Animal (2008), **2:7**, pp 991–998 © The Animal Consortium 2008
doi:10.1017/S1751731108002255

animal

Genetic diversity and admixture analysis of Sanfratellano and three other Italian horse breeds assessed by microsatellite markers

A. Zuccaro[1], S. Bordonaro[1], A. Criscione[1], A. M. Guastella[1], G. Perrotta[2], M. Blasi[2], G. D'Urso[1] and D. Marletta[1†]

[1]*DACPA, Sezione di Scienze delle Produzioni Animali, Università degli studi di Catania, Via Valdisavoia 5, 95123 Catania, Italy;* [2]*LGS – Laboratorio di Genetica e Servizi, Cremona, Italy*

(Received 26 October 2007; Accepted 25 March 2008)

Sanfratellano is a native Sicilian horse breed, mainly reared in the north east of the Island, developed in the 19th century from local dams and sires with a restricted introgression of Oriental, African and, more recently, Maremmano stallions. In this study, the genetic relationships and admixture among Sanfratellano, the other two Sicilian autochthonous breeds and Maremmano breed were assessed using a set of microsatellites. The main goals were to infer the impact of Maremmano breed in the current Sanfratellano horse and to provide genetic information useful to improve the selection strategies of the Sanfratellano horse. The whole sample included 384 horses (238 Sanfratellano, 50 Sicilian Oriental Purebred, 30 Sicilian Indigenous and 66 Maremmano), chosen avoiding closely related animals. A total of 111 alleles from 11 microsatellite loci were detected, from four at HTG7 to 15 at ASB2 locus. The mean number of alleles was the lowest in Oriental Purebred (6.7), the highest in Sanfratellano (8.3). All the breeds showed a high level of gene diversity (H_e) ranging from 0.71 ± 0.04 in Sicilian Oriental Purebred to 0.81 ± 0.02 in Sicilian Indigenous. The genetic differentiation index was low; only about 6% of the diversity was found among breeds. Nei's standards (D_S) and Reynolds' (D_R) genetic distances reproduced the same population ranking. Individual genetic distances and admixture analysis revealed that: (a) nowadays Maremmano breed does not significantly influence the current Sanfratellano breed; (b) within Sanfratellano breed, it is possible to distinguish two well-defined groups with different proportions of Indigenous blood.

Keywords: horse breeds, microsatellites, genetic relationships, genetic distances, admixture analysis

Introduction

The Sanfratellano horse is a native Sicilian breed mainly reared in extensive system in the Natural Park of Nebrodi. Well adapted to harsh conditions, this meso-dolicomorph saddle-horse, characterized by a uniform blackish or bay coat, is mainly employed in light draft and equestrian tourism. In 1995, Sanfratellano was enrolled in the Italian Herd Book of local equine breeds, although genealogical records are available only since the second half of the 19th century. Nowadays, the breed, using an open-nucleus breeding scheme, includes about 1600 horses, with an enrolled selection nucleus of 1430 dams and 80 sires. However, only 14 stallions are employed in planned natural mating, which is the most frequent practice occurring mainly in three stud farms in the breeding area. The origin of Sanfratellano horse can be traced back to the Middle Ages (700–1200 A.D.), when native Sicilian horses were crossed with North African, Oriental and, later, Iberian populations (Fogliata, 1910). Limited introgression of Thoroughbred and Oriental Stallions was practiced in 1925 to improve the morphological structure of Sanfratellano (Hendricks, 1995). More recently, in 1934 and 1958, five Maremmano stallions were used in planned mating to improve withers height and size (Liotta and Chiofalo, 2004), but Maremmano sires were used until the end of the century (Chiofalo *et al.*, 2003). Nowadays Sanfratellano is included in the FAO endangered populations list (http://dad.fao.org) and incentives by EU are assigned to the breeders in order to preserve this genetic resource.

In Sicily, two other horse breeds are reared historically, both originating from local mares and sires crossed with

† E-mail: d.marletta@unict.it

horses of diverse origin (Chiari, 1901). The Sicilian Oriental Purebred, an endangered breed with less than 70 animals remaining (included in the FAO endangered populations list), is a genetic type of Arab–Oriental matrix. A Stud Book has existed since 1875 and proves the accurate selection practiced in Sicily on this breed starting from Arab and Oriental horses (Balbo, 1995). Sicilian Oriental Purebred is included in the Stud Book of Italian Saddle Horse. The other autochthonous horse is the Sicilian Indigenous, a very heterogeneous and largely unmanaged population. It is probably derived from a primitive strain of Sicilian horse and from the Borbon Real Casa di Ficuzza breed (related to the Napoletano, Persano and Arab breeds), and today it can be considered an Anglo-Oriental half-breed (Borgioli, 1959).

Maremmano horse represents an important part of the equestrian tradition in Tuscany and is bred throughout Italy. According to one hypothesis, this breed originated in Tuscany from ancient local populations living along the Tyrrhenian coast during the Etruscan time and influenced by ages of crossing with other horse breeds (Gandini and Rognoni, 1996). A second hypothesis postulates that its early ancestors came from Numidia, in north Africa, because Maremmano shares many features with the present-day Barbary horse (Edwards, 1994). In the 19th century, four stallions, two of which were Thoroughbred, were chosen as founders of the current Maremmano breed.

Within the frame of breed conservation, genetic characterization is important with regard to breed integrity and is an essential prerequisite for handling genetic resources (Bjørnstad and Røed, 2002). During the last decade, microsatellite markers proved to be a reliable and frequently used tool to quantify genetic variation within and among breeds and useful for the conservation management of animal populations (Cañón *et al*., 2000; Bruford *et al*., 2003). A preliminary characterization of Sanfratellano horse was achieved in the context of Western Mediterranean horse biodiversity (Marletta *et al*., 2006).

In this study, a set of microsatellites was used to assess the genetic diversity within and among three Sicilian autochthonous horses (Sanfratellano, Sicilian Oriental Purebred and Sicilian Indigenous) and one Italian breed (Maremmano).

The main goals were (i) to infer the impact of Maremmano breed in the current Sanfratellano horse, taking into account that Maremmano was used in recent crossing and (ii) to gather genetic information that might be useful to improve the selection strategies of the Sanfratellano horse.

Material and methods

Samples

Blood samples were collected from four Italian horse breeds: three reared almost exclusively in Sicily (Sanfratellano, Sicilian Oriental Purebred and Sicilian Indigenous) and one (Maremmano) originated in Tuscany (central Italy) and reared all over Italy. The whole sample included 384 horses (238 Sanfratellano, 50 Sicilian Oriental Purebred, 30 Sicilian Indigenous collected in Sicily, and 66 Maremmano horses collected in Tuscany). Sanfratellano, Maremmano and Indigenous horses were chosen from large geographical area, according to their pedigree information, when available, avoiding first- and second-order relatives. In the case of the Sicilian Oriental Purebred, the collected sample ($n = 50$) consisted of nearly the entire Stud Book-registered population (68 heads). DNA extraction from leucocytes was carried out by standard methods.

Microsatellite analysis

A set of 11 microsatellite markers (recommended by the International Society of Animal Genetics) was included in this study: *HTG4, HTG7, HTG10, HMS2, HMS3, HMS6, HMS7, VHL20, ASB2, AHT4* and *AHT5*. Template DNA was amplified in a multiplex reaction according to standard protocols (Cañón *et al*., 2000) using a PE GeneAmp PCR 9600 or 9700 system thermocyclers (Applied Biosystems, Foster City, CA, USA). Fluorescent-labelled PCR products were diluted, mixed with an internal size standard (ROX 350, ABI PRISM®; Applied Biosystems) and analysed using an ABI PRISM® 377 genetic analyser equipped with GeneScan® and Genotyper® software (Applied Biosystems). DNA samples from ISAG comparison test were used as references to standardize allele sizes.

Statistical analysis

Unbiased estimates of gene diversity (expected or Hardy–Weinberg heterozygosity) (Nei, 1987), observed heterozygosity (Hedrick, 1983) and the number of alleles per breed were calculated using the Microsatellites Analyser (Dieringer and Schlötterer, 2003). Genotypic frequencies were tested for the Hardy–Weinberg equilibrium using the GENEPOP computer package (Raymond and Rousset, 1995), which performs a probability test using a Markov chain (dememorization = 5000; batches = 100; and iterations per batch = 1000).

The FSTAT program (Goudet, 2001) was used to calculate the allelic richness (AR) standardized for variation in sample size using the rarefaction method (Hurlbert, 1971); the principle is to estimate the expected number of alleles in a sub-sample of $2n$ genes, where n is fixed as the smallest number of individuals typed for a locus in a sample. When calculated in this way, the number of alleles between populations with different sample sizes can be compared. FSTAT was also used to estimate Wright's F statistics (F_{IT}, F_{IS} and F_{ST}) (Weir and Cockerham, 1984). Mean and standard error of F statistics were obtained for the entire sample by the jackknifing procedure over all loci.

Two classical measures of genetic distance were obtained by Population software (Langella, 2002): Nei's standard distance (D_S) (Nei, 1987), which increases linearly in time, and the Reynolds' (D_R) distance (Reynolds *et al*., 1983), which under a pure genetic drift model (excluding mutations and admixtures) also increases linearly in time. A bootstrap value of 1000 was adopted as a common criterion to estimate genetic distances. In addition, the proportion of shared alleles distance (Dps') (Bowcock *et al*., 1994),

calculated by MICROSAT 1.5b (Minch *et al*., 1998), was chosen as the distance among individuals.

Phylogenetic trees were built from D_R and Dps' matrices applying the neighbour-joining method (Saitou and Nei, 1987), in order to show the relationships among breeds and individuals, respectively. D_R was chosen to create the population phylogenetic tree, because this is the most appropriate measure for livestock population with short-term divergence (Reynolds *et al*., 1983). Both trees were displayed by MEGA 4 (Tamura *et al*., 2007).

In order to estimate the individual admixture proportions and to simultaneously identify the ancestral K (unknown) clusters, STRUCTURE 2.2 program (Pritchard *et al*., 2000) was adopted. This program implements a clustering method for inferring population structure using multilocus genotype data; the criterion for inferring homogenous clusters was to minimize the Hardy–Weinberg and the gametic phase disequilibriums between loci within groups. The population structure was assessed first, using the whole sample set ($n = 384$), and then using the Sanfratellano breed only.

The admixture model using no prior information and the option of correlated allele frequencies between populations were used, as this configuration is considered best by Falush *et al*. (2003) in case of subtle population structure. Posterior probabilities of K ($\ln \Pr(X|K)$) were estimated using 100 000 Markov chain Monte Carlo (MCMC) repetitions and the 100 000 burn-in period. The range of Ks we tested was from 1 to 11. The best number of clusters fitting our data was established by plotting the mean $\ln \Pr(X|K)$ over multiple independent runs (10) for each K, as suggested by Pritchard *et al*. (2000).

The 10 runs of each K were averaged using the cluster matching and permutation program (CLUMPP) program (Jakobsson and Rosenberg, 2007). The output of cluster analysis was visualized by the DISTRUCT program (http://rosenberglab.bioinformatics.med.umich.edu/distruct.html).

Results

Genetic diversity within breeds

A total of 111 alleles were detected across the 11 analysed loci. Table 1 shows the number of alleles per marker, which ranged from 4 to 15 (average 10.09). The observed (H_o) and expected heterozygosity (H_e) for each marker are also given: the value of H_o ranged between 0.56 (*HTG7*) and 0.837 (*HMS2*), while the H_e value varied from 0.559 (*HTG7*) to 0.806 (*VHL20*). The average of Wright's fixation indices and the unbiased coefficient of gene differentiation (G_{ST}), due to breed differences, are also reported. Mean number of alleles (MNA) per locus, allelic richness (AR), calculated on the minimum sample size of individuals, and the heterozygosities per breed are reported in Table 2. MNA ranged between 6.7 ± 1.85 and 8.3 ± 2.97 in Sicilian Oriental Purebred and Sanfratellano, respectively. For all the other parameters, the Sicilian Indigenous showed the highest values, the Sicilian Oriental Purebred the lowest. A total of 24 breed-specific alleles (21.6% of the total alleles) were observed (nine in Sanfratellano, seven in Sicilian Oriental

Table 1 *Number of alleles, gene differentiation value (G_{ST}), observed and expected Heterozygosity (H_o, H_e) and F statistics value (F_{IS}, F_{IT}, F_{ST}) estimate by locus*

Locus	No. of alleles	G_{ST}	H_o	H_e	F_{IS}	F_{IT}	F_{ST}
HTG10	11	0.059	0.684	0.801	0.156	0.221	0.077
VHL20	13	0.030	0.786	0.806	0.017	0.058	0.041
HTG7	4	0.055	0.560	0.559	−0.050	0.005	0.052
HTG4	9	0.091	0.694	0.671	−0.046	0.094	0.134
AHT5	9	0.032	0.745	0.796	0.059	0.117	0.062
AHT4	12	0.036	0.797	0.799	−0.009	0.038	0.046
HMS3	9	0.051	0.719	0.773	0.072	0.125	0.056
HMS6	8	0.029	0.637	0.726	0.098	0.127	0.032
HMS7	8	0.026	0.782	0.795	0.002	0.030	0.031
HMS2	13	0.055	0.837	0.774	−0.059	0.004	0.060
ASB2	15	0.079	0.779	0.796	−0.008	0.085	0.093
Mean ± s.e.	10.09	0.049	0.729	0.754	0.023 ± 0.02	0.083 ± 0.02	0.062 ± 0.01

Table 2 *Sample size, observed and expected Heterozygosity (H_o, H_e), mean number of alleles (MNA), allelic richness (AR) and F_{IS} values in four Italian horse breeds*

Breeds	Sample size	H_e	H_o	MNA	AR	F_{IS}
Sanfratellano	238	0.747 ± 0.03	0.729 ± 0.01	8.3 ± 2.97	7.0	0.023
Sicilian Oriental	50	0.712 ± 0.04	0.704 ± 0.02	6.7 ± 1.85	6.2	0.012
Sicilian Indigenous	30	0.810 ± 0.02	0.735 ± 0.02	8.1 ± 2.02	8.1	0.093
Maremmano	66	0.747 ± 0.02	0.746 ± 0.02	7.0 ± 1.90	6.4	0.001

MNA = mean number of alleles; AR = allelic richness.

Purebred, five in Sicilian Indigenous and three in Maremmano), but none of them were with a frequency of over 0.055 (data not shown). F_{IS} values calculated per breed are not far from zero, but for Indigenous (Table 2) in which a deficit of heterozygotes can be supposed.

Two markers (*HTG10* and *HMS3*) were not consistently in the Hardy–Weinberg equilibrium ($P < 0.005$) in Sanfratellano and Sicilian Indigenous, whereas it was only one (*HMS6*) in Sicilian Oriental Purebred. All the markers showing genetic disequilibrium were excluded from the clustering analysis.

Genetic distances and clustering

D_S and D_R distances reproduced the same ranking, from a minimum value between Sanfratellano and Maremmano to a maximum between Maremmano and Sicilian Oriental Purebred (Table 3).

A neighbour-joining dendrogram was built starting from Reynolds' (D_R) genetic distance matrix (Figure 1); the number at the node indicates the percentage of a group's occurrence out of 1000 bootstraps. The phylogenetic tree formed clearly defined clusters, displaying the short distance between Sanfratellano and Maremmano.

A second neighbour-joining tree, based on the Dps among individuals, was drawn (Figure 2). In this case, the Sicilian Oriental Purebred clustered well except for some individuals. Sanfratellano and Maremmano horses did not form well-defined groups, while a remarkable number of Indigenous animals grouped clearly. The Sanfratellano is further subdivided into a group of individuals (*SanfA*) representative of the common population (in red), and a group of individuals (*SanfO*) representative of the old population (in violet).

Clustering analysis using a Bayesian approach was performed on the entire set with increasing numbers of inferred clusters (from $K = 1$ to $K = 11$). The analysis was based on the assignment of an individual's genomes to clusters using no prior information. The most likely number of subdivisions within our population is six. As shown in Figure 3, the likelihood of the observed data increases steadily when the number of clusters increases, reaches a peak in correspondence of cluster 6 and then declines progressively. At $K = 6$, the Sicilian Oriental Purebred and Maremmano animals form their own clusters, with the remaining breeds distributed among the remaining four clusters. Sicilian Indigenous grouped mainly into three (1, 2, 3) whereas Sanfratellano split into four clusters (3, 4, 5, 6) (Figure 4a). Sicilian Indigenous and Sanfratellano shared cluster 5 with an estimated membership of 32% and 19%, respectively (Table 4).

In order to confirm this result, an additional clustering analysis was performed on the Sanfratellano sample only. The genetic structure of the 238 Sanfratellano horses is shown in Figure 4b; in this case, we considered the total population split in two, namely *SanfO* and *SanfA*. The genetic equilibrium was calculated again and no more disequilibrium was found at *HTG10* and *HMS3*, but for *HMS6* that remained out of the analysis. Cluster 1 mainly represents the *SanfO* group, with an estimated membership of 90% (Table 5).

Table 3 *Nei's standard* D_S *(above the diagonal) and Reynolds'* D_R *(below the diagonal) genetic distances among four Italian horse breeds*

Breed	Sanfratellano	Sicilian Oriental	Sicilian Indigenous	Maremmano
Sanfratellano	–	0.395	0.184	0.115
Sicilian Oriental	0.109	–	0.290	0.401
Sicilian Indigenous	0.042	0.072	–	0.180
Maremmano	0.032	0.110	0.039	–

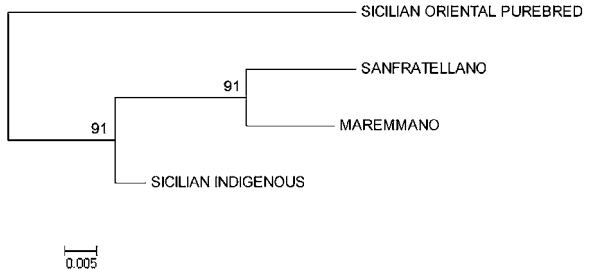

Figure 1 Neighbour-joining tree constructed from Reynolds' (D_R) distance among four Italian horse breeds.

Discussion

These four Italian horse breeds showed a good level of genetic diversity, in terms of MNA (average 7.5) being almost the same as that observed among 10 Western Mediterranean horse breeds including some breeds analysed here (Marletta *et al.*, 2006). MNA was greater than that observed in Lipizzan horse from seven national studs (Achmann *et al.*, 2004), in 11 Asian and two European populations (Tozaki *et al.*, 2003) and in Thoroughbred from four national studs but one (Cunningham *et al.*, 2001). Other European breeds showed the same trend (Aberle *et al.*, 2004; Glowatzki-Mullis *et al.*, 2006; Azor *et al.*, 2007). Allelic richness was high, particularly in Sicilian Indigenous.

Between breeds, genetic diversity (H_e) was never below 0.70 and was consistent with the values reported for other horse breeds (Cunningham *et al.*, 2001; Achmann *et al.*, 2004; Solis *et al.*, 2005; Azor *et al.*, 2007). The more intensive and effective the selection, the less the genetic variability. As expected, Sicilian Oriental Purebred showed the lowest gene diversity value ($H_e = 0.712 \pm 0.04$) probably because of the very small population size. The recorded pedigrees of Sicilian Oriental Purebred reveal a strong inbreeding and founder effect; however, in this threatened breed, heterozygosity values were higher than those of other European and Asiatic endangered populations, encouraging its preservation (Tozaki *et al.*, 2003; Morais

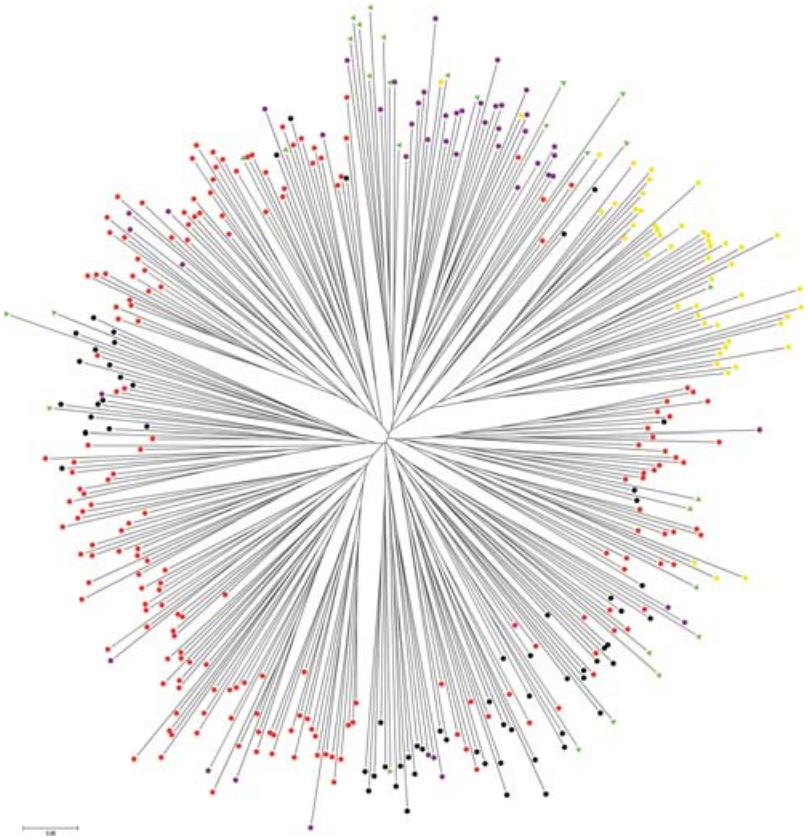

Figure 2 Neighbour-joining tree based on allele-sharing distances among 384 individuals of four Italian horse breeds: Sicilian Oriental Purebred (yellow), Sicilian Indigenous (green), Maremmano (black). The Sanfratellano population consists of two groups: *SanfA* (red) and *SanfO* (violet).

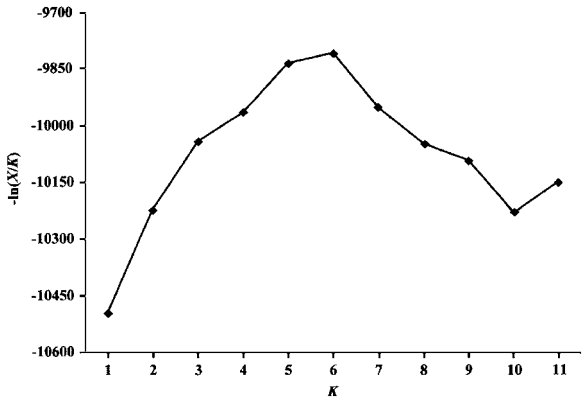

Figure 3 Plot of data likelihoods of 10 independent runs for 11 *Ks*: $\ln \Pr(X|K)$ *v. K.*

et al., 2005; Luis *et al.*, 2007). Sicilian Indigenous, in spite of the small sample ($n = 30$), showed the highest genetic variability ($H_e = 0.810 \pm 0.02$); these results probably reflect its wide genetic base and heterogeneity, even apparent in morphology.

For all the breeds, genetic diversity is in agreement with the allelic richness rank, a parameter that does not account for sample size.

The *F* statistics allow to study the genetic differentiation and analyse the deficit or excess of heterozygotes, per locus and breed, among animal breeds. The F_{IS} coefficient, which indicates the degree of departure from random mating, resulted close to zero for each breed, indicating low inbreeding levels (Table 2).

The average F_{ST} value for each locus and across the different loci (Table 1) revealed a relatively low gene flow within breeds: only 6.2% of the total genetic variation is explained by population differences, the remaining 93.8% corresponds to differences among individuals. The genetic differentiation value (F_{ST}) was higher than that observed in four Basque-Navarrese semi-feral breeds (Solis *et al.*, 2005) and lower than those obtained in European (Zabek *et al.*, 2005; Azor *et al.*, 2007) and South American breeds (Lippi and Mortari, 2003). On average, a deficit of heterozygotes of 2.3% (F_{IS}) exists for each analysed breed; this deficit was 8.3% in the whole sample (F_{IT}) (Table 1).

The tree based on Dps distance shows the relationship among individuals: the Sicilian Oriental Purebred is the group that clusters best (44/50). This result is not surprising if we consider that the whole breed counts less than 70 animals.

On the contrary, there was no definite differentiation for the other breeds. Within the Sanfratellano sample, it is possible to distinguish two groups: *SanfA* and *SanfO*. The largest consists of 191/238 heads (*SanfA*) and represents the most common Sanfratellano gene pool; the remaining 47/238 heads (*SanfO*) correspond to the old Sanfratellano

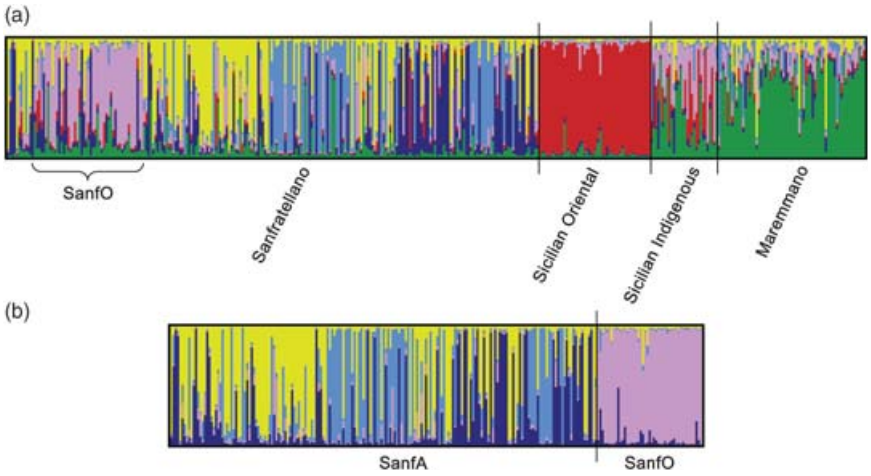

Figure 4 (a) Clustering assignment of 384 horses belonging to four Italian breeds. Each individual is represented on the graph by a vertical line divided into coloured segments corresponding to six genetic clusters. The length of each coloured segment is proportional to the individual membership in the cluster of corresponding colour. (b) Clustering assignment of 238 Sanfratellano horses grouped in two subgroups: *SanfO* and *SanfA*.

Table 4 *Average membership coefficient for each given breed for* K = *6 clustering result in Figure 4a*

Clusters Breeds	*K*1	*K*2	*K*3	*K*4	*K*5	*K*6
Sanfratellano	0.04	0.08	0.23	0.21	0.19	0.25
Sicilian Oriental	0.88	0.04	0.02	0.01	0.04	0.02
Sicilian Indigenous	0.21	0.30	0.08	0.04	0.32	0.06
Maremmano	0.04	0.61	0.05	0.10	0.10	0.09

Table 5 *Average membership coefficient for the two Sanfratellano subgroups for* K = *4*

Clusters Groups	*K*1	*K*2	*K*3	*K*4
SanfA	0.05	0.40	0.27	0.28
SanfO	0.90	0.03	0.05	0.02

breed according to their genealogical data. More than 60% of these animals (30 on 47 individuals) are well grouped, as displayed in the tree; furthermore, within the same cluster, a large number of Indigenous and only few individuals of the other breeds are present. The mixed clustering among 30 *SanfO* and 15 Indigenous could be due to their common origin; in fact these autochthonous breeds could represent, within a wide historical context, the old Sicilian horse gene pool.

The Bayesian approach has been proved to be a powerful method in studying population structure by the identification of genetic *K* clusters, also allowing a probabilistic estimation of the clusters' membership (*Q*).

Six inferred clusters fit our data best. Only Sicilian Oriental Purebred and Maremmano horse genomes are assigned to their own cluster, with an estimated membership of 88% and 61%, respectively; the membership value for the other two breeds is less than 35%. The Sicilian Indigenous seems to be very heterogeneous and spreads its gene pool over three groups. It shares part of the multilocus genotype with Maremmano ($Q = 0.30$) and with Sicilian Oriental Purebred ($Q = 0.21$). The presence of a remarkable percentage of Sicilian Indigenous horses clustering into the Maremmano group (cluster 2) can be explained by the assessed presence of Thoroughbred genetics into both populations (Balbo, 1995; Hendricks, 1995). Also, the Sicilian Oriental Purebred shares its multilocus genotype with the Sicilian Indigenous ($Q = 0.20$); in this case, the relationship is due to a strong presence of Oriental blood introduced into Sicily during the Arab domination.

The Sanfratellano breed appears to be the most heterogeneous and its genetic information spreads over four clusters: one of them (cluster 5) is in common with the Sicilian Indigenous, the others (cluster 3, 4 and 6) seem to equally represent the breed. Admixture between Sicilian Indigenous and Sanfratellano ($Q = 0.19$) occurs mainly in the *SanfO* group; this fact suggests a gene flow from Sicilian Indigenous to the native Sanfratellano horse and confirms the common origin of these two Sicilian horses, even strengthened by historical data. The other three clusters, within Sanfratellano, might be due to an unbalanced employment of a restricted number of stallions during mating, probably belonging to an old-type Maremmano horse. In fact, the role of the current Maremmano breed into the Sanfratellano population is different: the estimated membership of Sanfratellano to cluster 2 is 8%. A feeble connection between Sanfratellano and Maremmano may only be observed in clusters 4 and 6, notwithstanding the fact that Maremmano sires were used many times during the last century. Clustering analysis was also performed on the Sanfratellano sample only, highlighting the same result (Figure 4b): the *SanfO* group clusters clearly with an estimated membership of 90% to its own cluster.

In conclusion, all the Sicilian horse breeds showed high genetic variability, even if not so remarkable in Sicilian

Oriental Purebred. The information given by clustering analysis allowed us to understand more about the genetic structure of the Sanfratellano and the other Sicilian breeds. In particular, Sanfratellano appears well distinct from the two other autochthonous breeds, not significantly influenced by the crossing with Maremmano, which occurred during the last century, but is still rather genetically heterogeneous.

The results reported here could be useful for selection strategies as well as for conservation purpose of Sanfratellano. The identification of two sub-populations (*SanfO* and *SanfA*) may suggest to the breeders the appropriate management of mating. Stallions belonging to the *SanfO* group are the reference sires for those breeders interested in recovering the most original nucleus of Sanfratellano, well adapted to equestrian tourism. On the other hand, stallions of the *SanfA* group, which represent the current Sanfratellano lineage, will be the reference for addressing this breed to equestrian competitions.

In both cases, Sanfratellano horse should be maintained and preserved in that mountainous part in the north-east of Sicily, Natural Park of Nebrodi, in which it plays an important economic and social role.

Acknowledgements

Financial support was provided by 'Convenzione con Istituto Sperimentale Zootecncico per la Sicilia del. CdA del ISZS no. 31/05'.

References

Aberle KS, Hamann H, Drögemüller C and Distl O 2004. Genetic diversity in German draught horse breeds compared with a group of primitive, riding wild horses by means of microsatellite DNA markers. Animal Genetics 35, 270–277.

Achmann R, Curik I, Dovc P, Kavar T, Bodo I, Habe F, Marti E, Solkner J and Brem G 2004. Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. Animal Genetics 35, 285–292.

Azor PJ, Valera M, Gòmez MD, Goyache F and Molina A 2007. Genetic characterization of Spanish Trotter horse breed using microsatellite markers. Genetics and Molecular Biology 30, 37–42.

Balbo SM 1995. L'influenza dell'Arabo-Orientale sul cavallo Siciliano. In L'ASIL Arabo – Il cavallo nobile d'Arabia (ed. M Savier), pp. 169–175. R&R Editrice, Spoleto (PG) Italy.

Bjørnstad G and Røed KH 2002. Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. Animal Genetics 33, 264–270.

Borgioli E 1959. In Elementi di zootecnica speciale degli animali domestici agricoli, 6th edition (ed. G Barbèra), Firenze, Italy.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR and Cavalli-Sforza LL 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368, 455–457.

Bruford MW, Bradley DG and Luikart G 2003. DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics 4, 900–910.

Cañón J, Checa MI, Carleos C, Vega-Pla JI, Vallejo M and Dunner S 2000. The genetic structure of the Spanish Celtic horse breeds inferred from microsatellite data. Animal Genetics 31, 39–48.

Chiari E 1901. Trattato di ippologia, vol. 2. Unione Tipografico-Editrice, Torino, Italy.

Chiofalo L, Portolano B, Liotta L, Rundo Sotera A and Finocchiaro R 2003. Demographic characterization, inbreeding and genetic variability within Sanfratellano population horse from genealogical data. Italian Journal of Animal Science 2(Suppl. 1), 592–594.

Cunningham EP, Doley JJ, Splan RK and Bredley DG 2001. Microsatellite diversity, pedigree relatedness and the contribution of the founder lineage to thoroughbred horses. Animal Genetics 32, 360–364.

Dieringer D and Schlötterer C 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3, 167–169.

Edwards EH 1994. The encyclopedia of the horse. Dorling Kindersley Limited, London, UK.

Falush D, Stephens M and Pritchard JK 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164, 1567–1587.

Fogliata G 1910. Tipi e Razze Equine, 2nd edition. Tipografia Editrice. Cav. F. Mariotti, Pisa, Italy.

Gandini G and Rognoni G 1996. Atlante etnografico delle popolazioni equine ed asinine italiane. CNR IDVGA, Città Studi Edizioni, Milano, Italy.

Glowatzki-Mullis ML, Muntwyler J, Pfister W, Marti E, Rieder S, Poncet PA and Gaillard C 2006. Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. Animal Genetics 37, 33–39.

Goudet J 2001. FSTAT, a program to estimate and test gene diversity and fixation indices, version 2.9.3. Retrieved August 23, 2005 from http://www2.unil.ch/popgen/softwares/fstat.htm

Hedrick PW 1983. Genetics of populations. Science Books International, Boston, MA, USA.

Hendricks BL 1995. International encyclopedia of horse breeds. University of Oklahoma Press, Norman, OK, USA.

Hurlbert SH 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52, 577–586.

Jakobsson M and Rosenberg NA 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806.

Langella O 2002. Population 1.2.28. Logiciel de génétique des populations. http://www.legs.cnrs-gif.fr. Laboratoire Populations, génétique et évolution, CNRS UPR 9034, Gif-sur-Yvette, France.

Liotta L and Chiofalo L 2004. Evoluzione morfo-biometrica della popolazione cavallina Sanfratellana. Atti 6° Convegno Società Italiana di Ippologia 'New findings in equine practices' 7-8-9 Luglio 2004, Campobasso, Italy, pp. 51–55.

Lippi AS and Mortari N 2003. Studies of blood groups and protein polymorphisms in the Brazilian horse breed Mangalarga Marchador and Mangalarga (*Equus caballus*). Genetics and Molecular Biology 26, 431–434.

Luis C, Cothran EG and Oom MM 2007. Inbreeding and genetic structure in the endangered Sorraia horse breed: implication for its conservation and management. Journal of Heredity 98, 232–237.

Marletta D, Tupac-Yupanqui I, Bordonaro S, Garcia D, Guastella AM, Criscione A, Cañón J and Dunner S 2006. Analysis of genetic diversity and the determination of relationships among western Mediterranean horse breeds using microsatellite markers. Journal of Animal Breeding and Genetics 123, 315–325.

Minch E, Ruiz-Linares A, Goldstein DB, Feldman MW and Cavalli-Sforza LL 1998. Microsat2: A Computer Program for Calculating Various Statistics on Microsatellite Allele Data. Software available at http://hpgl.stanford.edu/projects/microsat/. Department of Genetics, Stanford University, Stanford, CA, USA.

Morais J, Oom MM, Malta-Vacas J and Luis C 2005. Genetic structure of an endangered Portuguese semiferal pony breed, the Garrano. Biochemical Genetics 43, 347–364.

Nei M 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY, USA.

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

Raymond M and Rousset F 1995. GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. Journal of Heredity 86, 248–249.

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

Saitou N and Nei M 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406–425.

Solis A, Jugo BM, Mériaux JC, Iriondo M, Mazòn LI, Aguirre AI, Vicario A and Estomba A 2005. Genetic diversity within and among four south European native horse breed on microsatellite DNA analysis: applications for conservation. Journal of Heredity 96, 670–678.

Tamura K, Dudley J, Nei M and Kumar S 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24, 1596–1599.

Tozaki T, Takezaky N, Hasegawa T, Ishida N, Kurosawa M, Tomita M, Saitou N and Mukoyama H 2003. Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a European horse outgroup. Journal of Heredity 94, 374–380.

Weir BS and Cockerham CC 1984. Estimating F statistics for the analysis of population structure. Evolution 38, 1358–1370.

Zabek T, Nogaj A, Radko A, Nagaj J and Slota E 2005. Genetic variation of Polish endangered Bilgaraj horses and two common horse breeds in microsatellite loci. Journal of Applied Genetics 46, 299–305.