

Phylogenetic Relationships of Three Italian Merino-Derived Sheep Breeds Evaluated through a Complete Mitogenome Analysis

Hovirag Lancioni¹, Piera Di Lorenzo², Simone Ceccobelli^{2,3}, Ugo A. Perego¹, Arianna Miglio⁴, Vincenzo Landi⁵, Maria T. Antognoni⁴, Francesca M. Sarti², Emiliano Lasagna^{2,*}, Alessandro Achilli^{1,*}

1 Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi di Perugia, Perugia, Italy, **2** Dipartimento di Biologia Applicata, Università degli Studi di Perugia, Perugia, Italy, **3** Dipartimento di Agronomia Animali Alimenti risorse Naturali e Ambiente, Università degli Studi di Padova, Padova, Italy, **4** Dipartimento di Patologia, Diagnostica e Clinica Veterinaria, Università degli Studi di Perugia, Perugia, Italy, **5** Departamento de Genética, Universidad de Córdoba, Córdoba, Spain

Abstract

In Italy, the crisis of the wool industry triggered the necessity to reconvert the two traditional Merino-derived breeds, Gentile di Puglia and Sopravissana, to meat production, by creating the Merinizzata Italiana. The aim of the present study was to assess the genetic diversity of these three Italian Merino-derived (IMd) breeds by examining the molecular information encoded in the maternally-inherited mitochondrial DNA (mtDNA). A parallel molecular investigation was performed on the putative paternal and maternal breeds, the Merino from Spain and the Appenninica from Italy, respectively, as well as on three unrelated dairy breeds (Sarda and Comisana from Italy, and Lacaune from France). Firstly, the mtDNA control region of 291 samples was analyzed. When comparing the overall genetic distances among the eight stocks, the three IMd breeds clustered together close to the Appenninica, thus confirming its parental role. Among the 90 IMd samples, 82 different haplotypes were observed, almost all belonging to haplogroup B, and only one to A. For 23 mtDNAs, including nine IMd, the analysis was then brought to the level of entire mitogenomes. Three distinct sub-haplogroups within B were found to encompass the IMd samples, with one clade (B1a2a1) apparently restricted to those sheep. Thus, despite experiencing a drastic reduction in number (mainly due to changes in breeding practices driven by the economy), the IMd breeds still represent a reservoir of distinctive mitochondrial variants, which could potentially contribute to the development of conservation and management programs of Italian sheep breeds.

Citation: Lancioni H, Di Lorenzo P, Ceccobelli S, Perego UA, Miglio A, et al. (2013) Phylogenetic Relationships of Three Italian Merino-Derived Sheep Breeds Evaluated through a Complete Mitogenome Analysis. PLoS ONE 8(9): e73712. doi:10.1371/journal.pone.0073712

Editor: David Caramelli, University of Florence, Italy

Received: April 10, 2013; **Accepted:** July 28, 2013; **Published:** September 9, 2013

Copyright: © 2013 Lancioni et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research received support from the Italian Ministry of the University: Progetti Ricerca Interesse Nazionale 2009 (to AA) and FIRB-Futuro in Ricerca 2008 and 2012 (Italian Ministry of the University) (to AA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors confirm that Alessandro Achilli is a PLOS ONE Editorial Board member. After having carefully read the journal's policy, the authors confirm that this does not alter their adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: elasagna @unipg.it (EL); alessandro.achilli@unipg.it (AA)

These authors contributed equally to this work.

Introduction

Following domestication in South West Asia approximately 8,000–9,000 years ago, sheep gradually populated a wide geographic range due to their ability to adapt to poor diets and their tolerance to extreme climatic conditions. However, quite recently, some local breeds have been lost by substitution or crossbreeding with commercial breeds. The loss of diversity in livestock species has important economic, ecological, and scientific implications, as well as social considerations. In Italy, sheep breeding has always played an important role, mainly in less-developed and rural areas, where local traditions and native breeds strongly influenced the production systems [1]. For about half a millennium, the two traditional Italian Merino-derived (IMd) breeds, Gentile di Puglia and Sopravissana, represented the most important sheep resource in central and southern Italy. The breed Gentile di Puglia (literally “Gentle Apulian”, in reference to its fine wool) originated from repeated introgression (starting in 1435) of original Spanish Merino rams into native ewes from

Apulia. The breed Sopravissana (literally “above Vissana”, see below) derived from sporadic introgression (1792) of Merino-Rambouillet rams (from France) into Vissana ewes (from Visso, in the Sibillini Mountains of the Apennines), a now extinct central Italian local breed that could be considered the ancestor of the extant Appenninica (from the Apennines) sheep [2,3]. In the last few decades, the increased demand for meat, the gradual abandonment of transhumance practices, and the advent of artificial fibers triggered the necessity to convert the IMd breeds, historically tied to transhumance in central-southern Apennines, to meat production. In many flocks, these sheep were used to create a large crossbred population that eventually resulted in the most popular Italian sheep for meat, officially established in 1989 and named Merinizzata Italiana [4–6]. An undesired consequence was the strong numerical reduction of the two local breeds (e.g. Sopravissana decreased from >1 million to a few thousand), which eventually became endangered, according to the European Council Regulation (EEC) No. 2078/92. Currently, a major

objective is the preservation of the residual genetic variability of the two threatened Merino-derived breeds (Sopravissana and Gentile di Puglia) and to avoid the improper registration of some Merinizzata Italiana animals within the two native breeds because of their similar morphology [6].

Molecular genetics has proven highly informative to better understand the (recent) evolutionary history of livestock breeds, as well as to determine the level of their genetic variability, which is an essential aspect to consider when defining conservation priorities and regional breed-specific programs [7–17]. The reported history of the three IMd breeds was recently confirmed by comparing their microsatellite profiles (based on 30 autosomal markers) with those of the Spanish Merino [18]. Here, we present additional information on the genetic history of these breeds obtained from the analysis of the maternally-inherited mitochondrial DNA (mtDNA). While the informative power of the mitogenome has been widely used to investigate human evolution [19], it has also brought new insights to the analysis of domesticated species [20–23]. Until now, molecular surveys of ovine mtDNA have been mostly restricted to the short hypervariable region, with the reported haplotypes resulting in the identification of a few maternal lineages, which are poorly correlated to specific geographical distributions [11,15,21,24–47]. The aim of the present study was to assess the matrilineal genetic diversity of the Italian Merino-derived breeds and to further investigate their origins through a refined analysis of their entire mitogenomes. The phylogenetic reconstruction was also enriched through a parallel analysis of two closely related breeds (Spanish Merino and Appenninica), and three unrelated dairy breeds (Lacaune, the most widely used dairy sheep in France; Comisana, a breed from Sicily mainly raised for milk; and Sarda, a Sardinian sheep primarily used to produce milk for distinctive cheeses, such as Pecorino Sardo).

Results

A total of 812 base pairs (bps) of the mitochondrial control region (from np 15,452 to np 16,263) were obtained for all the 291 samples (Table S1). Excluding ambiguous sites, the overall sequence alignment revealed 221 polymorphic sites (*S*), all represented by single nuclear polymorphisms (SNPs), with an overall nucleotide diversity (π) of 0.016 (Table 1). The average number of nucleotide differences (κ) between two randomly chosen sequences was 12.6. A total of 80 singletons (mutations from the reference sequence appearing only in a single animal) and 235 different haplotypes (nh) were identified, with an observed haplotype diversity (*Hd*) higher than 0.99 across all breeds (Table 1). When comparing haplotype and nucleotide diversity indices (calculated within each breed), no significant differences were observed. The highest haplotype diversity was detected in the Spanish Merino (*Hd* = 1.0), where all observed sequences were different, and the lowest in the Comisana (*Hd* = 0.920). However, the fewer Comisana haplotypes are characterized by a higher number of nucleotide variants, as indicated by their greatest variability in the nucleotide diversity (π = 0.028) (Table 1). As shown hereafter, the reason for this peculiarity becomes evident when analyzing the phylogenetic relationships of the control-region haplotypes. Using an analysis of molecular variance (AMOVA), we examined fixation indices in two (artificially created) population groups, one including the Merino related breeds (SM, AP, SO, GP, MI) and the other encompassing the remaining control-stocks (SA, CO, LA). The among-breeds component of genetic variation is higher for the control-stocks than for the “Merino-related” breeds (by ~14%). In fact, the vast

majority of the observed variance (98.93%) within those five breeds is attributable to differences among samples within breeds, with only 1.07% of differences represented among breeds (Table S2). The overall AMOVA analysis, which concomitantly incorporates a hierarchical grouping of populations, revealed no significant differentiation among the two groups (0.15%, Φ_{CT} = 0.0015, *P* = 0.34) and confirmed that the source of variation can be mostly (and significantly) attributed to differences among samples within breeds (91.08%, Φ_{ST} = 0.089, *P* < 0.01). According to the pairwise, population-by-population, genetic distances (Figure 1), Lacaune is closely related to all other breeds, whereas Comisana is the most genetically distant and it is also the most heterogeneous breed. A Principal Coordinate (PCo) analysis was also performed to better visualize the breed genetic relationships. Figure 2 illustrates the two PCos. In the third quadrant, where the Spanish Merino (SM) falls, the three IMd breeds (GP, SO and MI) and the Appenninica (AP) form a cluster of closely related breeds. We also observed that the first coordinate (PCo1) clearly separates the Comisana from the other breeds. The genetic differentiation of this breed is confirmed (and explained) by the network analysis of the control-region haplotypes (Figure 3, top). Most samples fall into a single clade with sub-branches radiating from a central node in a typical star-like fashion. The only notable exception is represented by few mtDNAs that cluster into a distinct branch separated by 35 mutations. This second cluster includes one Merinizzata Italiana and about a third of the Comisana samples (13 out of 45). The two distinct clades correspond to lineage affiliations that were previously reported as haplogroups B and A, respectively [24,35]. Of particular note is the network obtained by activating the frequency >1 option (Figure 3, bottom). This clearly shows that hardly any haplotype is shared among different breeds. The only two haplotypes found to be identical across the IMd breeds are observed in Sopravissana and Gentile di Puglia. Two additional haplotypes are shared between the Comisana and Lacaune breeds, and one more is shared between Gentile di Puglia and Lacaune. In this analysis, haplogroup A is represented by a single haplotype identical in 12 Comisana samples, while haplogroup B includes, among others, 28 shared haplotypes distributed among different breeds. As expected by the *Hd* value, the Spanish Merino are not represented in this network because their mtDNAs are all different.

On the basis of these control-region data, we selected for complete sequencing those samples with shared mtDNAs as well as the most divergent ones (with the aim of including the largest possible range of mtDNA variation). The 23 complete mitogenomes were compared with the sheep reference sequence (SRS, GenBank: NC_001941) along with 10 published records (GenBank: HM236174 - HM236183) (Table 2). Figure S2 illustrates the variation of nucleotide diversity (π) along the entire mitogenome. As expected [21], the highest diversity was observed around the control region (from np 15,437 to np 16,616), with a peak of ~0.04. The overall evolutionary history of the 33 mitogenomes was inferred by a parsimony approach, using the published argali (*Ovis ammon*) sequence (GenBank: HM236188) as an outgroup (Figure 4). As for samples sharing the same control-region haplotype (between nps 15,452 and 16,263), only one set (composed of Gentile di Puglia and Sopravissana) was found to be identical at the whole molecule level, while a single mutation (at np 11429) separates the Gentile di Puglia from the Sopravissana samples, and another set of mutations (at nps 9218 and 12305) distinguishes the Comisana from the Lacaune subjects. Lastly, several mutations separate the remaining samples that shared control-region haplotypes.

Table 1. Estimates of genetic diversities ^a and haplogroup frequencies ^b for each sheep breed.

Breed	ID	Origin	N ^a	π	nh	Hd	S	HgA ^b	HgB ^b
Sopravissana	SO	ITALY, CENTRE	30	0.012	29	0.998	80	0 (0.00)	30 (1.00)
Gentile di Puglia	GP	ITALY, SOUTH	30	0.012	28	0.995	69	0 (0.00)	30 (1.00)
Merinizzata Italiana	MI	ITALY, CENTRE-SOUTH	30	0.015	27	0.991	95	1 (0.03)	29 (0.97)
Appenninica	AP	ITALY, CENTRE	30	0.012	25	0.989	71	0 (0.00)	30 (1.00)
Spanish Merino	SM	SPAIN	26	0.012	26	1.000	70	0 (0.00)	26 (1.00)
Comisana	CO	ITALY, SICILY	45	0.028	25	0.920	85	13 (0.29)	32 (0.71)
Sarda	SA	ITALY, SARDINIA	30	0.011	22	0.977	56	0 (0.00)	30 (1.00)
Lacaune	LA	FRANCE, ROQUEFORT	70	0.012	63	0.997	112	0 (0.00)	70 (1.00)
TOTAL			291	0.016	235	0.997	221	14 (0.05)	277 (0.95)

^aN = number of analyzed samples; π = nucleotide diversity; nh = number of unique haplotypes; Hd = haplotype diversity; S = number of polymorphic sites.

^bRelative frequencies are in parentheses.

doi:10.1371/journal.pone.0073712.t001

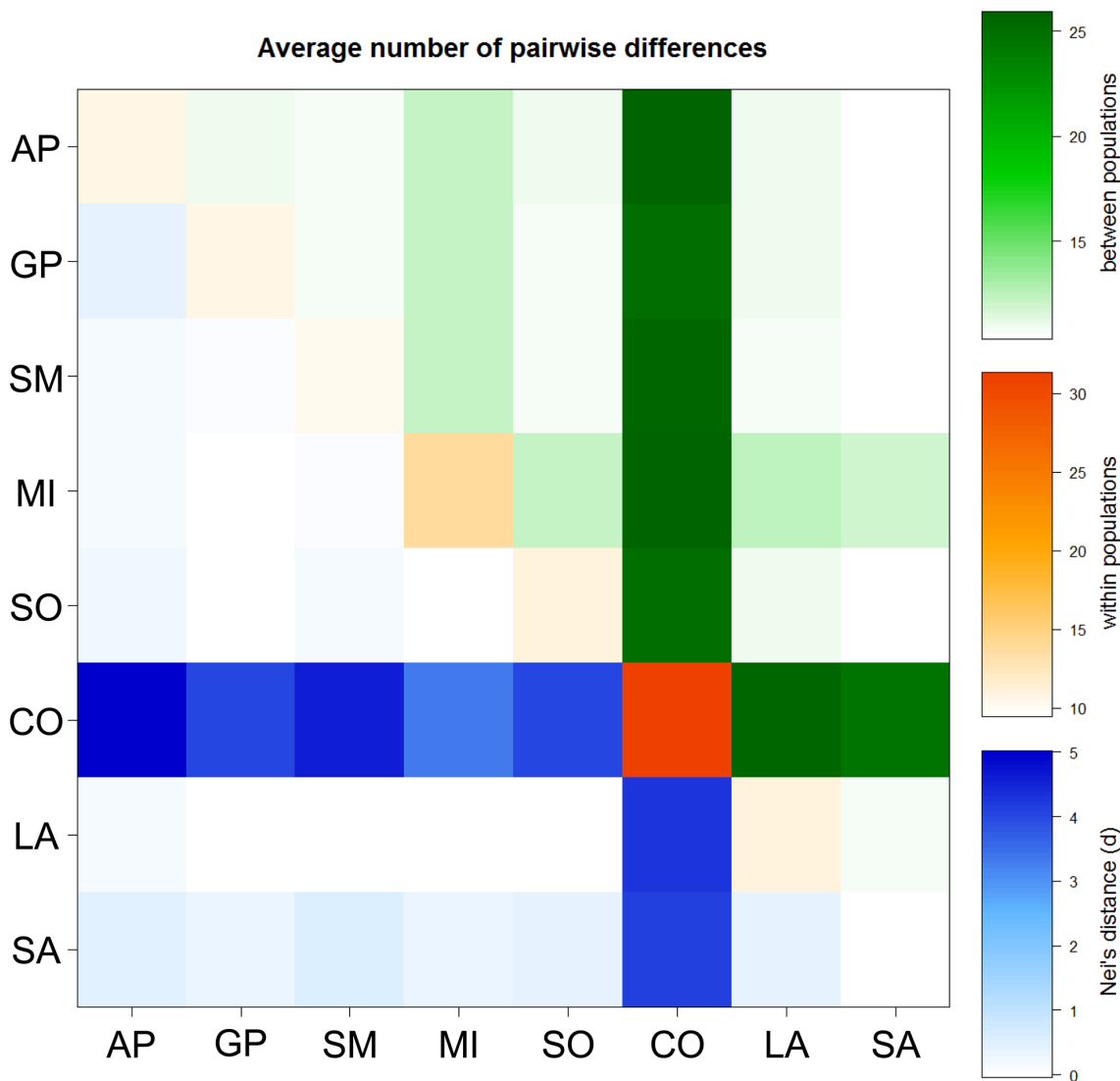


Figure 1. Plot of pairwise population genetic distances. Data were obtained by the concomitant analysis of all breeds subdivided into two groups: (1) SM, AP, SO, GP, MI; (2) SA, CO, LA.
doi:10.1371/journal.pone.0073712.g001

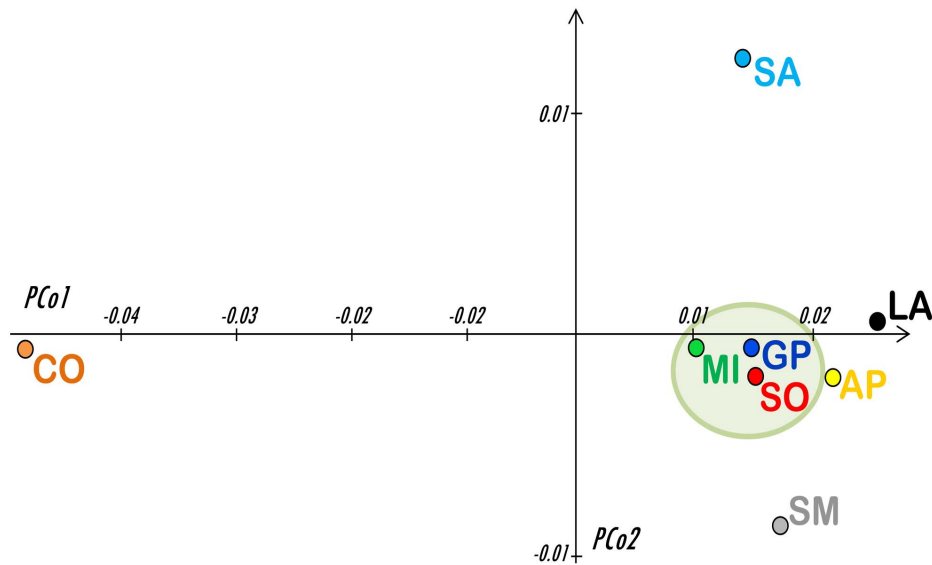


Figure 2. PCo analysis. The plot is based on pairwise genetic distances among the eight breeds (indicated by different colors) and based on 291 control-region haplotypes. The IMd cluster is indicated by the oval. doi:10.1371/journal.pone.0073712.g002

The tree topology reveals two major clusters (here named A'B and C'D'E) that radiate early in the tree and include all the five haplogroups known to date. Our samples are confirmed as belonging to the A and B branches, but spread across different derived sub-groups. Haplogroup A includes one Merinizzata Italiana and shows a major (most frequently represented) sub-clade, here named A1, encompassing six Comisana samples. Haplogroup A was first thought to be confined to animals with an Asiatic origin, [47], but it was later identified, although at a much lower frequency, in European breeds as well [15,21,34], such as the British Romney (HM236175), and breeds from the south western fringe of Europe including the Spanish Merino (HM236174). Haplogroup B, which is considered a typical European lineage, exhibits a more complex sub-structure with two early splits (one marked by a Gentile di Puglia) that predate a major sub-clade named B1a. Excluding two previously published Karakas ewes from Turkey, the B1a mitogenomes cluster into two sub-groups. B1a1 contains the SRS and three of our samples (grouped as B1a1a): two Appenninica and one Sopravissana. B1a2, which also includes various breeds, could be further subdivided into B1a2a and B1a2b. It is worth noting that a coding transition at np 7983 discriminates a clade (B1a2a1) made only of IMd mtDNAs. The remaining six published mtDNAs also allowed us to propose a preliminary full-mitogenome phylogeny of the three remaining sheep haplogroups (C, D and E) identified thus far. Even if the complete mitogenome of each haplogroup was previously published [21], through the updated phylogeny reported in this study, it is possible to define all unique clades by a specific motif (relative to SRS) encompassing both the coding and the control regions (Table S3).

Discussion

Currently, there are just a few genetic studies aiming to improve the knowledge of the genetic composition of Italian sheep breeds and often they are limited to the analyses of the nuclear genome [18,48–51]. Mitochondrial data for Italian breeds were previously reported only by Pariset and colleagues [45] that performed phylogenetic analyses on mtDNA control-region haplotypes and

nuclear polymorphisms of Gentile di Puglia and four other Italian sheep breeds. The aim of the present study was to assess the diversity of IMd breeds and to obtain more information on their maternal origins. A comparative mtDNA analysis included also two closely related breeds (Spanish Merino and Appenninica) and a control-group of two (insular) Italian and one French breeds (Comisana, Sarda and Lacaune, respectively). The mtDNA highest variable segment was preliminarily analyzed in order to identify molecular similarities and phylogenetic relationships. Genetic diversity considered in terms of number of haplotypes and nucleotide diversities showed the same high level as presented in previous ovine mtDNA studies [11,24,30,31]. When comparing the overall genetic distance of the eight analyzed stocks, the three IMd breeds form a closely related cluster near to their female ancestor breed (Appenninica), which might reflect recent maternal gene flow among them. Once distributed throughout a median-joining network, the numerous haplotypes resulted in two haplogroups (A and B), which are consistent with previous studies on domestic sheep breeds in Italy [45] and in Western Eurasia [15,24,30,31,35,45].

The predominant haplogroup B (95% in total; 99% in IMd breeds) shows a star-like pattern typical of livestock species and interpreted as recent population expansions (e.g. following domestication) [52]. Haplogroup A (5% in total; 1% in IMd breeds), herein identified mostly in Comisana and in one sample (#07) of the Merinizzata Italiana breed, has been previously found in a Merino sample and in (few) other Iberian breeds. While the mtDNA control region initially contributed predominantly to the description of sheep genetic diversity (Figure S2), the detection of several additional variants was made possible through the molecular analysis of the entire mitochondrial genome. Thus, the IMd mitogenomes were assigned to different sub-clades, one of them (B1a2a1) comprising only IMd samples. Moreover, the identical nature of just one of the five shared control-region haplotypes (shown by samples reported as #17 and #18, Figure 4) was observed on the entire mtDNA, while the remaining cases revealed at least one coding region mutation that discriminates between two samples. Thus, only at the maximum possible level of resolution can the extent of mtDNA information be fully exploited,

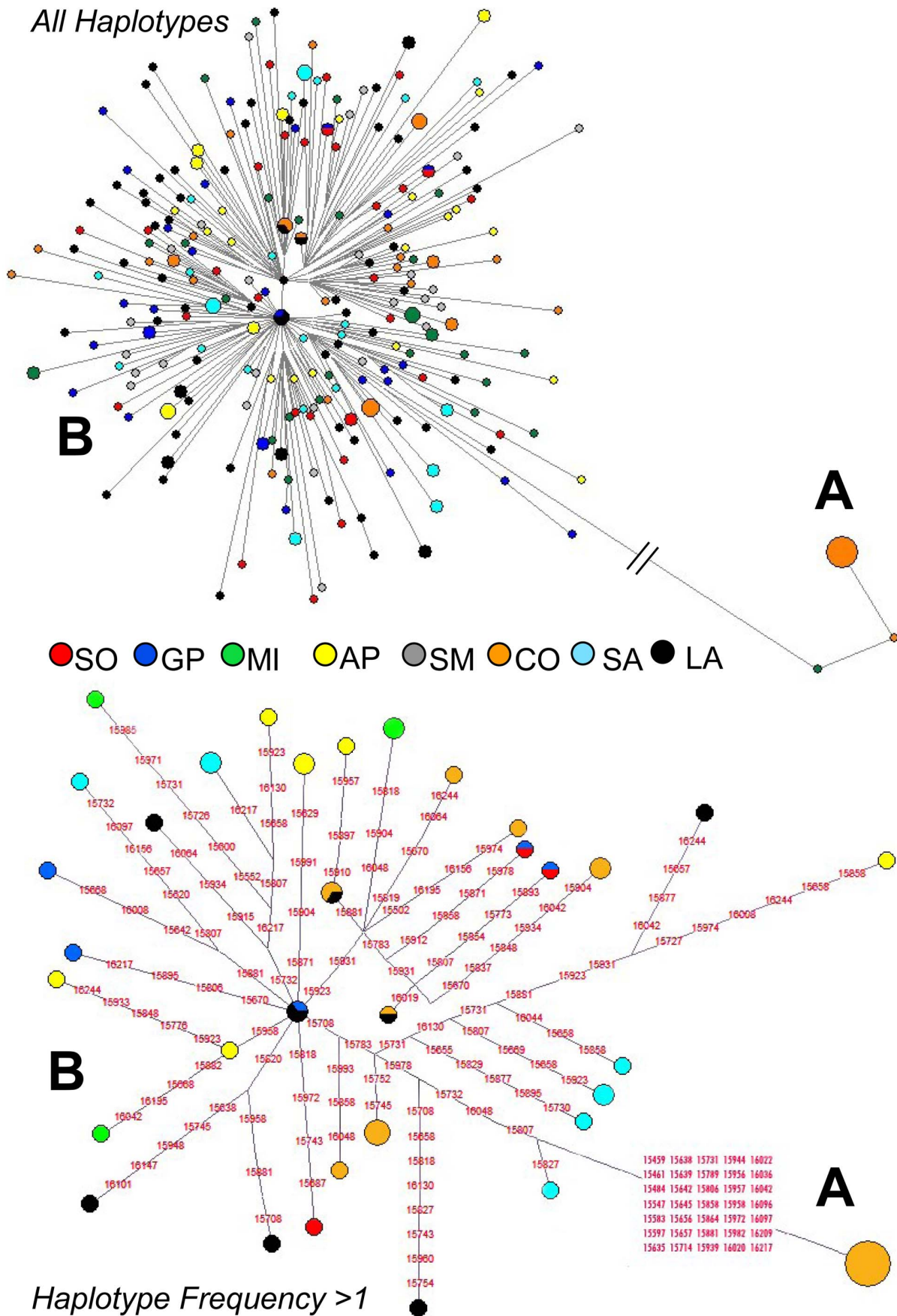


Figure 3. Median-joining networks of the sheep control-region sequences. The top network was obtained from the entire dataset, while the bottom one was acquired by activating the haplotype frequency>1 option.
doi:10.1371/journal.pone.0073712.g003

identifying unexpected genetic variants that would otherwise go undetected.

Conclusion

The present study is the first example of a mitochondrial phylogenetic analysis of Italian sheep breeds evaluated at the maximum level of resolution. Based on the overall genetic distances calculated from control-region haplotypes and graphed

through the PCo analysis, the IMd breeds appear deeply entangled within each other and genetically closer to the Appenninica (their female ancestors) and to the Spanish Merino, with respect to the control breeds analyzed here. In agreement with the first genetic records [18], our mitochondrial data demonstrate the influence of local Italian breeds (i.e. Appenninica) and support the historical origin of the Italian derived Merino breed from the Spanish populations mentioned in the chronicles.

Table 2. Sources and haplogroup affiliation of the sheep complete mtDNA sequences.

	<i>ID#^a</i>	<i>Original name</i>	<i>Haplogroup</i>	<i>Breed</i>	<i>GenBank</i>
	Outgroup	<i>Ovis ammon</i>			HM236188
Novel mitogenomes	01	CO014	A1	Comisana	KF302440
	02	CO015	A1	Comisana	KF302441
	03	CO020	A1	Comisana	KF302442
	04	CO021	A1	Comisana	KF302443
	05	CO026	A1	Comisana	KF302444
	06	CO017	A1	Comisana	KF302445
	07	MI066	A	Merinizzata Italiana	KF302446
	08	LA021	B	Lacaune	KF302447
	09	GP117	B1	Gentile di Puglia	KF302448
	10	SO006	B1a1a	Sopravissana	KF302449
	11	AP038	B1a1a	Appenninica	KF302450
	12	AP053	B1a1a	Appenninica	KF302451
	13	CO002	B1a2a	Comisana	KF302452
	14	LA068	B1a2a	Lacaune	KF302453
	15	SO012	B1a2a1	Sopravissana	KF302454
	16	GP092	B1a2a1	Gentile di Puglia	KF302455
	17	SO030	B1a2a1a	Sopravissana	KF302456
	18	GP105	B1a2a1a	Gentile di Puglia	KF302457
	19	SO009	B1a2a1a	Sopravissana	KF302458
	20	MI075	B1a2a1a	Merinizzata Italiana	KF302459
	21	LA047	B1a2b	Lacaune	KF302460
	22	CO027	B1a2b	Comisana	KF302461
	23	CO046	B1a2b	Comisana	KF302462
Published mitogenomes	HM236174	HM236174	A	Merino	HM236174
	HM236175	HM236175	A	Romney	HM236175
	HM236176	HM236176	B1a	Karakas	HM236176
	HM236177	HM236177	B1a	Karakas	HM236177
	HM236178	HM236178	C	Karakas	HM236178
	HM236179	HM236179	C	Morkaraman	HM236179
	HM236182	HM236182	E	Awassi	HM236182
	HM236183	HM236183	E	Tuj	HM236183
	HM236180	HM236180	D	Morkaraman	HM236180
	HM236181	HM236181	D	Morkaraman	HM236181
SRS	NC_001941	B1a1	Merinolandschaf	NC_001941	

^aLabeled as in Figure 4.
doi:10.1371/journal.pone.0073712.t002

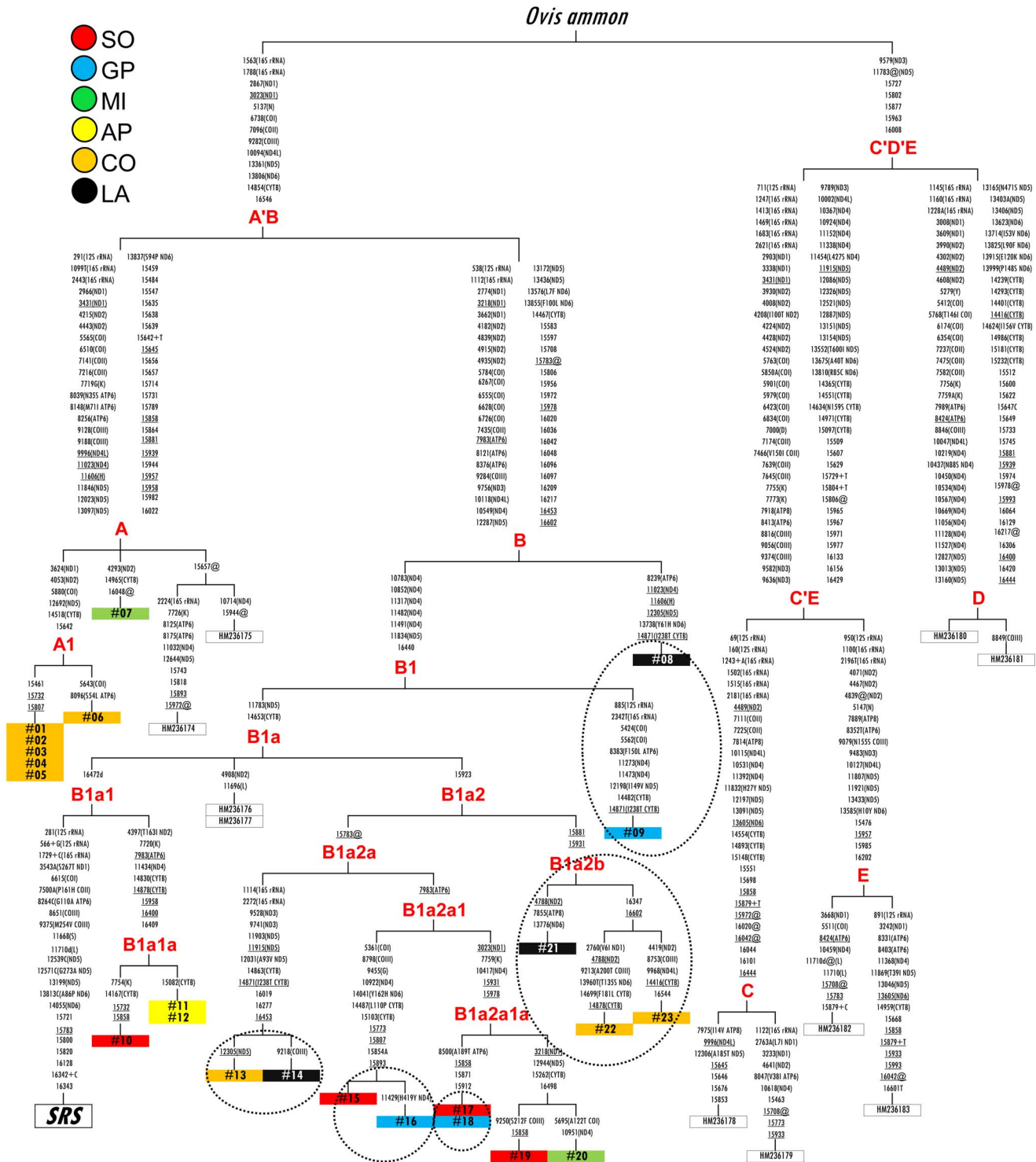


Figure 4. A parsimonious tree of complete mtDNA sequences from sheep. This tree was rooted by using the published argali (*O. ammon*) sequence (GenBank: HM236188). Mutations are shown on the branches and are numbered according to the sheep reference sequence (SRS, GenBank: NC_001941.1). Mutations are transitions unless a base is explicitly indicated. Suffixes indicate transversions (to A, G, C, or T), while “+” and “d” denote insertions and deletions, respectively. Recurrent mutations are underlined. The samples are highlighted with different colors to identify their breed of origin. Control-region shared haplotypes (from np 15,452 to np 16,263; Figure 3) are indicated by dotted ovals. Additional information regarding each mtDNA is available in Table 2. doi:10.1371/journal.pone.0073712.g004

These genetic similarities are also confirmed when looking at the haplogroup affiliation of these breeds that, with the only notable exemption of a unique Merinizzata Italiana (#07 in Table 2),

belong to the same (typically European) haplogroup (B). However, when dealing with the time frame of about a half millennium, since the origin of the Imd breeds, our analysis must shift to the

entire mitogenome in order to discriminate between different sub-haplogroups or even between single haplotypes. The nine IMd mitogenomes belong to different sub-clades, mostly to B1a2a1 (6 out of 9) and reveal only one haplotype shared among them and/or with the closely related Appenninica and Spanish Merino stocks. Thus, it is likely that, even if the IMd breeds share a common maternal origin, many different haplotypes, belonging to different haplogroups (A and B) and sub-haplogroups (B1*, B1a1, B1a2a1, B1a2a1a), could account for the present mitochondrial gene pool of these Italian stocks. It is apparent that mtDNA variation reflects only maternal inheritance and does not reveal any male-mediated gene flow [16]. However, the novel mitogenomic data of these IMd sheep, as reported here, represents a first step toward the complete genetic characterization of Italian breeds, as well as the ovine mitochondrial phylogeny.

Even though the first complete mtDNA sequence for sheep has been available since 1998 [41], all of the molecular and evolutionary studies of sheep mitochondrial genome have generally focused on a short fragment of the control-region sequence [30,35,45,53]. Very few authors have reported mtDNA phylogeny data based on larger segments encompassing coding regions [11,24,32,46]. It was not until very recently that Meadows and colleagues published the first sheep phylogeny based on a mitogenome panel [21], revealing that the control region was the mtDNA component, which contributed the greatest amount of support to the basal topology of the sheep phylogeny. However, our mitogenome comparison clearly shows the limitations of previous sheep mtDNA studies that confine analysis to a minor fragment of the control region. This approach could still be justified in cases where it was only desired to distinguish between very divergent haplotypes, i.e. belonging to different haplogroups. However, in cases where it is necessary to discriminate between different sub-haplogroups and/or single haplotypes, complete mitogenome analysis is the method of choice. Such enhanced resolution could eventually discriminate more recent evolutionary events, thus also providing information on whether the deliberate selection during the development of modern breeds (after initial domestication) affected mitochondrial genomes. In the present full-mitogenome study, besides the successful discrimination of the sample pairs (four out of five, 80%) sharing the same control-region mutational motif, we identified a typical IMd sub-clade (B1a2a1) and the novel A1 clade identified only in the Comisana breed.

Here, we also provide for the first time a list of coding- and control-region molecular markers for the previously defined haplogroups, as well as for a number of newly defined sub-haplogroups. This work contributes information that is potentially useful in conservation and management programs where genetic information plays a role of primary importance. Finally, this study shows that when employing entire mtDNA sequences, the phylogenetic resolution of animal mtDNA trees is improved; meaning that, to achieve sustainable management of genetic resources, these local breeds should be valued not only in economic terms but also by taking into account their worth as reservoirs of unique diversity [7,54].

Materials and Methods

Ethics Statement

All experimental procedures were reviewed and approved by the Animal Research Ethics Committee of the University of Perugia.

Animal Sampling

The IMd samples (30 Gentile di Puglia, 30 Sopravissana and 30 Merinizzata Italiana) were described in a previous paper [18], as well as Appenninica (30) and Spanish Merino (26) samples. By using the same procedures of sampling and DNA extraction, we included here three unrelated breeds used as control groups: Sarda (30), Comisana (45), and Lacaune (70) (Figure S1). They are commercial breeds strongly selected for milk production, as opposed to the IMd that are selected for their meat and wool.

All the animals were registered in the Herd Book of the corresponding breed. To ensure that the sampling was representative, the animals were randomly selected from diverse flocks. Additionally, in order to minimize the potential for sampling closely related individuals, we selected Merino-related animals that were unrelated for at least two generations and asked the shepherds about the genealogical relationships of the animals belonging to the control breeds.

Analysis of mtDNA Control-regions

The complete sheep mitochondrial sequence NC_001941 [41] was used to design the primer set 15346for-157rev (Table S4) for the amplification of the control region from np 15,346 to np 00157. DNA amplifications were performed by following previously reported protocols [18]. PCR products were first purified using the ExoSAP-IT[®] enzymatic system ExoSAP-IT (USB Corporation, Cleveland, OH, USA) and sent to BMR-Genomics s.r.l. (www.bmr-genomics.com) for standard direct Sanger di-deoxy-sequencing (in one direction) triggered by the internal primer 15393for (5'ACTATCAACACCCAAAGCTG3'). Electropherograms were visualized, edited and aligned using the Sequencher[™] 5.10 software (www.genecodes.com). New PCR amplifications and sequencing reactions were further performed to resolve sequence ambiguities. All sequences were trimmed to the shortest one. Fragments of 812 base pairs (from np 15,452 to np 16,263) eventually resulted from the standardization of the sequencing results.

Indices such as Haplotype diversity (H_d), nucleotide diversity (π) and average number of nucleotide differences (κ) were estimated with DnaSP 5.10 software [55]. Intra- as well as inter-population comparisons were performed based on the number of pairwise differences between sequences and molecular variance (AMOVA). Calculations were performed using the Arlequin v. 3.5 software package [56]. To minimize biases associated with a higher mutation rate for mtDNA, we calculated genetic distances using a Tamura-Nei distance with high among site rate heterogeneity ($\gamma = 0.22$), as in [57]. The PCo analysis on pairwise, population-by-population, genetic distances was performed using Excel software implemented by GenAlEx 6.4 software [58].

The evolutionary relationships were investigated through a median-joining network of control-region haplotypes constructed with the Network 4.6 software (www.fluxus-engineering.com) by using the reduced median algorithm ($\rho = 2$), followed by the median-joining algorithm ($\epsilon = 0$). Nucleotide weighting (ω) was adjusted to reflect the difference in mutational frequency among indels ($\omega = 30$), transversions ($\omega = 20$), and transitions ($\omega = 10$), where the least-common event received the highest value. The maximum parsimony (MP) calculation was used to remove unnecessary median vectors and to avoid reticulations, which could be switched off in the results display.

Analysis of Entire Mitogenomes

Control-region haplotypes were particularly useful to select mtDNAs for whole genome sequencing, which was performed as previously described for humans [59], cattle [22], and horse [23].

The oligonucleotides used to amplify and sequence sheep mitochondrial genomes are shown in Table S4. The GenBank BLAST tool was used to investigate if these oligonucleotides matched NuMtS (nuclear insertions of mitochondrial sequences) in order to avoid amplification of these regions [60]. The phylogeny including 34 sheep mitogenomes was built following a maximum parsimony approach, as described elsewhere [23,61], and rooted by using the published argali (*O. ammon*) sequence (GenBank: HM236188).

Accession Numbers

The GenBank accession numbers for the 291 mtDNA control-region sequences reported in this paper are KF228586-KF228853, while those for the 23 complete mitochondrial genomes are KF302440-KF302462.

Supporting Information

Figure S1 Geographical sampling areas of the 291 sheep samples.

(PDF)

Figure S2 Nucleotide diversity variation (π) along the entire mitogenome.

A schematic linearized genetic map of the mitogenome is presented on the top.

(PDF)

Table S1 Control-region haplotypes and haplogroup classification of the 291 sheep mtDNAs.

(XLSX)

Table S2 AMOVA results for the two breed groups considered.

(DOCX)

Table S3 Diagnostic coding- and control-region mutational motifs of sheep mtDNA haplogroups and sub-haplogroups.

(DOCX)

Table S4 Oligonucleotides used to amplify and sequence the entire sheep mitogenome.

(DOCX)

Acknowledgments

The authors are grateful to all farmers and collaborators for providing biological specimens, to Jayne E. Ekins for proofreading the manuscript, and to two anonymous reviewers for their helpful comments and suggestions.

Author Contributions

Conceived and designed the experiments: HL FMS EL AA. Performed the experiments: HL PDL SC AM. Analyzed the data: HL UAP EL AA. Contributed reagents/materials/analysis tools: VL MTA FMS EL AA. Performed the collection of biological samples: AM VL MTA EL. Wrote the manuscript with contributions from the other authors: HL UAP EL AA.

References

- de Rancourt M, Fois N, Lavin M, Tchakerian E, Vallerand F (2006) Mediterranean sheep and goats production: An uncertain future. *Small Rumin Res* 62: 167–179.
- Casasús I, Rosati A (2011) *Animal Farming and Environmental Interactions in the Mediterranean Region*: Wageningen Academic Publishers.
- Sarti DM (1996) Le razze derivate Merinos ad attitudine carne nel centro sud europeo. *L'Allevatore di Ovini e Caprini* 7/8: 1–2.
- Panella F, Sarti FM, Lasagna E, Renieri C, Antonini M (2006) The old Italian Merino-derived breeds and their role in landscape conservation in typical production and in the maintenance of traditional culture. *Livestock farming systems*. EAAP publication. 247–252.
- Pieramati C, Lasagna E, Panella F, Piro F, Giontella A, et al. (2011) Suitability of linear scoring in meat sheep: the practical case of Merinizzata Italiana breed. *Ital J Anim Sci* 10: 51–53.
- Sarti F, Lasagna E, Panella F, Lebboroni G, Renieri C (2006) Wool quality in Gentile di Puglia sheep breed as measure of genetic integrity. *Ital J Anim Sci* 5: 371–376.
- Mendelsohn R (2003) The challenge of conserving indigenous domesticated animals. *Ecol Econ* 45: 501–510.
- Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, et al. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4: e4668.
- Meadows JR, Hanotte O, Drögemüller C, Calvo J, Godfrey R, et al. (2006) Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. *Anim Genet* 37: 444–453.
- Tapio M, Tapio I, Grisli Z, Holm LE, Jeppsson S, et al. (2005) Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Mol Ecol* 14: 3951–3963.
- Chen SY, Duan ZY, Sha T, Xiangyu J, Wu SF, et al. (2006) Origin, genetic diversity, and population structure of Chinese domestic sheep. *Gene* 376: 216–223.
- Lawson Handley IJ, Byrne K, Santucci F, Townsend S, Taylor M, et al. (2007) Genetic structure of European sheep breeds. *Heredity (Edinb)* 99: 620–631.
- Kusza S, Nagy I, Sasvari Z, Stigel A, Nemeth T, et al. (2008) Genetic diversity and population structure of Tsigai and Zackel type of sheep breeds in the Central-, Eastern- and Southern-European regions. *Small Rumin Res* 78: 13–23.
- Peter C, Bruford M, Perez T, Dalamitra S, Hewitt G, et al. (2007) Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Anim Genet* 38: 37–44.
- Pereira F, Davis S, Pereira L, McEvoy B, Bradley D, et al. (2006) Genetic signatures of a Mediterranean influence in Iberian peninsula sheep husbandry. *Mol Biol Evol* 23: 1420–1426.
- Bruford MW, Bradley DG, Luikart G (2003) DNA markers reveal the complexity of livestock domestication. *Nat Rev Genet* 4: 900–910.
- Rege J, Gibson J (2003) Animal genetic resources and economic development: issues in relation to economic valuation. *Ecol Econ* 45: 319–330.
- Lasagna E, Bianchi M, Ceccobelli S, Landi V, Martinez A, et al. (2011) Genetic relationships and population structure in three Italian Merino-derived sheep breeds. *Small Rumin Res* 96: 111–119.
- Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22: 339–345.
- Achilli A, Bonfiglio S, Olivieri A, Malusà A, Pala M, et al. (2009) The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS One* 4: e5753.
- Meadows JR, Hiendleder S, Kijas JW (2011) Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity (Edinb)* 106: 700–706.
- Achilli A, Olivieri A, Pellicchia M, Uboldi C, Colli L, et al. (2008) Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr Biol* 18: R157–158.
- Achilli A, Olivieri A, Soares P, Lancioni H, Kashani BH, et al. (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc Natl Acad Sci U S A* 109: 2449–2454.
- Meadows JR, Cemal I, Karaca O, Gootwine E, Kijas JW (2007) Five ovine mitochondrial lineages identified from sheep breeds of the Near East. *Genetics* 175: 1371–1379.
- Niemi M, Bläuer A, Iso-Touru T, Nyström V, Harjula J, et al. (2013) Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (*Ovis aries*) in Finland: comparison with contemporary sheep breeds. *Genet Sel Evol* 45: 2.
- Ferencakovic M, Curik I, Pérez-Pardal L, Royo IJ, Cubric-Curik V, et al. (2012) Mitochondrial DNA and Y-chromosome diversity in East Adriatic sheep. *Anim Genet* 4: 184–192.
- Gonçalves J, Pereira F, Amorim A, van Asch B (2012) New method for the simultaneous identification of cow, sheep, goat, and water buffalo in dairy products by analysis of short species-specific mitochondrial DNA targets. *J Agric Food Chem* 60: 10480–10485.
- Olivieri C, Ermini L, Rizzi E, Corti G, Luciani S, et al. (2012) Phylogenetic position of a copper age sheep (*Ovis aries*) mitochondrial DNA. *PLoS One* 7: e33792.
- Zhao Y, Zhao E, Zhang N, Duan C (2011) Mitochondrial DNA diversity, origin, and phylogenetic relationships of three Chinese large-fat-tailed sheep breeds. *Trop Anim Health Prod* 43: 1405–1410.
- Cinkulov M, Popovski Z, Porcu K, Tanaskovska B, Hodzić A, et al. (2008) Genetic diversity and structure of the West Balkan Pramenka sheep types as

- revealed by microsatellite and mitochondrial DNA analysis. *J Anim Breed Genet* 125: 417–426.
31. Pedrosa S, Arranz J, Brito N, Molina A, San Primitivo F, et al. (2007) Mitochondrial diversity and the origin of Iberian sheep. *Genet Sel Evol* 39: 91–103.
 32. Pardeshi VC, Kadoo NY, Sainani MN, Meadows JR, Kijas JW, et al. (2007) Mitochondrial haplotypes reveal a strong genetic structure for three Indian sheep breeds. *Anim Genet* 38: 460–466.
 33. Loehr J, Worley K, Grapputo A, Carey J, Veitch A, et al. (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *J Evol Biol* 19: 419–430.
 34. Bruford MW, Townsend SJ (2006) Mitochondrial DNA diversity in modern sheep: Implications for domestication. In: Zeder MA, Bradley DG, Emshwiller E, Smith BD, editors. *Documenting Domestication: New Genetic and Archaeological Paradigm*. California, USA: University of California Press. 307–317.
 35. Tapio M, Marzanov N, Ozerov M, Cinkulov M, Gonzarenko G, et al. (2006) Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Mol Biol Evol* 23: 1776–1783.
 36. Luo YZ, Cheng SR, Batsuuri L, Badamdorj D, Olivier H, et al. (2005) [Origin and genetic diversity of Mongolian and Chinese sheep using mitochondrial DNA D-loop sequences]. *Yi Chuan Xue Bao* 32: 1256–1265.
 37. Guo J, Du LX, Ma YH, Guan WJ, Li HB, et al. (2005) A novel maternal lineage revealed in sheep (*Ovis aries*). *Anim Genet* 36: 331–336.
 38. Zhao X, Li N, Guo W, Hu X, Liu Z, et al. (2004) Further evidence for paternal inheritance of mitochondrial DNA in the sheep (*Ovis aries*). *Heredity (Edinb)* 93: 399–403.
 39. Tserenbataa T, Ramey RR, Ryder OA, Quinn TW, Reading RP (2004) A population genetic comparison of argali sheep (*Ovis ammon*) in Mongolia using the ND5 gene of mitochondrial DNA; implications for conservation. *Mol Ecol* 13: 1333–1339.
 40. Hiendleder S, Mainz K, Plante Y, Lewalski H (1998) Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. *J Hered* 89: 113–120.
 41. Hiendleder S, Lewalski H, Wassmuth R, Janke A (1998) The complete mitochondrial DNA sequence of the domestic sheep (*Ovis aries*) and comparison with the other major ovine haplotype. *J Mol Evol* 47: 441–448.
 42. Ramey RR (1995) Mitochondrial DNA variation, population structure, and evolution of mountain sheep in the south-western United States and Mexico. *Mol Ecol* 4: 429–439.
 43. Chen H, Leibenguth F (1995) Restriction endonuclease analysis of mitochondrial DNA of three farm animal species: cattle, sheep and goat. *Comp Biochem Physiol B Biochem Mol Biol* 111: 643–649.
 44. Upholt WB, Dawid IB (1977) Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution in the D loop region. *Cell* 11: 571–583.
 45. Pariset L, Mariotti M, Gargani M, Joost S, Negrini R, et al. (2011) Genetic diversity of sheep breeds from Albania, Greece, and Italy assessed by mitochondrial DNA and nuclear polymorphisms (SNPs). *Scientific World Journal* 11: 1641–1659.
 46. Meadows J, Li K, Kantanen J, Tapio M, Sipos W, et al. (2005) Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *J Hered* 96: 494–501.
 47. Hiendleder S, Kaupe B, Wassmuth R, Janke A (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc Biol Sci* 269: 893–904.
 48. Moiola B, Napolitano F, Orru L, Catillo G (2006) Analysis of the genetic diversity between Gentile di Puglia, Sopravivanna and Sarda sheep breeds using microsatellite markers. *Ital J Anim Sci* 5: 73–78.
 49. Ciani E, Ciampolini R, D'Andrea M, Castellana E, Cecchi F, et al. (2013) Analysis of genetic variability within and among Italian sheep breeds reveals population stratification and suggests the presence of a phylogeographic gradient. *Small Rumin Res* 112: 21–27.
 50. Dalvit C, De Marchi M, Zanetti E, Cassandro M (2009) Genetic variation and population structure of Italian native sheep breeds undergoing in situ conservation. *J Anim Sci* 87: 3837–3844.
 51. Passero A, Ciampolini R, D'Andrea M, Lacalandra G, Macciotta N, et al. (2012) A comprehensive survey of genetic variation of Italian sheep breeds at 50 K SNP loci: implications for breed management and conservation. *Reprod Domest Anim* 47: 509–509.
 52. Lavery S, Moritz C, Fielder D (1996) Genetic patterns suggest exponential population growth in a declining species. *Mol Biol Evol* 13: 1106–1113.
 53. Pedrosa S, Uzun M, Arranz JJ, Gutiérrez-Gil B, San Primitivo F, et al. (2005) Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proc Biol Sci* 272: 2211–2217.
 54. Taberlet P, Valentini A, Rezaei HR, Naderi S, Pompanon F, et al. (2008) Are cattle, sheep, and goats endangered species? *Mol Ecol* 17: 275–284.
 55. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
 56. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47–50.
 57. Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF (2004) Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat Genet* 36: 1122–1125.
 58. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288–295.
 59. Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, et al. (2012) Decrypting the mitochondrial gene pool of modern Panamanians. *PLoS One* 7: e38337.
 60. Nergadze SG, Lupotto M, Pellanda P, Santagostino M, Vitelli V, et al. (2010) Mitochondrial DNA insertions in the nuclear horse genome. *Anim Genet* 41 Suppl 2: 176–185.
 61. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75: 910–918.