autochthonous Turkey Populations: Homozygous Genomic Regions Occurrences Due to Artificial and Natural Selection. Animals 2020;10:1318. https://doi.org/10.3390/ani10081318.

[14] Szmatoła T, Gurgul A, Ropka-Molik K, Jasielczuk I, Ząbek T, Bugno-Poniewierska M. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. Livest Sci 2016;188:72–80. https://doi.org/10.1016/j.livsci.2016.04.006.

[15] Xu L, Zhao G, Yang L, Zhu B, Chen Y, Zhang L, et al. Genomic Patterns of Homozygosity in Chinese Local Cattle. Sci Rep 2019;9:16977. https://doi.org/10.1038/s41598-019-53274-3.

[16] Álvarez I, Fernández I, Traoré A, Pérez-Pardal L, Menéndez-Arias NA, Goyache F. Ancient Homozygosity Segments in West African Djallonké Sheep Inform on the Genomic Impact of Livestock Adaptation to the Environment. Anim Open Access J MDPI 2020;10. https://doi.org/10.3390/ani10071178.

[17] Druet T, Oleński K, Flori L, Bertrand AR, Olech W, Tokarska M, et al. Genomic Footprints of Recovery in the European Bison. J Hered 2020;111:194–203. https://doi.org/10.1093/jhered/esaa002.

[18] Solé M, Gori A-S, Faux P, Bertrand A, Farnir F, Gautier M, et al. Age-based partitioning of individual genomic inbreeding levels in Belgian Blue cattle. Genet Sel Evol GSE 2017;49:92. https://doi.org/10.1186/s12711-017-0370-x.

[19] Senczuk G, Mastrangelo S, Ciani E, Battaglini L, Cendron F, Ciampolini R, et al. The genetic heritage of Alpine local cattle breeds using genomic SNP data. Genet Sel Evol 2020;52:40. https://doi.org/10.1186/s12711-020-00559-1.

[20] Mastrangelo S, Tolone M, Gerlando RD, Fontanesi L, Sardina MT, Portolano B. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. Animal 2016;10:746–54. https://doi.org/10.1017/S1751731115002943.

[21] Fontanesi L, Scotti E, Samorè AB, Bagnato A, Russo V. Association of 20 candidate gene markers with milk production and composition traits in sires of Reggiana breed, a local dairy cattle population. Livest Sci 2015;176:14–21. https://doi.org/10.1016/j.livsci.2015.03.022.

[22] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4. https://doi.org/10.1186/s13742-015-0047-8.

[23] Dixit SP, Singh S, Ganguly I, Bhatia AK, Sharma A, Kumar NA, et al. Genome-Wide Runs of Homozygosity Revealed Selection Signatures in Bos indicus. Front Genet 2020;11. https://doi.org/10.3389/fgene.2020.00092.

[24] Meyermans R, Gorssen W, Buys N, Janssens S. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. BMC Genomics 2020;21:94. https://doi.org/10.1186/s12864-020-6463-x.

[25] Manunza A, Noce A, Serradilla JM, Goyache F, Martínez A, Capote J, et al. A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. Genet Sel Evol 2016;48:52. https://doi.org/10.1186/s12711-016-0229-6.

[26] scatterplot3d function | R Documentation n.d. https://www.rdocumentation.org/packages/scatterplot3d/versions/0.3-41/topics/scatterplot3d (accessed January 20, 2021).

[27] Biscarini F, Cozzi P, Gaspa G, Marras G. detectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes. 2019.

[28] Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet 2015;46:110–21. https://doi.org/10.1111/age.12259.

[29] Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, et al. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. J Anim Breed Genet 2013;130:286–93. https://doi.org/10.1111/jbg.12012.

[30] McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet 2008;83:359–72. https://doi.org/10.1016/j.ajhg.2008.08.007.

[31] Alemu SW, Kadri NK, Harland C, Faux P, Charlier C, Caballero A, et al. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. Heredity 2020. https://doi.org/10.1038/s41437-020-00383-9.

[32] Druet T, Bertrand A, Kadri N, Gautier M. The RZooRoH package n.d.:46.

[33] Fabbri MC, Gonçalves de Rezende MP, Dadousis C, Biffani S, Negrini R, Souza Carneiro PL, et al. Population Structure and Genetic Diversity of Italian Beef Breeds as a Tool for Planning Conservation and Selection Strategies. Anim Open Access J MDPI 2019;9. https://doi.org/10.3390/ani9110880.

[34] Fabbri MC, Dadousis C, Bozzi R. Estimation of Linkage Disequilibrium and Effective Population Size in Three Italian Autochthonous Beef Breeds. Animals 2020;10:1034. https://doi.org/10.3390/ani10061034.

[35] Cesarani A, Sorbolini S, Criscione A, Bordonaro S, Pulina G, Battacone G, et al. Genome-wide variability and selection signatures in Italian island cattle breeds. Anim Genet 2018;49:371–83. https://doi.org/10.1111/age.12697.

[36] Peripolli E, Stafuzza NB, Amorim ST, de Lemos MVA, Grigoletto L, Kluska S, et al. Genome-wide scan for runs of homozygosity in the composite Montana Tropical® beef cattle.

J Anim Breed Genet Z Tierzuchtung Zuchtungsbiologie 2020;137:155–65. https://doi.org/10.1111/jbg.12428.

[37] Addo S, Klingel S, Hinrichs D, Thaller G. Runs of Homozygosity and NetView analyses provide new insight into the genome-wide diversity and admixture of three German cattle breeds. PLOS ONE 2019;14:e0225847. https://doi.org/10.1371/journal.pone.0225847.

[38] Browett S, McHugo G, Richardson IW, Magee DA, Park SDE, Fahey AG, et al. Genomic Characterisation of the Indigenous Irish Kerry Cattle Breed. Front Genet 2018;9. https://doi.org/10.3389/fgene.2018.00051.

[39] Martikainen K, Tyrisevä AM, Matilainen K, Pösö J, Uimari P. Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. J Anim Breed Genet 2017;134:383–92. https://doi.org/10.1111/jbg.12285.

[40] Kim ES, Sonstegard TS, Van CT, Wiggans G, Rothschild MF. The Relationship between Runs of Homozygosity and Inbreeding in Jersey Cattle under Selection. PloS One 2015;10:e0129967–e0129967. https://doi.org/10.1371/journal.pone.0129967.

[41] Zhang Q, Guldbrandtsen B, Bosse M, Lund MS, Sahana G. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics 2015;16:542. https://doi.org/10.1186/s12864-015-1715-x.

[42] Xue H-L, Zhou Z-X. Effects of the MyoG Gene on the Partial Growth Traits in Pigs. Acta Genet Sin 2006;33:992–7. https://doi.org/10.1016/S0379-4172(06)60134-0.

[43] Wei Y, Zhang GX, Zhang T, Wang JY, Fan QC, Tang Y, et al. Myf5 and MyoG gene SNPs associated with Bian chicken growth trait. Genet Mol Res GMR 2016;15. https://doi.org/10.4238/gmr.15037043.

[44] Jiang W, Ting K, Lee S, Zara JN, Song R, Li C, et al. Fibromodulin reduces scar size and increases scar tensile strength in normal and excessive-mechanical-loading porcine cutaneous wounds. J Cell Mol Med 2018;22:2510–3. https://doi.org/10.1111/jcmm.13516.

[45] Lagor WR, Fields DW, Khetarpal SA, Kumaravel A, Lin W, Weintraub N, et al. The effects of apolipoprotein F deficiency on high density lipoprotein cholesterol metabolism in mice. PloS One 2012;7:e31616. https://doi.org/10.1371/journal.pone.0031616.

[46] Lee SH, Cho YM, Lim D, Kim HC, Choi BH, Park HS, et al. Linkage Disequilibrium and Effective Population Size in Hanwoo Korean Cattle. Asian-Australas J Anim Sci 2011;24:1660–5. https://doi.org/2011.24.12.1660.

[47] Magalhães AFB, Camargo GMF de, Junior GAF, Gordo DGM, Tonussi RL, Costa RB, et al. Genome-Wide Association Study of Meat Quality Traits in Nellore Cattle. PLOS ONE 2016;11:e0157845. https://doi.org/10.1371/journal.pone.0157845.

[48] Snijders T, Wall BT, Dirks ML, Senden JMG, Hartgens F, Dolmans J, et al. Muscle disuse atrophy is not accompanied by changes in skeletal muscle satellite cell content. Clin Sci 2014;126:557–66. https://doi.org/10.1042/CS20130295.

[49] Esmailizadeh AK, Bottema CDK, Sellick GS, Verbyla AP, Morris CA, Cullen NG, et al. Effects of the myostatin F94L substitution on beef traits. J Anim Sci 2008;86:1038–46. https://doi.org/10.2527/jas.2007-0589.

Supporting information

S1 Table. Number of ROH per chromosome in each breed, where CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda

BTA	CAL	CHA	LIM	MUP	PON	SAB	SAM	SAR
1	2,873	2,956	3,312	2,543	599	3,01	1,439	2,821
2	2,18	2,838	2,969	2,679	588	2,676	1,355	2,712
3	1,907	2,266	2,353	2,284	491	2,245	1,084	2,286
4	2,13	2,647	2,731	2,311	477	2,475	1,19	2,612
5	2,274	2,656	2,649	2,108	534	2,393	1,092	2,239
6	2,483	2,583	2,726	2,323	472	2,542	1,304	2,44
7	1,971	2,436	2,556	2,038	473	2,319	935	2,3
8	2,27	2,719	3,003	2,223	528	2,699	1,165	2,641
9	1,864	2,102	2,236	1,967	435	2,004	1,048	2,098
10	1,885	2,178	2,132	1,965	432	1,943	959	2,187
11	1,856	2,042	2,185	1,882	405	2,047	977	2,077
12	1,483	1,842	1,937	1,847	357	1,706	796	1,72
13	1,963	2,366	2,402	2,024	483	2,197	1,026	2,265
14	1,791	2,077	2,31	1,712	405	2,059	933	2,063
15	1,263	1,523	1,694	1,388	299	1,5	712	1,521
16	1,207	1,578	1,633	1,355	303	1,488	699	1,6

17	1,212	1,344	1,408	1,138	338	1,299	640	1,261
18	1,094	1,169	1,201	1,127	225	1,043	534	1,049
19	905	1,095	1,06	1,08	225	1,009	498	1,053
20	865	1,105	1,377	931	236	1,036	550	1,116
21	1,184	1,496	1,632	1,263	323	1,574	711	1,534
22	908	1,072	1,145	850	202	1,034	477	1,028
23	824	777	902	938	178	782	368	736
24	997	1,102	1,22	945	237	1,121	586	1,088
25	588	624	700	553	119	549	266	630
26	838	841	953	896	186	785	385	877
27	536	691	751	696	124	633	304	651
28	732	758	884	776	182	743	315	792
29	790	929	949	906	188	865	410	942

S2 Table. Characterization of genomic regions with frequency equal to 80% of ROH occurrence, where CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese; SAB = Sardo Bruna; SAM = Sardo Modicana; SAR = Sarda

Breed	CHR	Start SNP	End SNP	N. SNPs	from	to	Length (Mbp)
	3	RNF11	BovineHD0300027507	3	95,601,697	95,753,782	0.15
	6	BovineHD0600010715	Hapmap26233-BTA-75846	66	38,698,886	40,183,935	1.49
	6	BovineHD0600020167	BovineHD0600034501	3	72,625,498	72,991,427	0.37
	7	BovineHD0700014803	BovineHD0700015160	10	50,751,321	52,460,183	1.71
	13	Hapmap47850-BTA-118310	BovineHD1300017027	9	58,904,174	59,467,689	0.56
CAL	13	BovineHD4100010307	BTA-112783-no-rs	12	63,242,901	63,715,266	0.47
	13	ARS-BFGL-NGS-17925	BovineHD1300018419	9	63,907,611	64,683,488	0.78
	13	BovineHD1300018513	BovineHD1300018659	8	65,161251	65,721,796	0.56
	13	BovineHD1300018795	BovineHD1300018836	4	66,230,066	66,402,921	0.17
	14	BovineHD1400013371	BovineHD1400013441	3	47,279,060	47,535,651	0.26
	16	BovineHD1600000011	BovineHD1600000286	17	99,900	1,163,809	1.06
	5	BovineHD0500016070	BovineHD0500016088	18	56,625,841	56,716,286	0.09
CHA	5	BovineHD0500016090	BovineHD0500016469	52	56,722,571	58,464,570	1.74
	13	ARS-BFGL-NGS-4243	BovineHD1300003267	5	10,966,294	11,842,431	0.88

	21	BovineHD2100000012	BovineHD2100000219	18	83,766	1,786,020	1.70
	21	BovineHD2100021044	BovineHD2100000258	2	1,854,171	2,085,345	0.23
LIM	2	ARS-BFGL-NGS-21306	BTB-01111412	35	5,305,197	7,958,492	2.65
	13	BovineHD1300003108	BovineHD1300003267	4	11,397,610	11,842,431	0.44
	14	BovineHD1400007190	Hapmap40958-BTA-34312	29	24,754,549	26,450,034	1.70
	21	BovineHD2100000012	BovineHD2100000320	25	83,766	2,467,774	2.38
MUP	5	5_74257621	BovineHD0500021258	4	74,257,621	74,762,971	0.51
	5	5_74951342	5_75130860	19	74,951,342	75,130,860	0.18
	7	BovineHD0700014803	BovineHD0700015160	11	50,751,321	52,460,183	1.71
	7	BovineHD0700015212	BovineHD0700015266	1	52,728,318	53,007,998	0.28
	11	BovineHD1100017061	Hapmap41117-BTA-99065	9	59,628,354	60,548,547	0.92
	11	Hapmap48973-BTA-99093	BovineHD1100017358	3	60,706,511	60,928,551	0.22
	12	BovineHD1200027369	BovineHD1200013148	2	47,684,978	47,833,336	0.15
	21	BovineHD2100000012	BovineHD2100000283	23	83,766	2,256,102	2.17
PON	5	BovineHD0500021258	5_75130860	20	74,762,971	75,130,860	0.37
	13	BovineHD1300015452	BovineHD1300015609	5	54,453,339	54,979,764	0.53
	14	BovineHD1400013371	BovineHD1400013483	5	47,279,060	47,665,095	0.39
	21	BovineHD2100000012	BovineHD2100000219	18	83,766	1,786,020	1.70
SAB	21	BovineHD2100000012	BovineHD2100000219	17	83,766	1,786,020	1.70
SAM	21	BovineHD2100000012	BovineHD2100000258	20	83,766	2,085,345	2.00
SAR	21	BovineHD2100000012	BovineHD2100000258	20	83,766	2,085,345	2.00

Assembly Bos_taurus_UMD_3.1

S3 Table. List of genes within significant ROH (80% of occurrence) with a minimum of 15 SNPs, where CAL = Calvana; CHA = Charolaise; LIM = Limousine; MUP = Mucca Pisana; PON = Pontremolese

	CHR	Start_SNP	End_SNP	Genes in ROH
CAL	6	BovineHD0600010715	Hapmap26233-BTA-75846	*SLIT2, MIR218-1*
	16	BovineHD1600000011	BovineHD1600000286	*OR5L1, TRNAR-UCU, TMEM183A,*

				PPFIA4, MYOG, ADORA1, MYBPH, CHI3L1, CHIT1, BTG2, FMOD
CHA	5	BovineHD0500016070	BovineHD0500016088	*HSD17B6, TRNAG-CCC, PRIM1*
	5	BovineHD0500016090	BovineHD0500016469	*PRIM1, PTGES3, NACA, ATP5F1B, MIR677, BAZ2A, RBMS2, GLS2, SPRYD4, MIP, TIMELESS, APOF, ApoN, STAT2, IL23A, MIR2432, PAN2, CNPY2, CS, COQ10A, ANKRD52, MIR2433, SLC39A5, NABP2, RNF41, TRNAS-CGA, SMARCC2, MYL6, MYL6B, ESYT1, ZC3H10, PA2G4, ERBB3, RPS26, IKZF4, SUOX, RAB5B, CDK2, PMEL, DGKA, PYM1, MMP19, DNAJC14, ORMDL2, SARNP, GDF11, CD63, RDH5, BLOC1S1, ITGA7, METTL7B, OR10P1, OR6C4, OR2AP1, OR6C2, OR6C68, OR6C68*
LIM	2	ARS-BFGL-NGS-21306	BTB-01111412	*CYP27C1, BIN1, MIR2350 NAB1, NEMP2, MFSD6, INPP1, HIBCH, C2H2orf88, MSTN, PMS1, ORMDL1, OSGEPL1, ANKAR, ASNSD1, SLC40A1, WDR75,COL5A2, MIR2917, COL3A1, DIRC1,GULP1*
	14	BovineHD1400007190	Hapmap40958-BTA-34312	*SDCBP, NSMAF, TOX, TRNAC-GCA, CA8, RAB2A, CHD7*
MUP	5	5_74951342	5_75130860	*CACNG2*
PON	5	BovineHD0500021258	5_75130860	*MYH9, TXN2, FOXRED2, EIF3D, TRNAE-UUC, CACNG2*

Assembly Bos_taurus_UMD_3.1

Chapter 7

Conclusion

The thesis has the aim to analyze through pedigree and genomic information the population structure of six Italian local beef breeds at risk of extinction (Calvana, Mucca Pisana, Sarda, Sardo Bruna and Sardo Modicana). Nowadays, pedigree is used to calculate inbreeding and to estimate the individual breeding value. Results of pedigree investigation highlighted the medium - low pedigree quality which interferes with an accurate assessment of the genetic diversity and practically, on future matings plans in order to establish conservation schemes. To bridge this gap and to shed more light on the genomic background of the aforementioned breeds and the level of their genomic diversity, genomic analysis has been performed. Most accurate estimates of inbreeding were provided with Runs of Homozygosity analysis, which also allowed a more precise description of the present and historical population structure and better contextualize when matings between relatives have occurred in the past. Future studies are needed to focus on the genomic regions with high linkage disequilibrium, on significant runs of homozygosity detected and on genes related to these latter. Those genomic regions represent a useful tool for providing background information for the correct genetic management and conservation for the six local populations in order to maintain biodiversity.

The importance of genetic diversity in domestic species is directly related to the necessity for genetic improvement of selected breeds as well as to facilitate rapid adaptation to potential changes in breeding goals. Selection applied to livestock breeds tends, however, to reduce levels of genetic variation through two major processes. First, most domestic species are highly selected for a few economically important traits (e.g. milk or meat production), which decreases genetic diversity as a consequence of directional selection. Secondly, most breeds tend to be genetically uniform as a result of high levels of gene flow among populations and artificial selection of some reproductive individuals (e.g. through artificial insemination and embryo transfer). The high levels of artificial selection through the intensive use of specific sires and assisted reproduction have greatly reduced the effective population size of commercial domestic breeds. Most local breeds are unmanaged or managed through traditional husbandry; therefore, they are subjected to the process of selective pressure. Consequently, these breeds have become locally adapted to a wide range of environments,

showing high levels of phenotypic variability and increased fitness under natural conditions. For all these reasons, local breeds are considered as a *reservoir* of biodiversity and they are safeguard for future uses also for cross breeding purposes with selected breeds, in case genetic diversity will continue to decrease in specialized breeds. As it is known local populations are characterized by adaptability, resilient capability and resistance to disease which are generally missing in selected breeds. Nevertheless, the conservation programs used until now don't avoid the risk of extinction of those breeds which have a low census size as the Tuscan beef breeds here described. To elaborate conservation schemes the primary step is to identify the genetic, genomic and phenotypic features of these populations, and the measure most important and urgent is to coordinate matings plans for each breed and to supervise the correct application of farmers. To do that, inbreeding estimates, linkage disequilibrium picture and homozygous regions description should be the criteria from to start with. On the other hand, some local breeds have a great census size, for instance Sardinian beef breeds here included, but selection schemes are not allowed in Italy if applied to local populations at risk of extinction. The number of alive animals, especially for Sardo Bruna and Sarda, are comparable to other local breeds which are included in selection schemes, for instance the Rendena breed. Although Rendena suffered a sample size contraction in the last century, selection allowed the improvement of productions and consequently it leads to the increase in sample size. Consequently, in theory, genetic and genomic approaches to control inbreeding and to increase genetic gain may be used in breeds at risk of extinction to facilitate their spread. Cornerstones of this idea is the Optimal Contribution Selection (OCS) proposed by Meuwissen in 1997. He proposed to maximize the expected mean breeding value of the offspring while constraining its gene diversity to a predefined value. A related but not equivalent approach is to maximize the gene diversity in the offspring with or without constraining its expected mean breeding value to a predefined value. This latter approach seems more appropriate because the focus is on conservation. In general, the method consists of calculating an optimum contribution c_α (or the desired number of offspring) for each individual α such that the offspring population maximizes an appropriate objective function φ under some side conditions. In Rendena cattle this method has been just applied, the consequent matings schemes elaborated are based on genetic/genomic information and not only on pedigree records, which are, as demonstrated, too affected by incompleteness in local breeds.

However, the obtained results in these studies represent the first exhaustive genomic background description of Calvana, Mucca Pisana, Pontremolese, Sarda, Sardo Bruna, Sardo Modicana, Italian Limousine and Italian Charolaise breeds. Consequently, they may be used as a tool for preserving biodiversity of the aforementioned autochthonous populations, and may provide the guidelines for correct management and conservation schemes.