Volume 102, Issue 1, January 2012 ISSN 0921-4488

SMALL RUMINANT RESEARCH

The Official Journal of the International Goat Association

Genetic diversity and population structure of Sicilian sheep breeds using microsatellite markers

M. Tolone [a,b,∗,1], S. Mastrangelo [a,1], A.J.M. Rosa [c], B. Portolano [a]

[a] *Dipartimento DEMETRA, Sezione Produzioni Animali, Facoltà di Agraria, Università degli Studi di Palermo, Viale delle Scienze – Parco d'Orleans, 90128 Palermo, Italy*
[b] *Consorzio Regionale di Ricerca Bioevoluzione Sicilia, c/o DEMETRA – Sezione di Produzioni Animali, Facoltà di Agraria, Università degli Studi di Palermo, Viale delle Scienze – Parco d'Orleans, 90128 Palermo, Italy*
[c] *Brazilian Agricultural Research Corporation, Ministry of Agriculture, Livestock and Food Supply, Planaltina 73310-970, DF, Brazil*

ARTICLE INFO

ABSTRACT

Genetic diversity studies in domestic animals aim at evaluating genetic variation within and across breeds mainly for conservation purposes. In Sicily, dairy sheep production represents an important resource for hilly and mountain areas economy. Their milk is used for the production of traditional raw milk cheeses, sometimes protected designation of origin (PDO) cheeses. In some cases, the quality of these products is linked to a specific breed, i.e. mono-breed labelled cheeses and it is therefore important to be able to distinguish the milk of a breed from that of others, in order to guarantee both the consumer and the breed itself. In order to investigate the genetic structure and to perform an assignment test, a total of 331 individuals (Barbaresca, BAR $n = 57$, Comisana, COM $n = 65$, Pinzirita, PIN $n = 75$, Sarda, SAR $n = 64$, and Valle del Belice, VDB $n = 70$) were analysed using a panel of 20 microsatellite markers. A total of 259 alleles were observed with average polymorphic information content equal to 0.76, showing that the microsatellites panel used was highly informative. Estimates of observed heterozygosity ranged from 0.65 in the BAR breed to 0.75 in the COM breed. The low value of genetic differentiation among breeds ($F_{st} = 0.049$) may indicate that these breeds are little differentiated probably due to common history and breeding practices. The low F_{is} and F_{it} values indicated low level of inbreeding within and among breeds. The unrooted neighbor-joining dendrogram obtained from the Reynold's genetic distances, and factorial correspondence analysis revealed a separation between BAR and the other sheep breeds. Recent migration rates were estimated, showing that four out of the five breeds have not received a significant proportion of migrants. Only for the PIN breed a recent introgression rate from the VDB breed (7.2%) was observed. The Bayesian assignment test showed that BAR and SAR breeds had a more definite genetic structure (proportion of assignment of 92% and 86.6%, respectively), whereas the lowest assignment value was found in the PIN breed (67.1%). Our results indicated high genetic variability, low inbreeding and low genetic differentiation, except for BAR breed, and were in accordance with geographical location, history, and breeding practices. The low robustness of the assignment test makes it unfeasible for traceability purposes, due to the high level of admixture, in particular for COM, PIN and VDB.

∗ Corresponding author at: Dipartimento DEMETRA, Sezione di Produzioni Animali, University of Palermo, Viale delle Scienze – Parco d'Orleans, 90128 Palermo, Italy. Tel.: +39 091 23896070; fax: +39 091 23860814.
*E-mail address:* marco.tolone@unipa.it (M. Tolone).
[1] Equal contributors.

1. Introduction

In the past years, selection programs have mainly put high emphasis on production traits, which led to an increased specialization for traits such as milk yield and quality, meat, wool, etc. This happened sometimes even by crossbreeding the local breeds with exotic ones, to generate populations with the desired phenotypes. This hybridization process, however, has resulted in an increased reliance on a small number of breeds to meet the local's food requirements, which could lead to the disappearance of local breeds. Nevertheless, this aspect has received greater interest in the last years, based on the awareness that indigenous and locally developed sheep breeds are an important asset, because of the unique combinations of adaptive traits developed to respond effectively to the pressures of the local environment (Buduram, 2004). From these considerations and given the importance of the local genetic resources, it is easily understandable the considerable interest given nowadays to genetic diversity studies in domestic animals in general and, recently, in small ruminants (Baumung et al., 2004). Genetic diversity studies in domestic animals aim at evaluating genetic variation within and across breeds, since the breed is the management unit for which factors such as inbreeding are controlled (Tapio et al., 2005). However, the definition of a breed, as applied by FAO, frequently does not reflect the underlying genetic population structure. Therefore, a molecular genetics study of the population diversity and structure may improve the understanding of the actual genetic resources.

In Sicily, dairy sheep production represents an important resource for hilly and mountain areas economy, in which other economic activities are limited (Scintu and Piredda, 2007). Their milk is mainly used for the production of traditional raw milk cheeses, sometimes protected designation of origin (PDO) cheeses as laid down in the European Union legislation. In some cases, the quality of these products is linked to a specific breed, i.e. mono-breed labelled cheeses and it is therefore important to be able to distinguish the milk of a breed from that of others, in order to guarantee both the consumer and the breed itself. Assignment of individuals to a specific breed, especially when the phenotypic differentiation between breeds is difficult, is therefore of great importance both for biodiversity purposes and dairy products traceability.

Nowadays only four native sheep breeds are reared in Sicily: Barbaresca (BAR), Comisana (COM), Pinzirita (PIN), and Valle del Belice (VDB). These breeds present differences in both morphology and production traits and show excellent adaptability to the local environments. In particular, BAR and PIN breeds, due to their good adaptive traits and hardiness are raised on farms located in marginal areas, representing therefore an important genetic resource for present and future needs. In addition to these autochthonous breeds, the Sarda (SAR) breed, reared mainly in Sardinia, was considered due to its likely contribution to the origin of the VDB breed (Portolano, 1987). Based on historical, geographical, and morphological information, it is likely that the VDB breed derives from the PIN breed, to which is similar for the horned trait in the males,
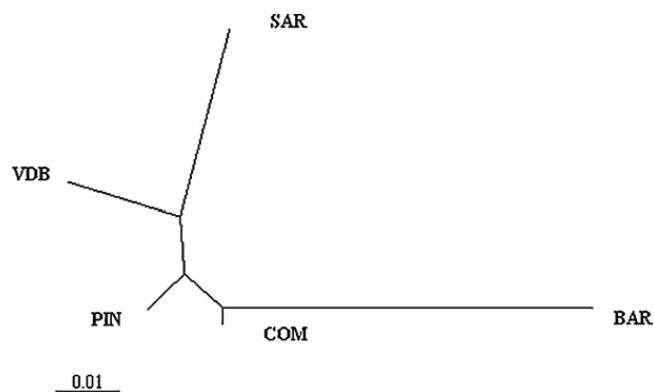
Fig. 1. Unrooted neighbor-joining dendrogram showing the genetic relationships among sheep breeds using Reynolds genetic distance: BAR = Barbaresca; COM = Comisana; PIN = Pinzirita; SAR = Sarda; VDB = Valle del Belice.

crossed with the COM breed, to which is similar for coat colour (i.e. white with red head) and milk production. Subsequently, the cross between these two breeds was likely crossed with the SAR breed (Portolano, 1987). In addition, the BAR breed derives from crosses between Tunisian Barbary breed from North Africa and the PIN breed, and posterior selection for growth performance (Sarti et al., 2002).

In recent years, several microsatellite studies on sheep genetic diversity, population structure, genetic differentiation, and phylogenetic reconstruction aiming at identifying endangered populations as well as developing genetic conservation strategies have been published (Dalvit et al., 2008; Ligda et al., 2009; Arora et al., 2011). Aim of this study was to explore the genetic structure of the four Sicilian autochthonous sheep and one Italian breed, through the analysis of the genetic diversity within and across breeds, and determine their genetic relationship, using microsatellite markers. Moreover, microsatellite markers were used for the proper assignment of an individual to a specific breed.

2. Materials and methods

2.1. Sampling and DNA extraction

A total of 331 blood samples were used for the analysis, 267 of which belonging to the four Sicilian breeds (BAR ($n = 57$), COM ($n = 65$), PIN ($n = 75$) and VDB ($n = 70$)); the remaining 64 samples for the SAR breed were provided by the AGRIS Sardegna. All samples were collected from randomly chosen unrelated individuals, in several farms of different areas of the two Islands, in order to get samples as representative as possible.

Genomic DNA was extracted from blood buffy coats of nucleated cells using a salting out method (Miller et al., 1988).

2.2. Microsatellites amplifications and analysis

A total of 20 microsatellite markers (Table S1) were selected as suggested by ISAG (http://www.isag.org. uk/Docs/2005_PanelsMarkersSheepGoats.

Table 1
Total number of alleles per locus (TNA), number of observed alleles per locus and per breed, and polymorphism information content (PIC).

Marker	TNA	BAR	COM	PIN	SAR	VDB	PIC
OarFCB128	7	5	6	6	7	7	0.68
IDVGA45	13	3	8	11	10	12	0.79
BM827B	8	6	7	7	6	7	0.73
ILSTS011	8	5	6	7	6	8	0.72
DU323541	17	7	11	17	10	12	0.86
OarCP49	24	10	16	18	17	13	0.87
INRA063A	17	8	14	15	13	14	0.83
SPS115	8	5	6	8	7	6	0.74
LSCV36B	14	7	10	13	12	12	0.84
MAF209	12	6	10	12	8	10	0.76
BRN	10	5	7	8	7	8	0.58
McM527	10	6	8	9	6	10	0.80
ILSTS005	11	6	8	9	10	9	0.63
TCRBV6	15	5	12	11	11	11	0.75
DU223896	17	7	8	14	10	11	0.82
APPO10	7	3	6	7	7	7	0.70
INRA132	15	7	11	13	12	11	0.84
CSRD247	16	7	12	12	10	10	0.80
DU194351	17	5	10	12	9	9	0.50
DU216028	13	9	12	13	11	12	0.89
Overall mean	12.95	6.15	9.40	11.05	9.45	9.95	0.76
s.d.	4.32	1.74	2.87	3.43	2.82	2.26	0.10

BAR = Barbaresca; COM = Comisana; PIN = Pinzirita; SAR = Sarda; VDB = Valle del Belice.

pdf) and FAO (http://www.fao.org/dad-is/). Moreover, some loci (i.e., DU323541, DU223896 and DU194351) were selected based on location and DNA sequence information available at NCBI website (http://www.ncbi.nlm.nih.gov/genome/guide/sheep/).

Genotypes for all 20 microsatellite markers were determined by means of four multiplex fluorescent PCR reactions and fragment lengths determined in a single semi-automated multiplex electrophoresis run by using an A3130 Genetic Analyzer and GeneMapper version 4.0 with recommended protocols (Applied Biosystems). Each reaction was performed in a total volume of 20 µl containing 50 ng template DNA, 1× Qiagen Multiplex PCR Master Mix, 1× PCR Master Mix, primer mix, and nuclease-free water. For the four multiplex reactions, the PCR program was: initial denaturation at 95 °C for 15 min; 32 cycles of 95 °C for 45 s, 58 °C for 1 min 50 s, and 72 °C for 1 min 20 s; and final extension at 60 °C for 30 min. Amplification was carried out using the GeneAmp PCR system 9700. A total of 3.7 µl of Multiplex PCR was mixed with 0.3 µl of LIZ 500 Size Standard and 6.0 µl of Hi-Di Formamide. The multiplex PCR/LIZ/Formamide mixture was denatured for 5 min at 95 °C and then immediately snap cooled on ice for 3 min before performing a capillary electrophoresis on an ABI 3130 Sequencer (Applied Biosystems).

2.3. Statistical analysis

Total number of observed alleles per locus (TNA), the mean number of alleles observed in a population over all loci genotyped (MNA) and its standard deviation, allelic richness, a measure of the number of alleles independent of sample size, (AR), observed and expected heterozygosity (H_o and H_s, respectively) per locus and overall loci in the whole sample (Nei, 1987), and Wright's fixation index (F_{is}, F_{it} and F_{st}) (Weir and Cockerham, 1984) were estimated using FSTAT 2.9.3 software (Goudet, 1995).

Moreover, FSTAT was used to estimate observed and expected heterozygosity (H_o and H_e respectively) within breed. Polymorphism information content (PIC) per locus was estimated with Cervus 2.0 software (Marshall et al., 1998) and the deviation from Hardy–Weinberg equilibrium (heterozygote deficiency) with the GENEPOP package version 4.0.11 (Raymond and Rousset, 1995) using a Markov Chain method (dememorization 10,000, batches 100, and iterations per batch 5000). In addition, GENEPOP software was used to assess genotypic linkage disequilibrium (LD) between each pair of loci in the whole sample. Nei's minimum distance (Nei, 1987) and Reynold's distance (Reynolds et al., 1983), both recommended for populations with short divergence time (Eding and Laval, 1999), were used to estimate pair-wise genetic distances among breeds. The Reynold's genetic distance was used to reconstruct a neighbor-joining consensus tree using the Phylip package version 3.69 (Felsentein, 2009), and the dendrogram was depicted using the software TreeView version 1.6.6 (Page, 1996). Tree robustness was evaluated by bootstrapping over loci (1000 replicates). Moreover, a factorial correspondence analysis (FCA) was performed based on the individual multilocus genotype using the GENETIX version 4.03 (Belkhir et al., 1996). In order to investigate the occurrence of recent migration among the breeds considered, the program BayesAss+ (Wilson and Rannala, 2003) was used, with 3×10^6 iterations, a burn-in of 999,999, and thin interval of 2000. This program simultaneously estimates the probability distribution of allelic frequencies for each locus, migration rates among populations (m) and inbreeding coefficient for each population (F). STRUCTURE version 2.3.1 (Pritchard et al., 2000) was used to analyse the genetic structure and identify the true number of populations (clusters) and assign the individuals to each cluster. The program estimates the natural logarithm of the probability that a given genotype (G) is part of a given population (K). The model used was based on an assumption of admixture

and correlated allele frequencies as suggested by several authors (Pritchard et al., 2000; Vicente et al., 2008; Zuccaro et al., 2008). The $\ln \Pr(G|K)$ was calculated for K ranging from 1 to 9, without prior information on the breed of origin, to estimate the most likely number of clusters in the dataset, with 50 independent runs for each K. All runs consisted of a burn-in period of 100,000 steps, followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations.

3. Results

In total, 259 alleles were observed for the 20 loci surveyed. The TNA per locus ranged from seven for Oar-FCB128 and APPO10 to 24 for OarCP49 with an average of 12.95 ± 4.32 (Table 1). The PIC considering all loci was equal to 0.76 ± 0.10, showing that the microsatellites panel used was highly informative. DU194351 was found to be the least informative marker (0.50), whereas DU216028 the most informative one (0.89) (Table 1). The number of alleles per locus observed in each breed ranged from 3 to 18. The BAR breed showed the lowest number of observed alleles (3) and the lowest overall mean number of alleles (6.15 ± 1.74). The highest number of observed alleles per locus was found in the PIN breed (18) with a mean number of alleles equal to 11.05 ± 3.43 (Table 1). Private alleles, i.e. alleles unique for a single population, were evidenced in all breeds with low frequency (<3%). However, two private alleles of the microsatellite INRA132 (161 and 179 bp alleles) were found with higher frequency (6.2% and 10.2%, respectively) in the SAR breed.

Hardy–Weinberg equilibrium was consistently rejected for DU223896 in all breeds (Table S2) and this marker was not further considered. The average observed (H_o) and expected (H_s) heterozygosities and the overall genetic diversity (H_t) for the 19 microsatellites are given in Table S3. Considering the whole sample, H_o ranged from 0.430 to 0.876, H_s from 0.504 to 0.866, and H_t from 0.517 to 0.899. DU194351 marker showed the lowest H_s (0.504), whereas DU216028 the highest one (0.866). The H_o was always higher than 0.50, except for the marker DU194351. However, up to seven markers were found to be not in Hardy–Weinberg equilibrium across breeds; in fact for these markers the observed heterozygosity was lower than the expected one. Heterozygote deficiency (F_{is}) analysis revealed that all five breeds exhibited deviations from HWE at some loci (Table S3). Considering the whole sample, the number of markers showing a significant deviation from Hardy–Weinberg equilibrium ranged from 2 (BAR) to 7 (COM) (Table S2). Significant linkage disequilibrium ($P < 0.001$) was found between BM827B and DU323541, INRA063A and LSCV36B, and OarFCB128 and DU216028.

The MNA and AR for each breed are showed in Table 2. MNA ranged between 6.05 ± 1.78 and 10.95 ± 3.45 for BAR and PIN breeds, respectively, whereas AR ranged between 5.66 ± 1.57 and 9.44 ± 2.84 for the same breed.

Table S3 also reports the population differentiation examined by fixation indices for each locus and across all loci and the mean estimates of F-statistic obtained were: $F_{it} = 0.080$, $F_{is} = 0.032$ and $F_{st} = 0.049$.

An overview of the genetic diversity parameters for each breed is given in Table 2. Estimates of H_o ranged

Table 2
Mean number of alleles (MNA), allelic richness (AR), observed (H_o) and expected (H_e) heterozygosity and standard deviation (s.d.).

Breed	MNA ± s.d.	AR ± s.d.	H_o ± s.d.	H_e ± s.d.
BAR	6.05 ± 1.78	5.66 ± 1.57	0.65 ± 0.165	0.67 ± 0.094
COM	9.47 ± 2.93	8.73 ± 2.71	0.75 ± 0.131	0.77 ± 0.095
PIN	10.95 ± 3.45	9.44 ± 2.84	0.72 ± 0.141	0.78 ± 0.092
SAR	9.42 ± 2.89	8.39 ± 2.32	0.71 ± 0.108	0.75 ± 0.102
VDB	9.89 ± 2.31	8.64 ± 1.95	0.71 ± 0.123	0.77 ± 0.097

BAR = Barbaresca; COM = Comisana; PIN = Pinzirita; SAR = Sarda; VDB = Valle del Belice.

Table 3
Reynolds (above the diagonal) and Nei-minimum (below the diagonal) genetic distance per pair of breeds.

Breed	BAR	COM	PIN	SAR	VDB
BAR	–	0.077	0.075	0.110	0.095
COM	0.061	–	0.025	0.057	0.036
PIN	0.058	0.020	–	0.050	0.036
SAR	0.087	0.045	0.040	–	0.054
VDB	0.075	0.028	0.028	0.043	–

BAR = Barbaresca; COM = Comisana; PIN = Pinzirita; SAR = Sarda; VDB = Valle del Belice.

from 0.65 in BAR to 0.75 in COM, whereas estimates of H_e from 0.67 in BAR to 0.78 in PIN. Table 3 shows the Nei and Reynolds genetic distance estimates. The lowest values were observed between COM and PIN breeds for both Reynolds (0.025) and Nei-minimum (0.020), whereas the highest ones between BAR and SAR breeds (0.110 and 0.087 for Reynolds and Nei-minimum, respectively). The BAR breed showed the highest genetic distance in relation to the other four breeds. The Reynold's genetic distances were also used to reconstruct the unrooted neighbor-joining dendrogram (Fig. 1), showing two clear clusters: VDB/SAR and PIN/COM, confirming that the BAR breed appears again to be more distant from the other breeds.

The factorial correspondence analysis was performed including all breeds and loci using the corresponding allele frequencies (Fig. 2). The first three components explained the 86.46% of the total variation, 38.52% of which explained by Axis 1 that clearly separates the BAR breed from the other breeds; 28.49% explained by Axis 2 that separates the SAR breed; and 19.45% explained by Axis 3 that separates the VDB breed.

Estimates of the recent migration rates (m) (up to the second generation of migrants) and inbreeding coefficients per breed are presented in Table 4. Values on the diagonal are defined as the proportion of individuals in each

Table 4
Means of the posterior distribution of the migration rates and means and standard deviations (s.d.) of the posterior distribution of the inbreeding coefficient (F) for each population.

	BAR	COM	PIN	SAR	VDB	F ± s.d.
BAR	0.9848	0.0239	0.0031	0.0024	0.0025	0.03 ± 0.015
COM	0.0023	0.9392	0.0221	0.0185	0.0051	0.03 ± 0.013
PIN	0.0053	0.0156	0.8991	0.0031	0.0266	0.08 ± 0.017
SAR	0.0053	0.0062	0.0036	0.9720	0.0110	0.04 ± 0.014
VDB	0.0024	0.0151	0.0721	0.0040	0.9548	0.07 ± 0.014

BAR = Barbaresca; COM = Comisana; PIN = Pinzirita; SAR = Sarda; VDB = Valle del Belice.
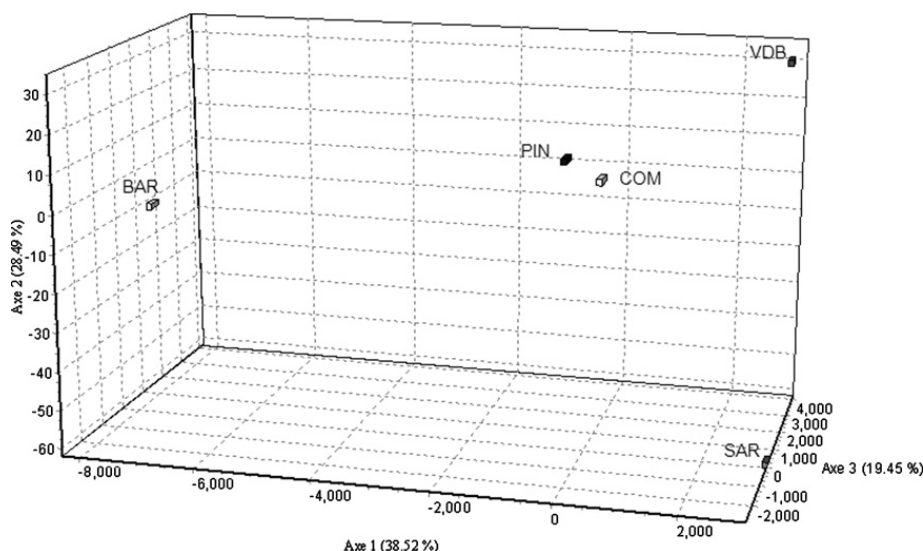
Fig. 2. Spatial representation of the breeds as defined by the factorial correspondence analysis: BAR = Barbaresca; COM = Comisana; PIN = Pinzirita; SAR = Sarda; VDB = Valle del Belice.

Table 5
Number of individuals (N) per breed and proportion of membership of each breed in each of the 5 clusters inferred in the most likely run of the program STRUCTURE.

	Inferred clusters					
Breed	1	2	3	4	5	n
BAR	0.031	0.016	0.014	0.020	0.919	57
COM	0.094	0.054	0.769	0.027	0.056	65
PIN	0.675	0.124	0.151	0.027	0.023	75
SAR	0.042	0.023	0.059	0.867	0.009	64
VDB	0.100	0.761	0.074	0.049	0.015	70

BAR = Barbaresca; COM = Comisana; PIN = Pinzirita; SAR = Sarda; VDB = Valle del Belice.

generation that are not migrants. All breeds but PIN have not received a significant proportion of migrants. Nevertheless, recent introgression was observed for the PIN breed from the VDB breed at a rate of 7.2%. The posterior distribution of inbreeding coefficients ranged from 0.03 (COM) to 0.08 (PIN).

Assignment test was performed using the program STRUCTURE with the number of expected population (K) ranging from 1 to 9. The Ln Pr($G|K$) increased from $K = 2$ to $K = 5$, reached a "plateau" at $K = 5$, while did not show a significant fluctuation from $K = 5$ to $K = 8$ and then decreased for $K = 9$ (Fig. S1). Therefore it was assumed that $K = 5$ is the most likely number of clusters. For $K = 2$ the BAR separates from other breeds, while for $K = 3$ it is the SAR that appears isolated and the Sicilian dairy breeds remained in the same cluster; at $K = 4$ the VDB separates from the other Sicilian dairy breeds (COM and PIN) and, finally, for $K = 5$ each breed tends to have their own distinct cluster (Fig. 3).

Table 5 shows the assignment proportion of each breed to the five most likely clusters inferred, choosing the iteration with the minimum variance. Clusters five and four included the BAR and the SAR individuals with 91.9% and 86.7%, respectively, showing a significant proportion of assignment for these breeds; cluster three and two included the COM and the VDB individuals with a proportion of assignment of 76.9% and 76.1%, respectively. The

lowest value of assignment was showed in the PIN breed with a proportion of assignment of 67.5% in the cluster one.

4. Discussion

In this study we investigated the genetic variability detected through microsatellite markers in the four native sheep breeds reared in Sicily and in the SAR breed. The majority of the markers were highly polymorphic and generally in Hardy–Weinberg equilibrium except for the marker DU223896 that showed the largest difference between observed and expected heterozygosity. The significant deviation observed for the marker DU223896 may be explained by unobserved null alleles leading
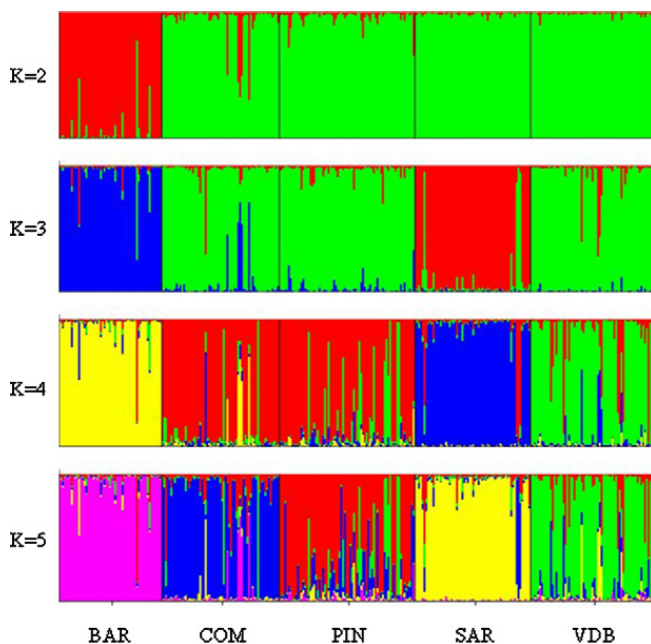
Fig. 3. Estimated population structure of the 5 sheep breeds for K ranging from 2 to 5: Barbaresca (BAR), Comisana (COM), Pinzirita (PIN), Sarda (SAR) and Valle del Belice (VDB).

to high within-breed F_{is} values, ranging from 0.3139 to 0.6415. Genetic linkage disequilibrium was found for some microsatellite markers and it can be due to a variety of factors, including physical linkage, epistatic selection, and genetic hitchhiking (Maudet et al., 2002). However, given that these markers have been mapped to different chromosomes, physical linkage was excluded.

Overall mean heterozygosity reflects a notably high variability, characteristic of microsatellites derived from a greater mutation level than other genetic markers, which makes them a valuable tool for genetic diversity analyses (Arranz et al., 2001). The high mean heterozygosity values could be attributed to low inbreeding levels found, low selection pressure, gene flow among Sicilian sheep breeds, and the large number of alleles present in all breeds. Another parameter indicative of the genetic variation is the PIC estimate, which was higher than 0.50 and therefore highly informative (Botstein et al., 1980). Estimates of observed heterozygosity over all loci confirmed the remarkable level of genetic variability in these breeds.

Most of the breeds considered in this study have never been genetically characterized before; therefore, it was not always possible to compare our results with what is reported in literature. Previous studies presented similar H_e estimates for the SAR breed (Pariset et al., 2003; Lawson Handley et al., 2007), but lower H_o (0.60 and 0.66, respectively); whereas our H_e and H_o estimates for the COM breed were higher than those reported by Lawson Handley et al. (2007). However, it has to be considered that differences in the estimates can be also attributed to the different number of markers and/or sampling scheme used. The similarities among breeds for the average number of alleles across loci and H_e, except for BAR, can be explained by the phylogenetic relationships; in fact as reported above, COM, PIN, SAR, and VDB seem to be genetically related among them. The lower number of alleles observed in the BAR breed is probably due to a reduced effective population size. This breed is, indeed, reared in a very restricted area of Sicily and nowadays about 1200 animals are enrolled in the herd book (ASSONAPA, 2010). The number of private alleles found in the SAR breed was probably due to geographic isolation, i.e. determining a reduced or absent gene flow with the Sicilian breeds.

The low F_{is} (within population inbreeding estimate) and F_{it} (total inbreeding estimate) values indicate low inbreeding level within and among breeds. The F_{st} (measurement of population differentiation) value may indicate that these breeds are not differentiated enough and that they may have a common history and breeding practices. To confirm this hypothesis another analysis was performed, by removing the SAR breed from the whole sample and F_{st} value decreased from 0.049 to 0.046, underlying the low differentiation among these breeds. Thus, a large part of the total genetic diversity can be explained by the variation within breeds (0.951) and to a smaller extent by the variation among breeds (0.049). Lack of differentiation between the Sicilian breeds could be due to geographic proximity, similarities in environment and breeding practices, but most likely due to past gene flow among them. These estimates of genetic differentiation are comparable to those reported by other authors for indigenous Ethiopian

(F_{st} = 0.046, Gizaw et al., 2007), Portuguese (F_{st} = 0.049, Santos-Silva et al., 2008), and Alpine (F_{st} = 0.057, Dalvit et al., 2008) sheep breeds and/or populations, but lower than those reported by Baumung et al. (2006) in Austrian sheep breeds and Arora et al. (2011) in Indian sheep breeds (0.080 and 0.111, respectively). Moreover, Lawson Handley et al. (2007) in a study on European sheep breeds, including the COM and the SAR, reported that the Southern breeds are characterized by higher within-breed diversity, lower genetic differentiation, and higher level of heterozygosity than Northern breeds.

In PIN and VDB breeds the higher genetic diversity (H_e and average number of alleles per locus), but at the same time the highest coefficient of inbreeding suggested that whereas PIN and VDB have a wide genetic base, individuals belonging to these two breeds were more inbred than those in the other breeds. The rather high level of genetic variability in PIN and VDB breeds is not surprising, as these are the most numerous sheep breeds reared in Sicily. The highest level of inbreeding could be affected by the farming system. In Sicily rams are reared together with ewes, therefore mating with close relatives can be quite frequent. Moreover, the reduced or absent exchange of rams between different flocks of the same breed may have induced the development of a substructure within the populations. The results of genetic distances are in agreement with a previous study conducted on the genetic structure and relationship between the VDB breed and the other sheep breeds considered as ancestors, using the genetic polymorphisms of seven protein systems (Di Stasio et al., 1992). The unrooted neighbor-joining dendrogram showed two clear clusters in accordance with the phylogenetic relationships among breeds. Moreover, the FCA analysis showed a relationship between COM and PIN breeds. The genetic closeness between these two breeds might be explained considering that these breeds are characterized by a common breeding system and geographical husbandry area, which might have led to genetic exchange between them. This seems to be partially confirmed by the structure analysis, showing that COM and PIN breeds shared a cluster.

The FCA results corroborate the findings based on the Nei and Reynolds genetic distances, which also indicated the isolation and a greater genetic distance of the BAR breed. In fact, the least variable breeds are usually the most distinct ones (Hedrick, 1999). It is important to highlight that this breed was mainly reared for meat production, unlike the other breeds reared exclusively for milk production; therefore, there was less gene exchange between the former and the other breeds.

We investigated population structure by varying K from 1 to 9. Although the analysis showed the highest probability of forming eight clusters, $K = 5$ was chosen as the best value to describe the genetic structure of our breeds, since the increase of $\mathrm{Ln}\,Pr(G|K)$ from $K = 5$ to $K = 8$ was low if compared with the increase from $K = 1$ to $K = 5$. However, as suggested by Pritchard et al. (2000), the inferred clusters are not necessarily the corresponding real ancestral population and they can be determined by sampling schemes. Assuming $K = 3$ the BAR and the SAR formed two distinct groups suggesting that admixture was nearly zero for these two breeds. The high average percentage of assignment of

individuals for BAR and SAR breeds pointed out the existence of clear genetic differences compared to other breeds and this result is also confirmed by the geographic distribution of these two breeds as well as the different breeding system (i.e., meat production) for the former. The COM, PIN and VDB breeds exhibited the presence of admixture, in fact the COM and the PIN clustered together up to $K = 4$, confirming the genetic closeness between these breeds. For $K = 5$, COM, PIN and VDB breeds formed three distinct clusters but with proportion of membership split in two or more clusters. This result could be probably due to the phylogenetic relationships among these breeds and/or to the migration of individuals among the several farms present in the area.

5. Conclusions

This study allowed exploring the genetic structure of the Sicilian autochthonous sheep breeds. Our results indicated that significant amounts of genetic variation are still maintained in these sheep breeds. The results are in accordance with their geographical location, origin, and breeding practices. The identification of genetic relationship and gene flow among livestock breeds/populations is important for breeders and conservationists. In order to maintain the existing genetic diversity, breeding strategies (i.e. oriented seedstock exchange) aiming at maintaining effective population size, minimizing inbreeding and genetic drift should be implemented for the different breeds, especially for the BAR that could be considered the most differentiated and endangered breed. In the case of BAR, besides the strategies suggested before, a conservation program is recommended and an *in vitro* (e.g. germplasm bank) and *in situ* (e.g. conservation population) approaches should be considered for future purposes. The robustness of assignment test depends on the genetic differentiation but, unfortunately, it is not the case for COM, PIN and VDB, and resulted in an unsatisfactory assignment performance and thus it cannot be used for traceability purposes.

Acknowledgement

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.smallrumres.2011.09.010.

References

Arora, R., Bhatia, S., Mishra, B.P., Joshi, B.K., 2011. Population structure in Indian sheep ascertained using microsatellite information. Anim. Genet. 42, 242–250.

Arranz, J.J., Bayon, Y., Primitivo, F.S., 2001. Genetic variation at microsatellite loci in Spanish sheep. Small Rumin. Res. 39, 3–10.

ASSONAPA, 2010. http://www.assonapa.it/Consistenze/.

Baumung, R., Simianer, H., Hoffmann, I., 2004. Genetic diversity studies in farm animals: a survey. J. Anim. Breed. Genet. 121, 361–373.

Baumung, R., Cubric-Curik, V., Schwend, K., Achmann, R., Solkner, J., 2006. Genetic characterisation and breed assignment in Austrian sheep breeds using microsatellite markers information. J. Anim. Breed. Genet. 123, 265–271.

Belkhir, K., Borsa, P., Goudet, J., Chikhi, L., Bonhomme, F., 1996. GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations. In: Laboratoire Génome Populations, Interactions CNRS UMR 5000. Université de Montpellier II, Montpellier, France.

Botstein, D., White, R.L., Skolnick, M., Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32, 324–331.

Buduram, P., 2004. Genetic characterization of Southern African sheep breeds using DNA narkers. In: Dissertation. University of the Free State, Bloemfontein, South Africa.

Dalvit, C., Sacca, E., Cassandro, M., Gervaso, M., Pastore, E., Piasentier, E., 2008. Genetic diversity and variability in Alpine sheep breeds. Small Rumin. Res. 80, 45–51.

Di Stasio, L., Rasero, R., Giaccone, P., Fiandra, P., 1992. Valle del Belice sheep: genetic structure and relationship with other sheep populations reared in Sicily. Agric. Medit. 122, 66–69.

Eding, J.H., Laval, G., 1999. Measuring genetic uniqueness in livestock. In: Oldenbroek, J.K. (Ed.), Genebanks and the Conservation of Farm Animal Genetic Resources. Institute for Animal Science and Health, Lelystad, pp. 33–58.

Felsentein, J., 2009. Phylogeny Inference Package PHYLIP, Version 3.69. Department of Genome Sciences and Department of Biology, University of Washington, Seattle, WA, USA.

Gizaw, S., Van Arendonk, J.A.M., Komen, H., Winding, J.J., Hanotte, O., 2007. Population structure, genetic variation and morphological diversity in indigenous sheep of Ethiopia. Anim. Genet. 38, 621–628.

Goudet, J., 1995. FSTAT (version 2.9.3): a computer programmeto calculate *F*-statistics. J. Hered. 8, 485–486.

Hedrick, P.W., 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53, 313–318.

Lawson Handley, L.J., Byrne, K., Santucci, F., Townsend, S., Taylor, M., Bruford, M.W., Hewitt, G.M., 2007. Genetic structure of European sheep breeds. Heredity 99, 620–631.

Ligda, C., Altarayrah, J., Georgoudis, A., 2009. Genetic analysis of Greek sheep breeds using microsatellite markers for setting conservation priorities. Small Rumin. Res. 83, 42–48.

Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity influence in natural populations. Mol. Ecol. 7, 639–655.

Maudet, C., Luikart, G., Taberlet, P., 2002. Genetic diversity and assignment among seven French cattle breeds based on microsatellite DNA analysis. J. Anim. Sci. 80, 942–950.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, USA.

Pariset, L., Savarese, M.C., Cappuccio, I., Valentini, A., 2003. Use of microsatellites for genetic variation and inbreeding analysis in Sarda sheep flocks of central Italy. J. Anim. Breed. Genet. 120, 425–432.

Page, R.D.M., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Portolano, N., 1987. La pecora della Valle Del Belice. In: Pecore e Capre Italiane. Edagricole, Bologna, Italia, pp. 117–124.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Raymond, M., Rousset, F., 1995. GENEPOP (version 4.0.11): population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.

Reynolds, J., Weir, B.S., Cockerham, C., 1983. Estimation of the coancestry coefficient basis for a short term genetic distance. Genetics 105, 767–779.

Santos-Silva, F., Ivo, R.S., Sousa, M.C.O., Carolino, M.I., Ginja, C., Gama, L.T., 2008. Assessing genetic diversity and differentiation in Portuguese coarse-wool sheep breeds with microsatellite markers. Small Rumin. Res. 78, 32–40.

Sarti, D.M., Lasagna, F., Panella, M., Pauselli, F., Sarti, F.M., 2002. Pecore e capre italiane. Edagricole, Bologna, Italia, p. 7.

Scintu, M.F., Piredda, G., 2007. Typicity and biodiversity of goat and sheep milk products. Small Rumin. Res. 68, 221–231.

Tapio, I., Tapio, M., Grislis, Z., Holm, L.E., Jeppsson, S., Kantanen, J., Miceikiene, I., Olsaker, I., Viinalass, H., Eythorsdottir, E., 2005. Unfolding of population structure in Baltic sheep breeds using microsatellite analysis. Heredity 94, 448–456.

Vicente, A.A., Carolino, M.I., Sousa, M.C., Ginja, C., Silva, F.S., Martinez, A.M., Vega-Pla, J.L., Carolino, N., Gama, L.T., 2008. Genetic diversity in native and commercial breeds of pigs in Portugal assessed by microsatellites. J. Anim. Sci. 86, 2496–2507.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Wilson, G.A., Rannala, B., 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163, 1177– 1191.

Zuccaro, A., Bordonaro, S., Criscione, A., Guastella, M., Perrotta, G., Blasi, M., D'Urso, G., Marletta, D., 2008. Genetic diversity and admixture analysis of Sanfratellano and three other Italian horse breeds assessed by microsatellite markers. Animal 2, 991–998.