

Proceedings of the World Congress on Genetics Applied to Livestock Production, 11. 236

A genome-wide perspective on the population structure of Italian cattle breeds

S. Mastrangelo^{1}, E. Ciani², P. Ajmone Marsan³, A. Bagnato⁴, L. Battaglini⁵, R. Bozzi⁶, A. Carta⁷, G. Catillo⁸, M. Cassandro⁹, S. Casu⁷, R. Ciampolini¹⁰, P. Crepaldi⁴, M. D'Andrea¹¹, R. Di Gerlando¹, L. Fontanesi¹², M. Longeri⁴, N.P.P Macciotta¹³, R. Mantovani⁹, D. Marletta¹⁴, D. Matassino¹⁵, M. Mele¹⁶, G. Pagnacco⁴, C. Pieramati¹⁷, B. Portolano¹, F.M. Sarti¹⁸, M. Tolone¹ & Fabio Pilla¹¹*

¹*University of Palermo, Dipartimento Scienze Agrarie, Alimentari e Forestali, 90128 Palermo, Italy*

²*University of Bari, Dipartimento di Bioscienze Biotechnologie e Biofarmaceutica, 70124 Bari, Italy*

³*University of Sacro Cuore, Istituto di Zootecnia, 20122 Piacenza, Italy*

⁴*University of Milano, Dipartimento di Medicina Veterinaria, 20133 Milano, Italy*

⁵*University of Torino, Dipartimento di Scienze Agrarie Forestali e Alimentari, 10095 Grugliasco, Italy*

⁶*University of Firenze, Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente, 50144 Firenze, Italy*

⁷*Agris Sardegna, Unità di Ricerca di Genetica e Biotechnologie, 07100 Sassari, Italy*

⁸*CREA, CREA Research Centre for Animal Production and Acquaculture, 00015 Monterotondo, Italy*

⁹*University of Padova, Dipartimento di Agronomia Animali Alimenti Risorse naturali e Ambiente, 35020 Legnaro, Italy*

¹⁰*University of Pisa, Dipartimento di Scienze Veterinarie, 56100 Pisa, Italy*

¹¹*University of Molise, Dipartimento Agricoltura, Ambiente e Alimenti, 86100 Campobasso, Italy*

¹²*University of Bologna, Dipartimento di Scienze e tecnologie Agroalimentari, 40127 Bologna, Italy*

¹³*University of Sassari, Dipartimento di Agraria, 07100 Sassari, Italy*

¹⁴*University of Catania, Dipartimento di Agricoltura, Alimentazione e Ambiente, 95125 Catania, Italy*

¹⁵*Consorzio per la Sperimentazione, Divulgazione e Applicazione di Biotechniche Innovative, 82100 Benevento, Italy*

¹⁶*University of Pisa, Dipartimento di Scienze Agrarie, Alimentari e Agro-ambientali, 56124 Pisa, Italy*

¹⁷*University of Perugia, Dipartimento di Medicina Veterinaria, 06126 Perugia, Italy*

¹⁸*University of Perugia, Dipartimento di Scienze Agrarie, Alimentari, Ambientali, 06121 Perugia, Italy*

pilla@unimol.it (Corresponding Author)

Summary

Despite the growing diffusion of cosmopolite specialized breeds, several autochthonous populations are still farmed in Italy. The aim of this study was to provide a high-resolution picture of the genome-wide diversity and population structure of Italian local cattle breeds using medium-density genome-wide SNP markers.

After data editing, the dataset included 800 samples from 32 breeds that were genotyped for 31,013 SNPs. For several breeds we observed a low level of polymorphism and genetic diversity that confirmed threat of extinction. Shared ancestry, admixture events, and reticulations observed on the phylogenetic tree between some breeds, all suggest high levels of gene flow. Clear clusters and relationships between breeds that originated from the same geographical area were detected. However, in spite of the complex admixture history, many of the local Italian cattle breeds have retained unique identities and are clearly differentiated breeds. Differences in their origin, in climate characteristics of farming areas, the genetic isolation, and the inbreeding can be the main reasons for such differentiation. This study represents the first exhaustive genome-wide analysis of Italian cattle diversity. The results largely agreed with the breeding history of the Italian cattle breeds. The population structure and the low genetic diversity presented here for several breeds should be evaluated in adopting conservation strategies. Thus, efforts should be made to improve genetic diversity in these breeds. Control of inbreeding, breeding stations development and improvement of recording system are strategies to conserve these breeds under in situ conservation situation, and in this context, genomic information may play a crucial role for the preservation and management of these populations.

Keywords: single nucleotide polymorphism, inbreeding, genetic diversity, population structure, Italian cattle breeds

Introduction

The complex origins of cattle, associated with natural and artificial selection to adapt bovine to a wide range of environments and to the different rearing system and production purpose, gave rise to numerous different breeds displaying a broad phenotypic variety over a short period of time (Gautier *et al.*, 2010). In the last few decades, there has been a severe decrease in cattle biodiversity. As a consequence, the monitoring of cattle genetic diversity has become an important issue in the management of cattle breeding programs (Hiemstra *et al.*, 2010). The recent advent of high-throughput and cost effective genotyping techniques makes it possible to provide a detailed genome wide assessment of the genome structure and relationships among cattle populations (Gautier *et al.*, 2010), opening new perspectives to livestock genetics. Despite the growing diffusion of the cosmopolite specialized breeds, several autochthonous cattle breeds and populations are still farmed in Italy. Nowadays, most of these breeds are fully adapted to a specific habitat or production system and represent a unique resource to meet present and future demands for a sustainable farming in a changing environment. Local cattle breeds have a dual role not only in food production, but also in the provision of other social and public goods, as territory preservation, consumer habits, tourists requests, and history and cultural aspects of link between breed and food. This study was thus undertaken to analyze the level of genetic diversity, population structure, admixture patterns and relationships among 30 Italian cattle breeds using medium-density genome-wide SNP markers.

Material and methods

Sample collection and genotyping

Genotyping data from the Illumina BovineSNP50 v2 BeadChip array were available for 814 animals (10-43 per breed) belonging to 30 Italian breeds and 2 cosmopolitan breeds (Charolais and Limousin) (Table 1). Genotyping data were merged in PLINK (Purcell *et al.*, 2007). SNPs were mapped using the *Bos taurus* UMD 3.1 genome assembly. Markers were filtered to exclude loci assigned to unmapped contigs and only SNPs on autosomes were considered. Additionally, the following filtering parameters were adopted before the generation of the pruned input file: (i) SNPs with call rate less than 98% and (ii) minor allele frequency (MAF) less than 5%, and (iii) animals with more than 10% of missing genotypes were removed.

Genetic diversity indices

PLINK was used to estimate the within-breed genetic diversity indices, including observed and expected heterozygosity (H_o and H_e , respectively), the genomic inbreeding coefficient calculated on the basis of the difference between the observed and expected numbers of homozygous genotypes (F_{HOM}), and average MAF (≥ 0.05).

Population genetic structure

Multidimensional scaling (MDS) plot analysis based on a matrix of genome-wide pairwise identity-by-state (IBS) distance was carried out using PLINK. The graphical representation was depicted using the statistical R software (<http://www.R-project.org/>). The presence of population substructures was evaluated through the model-based clustering algorithm implemented in ADMIXTURE (Alexander & Lange, 2011). The Reynold's genetic distances between breeds were calculated using ARLERQUIN (Excoffier *et al.*, 2010), and was followed by construction of a NeighborNet graph using SPLITSTREE (Huson & Bryant, 2006).

Results and discussion

Genetic diversity indices

The dataset used in the analysis included 31,013 SNPs from 800 samples (32 breeds). The genetic diversity parameters of the genetic variability within breed, are reported in Table 1. The set of 50K Illumina SNPs was found to be highly informative in all the considered breeds. H_o and H_e ranged from 0.297 ± 0.194 (Pontremolese) to 0.358 ± 0.167 (Piedmontese) and from 0.267 ± 0.187 (Mucca Pisana) to 0.353 ± 0.137 (Sarda), respectively. High accordance was observed between heterozygosity values and polymorphism level in different populations: Mucca Pisana showed the lowest mean MAF, whereas Sarda breed showed the highest value. Average MAF was homogeneous among the breeds, consistent with previous observations that concluded that on average, SNP are equally informative among breeds (Matukumalli *et al.*, 2009), even if at individual level SNP may show considerable variation in MAF. These results for H_e and MAF are in agreement with the range reported in a study on development and characterization of a high-density SNP genotyping assay for several cattle breeds (Matukumalli *et al.* 2009), and it was similar to those observed in European breeds (Gautier *et al.*, 2010). The highest F_{HOM} , was found in Pontremolese breed

(0.195 ± 0.101) and *Mucca Pisana* (0.183 ± 0.058). The lowest value was found in the *Sarda* breed (0.006 ± 0.063). Looking across all breeds, generally, the cosmopolitan breeds (Italian Holstein, Brown and Simmental, Limousin and Charolaise) showed moderate levels of genetic diversity. Moreover, despite the fact that some breeds (*Rossa Siciliana*, *Sardo-Bruna*, *Sardo-Modicana* and *Agerolese*) are endangered or with small census population size, they did not have values of high levels of consanguinity, likely because of their admixed origins. *Pontremolese* and *Mucca Pisana* showed a lower level of polymorphism and genetic diversity in comparison with other breeds. These results confirm the expectation as these two breeds, such as in general the local breeds, underwent a more extreme bottleneck in population size at the beginning of the twentieth century. If reproductive isolation within groups of the same herd and area is maintained for long time in subsequent generations, without specific mating plans as was suggested and occurring in some Italian populations (Gandini *et al.* 1993), the rate of inbreeding will increase dramatically, resulting in a decrease in the effective population size.

Population genetic structure

The MDS appeared to separate the breeds according to their geographical location with the notable exception of the Podolian-derived breeds. The first dimension (C1) distinguished the three Sicilian (*Cinisara*, *Rossa Siciliana*, *Modicana*) and the five Podolian-derived breeds (*Romagnola*, *Marchigiana*, *Chianina*, *Maremmana* and *Podolica*) from the other breeds (Figure 1). Moreover, in the same space area were positioned the *Sardo-Modicana* and the *Calvana* breeds. The *Sardo-Modicana*, derived by the cross of local *Sarda* cows with *Modicana* bulls. The Northern and Northern-central Italian populations formed a defined group, which was separated from the Sicilian and Podolian-derived breeds by a large gap on the first axis (C1). The plot obtained by ADMIXTURE shows the results for K ranging from 2 to 32 (Figure 2). The first two breeds to be differentiated from all of the others were *Calvana* and *Italian Brown* (at $K=2$). Other breed-specific clusters were *Mucca Pisana* and *Italian Holstein* at $K = 4$, *Italian Simmental*, *Garfagnina*, *Pontremolese* and *Modicana* at $K = 8$. Moreover, at $K = 8$ some genomic components appeared to be shared by several breeds, as for example among the Sicilian (*Cinisara*, *Modicana* and *Rossa Siciliana*) and *Sardo-Modicana* breeds, among the Podolian-derived breeds and among Northern and Northern-central Italian populations. This was consistent with the MDS results. When K was increased from 8 to 32, specific breeds were progressively assigned to distinct clusters. The analysis of cross-validation errors for different K values highlighted $K = 24$ as the most probable number of populations present in the total sample. At this K value, each breed tends to have its own distinct cluster but with some exception; in fact, *Agerolese*, *Varzese-Ottonese*, *Cinisara*, *Podolica*, *Barà-Pustertaler*, *Sardo-Bruna*, *Rossa Siciliana*, *Piedmontese*, *Sarda* and *Limousin* breeds showed less distinct clusters. As reported above, these breeds did not show signs of high levels of consanguinity due to their admixed origins and this suggests that crossbreeding with other breeds occurred. The results largely agreed with the breeding history of the Italian cattle breeds, given that some breeds are the result of crossbreeding (e.g. *Sardo-Modicana*, *Sardo Bruna*, *Rossa Siciliana*). However, in spite of the complex admixture history of most breeds at $K = 24$, many of the local Italian cattle breeds have retained unique identities and are differentiated breeds, and this can be due to differences in the origin, differences in climate, genetic isolation and inbreeding. Neighbor-network graphs account for gene flow among breeds (reticulation) and, thus, provide a more plausible reconstruction than linear representations. The graph showed some clear clusters and relationships between breeds that

originated from the same region or area and revealing the presence of an underlying geographical pattern of variation with the clusters of breeds, corresponding to MDS plot (Figure 3). A similar geographical pattern has been described in previous studies based on medium density SNPs in Italian goats (Nicoloso *et al.* 2015) and sheep breeds (Ciani *et al.* 2014). The reticulations towards the extremity of the graphs indicated increasing genetic relatedness between breeds and therefore past hybridization events among these populations. The shortest branch was observed for the Sarda population, whereas the longest one was found for the Mucca Pisana, according to the results of genetic diversity (Table 1). The long branch observed for Mucca-Pisana was typical of differentiated and isolated populations with small effective population sizes (Kijas *et al.* 2012).

All the conducted analyses revealed a clear genetic relationship, gene flow and admixture events for several Italian local cattle breeds. However, in spite the complex admixture history of most breeds, many of these breeds have retained unique identities and are well-differentiated breeds. Attention should be given to conservation of these populations because of their critical conservation status and in this context, genomic information may play a crucial role for their preservation and management.

Based on the use of a standardized genotyping array, it is possible to combine various datasets. Therefore, further studies will be conducted to provide an insight into the genetic composition and origin of some Italian cattle breeds, such as the Podolian-derived breeds, using data of other worldwide cattle population.

Table 1. Sample size after quality control (n -PostQC), observed (H_o) and expected (H_e) heterozygosity, average minor allele frequency (MAF), inbreeding coefficient (F_{HOM}) and standard deviation (s.d).

Breed	Breed code	n-PostQC	$H_o \pm s.d$	$H_e \pm s.d$	MAF \pm s.d	$F_{HOM} \pm s.d$
AGEROLESE	AGER	22	0.346 \pm 0.176	0.338 \pm 0.150	0.255 \pm 0.145	0.058 \pm 0.051
BARA'-PUSTERTALER	B-PUS	24	0.349 \pm 0.167	0.342 \pm 0.145	0.258 \pm 0.143	0.051 \pm 0.055
BURLINA	BURL	24	0.353 \pm 0.169	0.344 \pm 0.147	0.261 \pm 0.143	0.041 \pm 0.044
CABANNINA	CABN	22	0.347 \pm 0.174	0.336 \pm 0.151	0.254 \pm 0.146	0.056 \pm 0.032
CALVANA	CALV	24	0.307 \pm 0.198	0.294 \pm 0.175	0.221 \pm 0.158	0.167 \pm 0.063
CHAROLAIS	CHAR	25	0.353 \pm 0.166	0.346 \pm 0.144	0.262 \pm 0.142	0.039 \pm 0.066
CHIANINA	CHIA	23	0.327 \pm 0.177	0.323 \pm 0.158	0.242 \pm 0.149	0.111 \pm 0.048
CINISARA	CINI	30	0.343 \pm 0.155	0.348 \pm 0.141	0.263 \pm 0.140	0.068 \pm 0.070
GARFAGNINA	GARF	23	0.312 \pm 0.199	0.300 \pm 0.177	0.226 \pm 0.158	0.151 \pm 0.053
ITALIAN BROWN	IT-BR	32	0.307 \pm 0.187	0.299 \pm 0.171	0.223 \pm 0.154	0.166 \pm 0.033
ITALIAN HOLSTEIN	IT-HO	32	0.344 \pm 0.171	0.338 \pm 0.154	0.256 \pm 0.148	0.064 \pm 0.036
ITALIAN SIMMENTAL	IT-SI	32	0.340 \pm 0.168	0.332 \pm 0.152	0.249 \pm 0.145	0.079 \pm 0.023
LIMOUSIN	LIMO	20	0.345 \pm 0.177	0.335 \pm 0.152	0.253 \pm 0.146	0.062 \pm 0.024
MARCHIGIANA	MARC	22	0.339 \pm 0.173	0.333 \pm 0.151	0.250 \pm 0.145	0.078 \pm 0.023
MAREMMANA	MARE	24	0.325 \pm 0.192	0.311 \pm 0.167	0.234 \pm 0.154	0.118 \pm 0.049
MODENESE	MODE	23	0.341 \pm 0.174	0.332 \pm 0.153	0.251 \pm 0.147	0.073 \pm 0.045
MODICANA	MODI	29	0.329 \pm 0.171	0.328 \pm 0.156	0.247 \pm 0.148	0.105 \pm 0.067
MUCCA PISANA	M-PIS	23	0.301 \pm 0.225	0.267 \pm 0.187	0.200 \pm 0.162	0.183 \pm 0.058
PEZZATA R. D'OROPA	PRDO	23	0.333 \pm 0.178	0.327 \pm 0.158	0.246 \pm 0.149	0.096 \pm 0.054
PIEDMONTESE	PIED	21	0.358 \pm 0.167	0.347 \pm 0.141	0.262 \pm 0.141	0.027 \pm 0.011
PINZGAU	PINZ	24	0.349 \pm 0.174	0.337 \pm 0.151	0.254 \pm 0.145	0.051 \pm 0.030
PODOLICA	PODO	24	0.343 \pm 0.157	0.349 \pm 0.140	0.264 \pm 0.140	0.066 \pm 0.073
PONTREMOLESE	PONT	24	0.297 \pm 0.194	0.292 \pm 0.176	0.218 \pm 0.157	0.195 \pm 0.101
PUSTERTALER	PUST	24	0.339 \pm 0.185	0.323 \pm 0.161	0.243 \pm 0.151	0.078 \pm 0.028
REGGIANA	REGG	26	0.346 \pm 0.175	0.336 \pm 0.150	0.253 \pm 0.145	0.059 \pm 0.040
RENDENA	REND	24	0.332 \pm 0.178	0.325 \pm 0.158	0.244 \pm 0.149	0.096 \pm 0.024
ROMAGNOLA	ROMG	21	0.325 \pm 0.184	0.317 \pm 0.163	0.238 \pm 0.151	0.117 \pm 0.026
ROSSA SICILIANA	R-SIC	24	0.356 \pm 0.166	0.345 \pm 0.143	0.261 \pm 0.141	0.032 \pm 0.023
SARDA	SAR	30	0.346 \pm 0.151	0.353 \pm 0.137	0.267 \pm 0.139	0.060 \pm 0.063
SARDO-BRUNA	S-BRU	10	0.338 \pm 0.193	0.334 \pm 0.153	0.252 \pm 0.147	0.082 \pm 0.086
SARDO-MODICANA	S-MOD	28	0.344 \pm 0.168	0.338 \pm 0.149	0.255 \pm 0.145	0.065 \pm 0.031
VARZESE-OTTONESE	VR-OT	43	0.351 \pm 0.160	0.343 \pm 0.145	0.259 \pm 0.142	0.046 \pm 0.071

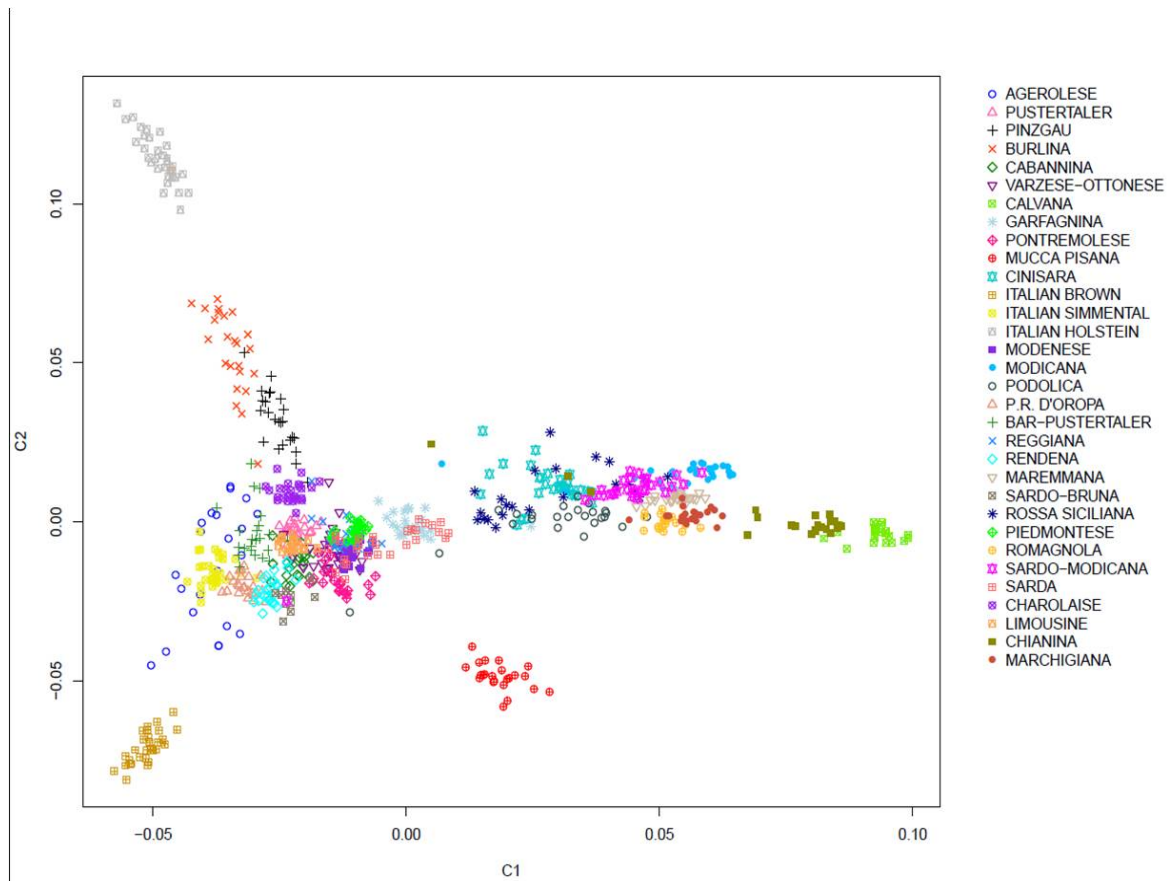


Figure 1. Genetic relationship defined with multidimensional scaling analysis among the Italian local cattle breeds.

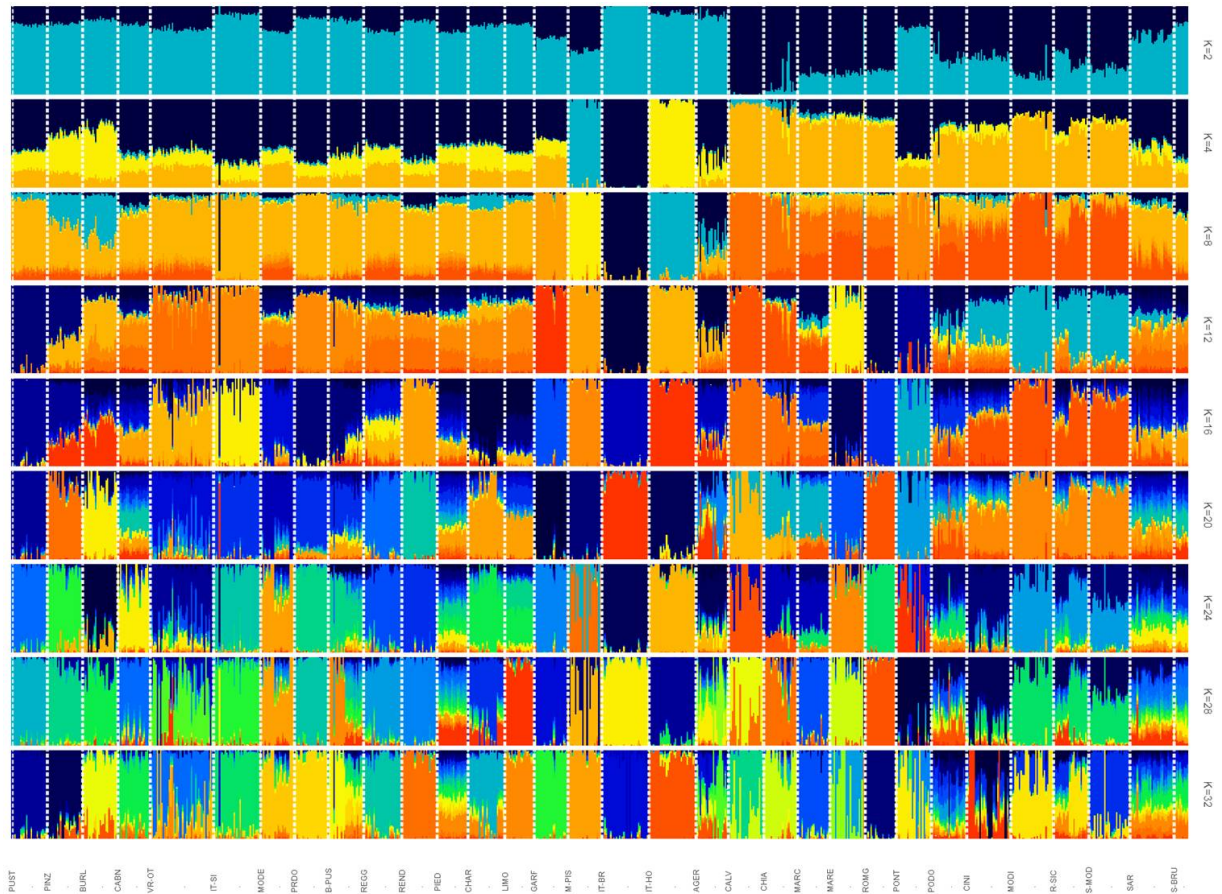


Figure 2. Model based clustering of the estimated membership fractions of individuals of the 32 breeds analyzed in each of the K inferred clusters, for $K = 2$ to $K = 32$. For full definition of breeds see Table 1.

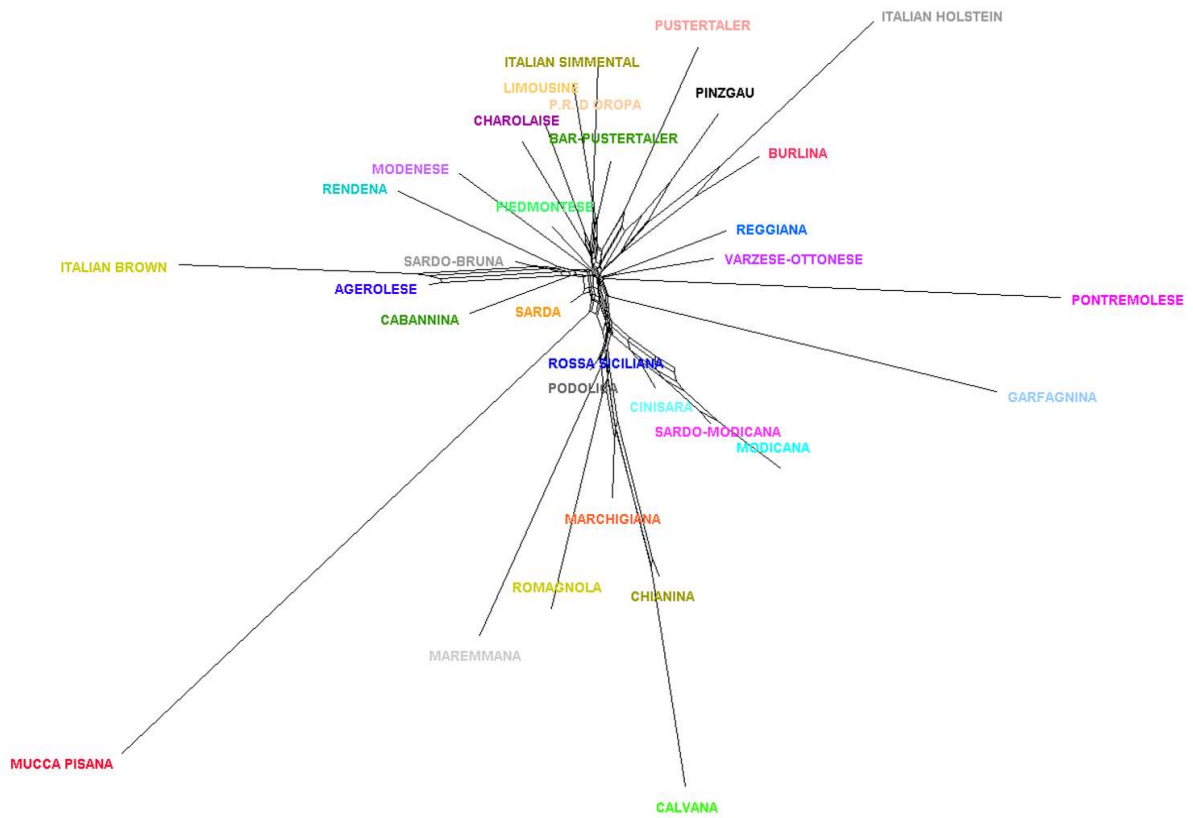


Figure 3. Relationship between breeds based on Reynold's genetic distance. An allele frequency-dependent distance metric (Reynolds) was used to construct a NeighborNet graph relating breeds.

List of References

- Alexander DH, Lange K, 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. 12:246.
- Excoffier L, Lischer HEL, 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 10:564-567.
- Ciani E, Crepaldi P, Nicoloso L, Lasagna E, Sarti FM, Moioli B, et al. 2014. Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. *Anim Genet*. 45:256–266.
- Gandini GC, Bagnato A, Leonarduzzi R, Pagnacco G, Cerutti F, 1993. Gestione della variabilità genetica in popolazioni bovine a numerosità fortemente ridotta: aspetti applicativi. *Proceeding 10th Conference ASPA, Bologna, 31 May / 3 June*, p. 621-625.
- Gautier M, Laloë D, Moazami-Goudarzi K, 2010. Insights into the Genetic History of French Cattle from Dense SNP Data on 47 Worldwide Breeds. *PLOS ONE*. 6(1):10.
- Hiemstra SJ, de Haas Y, Mäki-Tanila A, Gandini G, 2010. Local cattle breeds in Europe - Development of policies and strategies for self-sustaining breeds. Wageningen Academic Publishers. 2010;Wageningen 154. ISBN: 978-90-8686-144-6.
- Huson DH, Bryant D, 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 23:254–267.
- Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al, 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. 10, e1001258.
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al, 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*. 4:e5350.
- Nicoloso L, Bomba L, Colli L, Negrini R, Milanese M, Mazza R, et al, 2015. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet Sel Evol*. 47(1):62.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al, 2007. PLINK: A tool set for whole genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559–575.