Hierarchical structure of the Sicilian goats revealed by Bayesian analyses of microsatellite information

M. Siwek[*,1], R. Finocchiaro[+], I. Curik[‡] and B. Portolano[*]

*Department S.En.Fi.Mi.Zo., University of Palermo, Viale delle Scienze Parco d'Orleans 90128 Palermo, Italy. [+]Associazione Nazionale Allevatori Frisona Italiana, Via Bergamo, 292 - 26100 Cremona, Italy. [‡]Department of Animal Science, University of Zagreb, Faculty of Agriculture, Svetosimunska 25, 10000 Zagreb, Croatia

Summary

Genetic structure and relationship amongst the main goat populations in Sicily (Girgentana, Derivata di Siria, Maltese and Messinese) were analysed using information from 19 microsatellite markers genotyped on 173 individuals. A posterior Bayesian approach implemented in the program STRUCTURE revealed a hierarchical structure with two clusters at the first level (Girgentana vs. Messinese, Derivata di Siria and Maltese), explaining 4.8% of variation (AMOVA Φ_{ST} estimate). Seven clusters nested within these first two clusters (further differentiations of Girgentana, Derivata di Siria and Maltese), explaining 8.5% of variation (AMOVA Φ_{SC} estimate). The analyses and methods applied in this study indicate their power to detect subtle population structure.

Keywords biodiversity management, clustering, goat, hierarchical genealogical structure, microsatellite diversity.

Goat farming in Sicily occurs mainly in extensive and semi-extensive breeding systems. In addition to several crossbreed populations, four goat populations are reared in Sicily: Girgentana, Derivata di Siria, Maltese and Messinese. The typical goat breeding system is based on extensive farming with a natural mating system and, apart from the Girgentana population, animals are kept outdoors, and are indoors only during the nights. Given the importance of goat breeding for Sicily, the specific breed characteristics, the reductions in the number of goats that occurred in last decades, and potential interbreeding, there is a need to establish appropriate management programmes with respect to principles of conservation genetics. The use of microsatellite information combined with analytical methods has been successfully applied to resolve similar conservation issues (Sechi *et al.* 2007; Glowatzki-Mullis *et al.* 2008).

The main objective of this study was to infer the genetic structure and relationship amongst four defined Sicilian goat breeds. Furthermore, with a posterior Bayesian

approach implemented in the program STRUCTURE, we wanted to reveal potential recent admixture and to detect, if present, the 'hidden' structure and/or sub-structure caused by management strategies that are characteristic for Mediterranean goat husbandry.

Blood samples were taken randomly from distinct flocks originating from the four goat populations located in Sicily. Genomic DNA was extracted, using a GFX Genomic Blood DNA Purification Kit (Amersham) according to the provided protocol, from 173 blood samples. These were collected from Girgentana (54) in five farms, one of which is located out of Sicily in the Calabria region, Messinese (34) in three farms, Maltese (45) in five farms, and Derivata di Siria (40) goats in three farms. We started with 23 microsatellite markers, chosen from the goat linkage map (Vaiman *et al.* 1996), the ISAG panels, or provided by the ECONOGENE project (http://www.econogene.eu). Later, we excluded four microsatellite loci because of the strong deviations from Hardy–Weinberg equilibrium (see Table S1).

Population structure was inferred using STRUCTURE 2.2 (Pritchard *et al.* 2000, http://pritch.bsd.uchicago.edu/software/structure2_2.html). The choice of the most likely number of clusters (K) was made according to recommendations of Pritchard *et al.* (2000), and graphical presentations showing the rate of change in the $\ln \Pr(G|K)$ between successive K values as suggested by Evanno *et al.* (2005). Hierarchical analysis of variance: (i) between four Sicilian goat populations and their flocks nested within breeds and (ii) between STRUCTURE obtained clusters, was performed
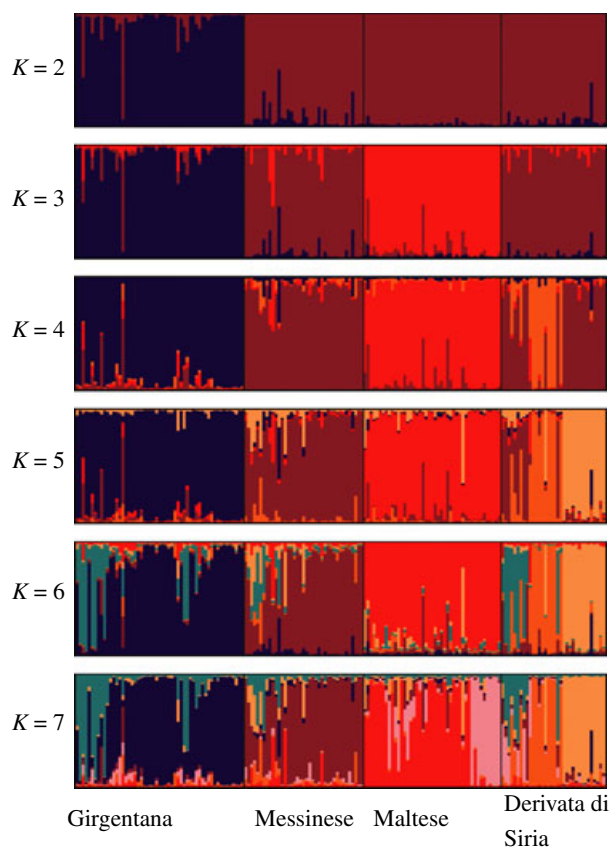
Figure 1 Graphical presentation of the hierarchical genetic structure for a sample of 173 Sicilian goats of four breeds; results were obtained by STRUCTURE analyses. Each goat is represented by a vertical line broken into K colour segments, with lengths proportional to the estimated membership of the inferred cluster. Only results from 'the best runs', i.e. runs with the largest ln Pr(G|K) values within each K (amongst 10 runs), are presented.

using the AMOVA method implemented in the ARLEQUIN 3.11 program (Excoffier *et al.* 2005; http://cmpg.unibe.ch/software/arlequin3/). The significance of the AMOVA Φ

statistics (F statistic analogues) was non-parametrically tested using 10 000 permutations.

Descriptive statistics of microsatellites, *a priori* defined populations and their genetic relationships can be found in Tables S1, S2 and S3 respectively. As suggested by the authors of STRUCTURE (Prichard *et al.* 2000; Falush *et al.* 2007), the most likely K value is that where ln Pr(G|K) is maximized or where the smallest value of K captures major structure in the data. The highest ln Pr(G|K) value in this study was obtained for K = 7, although smaller K values capturing the structure in the data were also possible. Evanno *et al.* (2005) pointed out that in populations with hierarchical island structure, suggestions based solely on maximized ln Pr(G|K) values might lead to less subtle ('erroneous') conclusions. According to Evanno *et al.* (2005), decisions based on ΔK, a quantity based on the second order rate of change with respect to K of the likelihood function, will in most situations be able to identify more subtle i.e. within-population structure (see Fig. S1 and Fig. 1). The pattern of ΔK values obtained in our analysis suggested the presence of a hierarchical structure with two clusters at the first level and with seven clusters at the second level, nested within the first level (see Fig. 1). Results obtained point to several characteristics of the typical Sicilian farming system, also noted during the blood sampling. In majority of flocks, individuals are exchanged within breeds and/or across breeds. While some of the goat farms are isolated, the others include different breeds in their flocks. We have detected the hierarchical population structure with two clusters (Girgentana vs. Messinese, Derivata di Siria and Maltese). The separation of Girgentana, an indigenous breed from the Agrigento area (Sicily) that most likely originates from Afghanistan and the Himalaya region (Portolano 1987), from the other goat breeds (mean pairwise $F_{ST} = 0.10$) is a consequence of differences in breeding system and origin. The Girgentana breeding traditionally took place in urban centres where the

Table 1 Hierarchical analysis of molecular variance (AMOVA) for 173 Sicilian goats with two subdivision levels. Analyses were based on 19 microsatellite loci.

Source of variation	Degrees of freedom	Variance components	Variation (%)	Fixation indices
A priori assumed structure				
Amongst breeds	3	0.488	10.3	$\Phi_{ST} = 0.155^*$
Amongst flocks within breeds	12	0.246	5.2	$\Phi_{SC} = 0.058^*$
Within flocks	328	4.005	84.5	$\Phi_{CT} = 0.103^*$
A posteriori derived structure[1]				
Amongst first level clusters (CL1)	1	0.2315	4.8	$\Phi_{ST} = 0.133^*$
Amongst CL2 within CL1	6	0.4099	8.5	$\Phi_{SC} = 0.089^*$
Within second level clusters (CL2)	340	4.1713	86.7	$\Phi_{CT} = 0.048^*$

*$P < 0.001$ (significance obtained on test with 10 000 permutations).
[1]A posteriori structure was derived from the results obtained by STRUCTURE and interpreted by Evanno *et al.* (2005). In a two-level AMOVA, first level (CL1) was defined according to the best run in K2 analysis while second level (CL2) was defined according to the best run in K7 analysis. Individual goats were assigned to clusters (CL1 and CL2) according to the highest estimated genomic membership.

animals were kept in the farmer's house. Some goats from Girgentana and Derivata di Siria as well as just few individuals from Messinese have shown the presence of an additional cluster (Fig. 1, see illustrations for $K = 6$ and $K = 7$). In a previous mtDNA analysis (Sardina *et al.* 2006), few Girgentana goats from this cluster had the mtDNA haplotype characteristic of wild goats. Sechi *et al.* (2007) have proven the occurrence of crossing between wild and domestic goats in Sardinia. Thus, although there is no explicit evidence, we speculate that this cluster might be related to wild goats.

The breeds from the second cluster were further separated into five populations in total. Derivata di Siria and Maltese differentiated into two clusters, while Messinese, with few individual exceptions, remained as a homogeneous cluster. The appearance of additional populations within the Maltese breed seems to be a consequence of mixed goat keeping (Maltese and Ionica breeds) at some farms.

When the AMOVA approach was applied to the hierarchical population structure, it was shown that variation at the first level explained 4.8% (AMOVA Φ_{ST} estimate), while variation at the second level explained 8.5% of variation (AMOVA Φ_{SC} estimate), (Table 1). The hierarchical population structure suggested by the analyses is biologically meaningful and does not conflict with our knowledge of historical events and breeding facts related to the breeds analysed.

In conclusion, Sicilian goats have shown strong population admixture structure caused by geographical location of the farms, influences of natural mating and traditional breeding systems where the flock is an important breeding unit. We further demonstrated the use of Evanno approach in explaining results obtained by STRUCTURE, which should not be ignored when hierarchical population structure is suspected.

Acknowledgements

References

Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Journal of Molecular Ecology* **14**, 2611–20.

Excoffier L., Laval G. & Schneider S. (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

Falush D., Stephens M. & Pritchard J.K. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574–8.

Glowatzki-Mullis M.L., Muntwyler J., Bäumle E. & Gaillard C. (2008) Genetic diversity measures of Swiss goat breeds as decision-making support for conservation policy. *Small Ruminant Research* **74**, 202–11.

Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–9.

Sardina M.T., Ballester M., Marmi J., Finocchiaro R., Van Kaam J.B.C.H.M., Portolano B. & Folch J.M. (2006) Phylogenetic analysis of Sicilian goats reveals a new mtDNA lineage. *Animal Genetics* **37**, 376–8.

Sechi T., Usai M.G., Miari S., Mura L., Casu S. & Carta A. (2007) Identifying native animals in crossbred populations: the case of the Sardinian goat population. *Animal Genetics* **38**, 614–20.

Vaiman D., Schibler L., Bourgeois F., Oustry A., Amigues Y. & Cribiu E.P. (1996) A genetic linkage map of the male goat genome. *Genetics* **144**, 279–305.

Supporting Information

Additional supporting information may be found in the online version of this article.

Figure S1 The choice of the representative number of clusters (K): A) Ln Pr(G|K) values presented as a function of the number of clusters.

Table S1 Characteristics of the 23 microsatellite loci analysed in four goat populations.

Table S2 Genetic diversity within the four goat populations (173 goats) based on 19 microsatellite loci.

Table S3 Population differentiation amongst four Sicilian goat breeds based on 19 microsatellite loci (173 goats), presented by F_{ST} estimates.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.