

## Mitochondrial D-loop analysis reveals single maternal lineage for coastal region goats of India

Shivani Rana<sup>1,2</sup>, Sumer Singh<sup>2</sup>, Vandana Dureja<sup>1,2</sup>, Jyoti Joshi<sup>1</sup>, Priyanka Banerjee<sup>1</sup>, Upasna Sharma<sup>1</sup>, Ramesh K. Vijn<sup>1</sup>

<sup>1</sup>National Bureau of Animal Genetic Resources, Indian Council of Agricultural Research, Karnal, Haryana, India

<sup>2</sup>Singhania University, Pachheri Beri, Jhunjhunu, Rajasthan, India

Corresponding Author: Ramesh Kumar Vijn

---

### ABSTRACT

D-loop is important for evaluating the diversity and relatedness among the breeds in different geographic locations. The D-loop region has not been studied and the relatedness is not known among Indian goat populations on this basis. In the present study, we sequenced and analyzed 835bp mitochondrial DNA D-loop sequences in 95 samples of 3 coastal regions and one semi-arid region of India as an out-group and identified 68 haplotypes with 92 polymorphic sites. All goat populations in this study were highly diverse on the basis of haplotype and nucleotide diversity. Analysis of molecular variance revealed large genetic differentiation within the populations. We implemented neutrality tests to assess recent historical demographic events. Tajima's D test and Fu's Fs test both suggested recent population growth for the populations. The network analysis revealed high levels of singleton haplotypes in the goat populations of coastal regions and semi-arid regions of India and deciphered a single maternal lineage.

**Keywords:** Mitochondrial D-loop, Goat, Neutrality test, Phylogenetic analysis.

### INTRODUCTION

Goat constitutes an important species of livestock in India and contributes greatly towards food, milk, skin, fibre and provides rural employment and to gross domestic product. India has 154.0 million goats (16.7% of the world goat population), which produces 4.3 million tonnes milk (25.8% of world's total goat milk produced) and 8.46 lakh tonnes of goat meat (which is 17.4% of the total meat produced) in India (BAHS, 2012).

Goats are an integral component of subsistence farming system and support a large rural population of landless and marginal farmers. Since goats thrive in varied climatic regions of India from cold desert of Ladakh in north to hot and dry desert of Rajasthan in West to humid climate of West Bengal (Singh *et al.*, 2008), they have become major concern for the conservation and sustainable utilization in vast and varied geographical location of the country.

Goat breeds in India are a result of migration among different breeds, geographical isolation natural selection and cross-breeding. Their names have been assigned on the basis of their morphological attributes and their natural habitat. This comparative study was done among goat populations of coastal regions of India (Ganjam, Black Bengal and Malabari goats) and 1 goat population of semi-arid region of Gujarat (Mehsana goat) was considered as an out-group. Ganjam goat is widely distributed in South Odisha and mostly used for meat purposes. Black Bengal goat is a precious germplasm of West Bengal and is reared because of its high prolificacy, high fertility, adaptability to hot humid conditions and superior quality meat and skin while Malabari goats inhabit the Calicut, Kannur, Waynad and Malappuram districts of Kerala, these goats are primarily reared for meat and for its high prolificacy rate. Mehshana goat which is a dual purpose breed, inhabit in Mehshana, Banaskantha, Gandhinagar and Ahmedabad districts of Gujarat. It is reared by small and marginal farmers for milk, meat and manure. The aim of this study was establishment of maternal lineage and determining genetic relationship among goat populations of coastal and semi-arid regions of India on the basis of mitochondrial D-loop region analysis.

## **MATERIAL AND METHOD**

Blood samples (about 10ml) of goats, were collected from Ganjam district of Odisha and southern part of West Bengal, in the Eastern Coastal region, Malabar area of Kerala from Western Coastal region of India. Blood samples of goat were also collected from Mehshana district of Gujarat which is semi-arid region. The isolation of the genomic DNA from the blood samples was carried out by Phenol-chloroform DNA isolation method using Proteinase K digestion (Sambrook *et al.*, 1989). Then, the quantification of the DNA was carried out by electrophoresis, using 0.6% agarose gel stained with Ethidium Bromide.

For mitochondrial D-loop sequence generation, we used 24 samples of goats each from Odisha, West Bengal and Gujarat, 23 samples from Kerala. Primer pairs were designed using Primer3 webserver and Accession no.AF533441 as reference sequence. The sequence of forward primer was 5'TTCCCACTCCACAAGCCTAC<sup>3'</sup> and that of reverse primer was 5'GGATGCATGATGAAATGCAA<sup>3'</sup>. We amplified the mitochondrial D-loop region in 25µl reaction, 1X buffer, 2.0mM MgCl<sub>2</sub>, 200mM dNTPs,

5pmol of each primer, 1U Taq polymerase and 50ng of genomic DNA. The amplified product was checked in 2% agarose gel and then was used for sequencing using Big Dye Terminator carried out on 3130XL Genetic Analyzer (Applied Biosystems).

D-loop sequences obtained were aligned and edited in BioEdit v7.1.3 software (Hall, 1999). DnaSP v5.10.01 (Rozas *et al.*, 2010) was used to identify and calculate haplotype diversity, polymorphic sites and nucleotide diversity. Mismatch distribution analysis (Schneider and Excoffier, 1999) was also used to analyse if expansion occurred in any population. In addition, Harpendings reggedness index and sum of squared deviations (SSD) were calculated for any deviation from population expansion model. Various statistical parameters like Tajima's D, Fu's Fs, gene flow,  $F_{ST}$  and analysis of molecular variance (AMOVA) were computed using Arlequin 3.5.1.2 (Schneider *et al.*, 2000).

A phylogenetic Neighbor-joining tree (Satiou and Nei, 1987) was constructed based on mtDNA D-loop sequences using MEGA5 (Tamura *et al.*, 2011). Median-joining network was drawn using the software Network 4.6.1.1 (Bandelt *et al.*, 1999) to investigate the possible relationships among haplotypes of 3 Coastal and 1 semi-arid region goat populations of India.

## RESULT AND DISCUSSION

A total of 95 partial D-loop sequences of 835bp thus obtained were found to be highly polymorphic, having 68 haplotypes. These haplotypes of mtDNA D-loop sequences of goats were deposited in the GenBank database [Accession no. are KC817843-KC817862 (Black Bengal), KC817907-KC817923 (Ganjam), KC817976-KC817991 (Malabari) and KC818011-KC818027 (Mehsana)]. 92 polymorphic sites were identified, out of which 34 were singleton variable sites and 58 were parsimony informative sites. The number of haplotypes varied from 16 to 20 as shown in Table 1. The highest polymorphic sites were observed in Mehana (59 polymorphic sites). The presence of high number of haplotypes is due to high substitution rate detected in D-loop region. The haplotype diversity or genetic diversity, which is the measure of uniqueness of particular haplotype, was high in the 4 populations studied, 0.975 being highest in West Bengal Coast, while the nucleotide diversity, which is the measure of genetic variation, was quite low in the all the populations (Table 1). High haplotype diversity and low nucleotide diversity values point towards demographic expansion (Mladineo *et al.*, 2013). The relative haplotype frequency was found to be high in H-21 which belongs to Ganjam goat, followed by H-58 and H-6.

Table 1: Sample size, number of haplotypes, no. of polymorphic sites, haplotype diversity, and nucleotide diversity of 4 goat populations of India.

| Regions           | Sample Size | No. of haplotypes | No. of polymorphic sites | Haplotype Diversity | Nucleotide Diversity |
|-------------------|-------------|-------------------|--------------------------|---------------------|----------------------|
| West Bengal Coast | 24          | 20                | 51                       | 0.975               | 0.010                |
| Odisha Coast      | 24          | 16                | 45                       | 0.931               | 0.013                |
| Malabar Coast     | 23          | 16                | 40                       | 0.964               | 0.011                |
| Mehsana           | 24          | 17                | 59                       | 0.949               | 0.012                |

The overall genetic variation within and between population was calculated using Analysis of molecular variance (AMOVA) (Excoffier *et al.*, 1992). The overall genetic variation within the populations (85.14 %) was much higher than the variation among population (14.86 %). Larger genetic variation within population suggests presence of high haplotype diversity. This large haplotype diversity within population may be due to the large number of singleton variations, wide spread distribution, environmental heterogeneity and life history traits that favour rapid population growth (Nei, 1987).

Several indirect methods exist for estimating gene flow from gene frequency data (Slatkin, 1985a; Slatkin and Barton, 1989) among which the  $F_{ST}$  method is the most widely used approach.  $F_{ST}$  and  $N_m$  values are useful measures of genetic differentiation and gene flow.  $N_m$  can be interpreted as the effective number of migrants exchanged between demes per generation (Wright, 1969). Values of  $F_{ST}$  and  $N_m$  estimated according to the methods of Slatkin, 1981 and Wright, 1943 are given in Table 2.

Table 2: Population pairwise  $F_{ST}$  values (below diagonal) and Number of migrants  $N_m$  (above diagonal), of 4 goat populations.

| $F_{ST}$ / $N_m$  | West Bengal Coast | Odisha Coast | Malabar Coast | Mehsana |
|-------------------|-------------------|--------------|---------------|---------|
| West Bengal Coast | 0.000             | 2.591        | 6.635         | 2.336   |
| Odisha Coast      | 0.162             | 0.000        | 4.513         | 2.130   |
| Malabar Coast     | 0.070             | 0.100        | 0.000         | 2.904   |
| Mehsana           | 0.176             | 0.190        | 0.147         | 0.000   |

Maximum differentiation according to  $F_{ST}$  values was observed 0.190 between Mehsana and Odisha Coast, while the least differentiation  $F_{ST}$  value was observed between Malabar Coast and West Bengal Coast (0.070). Same differentiation results were also depicted in Figure 1 based on the  $F_{ST}$  values, where dark blue colour reveals higher differentiation between populations of Mehsana and Odisha goats and light blue colour shows less differentiation between West Bengal and Malabar goats. The lowest  $N_m$  value was observed between Mehsana and Odisha Coast (2.130). The highest  $N_m$  value was

observed between West Bengal Coast and Malabar Coast (6.635). Both  $F_{ST}$  and  $Nm$  values support/complement each other.

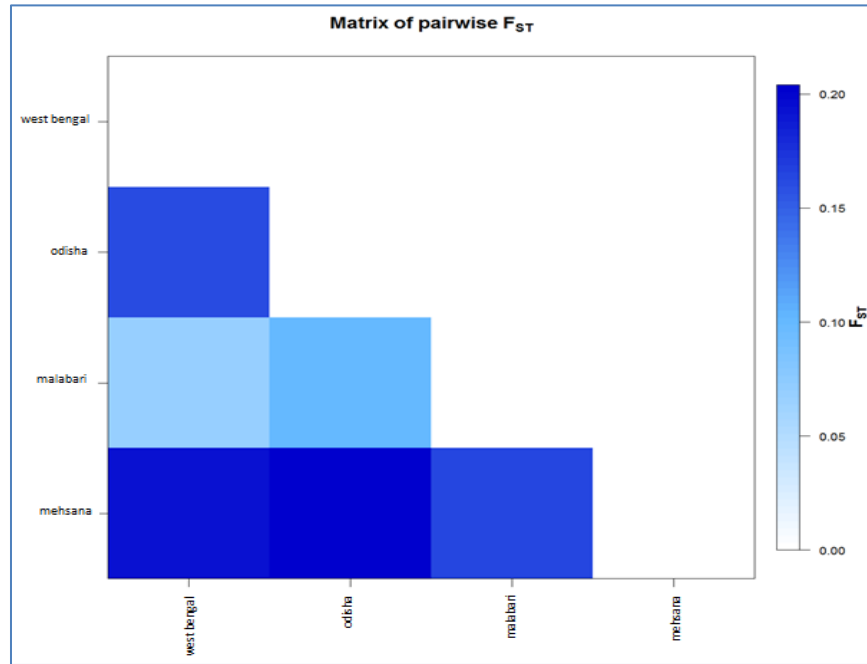


Figure 1: Diagrammatic representation of Matrix of Pairwise  $F_{ST}$  of 4 goat populations of India.

Tajima’s D (Tajima, 1989) and Fu’s  $F_s$  (Fu, 1997) tests were conducted to determine whether patterns of mitochondrial sequence variation were consistent with predictions of the neutral model (Santos *et al.*, 2010) and to infer the demographic history of goats of 4 goat populations of India (Table 3).

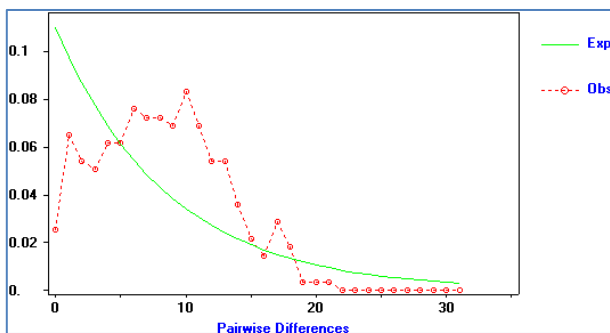
Table 3: Results of Tajima’s D, Fu’s  $F_s$  neutrality tests, SSD (sum of squares deviation) and Raggedness index including associated p-values.

| Regions                  | Tajima’s D (p)    | Fu’s $F_s$ (p)    | SSD (p)          | Raggedness index (p) |
|--------------------------|-------------------|-------------------|------------------|----------------------|
| <b>West Bengal Coast</b> | -1.582<br>(0.030) | -8.792<br>(0.001) | 0.002<br>(1.000) | 0.004<br>(1.000)     |
| <b>Odisha Coast</b>      | -0.365<br>(0.422) | -1.872<br>(0.202) | 0.016<br>(0.320) | 0.020<br>(0.470)     |
| <b>Malabar Coast</b>     | -0.470<br>(0.347) | -2.784<br>(0.141) | 0.010<br>(0.730) | 0.015<br>(0.800)     |
| <b>Mehsana</b>           | -0.602<br>(0.298) | -0.979<br>(0.338) | 0.017<br>(0.100) | 0.028<br>(0.180)     |
| <b>Mean</b>              | -0.755<br>(0.274) | -3.607<br>(0.170) | 0.011<br>(0.537) | 0.017<br>(0.613)     |

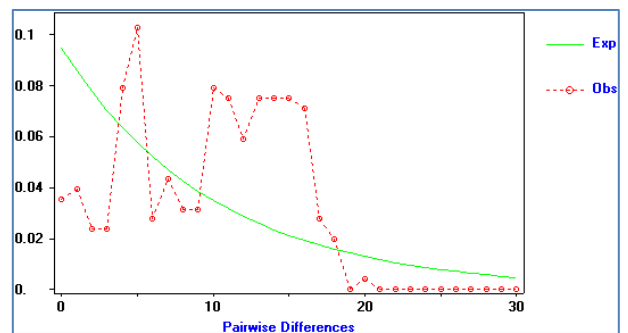
In this study, Tajima’s D values were negative for all the 3 coastal populations, and statistically non-significant, indicating an excess of rare nucleotide site variant compared to what is expected under

a neutral model of evolution, except West Bengal coast population which had significant value (0.030), based on distribution of haplotypes. The  $F_u$ 's  $F_s$  values were also observed to be negative and statistically non-significant in all populations except West Bengal Coast goat population which showed significant  $F_u$ 's  $F_s$  values (0.001). The overall negative values resulting from both tests indicate that there is evidence for an excess number of alleles as would be expected from a recent population or genetic hitchhiking, but the excess is statistically non-significant. This may also suggest that the expansion may have been restricted to separate local areas that resulted in the non-significant negative Tajima's  $D$  and  $F_u$ 's  $F_s$  values for most populations (Liao *et al.*, 2010). Non-significant negative Tajima's  $D$  indicates that the evolution within the populations have been independent of any selection, heterogeneity of mutation rates or major population perturbation (Liao *et al.*, 2010). This may be attributed to higher replacement rate in goats of West Bengal in conjunction with having high proliferacy.

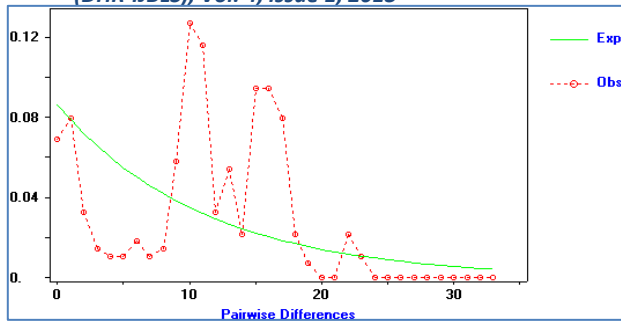
DnaSP v5.10.01 software (Rozas *et al.*, 2010) was used (under the assumption of selective neutrality) for mismatch distribution analysis of pairwise differences of all the 3 coastal goat populations and 1 semi-arid region goat population to evaluate possible historical events of population growth and decline. Mismatch distributions are plots showing the pattern of nucleotide (or restriction) site differences between pairs of individuals in a sample (Rogers, *et al.*, 1996). A population that is at equilibrium is expected to have a multimodal mismatch distribution, whereas populations that have experienced recent growth should have a unimodal mismatch distribution (Slatkin and Hudson, 1991: Rogers and Harpending, 1992). In this study, multimodal mismatch distribution graphs were observed for all the goat populations considered (Figure 2), which indicates stable population or a population in equilibrium.



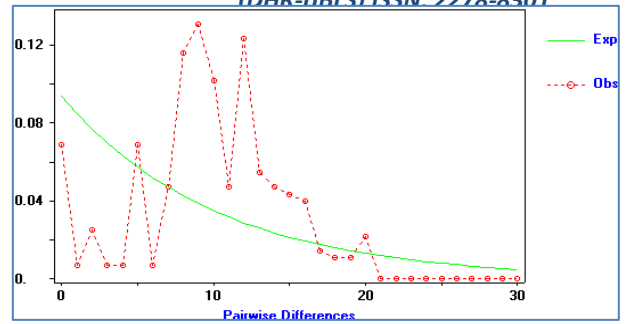
**West Bengal coastal goat**



**Malabar Coast goats**



**Ganjam Coast goats**



**Mehsana goats**

Figure 2: Mismatch distribution graphs for 4 goat populations of India. The x axis shows the number of pairwise differences, the y axis shows the frequency of the pairwise comparisons.

To assess the fit of our data, mismatch distribution under demographic expansion was simulated with SSD (sum of squared deviation) and Harpending's raggedness index ( $r$ ) where a non-significant test indicates a good fit and support of expansion. Non-significant SSD values validated the expansion model. Harpending's raggedness index (Harpending, 1994), based on mismatch distributions was used to test whether the sequence data from each population deviated from what is expected under a sudden expansion model and we observed low non-significant raggedness index (Table 3) which further supported population expansion. Although multimodal mismatch distribution graphs were observed, but low non-significant raggedness index and SSD values indicated population expansion. Sudden population growth results in a unimodal mismatch distribution in a geographically contiguous population, but this pattern can dramatically differ when there is geographical isolation/barrier (Hartl, 2004). With small migration rates between subpopulations, the mean of the mismatch distribution increases and thus the mismatch distribution can become multimodal (Hartl, 2004). Therefore, geographical subdivision might explain some of the anomalies in the mismatch distributions that were observed for the populations considered in this study. Thus, a good fit for demographic expansion model was inferred in the present study.

Despite very high replacement rate of goat species, one of the reasons for growth in goat population is its use as meat animal (17.4% of the total meat in India). High growth rate is also attributable to high fecundity in the goat species (giving birth to twins, triplets or in some cases up to quadruplets).

For better understanding of the population structure and to identify possible phylogenetic lineage of the 3 coastal goat populations and one semi-arid region goat of India, we constructed a phylogenetic tree based on mtDNA D-loop sequences where West Bengal, Malabar and Odisha goat populations clustered together while the Mehsana goat population forms a separate group (Figure 3), this may be because of the different geographical regions of the populations.

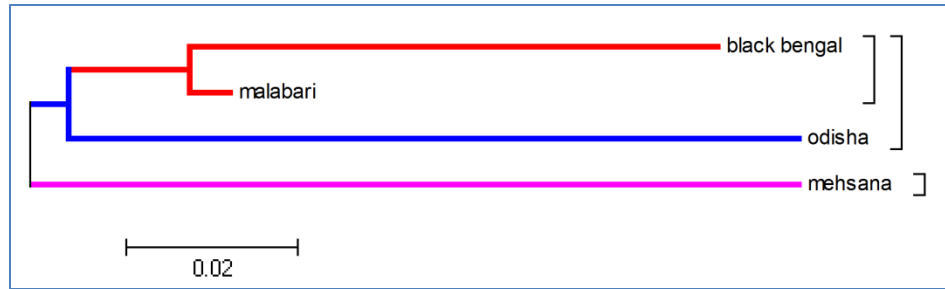


Figure 3: Neighbor-joining tree of 4 goat populations of India based on the mitochondrial D-loop sequences.

In the present study, the median joining network analysis depicted high ratio of singleton haplotypes indicating a high level of sequence diversity in the populations studied (Figure 4). According to Naderi *et al.*, 2007, goat mtDNA haplogroup A is the most diverse and widely distributed across all continents including India. Thus, these D-loop sequences were compared with the sequences given by Naderi *et al.*, 2007 and it was revealed that the Indian goats of 3 coastal regions and semi-arid region belong to the haplogroup A as reported in the study. Therefore, 4 goat populations considered have a single maternal lineage.

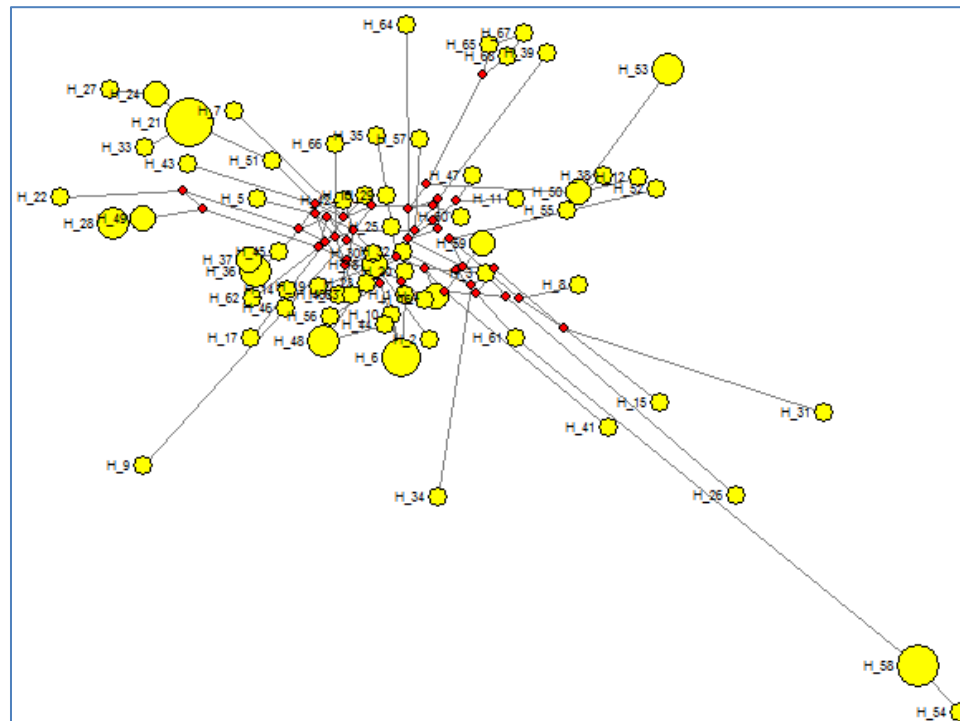




Figure 4: Median-Joining haplotype network of 4 goat populations of India. The circle size is relative to number of haplotypes copies present in the dataset. A branch represents a single nucleotide change. The red dots represent theoretical median vectors introduced by the network software.

## CONCLUSION

Substantial mtDNA diversity has been observed in 4 Indian goat populations. The goat populations have been found to fit the demographic expansion model. Mehsana goats belonging to semi-arid region were clearly separated from the 3 coastal goat populations (West Bengal, Malabar and Odisha) on the basis of  $F_{ST}$  values and Neighbor-joining tree while the phylogenetic inferences using network analysis revealed a single maternal lineage and high level of singleton haplotypes in all the goat populations. Mehsana goats have highest differentiation from rest of the 3 coastal populations on the basis of singleton haplotypes.

## ACKNOWLEDGEMENT

The study was funded by Indian Council of Agricultural Research (ICAR) grant C 3033 under National Agricultural Innovation Project Component IV, Basic and Strategic Research which is gratefully acknowledged.

## REFERENCES

1. A.R. Rogers, A.E. Fraley, M.J. Bamshad, W.S. Watkins and L.B. Jorde, *Mol. Biol. Evol.*, 1996, 13: 895–902.
2. A.R. Rogers, and H. Harpending, *Science*, 1992, 253: 1503–1507.
3. Basic Animal Husbandry Statistics, India 2012 (<http://dahd.nic.in/dahd/bahs-2012.aspx>).
4. D.L. Hartl, (Volume 2), *Nature Reviews*, 2004, doi: 10.1038/nrmicro795, 15-22.
5. F. Tajima, *Genetics*, 1989, 123 (3): 585–9.
6. H. Harpending, *Human Biol.*, 1994, 66:591–600.
7. H.J. Bandelt, P. Forster and A. Röhl, *Mol. Biol. Evol.*, 1999, 16:37-48.
8. I. Mladineo, T. Š. Bubić, R. Stanić and Y. Desdevises, *PLoS One*, 2013, 8(4): e62011.
9. J. Rozas, P. Librado, J. C. Sánchez-DelBarrio, X. Messeguer and R. Rozas, DnaSP v5.10.01, Universitat de Barcelona, Spain 2010.
10. J. Sambrook, E.F. Fritsch and T. Maniatis, (Vol. 3), *Molecular cloning: a laboratory Manual*, (Cold Spring Harbor Laboratory Press, New York 1989)
11. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, *Mol. Biol. Evol.*, 2011, 28, 2731–2739.

12. L. Excoffier, P.E. Smouse, and J.M. Quattro, *Genetics*, 1992, 131: 479-491.
13. M. Nei, *Molecular Evolutionary Genetics*, (Columbia University Press, New York, 1987).
14. M. Slatkin and N.H. Barton, *Evolution*, 1989, 43:1349-1368.
15. M. Slatkin, and R.R. Hudson, *Genetics*, 1991, 129: 555–562.
16. M. Slatkin, *Annu. Rev. Ecol. Evol. Syst.*, 1985a, 16:393-430.
17. M. Slatkin, *Evolution*, 1981, 35: 477–488.
18. M.K. Singh, B. Rai, P. Singh, P.K. Singh and N.P. Singh, *Indian J. Small Ruminants*, 2008,14: 48-70.
19. N. Saitou and M. Nei, *Mol. Biol. Evol.*, 1987, 4:406–425.
20. P.C. Liao, D.C. Kuo, C.C. Lin, K.C. Ho, T.P. Lin and S.Y. Hwang. *BMC Evol. Biol.*, 2010, 10:124.
21. S. Naderi, H.R. Rezaei, P. Taberlet, S. Zundel, S.A. Rafat, H.R. Naghash, M.A. A. El Barody, O. Ertugrul, F. Pompanon, *PLoS ONE*, 2007, 2(10): e1012. doi: 10.1371 /journal.pone.0001012.
22. S. Schneider and L. Excoffier, *Genetics*, 1999, 152: 1079-1089.
23. S. Schneider, D. Roessli, and L. Excoffier, Arlequin Ver 2.001: software for population genetics data analysis, University of Geneva, Switzerland, 2000.
24. S. Wright, *Evolution and the Genetics of Populations: The Theory of Gene Frequencies*, (The University of Chicago Press, Illinois, 1969).
25. S. Wright, *Genetics*, 1943, 28: 114–138.
26. S.R. Santos, Y. Xiang and A.W. Tagawa, *J. of Heredity*, 2010, doi:10.1093/jhered/esq101.
27. T.A. Hall, *Nucl. Acids. Symp. Ser.* 1999, 41:95-98.
28. Y. X. Fu, *Genetics*, 1997, 147: 915–925.