



Assessment of genetic diversity and structure of major sheep breeds from Pakistan



A.N. Naqvi^{a,*}, S. Mahmood^b, S.M.F. Vahidi^c, S.M. Abbas^d, Y.T. Utsunomiya^e, J.F. Garcia^e, Kathiravan Periasamy^f

^a Faculty of Life Sciences, Karakoram International University, Gilgit, Baltistan, Pakistan

^b Pakistan Agricultural Research Council, Islamabad, Pakistan

^c Agricultural Biotechnology Research Institute of Iran (ABRII), North Branch, Rasht 41635-4115, Iran

^d National University of Computer and Emerging Sciences, Islamabad, Pakistan

^e UNESP—Univ Estadual Paulista, Faculdade de Medicina Veterinária de Araçatuba, Araçatuba, São Paulo 16050-680, Brazil

^f Animal Production and Health Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency, Vienna, Austria

ARTICLE INFO

Article history:

Available online 21 December 2016

Keywords:

DNA
Microsatellite
Genetic diversity
Sheep
Pakistan

ABSTRACT

The present study aimed at the genetic characterization of major economically important sheep (*Ovis aries*) breeds from Pakistan. A total of 155 sheep from five breeds, Balkhi (PBS), Kaghani (PKS), Damani (PDS), Salt Range (PSS) and Thalli (PTS) were analyzed for measures of genetic variability, differentiation and population structure. All the five breeds exhibited moderate to high levels of genetic diversity, with an average observed heterozygosity of 0.631 ± 0.043 . The mean estimates of measures of genetic differentiation between pairs of breeds were low ($F_{ST} = 0.050 \pm 0.015$ and $D_A = 0.159 \pm 0.028$), and the analysis of molecular variance showed that most of the total genetic variation corresponded to differences among individuals within breeds (94.41%). Despite the low level of differentiation, Bayesian clustering suggested Thalli and Damani sheep were genetically distinct from Balkhi, Kaghani and Salt Range sheep. The clustering of Pakistani sheep did not conform to their geographical locations while it was broadly consistent with their utilities (meat and coarse wool vs milk, meat and coarse wool). The results also revealed migration and admixture between Balkhi, Salt Range and Kaghani sheep. This is the first report on genetic diversity of Pakistani sheep populations of economic importance and the information generated will have implications on genetic resources management and implementation of breeding programs for these sheep breeds.

© 2017 International Atomic Energy Agency. Published by Elsevier B.V. All rights reserved.

1. Introduction

Pakistan is endowed with diverse farm animal genetic resources and has 28.8 million sheep (*Ovis aries*) well adapted to the local environment (FAOSTAT, 2013). The distribution of sheep across different provincial and agro-ecological regions of Pakistan is heterogeneous with about 48% in Balochistan, while Punjab, Sindh and North West Frontier (NWFP) provinces share 24%, 15% and 13% of the total population respectively (Khan et al., 2007). Most of these sheep are reared in low input, extensive transhumant production system where post-harvest crop residues and naturally grown shrubs and grasses are the main sources of feeding. The majority of these sheep breeds produce coarse wool and their meat is

consumed as mutton (Qasim et al., 2011). The diversity of sheep in Pakistan is represented by at least 25 well recognized breeds (Hasnain, 1985) although 42 breeds have been officially listed in FAO's Domestic Animal Diversity-Information System (DAD-IS FAO, 2014). Sheep breeds in Pakistan can be grouped into thin-tailed (Baltistani, Buchi, Cholistani, Damani, Hissardale, Kaghani, Kail, Kajli, Kali, Kooka, Lohi, Pahari, Poonchi, Sipli and Thalli) and fat-tailed (Balkhi, Balochi, Bibrik, Dumbi, Gojal, Harnai, Hashtnagri, Khijloo, Kohai, Ghizar, Latti, Michni, Rakhshani, Tirahi and Waziri) sheep. About 30% of sheep breeds in Pakistan are found in government owned livestock farms that also serve as important centers of conservation.

Among the sheep available in Pakistan, Salt Range, Balkhi, Damani, Kaghani, Baluchi, Lohi, Hastnagri, Thalli and Kachhi are the most important breeds in terms of economic relevance. They are generally raised in mixed flocks with goats, and milk is important in certain breeds like Damani and Kachhi with production

* Corresponding author.

E-mail address: annaqvi@yahoo.co.uk (A.N. Naqvi).

of about one liter per day, for a lactation period of 4–6 months (Khan et al., 2007). However, there are little or no efforts for genetic improvement of these breeds except for up-gradation of Kaghani with Rambouillet in NWFP (Khan et al., 2007). Moreover, information on characteristics and genetic diversity of most sheep breeds are limited (Ibrahim et al., 2010; Ahmed et al., 2014; Wajid et al., 2014). Considering the decline in population size of some of these breeds, genetic characterization is an important first step to assess the existing diversity in them. The loss of diversity in domestic sheep breeds has important economic, ecological and scientific implications as well as social considerations. In response to these threats, the Joint FAO/IAEA Division of International Atomic Energy Agency initiated the coordinated research program on characterization of small ruminant genetic resources in Asia. As part of this program, the present study was undertaken with the objectives of evaluating diversity and genetic relationship of five economically important sheep breeds of Pakistan, Balkhi, Damani, Kaghani, Salt Range and Thalli. The results of the present work will help to better understand the current genetic diversity in these breeds, and contribute towards formulating effective strategies for prioritization, conservation and genetic improvement of indigenous sheep breeds in Pakistan (Gaouar et al., 2015; Guang-Xin et al., 2016; Sharma et al., 2016).

2. Materials and methods

2.1. Samples and microsatellite genotyping

Field survey was carried out in respective native habitats of five indigenous sheep breeds for collecting information on distribution pattern and production systems from local livestock officers and farmers (Supplementary Table ST1; Supplementary Fig. SF1). Blood samples were collected from five sheep breeds: Balkhi (PBS, $n = 37$), Damani (PDS, $n = 31$), Kaghani (PKS, $n = 28$), Salt Range (PSS, $n = 27$) and Thalli (PTS, $n = 32$). During sample collection, information on morphology of individuals and breed characteristics of each breed were recorded. Farmers were interviewed in detail to ensure the unrelatedness of sampled sheep. Jugular blood was collected in tubes containing EDTA. DNA purification was performed using salting out protocol (Miller et al., 1988). DNA samples were stored at -20°C until further utilization. The laboratory work flow consisted of the following steps: (1) DNA purification; (2) DNA quality and quantity estimation by agarose gel electrophoresis and spectrophotometry; (3) PCR amplification of ISAG/FAO recommended sheep microsatellite markers; (4) PCR product visualization in agarose gel electrophoresis; (5) preparation of PCR product and multiplex microsatellite genotyping using a capillary sequencer (ABI 3730 DNA Analyzer – Applied Biosystems, USA); (6) electropherogram analysis using GENEMAPPER software (Applied Biosystems, USA) for extraction of genotype data and (7) population genetic analysis using different software packages. Eighteen microsatellite markers recommended by ISAG/FAO for diversity analysis in sheep were selected for the present study: HUIJ616, ILSTS005, MCM527, OarFCB128, OarHH47, SRCRSP5, BM8125, DYMS1, ILSTS11, OarAE129, OarFCB226, OarJMP29, MAF209, MAF214, MAF65, OarCP34, OarFCB304, and SRCRSP9 (FAO, 2011).

2.2. Statistical analysis

The presence of null alleles in the dataset was checked using Microchecker version 2.2.3 (Oosterhout et al., 2004). Basic measures of genetic variability including observed number of alleles, identification of private alleles, allele frequencies, allelic richness, observed (H_o) and expected (H_e) heterozygosities were calculated using GENEPOP software (Raymond and Rousset, 1995), available at <http://genepop.curtin.edu.au/>. Polymorphism information content (PIC) estimates were obtained using an in-house PERL script, following methodology described by Botstein et al. (1980). Deviations of heterozygosities from Hardy-Weinberg Equilibrium (HWE) were estimated by: (1) calculating the degree of within-population reduction in heterozygosity due to inbreeding (F_{IS}) following Wright (1951); and (2) exact tests of heterozygote excess and deficit for each marker and in each breed, as implemented in GENEPOP. Balkhi sheep was tested for mutation drift equilibrium using BOTTLENECK program (Piry et al., 1999). Four different tests were performed using allele frequency data viz. sign test, standardized differences test, Wilcoxon sign rank test and a qualitative test of mode shift to find out whether the population was in mutation drift equilibrium. The neutrality of microsatellite loci used in this study was evaluated by comparing the markers against neutral expectations in a distribution of F_{ST} vs. heterozygosities under an island model of migration with neutral markers (Wright, 1951), i.e., detection of neutral and outlier loci with excessively high (putative positive selection) or low (putative balancing selection) F_{ST} . This analysis was conducted using LOSITAN version 1 (Antao et al., 2008), with 95% confidence intervals.

Table 1

Mean basic diversity indices in five indigenous Pakistani sheep breeds n_o = mean observed number of alleles per locus; r^a = allelic richness; NPA = number of observed private alleles; H_o = observed heterozygosity (H_o); H_e = Expected heterozygosity; F_{IS} = Within-population inbreeding estimate.

Breed	n_o	r^a	NPA	H_o	H_e	F_{IS}
PBS	6.83	6.429	9	0.692	0.698	0.009
PDS	5.94	5.727	1	0.579	0.641	0.098
PKS	7.89	7.611	12	0.643	0.724	0.115
PSS	6.83	6.621	5	0.638	0.655	0.026
PTS	6.33	5.972	8	0.602	0.654	0.079

^a Based on a minimum sample size of 23 diploid individuals.

Pairwise estimations of F_{ST} were obtained using the SPAGeDi 1.3 software (Hardy and Vekemans, 2002). Nei's D_A genetic distances (Nei et al., 1983) and Neighbor-Joining (NJ) tree were obtained using DISPAN software (available at <https://homes.bio.psu.edu/people/faculty/nei/lab/dispans2.htm>). Percent bootstrap values were calculated from 1000 replications of re-sampled datasets. Analysis of molecular variance (AMOVA) was computed using the ARLEQUIN package version 3.5.1.2 (Excoffier and Lischer, 2010). The effective number of migrants between pairs of breeds per generation (N_m) was calculated based on private alleles and corrected for sample size using GENEPOP. To ascertain the recent history of each of the five Pakistani sheep breeds, probabilities of assignment were computed for each animal using the Bayesian method with prior population information (Rannala and Mountain, 1997) as implemented in GENECLASS 2.0 (Cornuet et al., 1999; Piry et al., 2004). Detection of first generation migrants in Pakistani sheep populations was carried out following Paetkau et al. (2004) using GENECLASS 2.0. The Bayesian clustering analysis without prior population information was performed using STRUCTURE version 2.3.4 (Pritchard et al., 2000), which inferred fractions in genetic ancestry for individuals and populations assuming a given number (K) of clusters. A Monte Carlo Markov chain was run for $K = 2$ to $K = 7$ with a burn-in period of 200,000 and a run length of 200,000 iterations. For each K, ten replicates were performed to calculate the mean L (K). An admixture model with correlated allele frequencies was used in all runs. The optimal 'K' was identified based on ΔK , the second order rate of change in L (K) following the procedure of Evanno et al. (2005). As a second approach for breed clustering, principal component analysis was performed using SPSS version 13.0. Inter-individual allele sharing distance matrix (155×155 distance matrix) calculated from MICROSATELLITE ANALYZER (MSA; Dieringer and Schlötterer, 2003) was utilized to derive principal components that could describe the geometric relationship between sheep belonging to different breeds. The first three largest principal components were plotted in a three dimensional scatter gram using SPSS version 13.0.

3. Results and discussion

3.1. Measures of genetic variability and heterozygosity deficit

The basic information (region of distribution, utility, estimated population size and morphological characteristics) on sheep breeds assessed in the present study is presented in Supplementary Table ST1. A total of 2790 genotypes were generated across 18 microsatellite marker loci in five indigenous sheep breeds. Overall, 193 alleles were observed across the investigated loci with a mean of 10.72 per locus. The mean observed number of alleles varied from 5.94 (Damani) to 7.89 (Kaghani) while allelic richness ranged from 5.727 (Damani) to 7.611 (Kaghani) (Table 1). A total of 35 private alleles were observed across 18 microsatellite loci ranging from 1 (Damani) to 12 (Kaghani) (Supplementary Table ST2). Among different loci, private alleles were observed at MAF214 in four of the five sheep breeds analyzed while loci SRCRSP5 and SRCRSP9 did not show private alleles in all the studied breeds. The mean polymorphism information content overall breeds and loci was 0.662 and varied between 0.427 (SRCRSP5) and 0.815 (DYMS1). Among different loci analyzed, fifteen were highly informative ($\text{PIC} > 0.50$) and three were moderately informative ($0.25 < \text{PIC} \leq 0.50$) in Pakistani sheep breeds analyzed (Supplementary Table ST2; Botstein et al., 1980). The mean polymorphism information content varied from 0.634 (Damani and Salt Range) to 0.709 (Kaghani) across different breeds. The mean basic diversity measures, observed heterozygosity varied from 0.579 (Damani) to 0.692 (Balkhi) while expected heterozygosity ranged between 0.641 (Damani) and

Table 2
Analysis of molecular variance (AMOVA) at 18 microsatellite marker loci for the sheep breeds Balkhi (PBS), Damani (PDS), Kaghani (PKS), Salt Range (PSS) and Thalli (PTS).

Grouping	Source of variation	Degrees of freedom	Sum of squares	Squared value	Percentage of variation	P-value
No grouping	Among populations	4	96.09	0.31	5.59	0.000
	Within populations	305	1571.44	5.15	94.41	0.000
Grouping I (By Utility) 1-PBS, PSS, PKS; 2-PDS, PTS	Among groups	1	27.19	0.26	0.47	0.396
	Among populations within group	3	68.90	0.29	5.30	0.000
Grouping II (By Geography) 1-PBS, PDS; 2-PKS; 3-PSS; 4-PTS	Within populations	305	1571.44	5.15	94.23	0.000
	Among groups	3	59.87	-0.18	-3.24	0.829
	Among populations within group	1	36.23	0.46	8.47	0.000
	Within populations	305	1571.44	5.15	94.77	0.000

0.724 (Kaghani). Among different loci, the observed heterozygosity was lowest at SRCRSP5 in all the five investigated sheep breeds. The observed allelic diversity in the sheep breeds under study were much higher than reported for other Pakistani sheep breeds from Balochistan and Khyber Pukhtunkhwa (Hashtnagri and Michni sheep; Ibrahim et al., 2010), Azad Jammu and Kashmir (Kail sheep; Ahmed et al., 2014) and Balochistan (Balochi and Rakhshani sheep; Wajid et al., 2014). Similarly, the mean basic diversity measures observed in the present study were higher than all the above mentioned breeds from different regions of Pakistan except for Kail sheep in Azad Jammu and Kashmir which showed higher observed and expected heterozygosity (Ahmed et al., 2014). Genetic variability measures observed in five Pakistani sheep breeds studied were comparable to European sheep (Handley et al., 2007; Loukovitis et al., 2016), Chinese sheep (Guang-Xin et al., 2016), Algerian sheep (Gaouar et al., 2016), but relatively lower than Indian (Arora et al., 2011) and Eurasian (Tapio et al., 2010) sheep.

Comparison of observed and expected heterozygosity revealed deficit of heterozygotes in all investigated sheep breeds except Balkhi. The number of loci with heterozygosity deficiency was 14, 13, 10 and 12 in Damani, Kaghani, Salt Range and Thalli sheep, respectively, while 11 loci showed heterozygosity excess in Balkhi sheep (Supplementary Table ST3). Populations that have experienced a recent reduction in their effective population size exhibit a correlative reduction of allele numbers and heterozygosity (H_e) at polymorphic loci. But the allele numbers is reduced faster than the heterozygosity. Thus in a recently bottlenecked population, the observed heterozygosity is higher than the expected equilibrium heterozygosity (H_{eq}) which is computed from the observed number of alleles, under the assumption of a constant size population (Luikart et al., 1998). Allele frequency data of Balkhi sheep was utilized to test for the possible occurrence of genetic bottleneck event in the recent past. Three mutation models of microsatellite evolution were assumed, Infinite alleles model (IAM), stepwise mutation model (SMM) and two phase model (TPM). All the three statistical tests (sign test, standardized differences test, Wilcoxon sign rank test) did not detect significant heterozygosity excess when assumed under SMM and TPM (Supplementary Table ST4). When assumed under IAM, although standardized differences test and Wilcoxon sign rank test detected significant heterozygosity excess, the qualitative test for mode shift revealed normal L-shaped distribution indicating no evidence for genetic bottleneck in Balkhi sheep in the recent past.

The F_{IS} values averaged overall loci were not significantly different from zero in the sheep breeds sampled, suggesting that these populations are not inbred. The exact tests revealed significant deficit of heterozygotes at five investigated loci (OarFCB128 ($P < 0.01$), SRCRSP5 ($P < 0.01$), OarAE129 ($P < 0.01$), OarCP34 ($P < 0.05$) and OarFCB304 ($P < 0.05$)) in four sheep breeds studied (Damani, Kaghani, Salt Range and Thalli). Deviations from Hardy Weinberg equilibrium could be due to several reasons: (a) Wahlund effect – In order to assess the overall genetic diversity of Pakistani sheep breeds, different flocks were sampled for each breed.

Wahlund effect in the multi-flock samples could be a possible cause for the observed heterozygote deficit, where the flocks have been separated long enough. However, heterozygote deficit resulting from the Wahlund effect is expected to show across all loci for each affected breed, which is not the case in the present study. Only a few marker loci were deviating from HWE significantly and hence multi-flock sampling scheme does not fully account for the observed heterozygote deficit (Santucci et al., 2007). (b) The presence of null alleles could also be an important factor for the observed deficiency of heterozygosity. However, the Microchecker analysis did not detect any significant frequency of null alleles in the dataset. (c) Use of one or few rams for breeding ewes across different flocks over several generations in a small holder production setting could also result in heterozygote deficit. Although, this could be a possible reason, most of the samples used in the present study were obtained from flocks reared at farther locations. However, considering the distance covered by these sheep for grazing every day, this factor could not be completely rejected. (d) Influence of selection on marker loci could be another potential cause for the heterozygosity deficit observed at these loci. In the present study F_{ST} outlier approach was used to detect selection at the investigated loci. Generally, population differentiation (F_{ST}) based approaches can detect many types of selection unlike other methods. Large values of F_{ST} at a locus (relative to neutral regions) reveal stark differentiation between populations, which is indicative of directional selection. While, small F_{ST} values reveal the populations compared are homogenous, but also suggest balancing or directional selection (Vitti et al., 2013). The selection detection strategy implemented in LOSITAN evaluates relationship between F_{ST} and expected heterozygosity in an island model, describing the expected distribution of Wright's inbreeding coefficient F_{ST} vs H_e under an island model of migration with neutral markers. This distribution is used to identify outlier loci that have excessively high or low F_{ST} compared to neutral expectations (Antao et al., 2008). The results revealed three microsatellite loci as significant outliers (OarFCB304, HUI616 and OarFCB226) with high F_{ST} values indicating positive selection in force (Fig. 1). Among the five loci with significant heterozygosity deficit, OarFCB304 was found to be under the influence of positive selection. Hence the remaining four markers (OarFCB128, SRCRSP5, OarAE129 and OarCP34) were retained for further analysis while the markers under significant positive selection were not considered for further genetic differentiation and population structure analysis. Despite the availability of approaches for estimating adaptive diversity from outlier loci (see Adaptive Index (PAI), described by Bonin et al., 2007), we focused only on neutral diversity, as those methods require a larger number of genotyped loci to achieve considerable reliability. For instance, Bonin et al. (2007) screened 376 AFLP markers in common frog populations and demonstrated that convergence of population ranking results started to appear when at least 150 markers were considered. In terms of microsatellite markers, Medugorac et al. (2009) estimated PAI for 16 European cattle breeds using 105 markers, which were roughly six times more markers than we investigated in the present study.

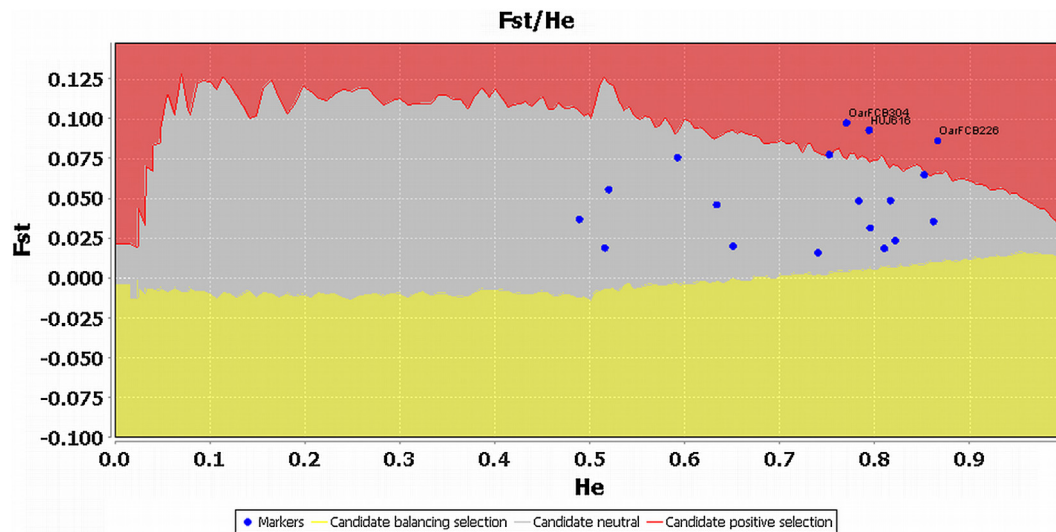


Fig. 1. Outlier detection at 18 microsatellite marker loci in the sheep breeds of Balkhi, Damani, Kaghani, Salt Range and Thalli. Blue dots represent markers distributed according to their F_{ST} values (y axis) as a function of expected heterozygosity (x axis). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Pairwise F_{ST} (upper diagonal), effective number of migrants per generation (within parenthesis), and Nei's D_A genetic distance (lower diagonal) between five Pakistani sheep breeds.

Breed	PBS	PDS	PKS	PSS	PTS
PBS	–	0.077 (0.97)	0.041 (1.45)	0.040 (2.14)	0.050 (0.94)
PDS	0.198	–	0.043 (1.52)	0.044 (1.71)	0.060 (1.21)
PKS	0.151	0.157	–	0.025 (1.90)	0.049 (1.35)
PSS	0.128	0.133	0.126	–	0.050 (1.64)
PTS	0.201	0.151	0.189	0.154	–

3.2. Genetic differentiation and phylogeny

Analysis of molecular variance (AMOVA) was performed to partition the total variance to assess within and between breed differences. AMOVA revealed 94.41% of total genetic variation was due to differences among individuals within breeds, and only 5.59% was accounted for by differences between breeds (Table 2). N_m estimates (effective numbers of migrants per generation) ranged from 0.94 (Balkhi-Thalli) to 2.14 (Balkhi-Salt Range), with an average of 1.48 ± 0.39 (Table 3). The pair-wise F_{ST} estimates ranged from 0.025 (Kaghani-Salt Range) to 0.077 (Balkhi-Damani) while pair-wise D_A ranged from 0.126 (Salt Range-Kaghani) to 0.201 (Balkhi-Thalli), respectively (Table 3). The average for the differentiation index and genetic distances between pairs of breeds were low ($F_{ST} = 0.050 \pm 0.015$, $D_A = 0.159 \pm 0.028$). The estimated level of differentiation of Pakistani sheep breeds was lower than reports on other sheep breeds in bordering countries, like India (Arora et al., 2011) and China (Niu et al., 2012). However, the estimated F_{ST} in the present study was relatively higher than the values reported for Algerian (Gaouar et al., 2015) and Iranian (Vahidi et al., 2016) sheep. Among many others, causes of genetic differentiation include (a) different origins; (b) geography and reproductive isolation; (c) genetic drift and founder effect; (d) different production goals and (e) different preferences for type traits. In the present study, low genetic differentiation could be due to the presence of fewer agro-ecological zones and shorter geographical distances in Pakistan, allowing for easier gene flow among different breeds. To illustrate this, a map indicating the distribution of sheep breeds under study is provided in Supplementary Fig. SF2. Furthermore, an estimated average of 1.483 effective migrants per generation between pairs of breeds showed considerable exchange of germplasm among

the populations. One of the main production systems of sheep rearing in Pakistan is migratory farming that includes nomadic, agro-pastoral, transhumant and sedentary types (Afzal and Naqvi, 2004). Farmers regularly or seasonally migrate their flocks in search of grazing areas and pasture lands. The pasture lands situated at altitudes in the northern areas of Khyber Pakhtunkhwa of Pakistan provides grazing to sheep during the hot and dry summer days (June to September). Further, to test the influence of geography and habitat distribution on genetic differentiation, AMOVA was conducted after grouping the breeds as follows: 1-PKS (Kaghan valley); 2-PBS and PDS North West Frontier Province (NWFP); 3-PSS (Salt Range Area); 4-PTS (Thal desert in Punjab) (Supplementary Table ST1). The results showed among group variance was negative while among population within group variance was positive and significant ($P < 0.01$) (Table 2). Negative among group variance clearly indicated that genes from different groups were more related to each other than genes from the same group. Thus, migrations together with absence of systematic breeding schemes might have contributed to the low level of genetic differentiation among the sheep breeds investigated. To assess the phylogeny based on microsatellite genotype data, Neighbour-Joining (NJ) tree was constructed using pairwise D_A distances between breeds which showed two major clusters. Damani and Thalli breeds were found to cluster together with a bootstrap support of 100% while the remaining three breeds (Salt Range, Kaghani and Balkhi) clustered together with a bootstrap support of 93% (Fig. 2). Within this cluster, Balkhi and Kaghani formed a distinct sub-cluster with a bootstrap support of 61%. The clustering of Pakistani sheep breeds under study was consistent to their utility/purpose. Damani and Thalli breeds are predominantly milk animals along with meat and wool utilities while Kaghani, Salt Range and Balkhi are mainly meat and wool type animals (Supplementary Table ST1). However, analysis of molecular variance after grouping the breeds based on utility showed positive among group variance but small and statistically not significant ($P > 0.05$) (Table 2).

3.3. Population structure in Pakistani sheep

Principal components analysis (PCA) was conducted to assess the genetic structure of five investigated Pakistani sheep breeds. PCA is a way of identifying patterns in data, and expressing the data in such a way to highlight their similarities and differences. Since it

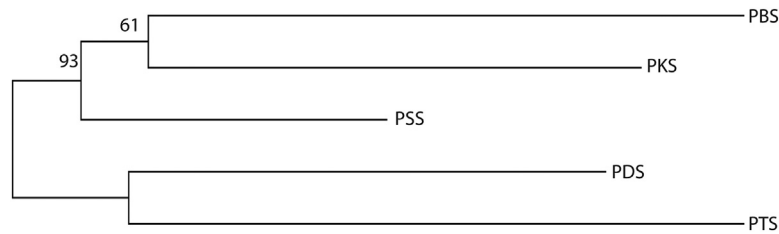


Fig. 2. Neighbour-Joining tree based on Nei's D_A genetic distances for the sheep breeds of Balkhi, Damani, Kaghani, Salt Range and Thalli. Numbers at the nodes indicate percent bootstrap values, based on 1000 permutations.

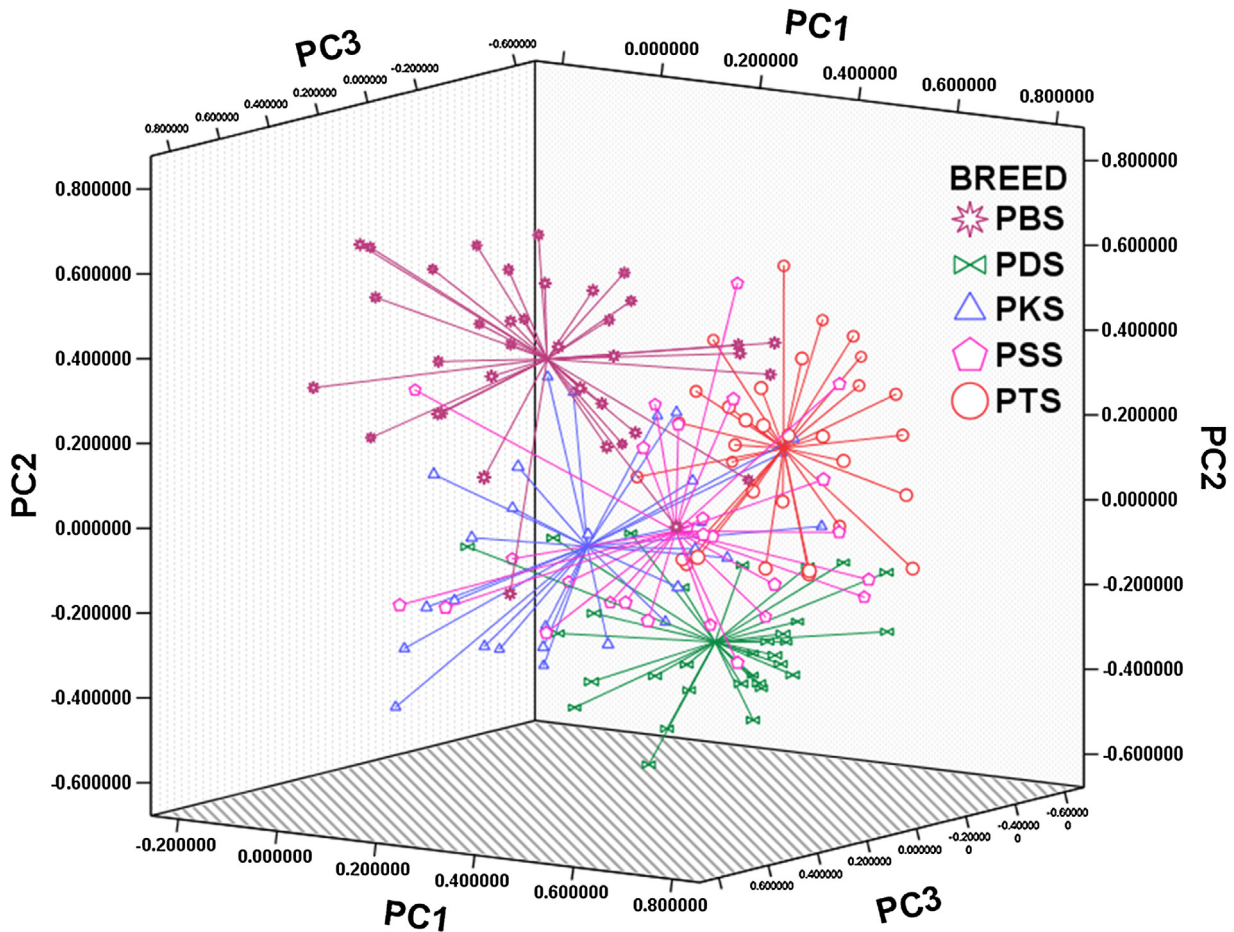


Fig. 3. Three dimensional scatter plot of three largest principal components (PCs) derived from pairwise inter-individual allele sharing distance between different sheep breeds of Pakistan (Three PCs explained 33.5% of total variance in the dataset; PC1 = 17.5%; PC2 = 8.8%; PC3 = 7.1%).

is difficult to find patterns in data with high dimensionality, PCA is powerful in reducing the dimensions of data without much loss of information and visualizing the relationship in a geometric space. Principal components analysis of pair-wise inter-individual allele sharing distance among Pakistani sheep breeds resulted in the extraction of 24 PCs each with eigen values more than one, which collectively explained 84.7% of total variance in the dataset. The first three largest PCs together explained 33.5% of total variance with PC1 explaining 17.5% while PC2 and PC3 explaining 8.8% and 7.1% respectively. The first three largest principal components (PC1, PC2 and PC3) were used to draw a three dimensional scatter gram which revealed distinct clustering of Balkhi, Thalli and Damani sheep. However, poor separation was observed among Salt Range and Kaghani sheep with significant overlapping among their centroids (Fig. 3). This is understandable as the pairwise genetic differentiation was lowest between Salt Range and Kaghani ($F_{ST} = 0.025$).

Although considerable overlapping of Damani sheep with Kaghani and Salt Range was observed, many of these animals were located distinctly. The results of PCA provided further evidence that genetic relationship between the investigated sheep breeds might not be influenced by geographical location, but could be due to potential migration and inter-mixing among different breeds (Kijas et al., 2009; Guang-Xin et al., 2016).

The recent history of Pakistani sheep breeds was assessed by Bayesian approach with prior population information (Rannala and Mountain, 1997). Genotype assignment following this procedure involved computation of likelihood of a genotype in a given population under the assumption of equal prior probability density to the allelic frequencies of each locus in each population. This method showed better assignment performance than frequency and distance based methods in simulated as well as real populations (Cornuet et al., 1999; Arranz et al., 2001; Legaz et al., 2008).

Table 4Assignment of genotypes to different sheep breeds using Bayesian method (Rannala and Mountain, 1997) with *a priori* population information.

Sheep Breed	N	No. Assigned correctly	% Correct Assignment	Most likely origin of incorrectly assigned sheep
Balkhi (PBS)	37	34	91.9	3PSS
Salt Range (PSS)	27	24	88.9	2PBS, 1PKS
Damani (PDS)	31	27	87.1	1PBS, 1PSS, 2PKS
Kaghani (PKS)	28	25	89.3	2PSS, 1PBS
Thalli (PTS)	32	32	100.0	–
Total	155	142	91.6	–

Table 5

Proportion of membership coefficient of individuals from five sheep breeds in different inferred clusters after STRUCTURE analysis.

Breed	N	Inferred clusters														
		K=2		K=3			K=4				K=5					
		1	2	1	2	3	1	2	3	4	1	2	3	4	5	
Balkhi	37	0.071	0.929	0.039	0.149	0.812	0.785	0.036	0.158	0.02	0.676	0.088	0.198	0.024	0.014	
Damani	31	0.887	0.113	0.833	0.131	0.036	0.033	0.047	0.082	0.839	0.025	0.064	0.063	0.045	0.803	
Kaghani	28	0.207	0.793	0.054	0.803	0.143	0.149	0.042	0.753	0.055	0.108	0.609	0.205	0.04	0.038	
Salt Range	27	0.398	0.602	0.164	0.726	0.111	0.114	0.051	0.649	0.187	0.066	0.171	0.642	0.035	0.086	
Thalli	32	0.965	0.035	0.954	0.031	0.015	0.018	0.915	0.026	0.042	0.016	0.019	0.031	0.899	0.035	

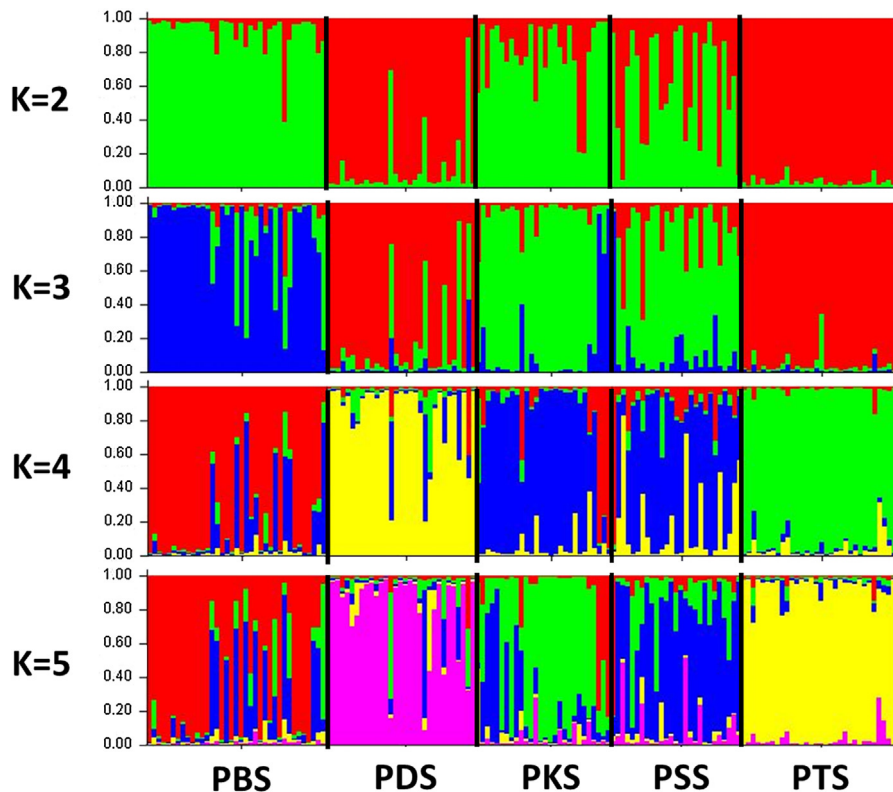


Fig. 4. Bayesian clustering analysis for the five Pakistani sheep breeds surveyed, applying an admixture model with independent allele frequencies. A vertical bar that can be partitioned into colored fragments, with its length proportional to cluster contribution, represents each individual. Black lines separate breeds. K values represent number of ancestral populations assumed. The analyses were conducted with a burn-in period of 200,000 and 200,000 iterations.

The results revealed an overall percent correct assignment of 91.6% in Pakistani sheep that ranged between 87.1% (Kaghani) and 100% (Thalli) (Table 4). Three of the incorrectly assigned Balkhi sheep were considered most likely to be Salt Range individuals. Among three wrong assignments in Salt Range sheep, two were most likely to be Balkhi while the other one was Kaghani sheep. In case of Kaghani sheep, two of the incorrect assignments were likely to be Salt Range while one sheep was considered likely to be Balkhi. Additionally, the same incorrectly assigned individuals were detected as first generation migrants from the respective sheep breeds. To further assess the cryptic genetic structure and degree of admixture

among the investigated sheep breeds, Bayesian clustering analysis without *a priori* population information was performed with K = 2 to K = 7. When K = 2 was assumed, most Thalli (96.5%) and Damani (88.7%) sheep were assigned to cluster 1 while most Balkhi sheep (92.9%) were assigned to cluster 2 (Table 5). Although Kaghani and Salt Range were predominantly assigned to cluster 2, considerable admixture was noticed in these two breeds (Fig. 4). When K = 3 was assumed, most Balkhi sheep (81.2%) was assigned to a distinct cluster 3 while most of Kaghani (80.3%) and Salt Range (72.6%) sheep were assigned to the cluster 2. Similarly, Thalli (95.4%) and most Damani (83.3%) sheep were assigned to cluster 1. When K = 4 was

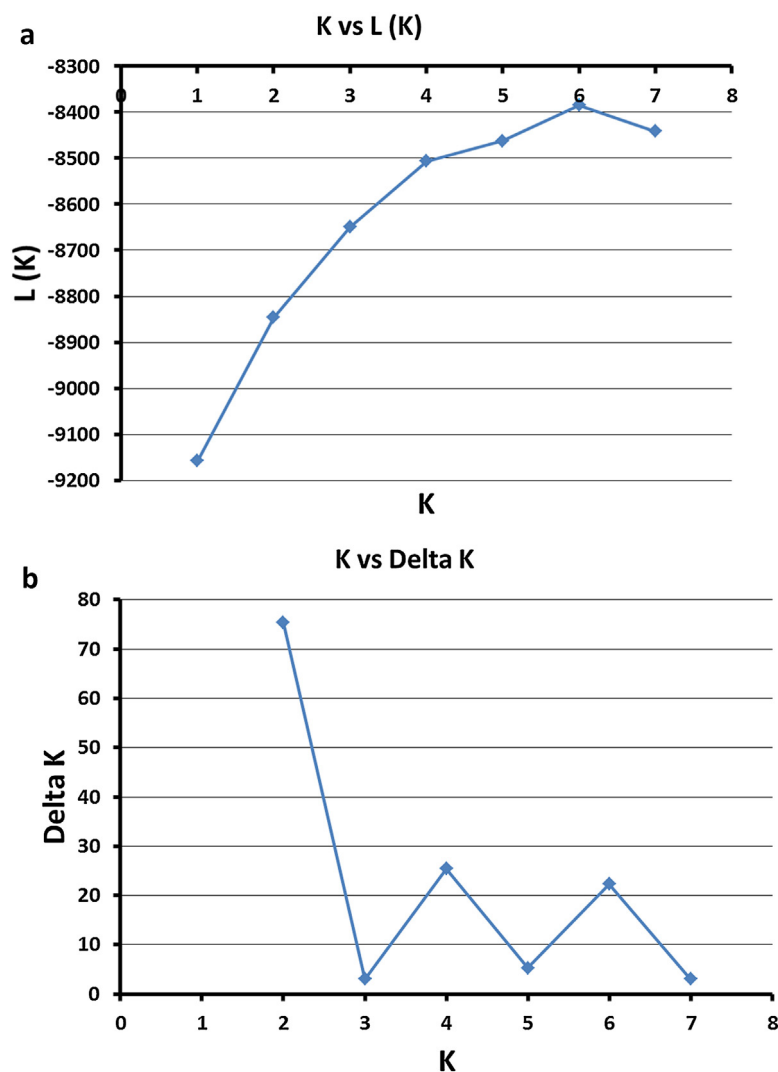


Fig. 5. Determination of correct number of clusters in Bayesian STRUCTURE analysis (Evanno et al., 2005) (a) Mean $L(K)$ over 10 runs for each K value of 1–7 (b) Distribution of ΔK with the modal value ($K=2$) indicating the true K or the uppermost level of structure.

assumed, Thalli and Damani sheep clustered distinctly into two separate clusters. When $K=5$ was assumed, more than 60% animals of each of Kaghani and Salt Range sheep clustered separately into two distinct clusters. Also, about 20% of Balkhi sheep were assigned to the predominantly Salt Range sheep cluster. The mean $L(K)$ over 10 runs for each value of K was highest at $K=6$ (Fig. 5a). To identify the optimal K , the second order rate of change in $L(K)$ i.e. delta K was estimated following the procedure given by Evanno et al. (2005). ΔK was based on the rate of change in the log probability of data between successive K values and has a mode at the true K . The results revealed $K=2$ as the optimal clustering solution for the given dataset (Fig. 5b), consistent with phylogeny and utility of studied sheep breeds. Kaghani and Salt Range, despite being geographically separated appear to share genetic background and form a single genetic group. Although, Balkhi sheep showed distinctness upon principal components based clustering, Bayesian analysis (with and without prior population information) revealed migration and admixture with Salt Range and Kaghani sheep.

4. Conclusion

In conclusion, the present work showed moderate to high genetic diversity among the five investigated sheep breeds of

Pakistan. Although the overall population differentiation was low, two major genetic groups were observed with Balkhi, Kaghani and Salt Range forming a single cluster while Thalli and Damani clustered distinctly. The clustering of Pakistani sheep did not conform to their geographical locations while it was broadly consistent with their utilities (meat and coarse wool vs milk, meat and coarse wool). The results also revealed migration and admixture between Balkhi, Salt Range and Kaghani sheep. This is among the first report on genetic diversity of Pakistani sheep populations of economic importance, and the information generated will have implications on genetic resources management and implementation of breeding programs for these sheep populations. Additionally, the genetic diversity of Pakistani sheep populations revealed by this study represents a valuable opportunity for improving breeding systems to enhance local people's livelihood.

Acknowledgements

The present work is part of the coordinated research project (CRPD3.10.25) of International Atomic Energy Agency (IAEA). The authors thank International Atomic Energy Agency for funding and technical support during the project. Also thanks to International Livestock Research Institute (ILRI), Nairobi, Kenya and the Chinese

Academy of Animal Sciences (CAAS), Beijing, China for the support during training on molecular genetic data analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.smallrumres.2016.12.032>.

References

- Afzal, M., Naqvi, A.N., 2004. Livestock resources of Pakistan: present status and future trends. *Q. Sci. Vis.* 9, 1–14.
- Ahmed, Z., Babar, M.E., Hussain, T., Nadeem, A., Awan, F.I., Wajid, A., Shah, S.A., Ali, M.M., 2014. Genetic diversity analysis of Kail sheep by using microsatellite markers. *J. Anim. Plant Sci.* 24, 1329–1333.
- Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A., Luikart, G., 2008. Lositan: a workbench to detect molecular adaptation based on a F_{ST} -outlier method. *BMC Bioinform.* 9, 323.
- Arora, R., Bhatia, S., Mishra, B.P., Joshi, B.K., 2011. Population structure in Indian sheep ascertained using microsatellite information. *Anim. Genet.* 42 (3), 242–250.
- Arranz, J.J., Bayón, Y., San Primitivo, F., 2001. Differentiation among Spanish sheep breeds using microsatellites. *Genet. Sel. Evol.* 33, 529–542.
- Bonin, A., Nicole, F., Pompanon, F., Miaud, C., Taberlet, P., 2007. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conserv. Biol.* 21, 697–708.
- Botstein, D., White, R.L., Skolnick, M., Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.
- Cornuet, J.M., Piry, S., Luikart, G., Estoup, A., Solignac, M., 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153, 1989–2000.
- DAD-IS, FAO, 2014. Domestic Animal Diversity Information System, Food and Agriculture Organization, Rome, Italy. <http://dad.fao.org/>.
- Dieringer, D., Schlötterer, C., 2003. MICROSAATELLITE ANALYSER (MSA): a platform independent analysis tool for larger microsatellite data sets. *Mol. Ecol. Resour.* 167–169.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620.
- Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10 (3), 564–567.
- FAO, 2011. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. No. 9. Rome, Italy.
- FAOSTAT, 2013. <http://faostat3.fao.org/home/index.html>.
- Gaouar, S.B.S., Da Silva, A., Ciani, E., Kdidi, S., Aouissat, M., Dhimi, L., Lafri, M., Mafah, A., Mehtar, N., 2015. Admixture and local breed marginalization threaten algerian sheep diversity. *PLoS One* 10 (4), e0122667, <http://dx.doi.org/10.1371/journal.pone.0122667>.
- Gaouar, S.B.S., Kdidi, S., Ouragh, L., 2016. Estimating population structure and genetic diversity of five Moroccan sheep breeds by microsatellite markers. *Small Rumin. Res.* 144, 23–27.
- Guang-Xin, E., Zhong, T., Ma, Y.-H., Gao, H.-J., He, J.-N., Liu, N., Zhao, Y.-J., Zhang, J.-H., Huang, Y.-F., 2016. Conservation genetics in Chinese sheep: diversity of fourteen indigenous sheep (*Ovis aries*) using microsatellite markers. *Ecol. Evol.* 6, 810–817.
- Handley, L.L.-J., Byrne, K., Santucci, F., Townsend, S., Taylor, M., Bruford, M.W., Hewitt, G.M., 2007. Genetic structure of European sheep breeds. *Heredity* 99, 620–631.
- Hardy, O.J., Vekemans, X., 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620.
- Hasnain, H.U., 1985. FAO Animal Production and Health Paper (FAO), no. 56 /Rome (Italy), FAO, 145 p. Accession No: 261197, Fiche No: 86M01761, ISBN 92-5-102346-8.
- Ibrahim, M., Ahmad, S., Swati, Z.A., Khan, M.S., 2010. Genetic diversity in Balkhi: Hashtnagri and Michni sheep populations using SSR markers. *Afr. J. Biotechnol.* 9, 7617–7628.
- Khan, M.S., Khan, M.A., Ahmad, S., Mahmood, S., 2007. Genetic resources and diversity in Pakistani sheep. *Int. J. Agric. Biol.* 9 (6), 941–944.
- Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS* 1 (4), e4668, <http://dx.doi.org/10.1371/journal.pone.0004668>.
- Legaz, E., Álvarez, I., Royo, L.J., Fernández, I., Gutiérrez, J.P., Goyache, F., 2008. Genetic relationships between Spanish Assaf (Assaf.E) and Spanish native dairy sheep breeds. *Small Rumin. Res.* 80, 39–44.
- Loukovitis, D., Siasiou, A., Mitsopoulos, I., Lymberopoulos, A.G., Laga, V., Chatziplis, D., 2016. Genetic diversity of Greek sheep breeds and transhumant populations utilizing microsatellite markers. *Small Rumin. Res.* 136, 238–242.
- Luikart, G., Allendorf, F.W., Cornuet, J.M., Sherwin, W.B., 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *J. Hered.* 89, 238–247.
- Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C.E., Taberlet, P., Luntz, B., Mix, H.M., Foster, M., 2009. Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Mol. Ecol.* 18, 3394–3410.
- Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16 (3), 1215.
- Nei, M., Tajima, F., Tateno, Y., 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19, 153–170.
- Niu, L.L., Li, H.B., Ma, Y.H., Du, L.X., 2012. Genetic variability and individual assignment of Chinese indigenous sheep populations (*Ovis aries*) using microsatellites. *Anim. Genet.* 43 (1), 108–111.
- Oosterhout, C.V., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4, 535–538.
- Paetkau, D., Slade, R., Burden, M., Estoup, A., 2004. Direct, real-time estimation of migration rate using assignment methods: a simulation-based exploration of accuracy and power. *Mol. Ecol.* 13, 55–65.
- Piry, S., Luikart, G., Cornuet, J.M., 1999. Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J. Hered.* 90, 502–503.
- Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L., Estoup, A., 2004. GeneClass 2: a software for genetic assignment and first generation migrant detection. *J. Hered.* 95, 536–539.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Qasim, M., Ahmad, H., Ghafoor, S., Afridi, S.G., Muhammad, I., Imtiaz, A.K., 2011. Estimation of genetic diversity in sheep (*Ovis aries*) using randomly amplified polymorphic DNA. *Int. J. Anim. Vet. Adv.* 3 (1), 6–9.
- Rannala, B., Mountain, J.L., 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9197–9201.
- Raymond, M., Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249.
- Santucci, F., Ibrahim, K.M., Bruzzone, A., Hewitt, G.M., 2007. Selection on MHC-linked microsatellite loci in sheep populations. *Heredity* 99, 340–348.
- Sharma, R., Kumar, B., Arora, R., Ahlawat, S., Mishra, A.K., Tandia, M.S., 2016. Genetic diversity estimates point to immediate efforts for conserving the endangered Tibetan sheep of India. *Meta Gene* 8, 14–20.
- Tapio, M., Ozerov, M., Tapio, I., Toro, M.A., Marzanov, N., Cinkulov, M., Goncharenko, G., Kiselyova, T., Murawski, M., Kantanen, J., 2010. *BMC Genet.* 11:76.
- Vahidi, S.M.F., Faruque, M.O., Falahati Anbaran, M., Afraz, F., Mousavi, S.M., Boettcher, P., Joost, S., Han, J.L., Colli, L., Periasamy, K., Negrini, R., Ajmone-Marsan, P., 2016. Multi locus genotypic data reveal high genetic diversity and low population genetic structure of Iranian indigenous sheep. *Anim. Genet.* 47, 463–470.
- Vitti, J.J., Grossman, S.R., Sabeti, P.C., 2013. Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120.
- Wajid, A., Wasim, M., Yaqub, T., Firyal, S., Tayyab, M., Siddique, S., Hussain, T., 2014. Assessment of Genetic diversity in Balochi and Rakhshani sheep breeds of Balochistan using microsatellite DNA markers. *J. Plant Anim. Sci.* 24, 1348–1354.
- Wright, S., 1951. The genetical structure of populations. *Ann. Eugen.* 15, 323–354.