Animal, page 1 of 10 © The Animal Consortium 2017
doi:10.1017/S1751731116002780

animal

Genome-wide analysis in endangered populations: a case study in Barbaresca sheep

S. Mastrangelo[1†], B. Portolano[1], R. Di Gerlando[1], R. Ciampolini[2], M. Tolone[1], M. T. Sardina[1] and The International Sheep Genomics Consortium

[1]*Dipartimento Scienze Agrarie e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy;* [2]*Dipartimento di Scienze Veterinarie, Università degli Studi di Pisa, 56100 Pisa, Italy*

(Received 29 April 2016; Accepted 1 December 2016)

Analysis of genomic data is becoming increasingly common in the livestock industry and the findings have been an invaluable resource for effective management of breeding programs in small and endangered populations. In this paper, with the goal of highlighting the potential of genomic analysis for small and endangered populations, genome-wide levels of linkage disequilibrium, measured as the squared correlation coefficient of allele frequencies at a pair of loci, effective population size, runs of homozygosity (ROH) and genetic diversity parameters, were estimated in Barbaresca sheep using Illumina OvineSNP50K array data. Moreover, the breed's genetic structure and its relationship with other breeds were investigated. Levels of pairwise linkage disequilibrium decreased with increasing distance between single nucleotide polymorphisms. An average correlation coefficient <0.25 was found for markers located up to 50 kb apart. Therefore, these results support the need to use denser single nucleotide polymorphism panels for high power association mapping and genomic selection efficiency in future breeding programs. The estimate of past effective population size ranged from 747 animals 250 generations ago to 28 animals five generations ago, whereas the contemporary effective population size was 25 animals. A total of 637 ROH were identified, most of which were short (67%) and ranged from 1 to 10 Mb. The genetic analyses revealed that the Barbaresca breed tended to display lower variability than other Sicilian breeds. Recent inbreeding was evident, according to the ROH analysis. All the investigated parameters showed a comparatively narrow genetic base and indicated an endangered status for Barbaresca. Multidimensional scaling, model-based clustering, measurement of population differentiation, neighbor networks and haplotype sharing distinguished Barbaresca from other breeds, showed a low level of admixture with the other breeds considered in this study, and indicated clear genetic differences compared with other breeds. Attention should be given to the conservation of Barbaresca due to its critical conservation status. In this context, genomic information may have a crucial role in management of small and endangered populations.

Keywords: OvineSNP50K, sheep, population structure, linkage disequilibrium, livestock conservation

Implications

An investigation of the genomic variation within a breed is an important prerequisite in maintaining the breed's integrity and in the design of appropriate conservation plan. The aim of this work was to study the genomic structure and the population evolution of Barbaresca sheep. All the investigated parameters showed a comparatively narrow genetic base, indicating endangered status for Barbaresca. Attention should be given to conservation of this breed because of its critical conservation status. In this context, genomic information may play a crucial role in the management of small and endangered populations.

† E-mail: salvatore.mastrangelo@unipa.it

Introduction

An investigation of genomic variation within a breed is an important prerequisite to maintain its integrity and to ensure appropriate conservation. The application of recently developed genomic technology has great potential to increase our understanding of the genetic architecture of complex traits, to improve selection efficiency in domestic animals through genomic selection, and to conduct association studies (García-Gámez *et al.*, 2012). The success of these applications depends on the extent of linkage disequilibrium (LD) across the genome. Linkage disequilibrium describes the non-random association of alleles at different loci and can result from processes such as migration, selection and genetic drift in finite populations (Wang, 2005). Linkage disequilibrium is

often used to determine the optimal number of markers required for fine mapping of quantitative trait loci, the evolutionary history of the populations (Hayes *et al.*, 2003), and the regions influenced by natural selection. Today, the extent of LD has been reported for common and important sheep breeds, but there is little information about the degree of genome-wide LD in local sheep breeds (Mastrangelo *et al.*, 2014), especially for small populations. Moreover, the study of LD could be used to infer effective population size (N_e) (Hayes *et al.*, 2003). The N_e is a general indicator of the risk of genetic erosion, contains relevant information for the monitoring of the genetic diversity, and helps to explain how populations evolved (Tenesa *et al.*, 2007). The pattern of historical N_e can also help us to understand the impact of selective breeding strategies on the genetic variation present in populations and can provide insight into the level of inbreeding in populations for which pedigree data are incomplete or unavailable (Corbin *et al.*, 2010). Moreover, other methods can be used to estimate genetic diversity using marker data, such as observed and expected heterozygosity, runs of homozygosity (ROH) and Wright's *F* statistic (F_{ST}). In particular, ROH are contiguous lengths of homozygous genotypes that are present in an individual due to parental transmission of identical haplotypes to their offspring (Gibson *et al.*, 2006). Because recombination events interrupt long chromosome segments, long ROH (~10 Mb) will arise as a result of recent inbreeding (up to five generations ago), whereas shorter ROH (~1 Mb) can indicate a more distant ancestral effect (up to 50 generations ago).

During the last century, erosion of livestock genetic resources was observed as the result of massive replacement of low-productivity local breeds with highly productive ones. Italians have a long history of sheep breeding and, despite a dramatic contraction in numbers, still raise several local sheep breeds that may represent a unique source of genetic diversity (Ciani *et al.*, 2013a). In fact, most local breeds are the result of particular adaptation to a singular, sometimes harsh environment, and in many cases no other breed could survive in that habitat (Fernández *et al.*, 2011). Therefore, the conservation and monitoring of the genetic diversity of these local breeds are fundamental to meet future breeding needs, especially in the context of global climate change.

An interesting situation is represented by Barbaresca, an ancient Sicilian sheep breed that is listed as endangered. The breed seems to originate from crosses between Tunisian Barbary sheep from North Africa and the Pinzirita breed. The endangered status of the Barbaresca breed is linked to the following peculiarities: in the year 1980, the population consisted of 35 000 heads, but by 20 years later almost 70% of the pre-existing population had disappeared. Nowadays, about 1300 animals in 20 herds are recorded in the Herd Book (ASSONAPA, 2014). Barbaresca sheep are reared in a very restricted area in central Sicily on small- and medium-sized farms under a semi-extensive farming system. Barbaresca is a dual-purpose breed; it is reared both for its milk (and the resultant local dairy products) and for its meat.

The aim of this work was to study the genomic diversity and population evolution in Barbaresca sheep using data generated from an Illumina OvineSNP50K array. Moreover, genetic diversity within the breed and its relationship with other breeds were investigated. This paper represents a case study that could have relevance for the conservation of other local endangered livestock breeds and highlights how genomic data may be used to ascertain population structures of small and endangered breeds.

Material and methods

DNA sampling, genotyping and quality control

Blood samples were collected from 40 individuals of the Barbaresca breed (36 ewes and four rams) from 10 different flocks to capture a representative sample of within-breed genetic diversity. For this breed, pedigree data were not available. All animals were genotyped for 54 241 single nucleotide polymorphisms (SNPs) using the Illumina OvineSNP50K array. Genotyping was performed by Dipartimento Scienze Agrarie e Forestali, University of Palermo, following standard operating procedures recommended by the manufacturer. Markers were filtered to exclude loci assigned to unmapped contigs. Therefore, only SNPs located on autosomes were considered for further analyses. Moreover, quality control included: call frequency ≥ 0.95, minor allele frequency (MAF) ≥ 0.05 and Hardy–Weinberg equilibrium (*P*-value > 0.001). Single nucleotide polymorphisms that did not satisfy these quality criteria were discarded. Animals with more than 5% missing SNPs were also removed from the analysis.

Linkage disequilibrium

As measurement of LD, the squared correlation coefficient of allele frequencies at a pair of loci (r^2) was estimated for all pairwise combinations of syntenic SNPs using the LD plot function in HAPLOVIEW v. 4.2 software (Barrett *et al.*, 2005). For each chromosome, a pairwise r^2 was calculated for SNPs between 0 and 50 Mb apart. To examine the decay of LD with physical distance, SNP pairs on autosomes were sorted into bins based on their inter-marker distance, and average values of r^2 were calculated for each bin. Moreover, r^2 was also estimated between adjacent SNPs on each chromosome using the PLINK command *–chr x –r² –ld-window 2 –ld-window-r² 0* (Purcell *et al.*, 2007). Because a relatively small number of genotyped individuals was surveyed, a corrected r^2 was calculated as $(r^2 - 1/N)/(1 - 1/N)$, where N is twice the number of individuals (Hill and Robertson, 1968).

Effective population size

The r^2 values combined with the distance between markers can be used to estimate the approximate effective population size (N_t) at a given point in time (t). The N_t was determined on the basis of r^2 values at different distances and assuming a model without mutation, as described by Sved (1971) and rearranged by Corbin *et al.* (2010), $N_t = (1/4c) \times (1/r^2 - 1)$, in which r^2 is the mean value of LD at a given distance and c the distance between SNPs in Morgans (assuming 1 Mb = 0.01 Morgans). Each genetic distance c corresponds to a value

of t generations in the past, and this value was calculated as $t = 1/(2c)$. The N_t was estimated at nine-time points from 250 to five generations ago. Moreover, the contemporary effective population size (N_e) was estimated using N_EESTIMATOR v. 2 (Do *et al.*, 2014) according to the random mating model of the LD method. We used a method described by Waples and Do (2008) to add a bias correction into the original LD method and a threshold of 0.05 as the lowest allele frequency to derive the least biased results. We reported our estimates with ±95% confidence intervals.

Runs of homozygosity detection

Runs of homozygosity were defined using PLINK with the following criteria: (i) the minimum length that constituted an ROH was set to 1 Mb to ensure the exclusion of very short and common ROH that occur prevalently throughout the genome due to LD; (ii) two missing SNPs were allowed in the ROH; (iii) one heterozygous SNP was allowed; (iv) minimum density was set at 1 SNP for every 100 kb; and (v) the maximum gap between consecutive SNPs was 1 Mb. To minimize the number of false-positive findings, the minimum number of SNPs that constituted each ROH (l) was calculated by a method similar to that proposed by Lencz *et al.* (2007)

$$l = \frac{\log_e \frac{\alpha}{n_s \cdot n_i}}{\log_e (1 - \text{het})}$$

where n_s is the number of SNPs per individual, n_i the number of individuals, α the percentage of false-positive ROH (set to 0.05 in the present study) and het the heterozygosity across all SNPs. Runs of homozygosity were estimated for each individual and then categorized on the basis of ROH length (1 to 5, 5 to 10, 10 to 15, 15 to 20 >20 Mb). The mean number of ROH per individual (M_{NROH}), the average length of ROH (L_{MROH}) and the sum of all ROH segments per animal were estimated. Moreover, the percentage of chromosomes covered by ROH was also calculated.

Genetic diversity and genomic inbreeding coefficients

PLINK was used to estimate basic genetic diversity indices, including observed and expected heterozygosity (H_o and H_e, respectively) and MAF (≥0.05).

The inbreeding coefficient (F) on the basis of ROH (F_{ROH}) for each animal was calculated as follows

$$F_{\text{ROH}} = \frac{L_{\text{ROH}}}{L_{\text{aut}}}$$

where L_{ROH} is the total length of all ROH in the genome of an individual, L_{aut} the specified length of the autosomal genome covered by SNPs on the chip (2452.06 Mb). Alternative estimates of inbreeding and coancestry coefficients were also calculated as follows: (i) the genomic inbreeding coefficient, on the basis of difference between the observed and expected numbers of homozygous genotypes (F_{HOM}) using PLINK; (ii) the molecular coancestry coefficient (f_{ij}) between individuals i and j (Caballero and Toro, 2002); and (iii) the molecular inbreeding coefficient (F_i) of individual i, calculated as $F_i = 2\ f_{ii} - 1$ (f_{ii} is the molecular self-coancestry).

Analysis of genetic relationships and population structure among breeds

Genotypes from 10 other sheep breeds were considered in these analyses to clarify the genetic relationships, levels of gene flow and the admixtures within and between them. Data of six breeds were taken from the International Sheep Genomics Consortium (Kijas *et al.*, 2012) (Castellana $n = 23$, Chios $n = 23$, Lacaune $n = 64$, Leccese $n = 24$, Merinos $n = 50$ and Sarda Black $n = 20$); data of Sicilian breeds (Comisana $n = 48$, Pinzirita $n = 70$, Valle del Belice $n = 47$) were taken from Mastrangelo *et al.* (2014); and samples of Sarda White ($n = 30$) were genotyped by the Dipartimento Scienze Agrarie e Forestali, University of Palermo.

PLINK software was used to calculate pairwise identical-by-state (IBS) distances between breeds, graphically represented by multidimensional scaling (MDS) analysis. The graphical representation was depicted using statistical R software (R Development Core Team, 2013) with the RColorBrewer package. The extents of all population substructures were evaluated through the model-based clustering algorithm implemented in ADMIXTURE (Alexander and Lange, 2011). The most probable number of populations in the data set (K) was estimated using the default (fivefold) cross-validation procedure in ADMIXTURE, by which estimated prediction errors are obtained for each K value by adopting a kind of 'leave one out' approach. The K value that minimizes the estimated prediction errors is then assumed to be the most suitable. Graphical representations were visualized using R software (R Development Core Team, 2013). GENEPOP software (Raymond and Rousset, 1995) was used to estimate population relatedness using pairwise estimates of F_{ST} among all breeds. Phylogenetic relationships among populations were also explored using Reynolds genetic distances. Neighbor networks were constructed from the estimated genetic distances using SPLITSTREE (Huson and Bryant, 2006).

To detect common ancestry and/or gene flow among populations, we investigated the extent of pairwise haplotype sharing between Barbaresca and the other sheep breeds reared in Sicily (Valle del Belice, Comisana, Pinzirita and Sarda White). The approach is based on the estimated Pearson correlation of signed r values for the same pair of SNPs in two breeds as measures of LD (de Roos *et al.*, 2008). The correlation of r between breeds was estimated for SNPs separated by <10, 10 to 25, 25 to 50, 50 to 100, 100 to 150 and 150 to 250 kb.

Results

Distribution of single nucleotide polymorphisms

Of a total of 54 241 genotyped SNPs, 378 were unmapped and 1450 were located on sex chromosomes. Thus, 52 413 SNPs mapped onto 26 sheep autosomes were used; after filtering, the final number of individuals and the number of SNPs retained for analyses were 36 and 41 857, respectively. The major point for SNP exclusion was MAF, for which 10 102 SNPs were rejected. The distribution of SNPs per chromosome (OAR) is reported in the Supplementary Table S1 and ranged from 593 on OAR24 to 4656 on OAR1.

Linkage disequilibrium

To examine the decay of LD with physical distance, SNP pairs on autosomes were sorted into bins on the basis of their inter-marker distance; then, average values of r^2 were calculated for each bin and plotted as a function of genomic distance between markers. Levels of pairwise LD decreased with increasing genomic distance between SNPs, and the most rapid decline was seen over the first 0.05 Mb (Figure 1). Short-range LD was observed in Barbaresca sheep. Overall autosomes, the average r^2 values were 0.241, 0.133 and 0.062 for SNPs up to 50 kb, SNPs separated by 200 to 500 kb and SNPs separated by >5000 kb, respectively (Table 1). The extent of LD was also evaluated for each adjacent SNP pair per chromosome (Supplementary Table S1). The mean average r^2 for all pairwise adjacent SNP combinations pooled overall autosomes was 0.215 ± 0.018, with an average distance between adjacent SNP pairs of 63.93 kb. The r^2 ranged from 0.189 ± 0.231 for OAR24 to 0.259 ± 0.286 for OAR10. When the decay of r^2 with distance was plotted separately for each chromosome, the highest value was observed in OAR6 ($r^2 = 0.270$) for the 0- to 50-kb inter-marker distance; when moving from 50 to 500 kb, the chromosome with the highest mean values for r^2 was OAR10 ($r^2 = 0.150$).

Effective population size

A graphical representation of the approximate N_t values at each time point t, from 250 to five generations ago, is shown in Figure 2. Moreover, to enable comparison, the N_t values for other Sicilian sheep breeds (Mastrangelo *et al.*, 2014) are also shown. The estimated N_t for Barbaresca ranged from 747 animals 250 generations ago to 28 animals five generations ago. Among the breeds examined, Barbaresca showed the lowest values. The contemporary effective population size (N_e) was 25.

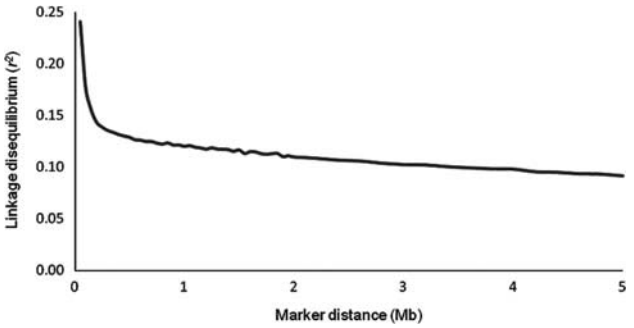

Figure 1 Distribution of r^2 between syntenic single nucleotide polymorphism pairs across the genome as a function of physical distance (Mb) in Barbaresca breed.

Table 1 *Average linkage disequilibrium measured as the squared correlation coefficient of allele frequencies at a pair of loci (r^2) and corrected r^2 (Hill and Robertson, 1968) among syntenic single nucleotide polymorphisms over different map distances (kb) in Barbaresca sheep breed*

Distance (kb)	r^2	Corrected r^2
<50	0.241	0.231
50 to 100	0.177	0.167
100 to 200	0.149	0.138
200 to 500	0.133	0.122
500 to 1000	0.123	0.112
1000 to 2000	0.116	0.104
2000 to 5000	0.100	0.089
>5000	0.062	0.050

Detection of runs of homozygosity

A total of 637 ROH were identified, and all individuals displayed at least six ROH. The mean number of ROH per individual was 15.9 and the average length of ROH was 10.03 Mb. There were considerable differences among individuals in the number of ROH and the length of the genome covered by ROH segments (Figure 3). The three animals with the highest level of

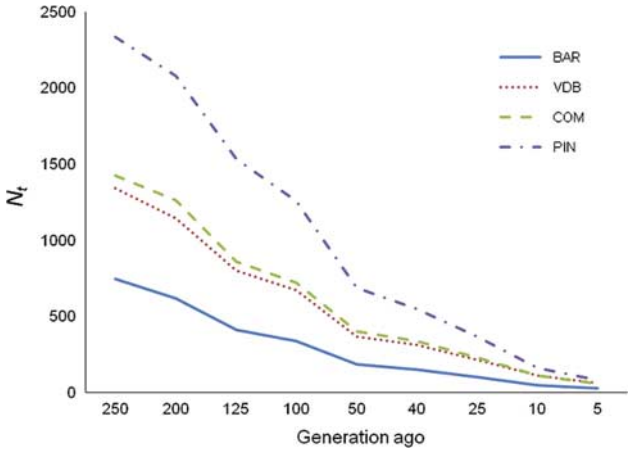

Figure 2 Past effective population size (N_t) over the past generations based on linkage disequilibrium from 26 autosomes in Barbaresca (BAR) and other Sicilian sheep breeds (COM = Comisana; PIN = Pinzirita; VDB = Valle del Belice).

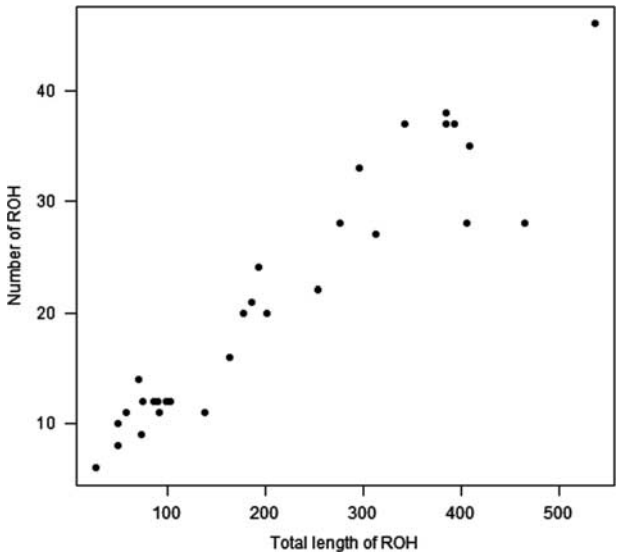

Figure 3 Relationship between the total number of runs of homozygosity (ROH) > 1 Mb and the total length (Mb) of genome in such ROH for individuals.

homozygosity belonged to different farms and showed 408.2, 464.5 and 536.7 Mb of their genome classified as ROH, respectively, representing close to 20% of the total genome length. Most of the ROH were short (67%) and ranged from 1 to 10 Mb (Supplementary Table S2). The number of ROH was greater in the first three chromosomes and tended to decrease with chromosome length. The maximum size of a ROH was 59.4 Mb with 956 SNPs in OAR1. The largest percentage of the genome in ROH was observed on OAR11 (at 25.3%), whereas the lowest coverage was observed on OAR6 (7.9%).

Genetic diversity indices and genomic inbreeding coefficients

Genetic diversity indices and inbreeding and coancestry coefficients, which were estimated using different approaches and which are key parameters in the genetic management of populations, were used to determine the levels of genetic variability in the breed (Table 2). The results revealed that Barbaresca tended to display lower variability than other Sicilian breeds. In fact, Mastrangelo *et al.* (2014) reported higher values of MAF (from 0.290 to 0.301), N_e (from 369 to 685) and gene diversity (H_e) (from 0.379 to 0.390), and lower inbreeding (from 0.016 to 0.055) in these breeds.

Analysis of genetic relationship and population structure among breeds

Genotyping data from Barbaresca samples were merged with those of other sheep breeds. The whole data set consisted of 435 individuals from 11 populations genotyped for 42 351 SNPs spread over all autosomal chromosomes.

We used an MDS plot of the pairwise IBS distance to compare the Barbaresca individuals with the other populations. The results showed that most sheep breeds formed non-overlapping clusters and were clearly separate populations (Figure 4a), with the breeds separated according to their geographic origin. In particular, the first dimension (C1) distinguished Barbaresca from other breeds. Moreover, the Barbaresca individuals showed more spread clusters, which is typical of breeds that show admixture with other breeds. To explore in detail the relatedness, an MDS plot was also created for the Sicilian and Sarda White sheep breeds (Figure 4b). The results indicated isolation and greater genetic distance among Barbaresca and the other breeds reared in Sicily and an absence of gene flow among them.

Results from within the population substructure, according to admixture analysis and considering a range of 2 through 20 potential clusters (K), indicated that the most probable number of inferred populations was $K = 9$. A graphic representation of the estimated membership coefficients for the nine clusters is shown in Figure 5, where model-based clustering partitions the genome of each sample into a predefined number of components. The first breed to be differentiated from the others was Barbaresca ($K = 2$). These results corroborate the findings on the basis of MDS plot for all breeds. At $K = 9$, each breed tends to have its

Table 2 *Estimates of genetic diversity indices and genomic inbreeding coefficients of Barbaresca breed*

MAF ± SD	H_o ± SD	H_e ± SD	N_e	$F_{ROH>1\,Mb}$ ± SD	F_{HOM} ± SD	F_i ± SD	f_{ij} ± SD
0.282 ± 0.131	0.394 ± 0.149	0.371 ± 0.122	25	0.087 ± 0.060	0.062 ± 0.067	0.606 ± 0.009	0.629 ± 0.010

MAF = average minor allele frequency; H_o = observed heterozygosity; H_e = expected heterozygosity; N_e = effective population size; $F_{ROH>1\,Mb}$ = inbreeding coefficient based on runs of homozygosity; F_{HOM} = inbreeding coefficient on the basis of the difference between observed *v.* expected number of homozygous genotypes; F_i = molecular inbreeding coefficient of individual *i*; f_{ij} = molecular coancestry coefficient between individuals *i* and *j*.

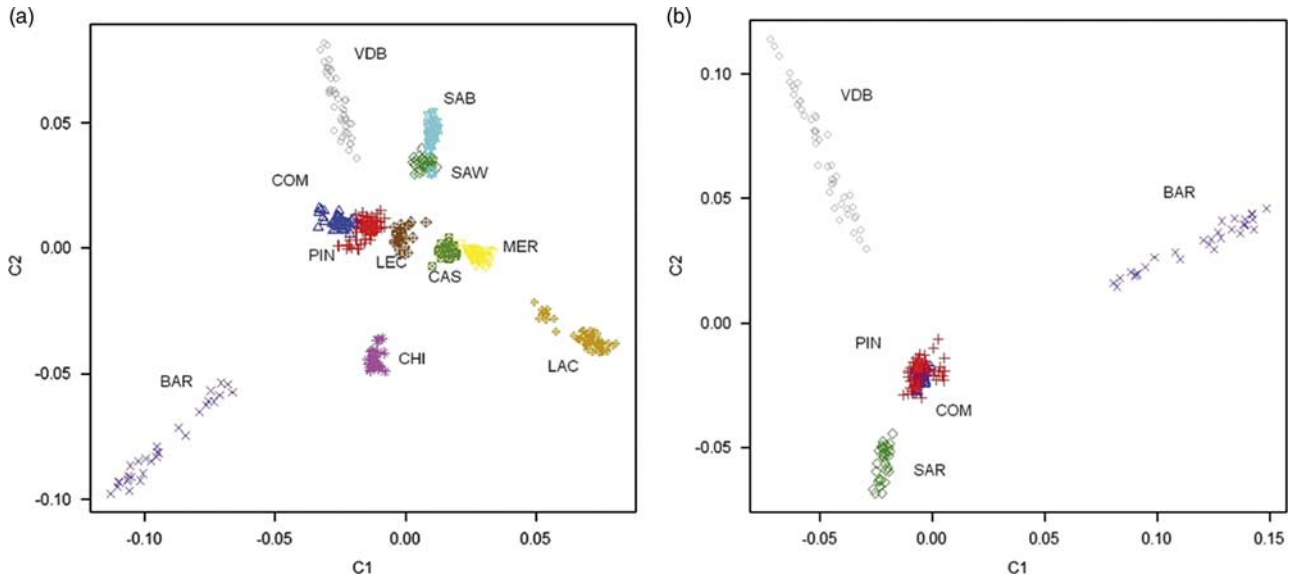

Figure 4 (a) Genetic relationship defined with multidimensional scaling analysis among 11 sheep breeds and (b) among Barbaresca and the sheep breeds reared in Sicily. BAR = Barbaresca; CAS = Castellana; CHI = Chios; COM = Comisana; LAC = Lacaune; LEC = Leccese; MER = Merinos; PIN = Pinzirita; SAB = Sarda Black; SAR = Sarda; SAW = Sarda White; VDB = Valle del Belice.

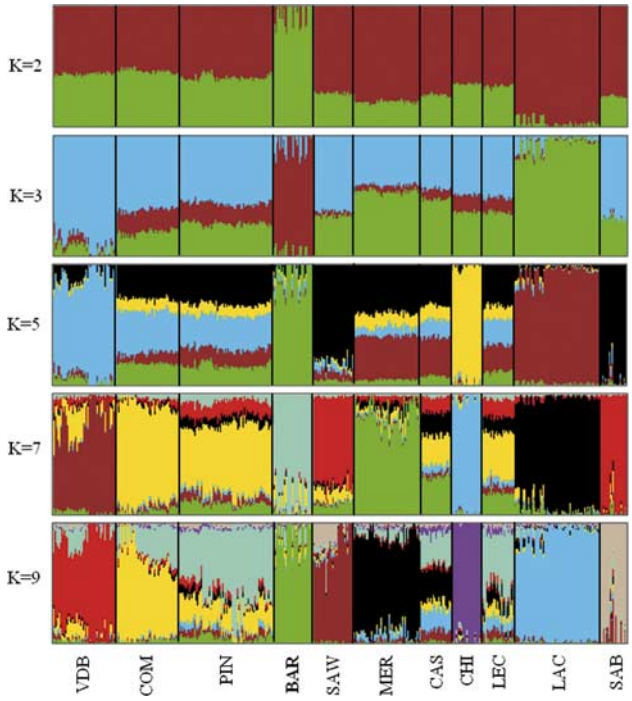

Figure 5 Model-based clustering of 11 sheep breeds analyzed in each of the inferred clusters (*K*), from $K = 2$ to $K = 9$. BAR = Barbaresca; CAS = Castellana; CHI = Chios; COM = Comisana; LAC = Lacaune; LEC = Leccese; MER = Merinos; PIN = Pinzirita; SAB = Sarda Black; SAW = Sarda White; VDB = Valle del Belice.

own distinct cluster, but some differences persist. In fact, some breeds showed a complex admixture-like pattern, notably, the Sicilian breeds (Comisana and Pinzirita), Leccese and Castellana.

The degree of genetic differentiation between pairs of breeds is reported in the Supplementary Table S3. With respect to pairwise F_{ST} among all populations, Barbaresca was the most divergent breed. The highest F_{ST} value was seen between Barbaresca and Chios ($F_{ST} = 0.145$) and the lowest value was for Comisana *v.* Pinzirita ($F_{ST} = 0.021$). To provide additional insight into the relationships and the origin of Barbaresca, we constructed a neighbor-net graph based on Reynolds genetic distances (Supplementary Figure S1) between pairs of breeds. The Pinzirita showed the lowest value for genetic distance from Barbaresca (Supplementary Table S3). In agreement with the MDS plot, the graph showed some clear clusters and relationships between breeds that originated from the same region or area, notably, the Sicilian, Sardinian and European ones. Barbaresca branched between Chios and Valle del Belice. The shortest branch was observed for Pinzirita, whereas the longest one was found for Barbaresca and Chios.

The Pearson correlations for the signed *r* values among Sicilian breed pairs are shown in Supplementary Table S4. In particular, the table reports the correlations for pairs of SNPs separated by <10 kb and by 150 to 250 kb. Similar to r^2, *r* decreased as the distance between markers increased. The haplotype sharing analysis (0 to 10 kb), which presumably reflects the oldest historical relationship between breeds, showed that Barbaresca had a consistently high degree of haplotype sharing with Pinzirita and a low degree of haplotype sharing with Sarda White. Among all breeds, the highest level of haplotype sharing was found between Pinzirita and Valle del Belice. As expected due to their history, haplotype sharing by the Sicilian dairy sheep breeds was by far the highest. The same trends were reported for pairs of SNPs separated by 150 to 250 kb.

Discussion

The advent of high-throughput genotyping arrays has greatly facilitated the study of genetic structure in livestock species, but whereas much effort has been devoted to uncovering genetic differences and population structure within dominant commercial breeds, local breeds, which are less widely used, are generally understudied (Beynon *et al.*, 2015). This study provides the first overview of population structure estimates of the Barbaresca sheep from a genome-wide perspective.

The average LD estimates in Barbaresca breed were quite variable. The variation for the average r^2 between chromosomes (Supplementary Table S1) was probably attributable to variations in recombination rates between and within chromosomes, genetic drift and selection (Qanbari *et al.*, 2010). A typical LD pattern was observed, with higher LD for markers close to each other that quickly decayed with increasing inter-marker distance. Our findings can be compared directly with results from other studies conducted using a similar LD measure and SNP panel. For example, García-Gámez *et al.* (2012), in Spanish Churra sheep, reported an average r^2 of 0.061 for SNPs separated by 200 to 500 kb. Mastrangelo *et al.* (2014) reported an average r^2 for adjacent SNP pairs of 0.155, 0.156 and 0.128 in the Valle del Belice, Comisana and Pinzirita breeds, respectively. These studies of LD in dairy sheep breeds showed lower values than those seen for Barbaresca, indicating that LD decay was breed-specific. Recently, Kijas *et al.* (2014), in a study of five sheep breeds that used a high-density SNP chip, found an average r^2 ranging from 0.080 to 0.224 for SNPs up to 70 kb apart and reported that LD decayed faster and at a much shorter genomic distance. Consistent with our results, Al-Mamun *et al.* (2015) reported that the maximum average LD was located between adjacent SNPs on OAR10 and a higher average LD was seen in meat sheep breeds. Values of r^2 for adjacent SNP pairs were also similar to the values observed in other livestock species such as cattle ($r^2 = 0.20$ for SNPs with an average distance of 68 Kb) (Bohmanova *et al.*, 2010) and goats ($r^2 = 0.11$ to 0.29 for SNPs with physical distance ranging from 53 to 73 kb) (Brito *et al.*, 2015), although these were lower than those observed in pigs, where one study reported an average r^2 of 0.41 to 0.46 for SNPs <50 kb apart (Veroneze *et al.*, 2013), and in horses (Corbin *et al.*, 2010), which showed r^2 values remaining >0.3 for distances up to 185 kb. The relatively small number of genotyped Barbaresca individuals may have resulted in an

overestimation of r^2, which could explain the differences in the LD values *v.* the other Sicilian sheep breeds (Mastrangelo *et al.*, 2014). The same consideration was reported by Veroneze *et al.* (2013) to explain the differences in the r^2 values in pig breeds. However, it should be emphasized that several studies have reported LD analysis in sheep (Kijas *et al.*, 2012; Ciani *et al.*, 2013b) in which, despite a higher number of breeds, the number of individuals per breed was similar to those involved in our study. To address this concern, a corrected r^2 was calculated; the differences between estimated and corrected r^2 values were small (about 0.01 units; Table 1 and Supplementary Table S1). Therefore, we decided to present the non-corrected values in the main text. The decay of LD in a genome determines the power of quantitative trait loci detection in association mapping studies and helps to determine the number of markers required for successful association mapping and genomic selection. In fact, species with extensive LD will require lower marker density to capture most of the genetic variation than species with low levels of LD. Qanbari *et al.* (2010) considered an r^2 threshold of 0.25 as a useful LD value for association studies. Corbin *et al.* (2010) reported that an average r^2 above 0.3 should be employed for genome-wide association study. An average r^2 below 0.25 was found in the Barbaresca breed for markers located up to 50 kb apart, whereas when the average distance between adjacent SNP pairs was 63 kb, the average r^2 was 0.215. Our results support the need to use denser SNP panels to ensure high power association mapping in future breeding programs for this local sheep breed. However, the relatively low levels of LD detected in Barbaresca might be explained by an effect of ascertainment bias, given that this breed was not used to set up the Illumina OvineSNP50 BeadChip.

The r^2 values, combined with marker distances, were used to infer approximate N_t at any given point t in the past. This important parameter helps predict rates of loss of neutral genetic variation, fixation of deleterious and favorable alleles, and any increase in inbreeding experienced by a population (England *et al.*, 2006). As pointed out in Hayes *et al.* (2003), LD at small distances reflects N_t in the distant past, whereas LD at large distances reflects N_t in the recent past. In fact, the higher r^2 observed in Barbaresca could be the result of a smaller N_t in the more recent past compared with other Sicilian sheep breeds (Mastrangelo *et al.*, 2014) (Figure 2). The N_t values have declined rapidly over time. The estimate of N_t 50 generations ago was ~183 individuals. Consistent with our results, Ciani *et al.* (2013b), in a genome-wide analysis of Italian sheep breeds, reported a low past N_e value of 224 in Altamurana, which, like Barbaresca, is an endangered breed that has experienced a dramatic census contraction. Kijas *et al.* (2012), in a genome-wide analysis of the world's sheep, reported high previous N_e for many breeds, with 25 breeds exceeding 500 individuals; only six breeds (Wiltshire, Soay, Dorset Horn, Valais Red, East Friesian Brown and Sakiz) showed evidence of a comparatively narrow genetic base (<200 individuals) like that displayed in Barbaresca. It should be emphasized that the estimates of N_e might be strongly biased when the sample size is small, probably as a result of the LD generated by the sampling process. England *et al.* (2006) showed that N_e estimates from the most commonly used LD-based estimator will be substantially biased with small sample sizes, unless the true N_e is smaller than the sample size used to estimate it. Actually, none of the formulas proposed to date for estimating past N_e from r^2 provides reliable predictions (Corbin *et al.*, 2012), probably because they all rely on simplifications or assumptions. For example, in all models, estimates of N_e were highly sensitive to thresholds imposed upon MAF and to *a priori* assumptions about the expected r^2 for adjacent markers (Corbin *et al.*, 2012). To infer N_e in the present study, we also used a method described by Waples and Do (2008), which is considered to correct for bias in estimates of N_e based on LD data. The N_e reported for Barbaresca was very low (25). This could have resulted from population bottlenecks caused by the geographic isolation of some farms and reduced interest from breeders. Recently, Al-Mamun *et al.* (2015), in a study of Australian sheep on the basis of SNP data and using the same method, showed higher values of N_e for all investigated breeds that ranged from 140 to 348. Generally, high selection pressure and the use of artificial insemination are the main reasons for the low N_e, but in local breeds, where selection is absent and uncontrolled mating of related animals is common, inbreeding and low genetic diversity, as observed in Barbaresca, are a consequence of the low N_e. As already mentioned, the reduction of N_e in this breed could also have been caused by reduced interest from breeders. In the previous years, selection programs have strongly emphasized production traits; this has led to increased specialization for traits such as milk yield and quality, meat, and wool (Tolone *et al.*, 2012). This has led to greater reliance on a small number of sheep breeds and a decline in the genetic diversity of livestock. Frankham *et al.* (2014) recommended that short-term and long-term N_e should be equal to 100 and 1000, respectively, individuals to control the inbreeding rate and maintain the evolutionary potential of the population. Therefore, if we consider 100 as the minimum acceptable N_e to conserve a population, the estimate for Barbaresca is below the critical value, meaning that the viability of the population may be compromised.

According to the ROH analysis, evidence of recent inbreeding was strong in this breed. In fact, although most of the ROH detected in Barbaresca were in the shorter length categories (1 to 10 Mb), as also reported in studies of cattle populations (Mastrangelo *et al.*, 2016), the individuals showed a total length of ROH characterized by the presence of large segments (23% of ROH > 10 Mb). In general, estimates of ROH are rare in local breeds, and to the best of our knowledge, few studies have been conducted in sheep. However, similar results were reported by Al-Mamun *et al.* (2015) for the mean number and length categories of ROH. F_{ROH} provided a good measure of individual genome-wide autozygosity, and compared with the alternative estimates of inbreeding and coancestry coefficients, it allows researchers to distinguish between recent and more distant inbreeding (McQuillan *et al.*, 2008). Estimates of inbreeding coefficients depend on the methods used. In fact, F coefficients estimated using allele frequencies (F_{HOM}) showed

considerable variation within the breed with respect to the other estimates (F_{ROH} and F_i) and had the highest CV. Molecular inbreeding (F_j) and molecular coancestry (f_{jj}) coefficient values were much higher than the other coefficients because these two methods (which are obtained on a SNP-by-SNP basis) do not discriminate between alleles that are identical by descent (IBD) or IBS (Rodríguez-Ramilo *et al*., 2015). Moreover, the strong correlation between the pedigree inbreeding coefficient and the sum of ROH reported by several authors (Purfield *et al*., 2012) indicates that the estimates of F_{ROH} are a good reflection of IBD and suggests that, in the absence of an animal's pedigree data, as may be the case for local endangered breeds, the extent of a genome under ROH may be used to infer aspects of recent population history, even from relatively few samples. In agreement with the low N_e, relatively higher genomic F coefficients and less gene diversity (H_e) were found in Barbaresca than in other Sicilian sheep breeds (Mastrangelo *et al*., 2014). All values were consistent with the range observed by other authors in other endangered sheep breeds (Kijas *et al*., 2012; Ciani *et al*., 2013b). The results were expected given the decline in population size during the past 30 years and confirmed the findings of a previous study conducted on the genetic structure of Sicilian sheep breeds based on microsatellite markers (Tolone *et al*., 2012) that reported the lowest values of genetic diversity for Barbaresca. In sum, all the investigated parameters indicated an endangered status and confirmed the threat of extinction for this local sheep breed.

Determination of genomic structure can be valuable in prioritizing populations for conservation and for developing suitable management practices. The results indicated that Barbaresca was the most distant group and emphasized the clear genetic differences *v.* the other breeds considered in this study. In fact, the MDS clearly separated Barbaresca from the other breeds, and this was in agreement with neighbor-networks, F_{ST}, and model-based clustering ($K = 2$). The breed showed a low level of admixture with other breeds, indicating that there is less genetic data remaining from any other ancestral breed that may have interacted with it, which may be considered as a typical signal of inbreeding (Ciani *et al*., 2015). Recently, Falush *et al*. (2016) reported a simulation with human data where one of the populations was admixed and just one of the ancestral populations of the mixture was sampled, and the pattern reported by ADMIXTURE for $K = 2$ was the same as the one found in our breed. The authors showed that if admixture events or genetic drift affect all members of the sample equally, the ADMIXTURE algorithms simply shifts the allele frequencies of the inferred ancestral population to reflect the fraction of admixture that is shared by all individuals and this can result in misinterpretation of the true admixture history, particularly when applied to data sets where there is little prior knowledge on the relationships between groups. Therefore, the results of ADMIXTURE and the MDS plot, where the Barbaresca showed a diagonal pattern against the center of the plot, could be misinterpreted and the amount of admixture may be not small at all. Moreover, considering the Barbaresca as an admixed breed, the estimation of F_{ST} can be biased because computed between a pair of breeds of which one of them is inbred with a distant breed. Consequently, high values of F_{ST} will not necessary mean that the breed has high inbreeding, but also that one of the ancestral populations is genetically far away from the group (Peter, 2016). It should be mentioned the possibility of admixture with other populations not investigated in this study. In fact, it may be that we found no admixture because the used sample was imperfect and because there were not sheep breeds from North Africa included (Barbarian breed) in the studied populations, but just European ones. Therefore, despite there were several results that showed the high inbreeding of the Barbaresca breed such as ROH analysis, it is important to underline that some of the statistics used in this study can be biased because the Barbaresca is an admixed breed.

The long branch observed for Barbaresca was typical of differentiated and isolated populations with small effective population sizes (Kijas *et al*., 2012). To evaluate the degree of haplotype sharing between breeds, the extent of LD was characterized by signed *r*. In a previous study of sheep breeds (Kijas *et al*., 2012), a high degree of conservation of LD was observed between the Sicilian breeds for SNP pairs separated by 10 kb or less. The results of haplotype sharing analysis between the Barbaresca and Pinzirita breeds, for SNP pairs separated by 150 to 250 kb, indicated that the phase relationship among alleles was similar in both populations. High correlation of *r* between breeds within long intervals implies that they share the same long haplotypes (de Roos *et al*., 2008). The results were consistent with their relationships and evolutionary history. In fact, the Barbaresca breed seems to derive from crosses between the Tunisian Barbary from North Africa and Pinzirita, along with posterior selection for growth performance (Tolone *et al*., 2012). These results suggest a common ancestry between these breeds. In fact, although the F_{ST} values indicated very strong differentiation between Barbaresca and other breeds, the lowest F_{ST} was found between Barbaresca and Pinzirita (0.071). Moreover, a hypothesis of coancestry between these two breeds was also supported by the lowest Reynolds genetic distance. However, there are robust procedures which reveal key information about ancestral relationships among individuals (Lawson *et al*., 2012; Maples *et al*., 2013) that, if needed, can be used to estimate the proportions of admixture of the two breeds with high accuracy.

Another interesting result was found for the Pinzirita. This breed always displays the lowest Reynolds genetic distance with all breeds and seems to share some ancestry with most of them (Figure 5). This would indicate that Pinzirita is ancestral to other Italian breeds.

The consistency of the results among the several approaches used (MDS, Bayesian clustering, F_{ST}, neighbor networks and haplotype sharing) suggests that our conclusions are robust.

The population structure and the low genetic diversity presented here should be useful in creating conservation strategies. The term conservation management implies a compromise between control of global genetic diversity, avoidance of high inbreeding levels and maintenance of a certain degree of differentiation (Fernández *et al*., 2008). The studied breed showed a defined genetic structure, a relatively lower

genetic diversity (0.371) and contemporary effective population size (25), and a higher level of inbreeding (0.062) compared with other Sicilian sheep breeds. If reproductive isolation within groups of the same flock and area is maintained in subsequent generations, the short-term rate of inbreeding will increase dangerously, resulting in a decrease in the effective population size and seriously damaging the breed's future. Thus, efforts should be made to improve genetic diversity in this breed. In particular, mating decisions will play an important role in limiting inbreeding and will increase the size of this breed. However, since the two ancestral populations (Tunisian Barbary and Pinzirita) still exist, the genetic heritage of the Barbaresca breed could perhaps be reconstructed, although the blood levels of the two breeds in Barbaresca are unknown.

Additional factors should be taken into account when defining conservation priorities, including economic and ecological aspects. Local populations play an important role in the utilization of marginal agricultural resources and contribute to environmental and socio-economic stability, but local farmers struggle because they cannot take advantage of economies of scale in breeding and marketing programs. Many typical high-quality dairy (ricotta and pecorino) and meat products can be obtained from the Barbaresca breed, and their use represents an important chance of survival for this breed. Consequently, another important aspect of conservation is recognition of the special value of these typical products. To ensure profitability, this marketing link between Barbaresca and its products should be defended against fraudsters who include milk or meat from other populations.

There are numerous potential uses for genomic information that make routine genotyping desirable for the management of small populations. However, whereas our study describes an example of the Barbaresca breed, the applied genome-wide analyses are valid for all small and endangered breeds. Therefore, we recommend the continuous collection of genotypes and their use in breed monitoring and improvement, particularly for local breeds.

Acknowledgments

The authors would like to thank four anonymous referees for valuable comments, which helped to improve the manuscript.

This research was financed by PON02_00451_3133441, CUP: B61C1200076005 funded by MIUR.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731116002780

References

Alexander DH and Lange K 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12, 246.

Al-Mamun HA, Clark SA, Kwan P and Gondro C 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genetics Selection Evolution 47, 1–14.

ASSONAPA 2014. Associazione Nazionale della Pastorizia. Available from http://www.assonapa.it/norme_ecc/Consistenze_Caprini.htm.

Barrett JC, Fry B, Maller J and Daly MJ 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.

Beynon SE, Slavov GT, Farré M, Sunduimijid B, Waddams K, Davies B, Haresign W, Kijas J, MacLeod IM, Newbold CJ, Davies L and Larkin DM 2015. Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. BMC Genetics 16, 65.

Bohmanova J, Sargolzaei M and Schenkel FS 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics 11, 421.

Brito LF, Jafarikia M, Grossi DA, Kijas JW, Porto-Neto LR, Ventura RV, Salgorzaei M and Schenkel FS 2015. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. BMC Genetics 16, 67.

Caballero A and Toro MA 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conservation Genetics 3, 289–299.

Ciani E, Ciampolini R, D'Andrea M, Castellana E, Cecchi F, Incoronato C, D'Angelo F, Albenzio M, Pilla F, Matassino D and Cianci D 2013a. Analysis of genetic variability within and among Italian sheep breeds reveals population stratification and suggests the presence of a phylogeographic gradient. Small Ruminant Research 112, 21–27.

Ciani E, Crepaldi P, Nicoloso L, Lasagna E, Sarti FM, Moioli B, Napolitano F, Carta A, Usai G, D'Andrea M, Marletta D, Ciampolini R, Riggio V, Occidente M, Matassino D, Kompan D, Modesto P, Macciotta N, Ajmone-Marsan P and Pilla F 2013b. Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. Animal Genetics 45, 256–266.

Ciani E, Lasagna E, D'Andrea M, Alloggio I, Marroni F, Ceccobelli S, Delgado Bermejo JV, Sarti FM, Kijas J, Lenstra JA and Pilla F, International Sheep Genomics Consortium 2015. Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. Genetics Selection Evolution 47, 1–12.

Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC and Woolliams JA 2010. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. Animal Genetics 41, 8–15.

Corbin LJ, Liu AYH, Bishop SC and Woolliams JA 2012. Estimation of historical effective population size using linkage disequilibria with marker data. Journal of Animal Breeding and Genetics 129, 257–270.

de Roos APW, Hayes BJ, Spelman RJ and Goddard ME 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus Cattle. Genetics 179, 1503–1512.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ and Ovenden JR 2014. NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size Ne from genetic data. Molecular Ecology Resources 14, 209–214.

England PR, Cornuet JM, Berthier P, Tallmon DA and Luikart G 2006. Estimating effective population size from linkage disequilibrium: severe bias in small samples. Conservation Genetics 2, 303–308.

Falush D, van Dorp L and Lawson D 2016. A tutorial on how (not) to over-interpret structure/admixture bar plots. bioRxiv, 066431.

Fernández J, Toro MA and Caballero A 2008. Management of subdivided populations in conservation programs: development of a novel dynamic system. Genetics 179, 683–692.

Fernández J, Meuwissen THE, Toro MA and Mäki-Tanila A 2011. Management of genetic diversity in small farm animal populations. Animal 5, 1684–1698.

Frankham R, Bradshaw CJA and Brook BW 2014. Genetics in conservation and management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation 170, 56–63.

García-Gámez E, Sahana G, Gutiérrez-Gil B and Arranz JJ 2012. Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC Genetics 13, 43.

Gibson J, Morton N and Collins A 2006. Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics 15, 789–795.

Hayes BJ, Visscher PM, McPartlan HC and Goddard ME 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research 13, 635–643.

Hill WG and Robertson A 1968. Linkage disequilibrium in finite populations. Theoretical Applied Genetics 38, 226–231.

Huson DH and Bryant D 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology Evolution 23, 254–267.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto-Neto LR, Cristobal MS, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J and Dalrymple B, International Sheep Genomics Consortium 2012. Genome-Wide Analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLos Biology 10, e1001258.

Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, Hayes BJ, Brauning R and McEwan J, International Sheep Genomics Consortium 2014. Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal Genetics 45, 754–757.

Lawson DJ, Hellenthal G, Myers S and Falush D 2012. Inference of population structure using dense haplotype data. PLoS Genetics 8, e1002453.

Lencz T, Lambert C, De Rosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R and Malhotra AK 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proceedings of the National Academy of Sciences 104, 19942–19947.

Maples BK, Gravel S, Kenny EE and Bustamante CD 2013. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. The American Journal of Human Genetics 93, 278–288.

Mastrangelo S, Di Gerlando R, Tolone M, Tortorici L, Sardina MT and Portolano B, International Sheep Genomics Consortium 2014. Genome wide linkage disequilibrium and genetic structure in sicilian dairy sheep breeds. BMC Genetics 15, 108.

Mastrangelo S, Tolone M, Di Gerlando R, Fontanesi L, Sardina MT and Portolano B 2016. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. Animal 10, 746–754.

McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H and Wilson JF 2008. Runs of homozygosity in European populations. The American Journal of Human Genetics 83, 359–372.

Peter BM 2016. Admixture, population structure, and F-statistics. Genetics 202, 1485–1501.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–575.

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

Qanbari S, Pimentel EC, Tetens J, Thaller G, Lichtner P, Sharifi AR and Simianer H 2010. The pattern of linkage disequilibrium in German Holstein cattle. Animal Genetics 41, 346–356.

R Development Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raymond M and Rousset F 1995. GENEPOP population genetics software for exact tests and ecumenicism. Journal of Heredity 86, 248–249.

Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D and Villanueva B 2015. Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. PLoS One 10, 4.

Sved JA 1971. Linkage disequilibrium of chromosome segments. Theoretical Population Biology 141, 125–141.

Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME and Visscher PM 2007. Recent human effective population size estimated from linkage disequilibrium. Genome Research 17, 520–526.

Tolone M, Mastrangelo S, Rosa AJM and Portolano B 2012. Genetic diversity and population structure of Sicilian sheep breeds using microsatellite markers. Small Ruminant Research 102, 18–25.

Veroneze R, Lopes PS, Guimaraes SEF, Silva FF, Lopes MS, Harlizius B and Knol EF 2013. Linkage disequilibrium and haplotype block structure in six commercial pig lines. Journal of Animal Science 91, 3493–3501.

Wang JL 2005. Estimation of effective population sizes from data on genetic markers. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 1395–1409.

Waples RS and Do C 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8, 753–756.