Brief communication

Genomic analysis suggests *KIT*LG is responsible for a roan pattern in two Pakistani goat breeds

Andrea Talenti[1,*], Francesca Bertolini[2,*], Jamie Williams[2], Muhammad Moaeen-ud-Din[3], Stefano Frattini[1], Beatrice Coizet[1], Giulio Pagnacco[1], James Reecy[2], Max F. Rothschild[2], Paola Crepaldi[1+] and the Italian Goat Consortium

[1] Department of Veterinary Medicine, Università degli Studi di Milano, Milan, Italy.

[2] Department of Animal Science, Iowa State University, Ames, Iowa, USA.

[3] Laboratories of Animal Breeding & Genetics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan.

[*] These authors contributed equally to the work

[+]**Correspondence**

Paola Crepaldi

Phone: +39-025 031-6448

Fax:

Mail paola.crepaldi@unimi.it

Running head: *KIT*LG suggested for roan in Pakistani goat breeds

Abstract

The roan coat color pattern is described as the presence of white hairs intermixed with pigmented hairs. This kind of pigmentation pattern has been observed in many domestic species, including the goat. The molecular mechanisms and inheritance that underlie this pattern are known for some species and the *KITLG* gene has been shown associated with this phenotype. To date, no research effort has been done to find the gene(s) that controls roan coat color pattern in goats. In the present study, after genotyping with the GoatSNP50 BeadChip, 35 goats that showed a roan pattern and that belonged to two Pakistan breeds (Group A) were analyzed and then compared to 740 goats of 39 Italian and Pakistan goats breeds that did not have the same coat color pattern (Group B). Run of homozygosity-based and XP-EHH analyses were used to identify unique genomic regions potentially associated with the roan pattern**.** A total of 3 regions on chromosomes 5, 6 and 12 were considered unique among the group A versus B group comparisons. The A region > 1.7 Mb on chromosome 5 was the most divergent between the two groups. This region contains six genes, including the *KITLG* gene. Our findings support the hypothesis that the *KITLG* gene may be associated with the roan phenotype in goats.

Keywords: Coat color, pigmentation, *KIT*LG, Selection signatures

1. Introduction

Genetic studies of coat color and color patterns have been conducted in many domestic animal species. The genetic basis of pigmentation is complex and may involve several genes that can act on several steps of melanogenesis. One of the results of the modulation of the melanocyte activity is the roan pattern, described as presence of white hairs intermixed with pigmented hairs. This pattern gives rise to individual phenotypes that range from almost white to almost completely pigmented. The roan pattern can be confused with ticking pattern, which is characterized by flecks or spot of

color on white area. So far, for the roan pattern, both the *KITLG* gene in Belgian blue cattle (Seitz et al., 1999; Li et al., 2016) and the *KIT* gene in horses (Dürig et al., 2017; Brooks and Bailey, 2005; Marklund et al., 1999) and pigs (Fontanesi et al., 2010; Cho et al., 2011) have been reported, while no genes have been detected as candidate genes for the ticking pattern.

Goats are one of the most important livestock species in developing countries because of their adaptability to different environments. Goats also display a high degree of genetic variability in coat color patterns. To date, only a few SNPs in 20 genes involved in coat color have been considered for their ability to discriminate among goat breeds (Nicoloso et al., 2012) and associations of SNPs and CNVs (Copy Number Variations) to red, black and white colors phenotypes have been reported for the *ASIP* and *MC1R* genes (Fontanesi et al., 2009a; b). No study has been tailored to detect genomic regions associated with the roan pattern in goats.

The present study focuses on the roan pattern seen in the Barri and Beetal Muki Cheni Pakistani breeds. The primary purpose of this investigation was to uncover the genomic region(s) that could be responsible for this pattern and to determine the gene(s) likely responsible for these phenotypes.

2. Material and Methods

2.1. *Sample collection and genotyping*

No animals were sacrificed for this study. Blood samples for Italian goats were collected according to the recommendations of the European Council (1986) concerning animal care. Animals were sampled from different farms in Italy and Pakistan. Italian goats were sampled in order to be as much unrelated as possible and offer a wider sample. Pakistani goats were sampled by PMAS University staff lead by Dr. Mouaeen-ud-din within the villages of Pakistanuder the approved Iowa State University animal care protocol. A total of 929 animals belonging to 41 different breeds were considered in this study (Table S1): 1) 369 animals and 15 breeds from Italy, described by Nicoloso

et al. (Nicoloso et al., 2015). 2) 560 animals of 26 breeds from Pakistan including the small roan like spotted coat pattern Barri and Beetal Muki Cheni breeds (Figure 1a and 1b).

DNA was extracted using commercially available kits and DNA samples were genotyped using the GoatSNP50 BeadChip (Illumina Inc., San Diego, CA) (Tosser-Klopp et al., 2014). SNPs with low call rate (< 95 %), non-polymorphic in all populations (minor allele frequency of 0 %) and not mapped to the assembly or on the X chromosome were excluded from subsequent analyses using Plink v1.9 (Chang et al., 2015). After the exclusion of low quality markers, all animals with call rates < 95 % were excluded from the dataset. Multi-dimensional scaling (MDS) was calculated using the cluster algorithm of Plink v1.9 (Chang et al., 2015). Duplicated and related animals were identified using an in-house script and removed if Identity by State (IBS) > 99% and Discordant Homozygotes (also known as Mendelian Errors, ME) <= 100. Analyses were performed by dividing all breeds into two groups: Group A, which was composed of the two breeds that showed the roan pattern (Barri and Beetal Muki Cheni), and the Group B, which was composed of Italian and Pakistan breeds with different coat color patterns. Genotypes in the resulting dataset were imputed and phased using Beagle v3.3.2 software considering one breed at a time (Browning, 2011).

2.2. *Runs of Homozygosity and ROH comparison*

The screening for the low heterozygosity regions across the genome was conducted for the two groups using the plink V1.9 software (Chang et al., 2015). The analysis was performed using 20 SNPs sliding windows (command --homozyg-snp), allowing no heterozygotes in each (--het 0). The number of homozygous animals at each marker was then normalized dividing that number by the total number of goats included in the analysis, obtaining a locus homozygosity score (H) ranging from 0 (0 %) to 1 (100 %). . The H score at each SNP calculated for the Group A was then compared with the H scores of all other populations calculated separately following the same

parameters previously described (dROH analysis). This comparison was performed applying the formula reported in Bertolini et al., 2016, obtaining a dH score that was then used to find the markers that most distinguish the Group A against all other breeds in Group B. Only regions with an H score value > 0.46 for the group A and a dH score value > 3.39 for the ROH comparison, which represented the top 0.2% of all SNP, were considered highly homologous for the ROH analysis and highly divergent for the ROH comparison respectively. These regions were then investigated for genes using most recent gene annotation available for *Capra hircus* genome (ARS1; Bickhart et al., 2017) and screened for the region of interest using the Bedtools software (Quinlan and Hall, 2010).

2.3. *Haplotype analysis and gene annotation*

Single SNP Cross-Population Extended Haplotype Homozygosity (XP-EHH) was calculated using the Selscan software (Szpiech and Hernandez, 2014). This method compared the integrated EHH at each marker between two different populations, allowing the detection of strong, directional selection of one allele in one of the two populations while remaining polymorphic in the other (Sabeti et al., 2007). Ancestral alleles were determined considering 15 wild animals of other *Capra* species including 8 *Capra ibex* and 7 *Capra Aegagrus* genotyped with the same SNPchip (data not shown).

All markers with normalized XP-EHH score above 3.09 (top 0.2 %) were considered as relevant and only the regions above the threshold that were concordant with the regions detected through the ROH analyses were considered for further investigations.

3. Results

The initial dataset contained 53,347 SNPs: among the 50,619 autosomal SNPs, 3,213 did not pass the quality-filtering steps. Animal filtering removed 72 individuals due to low call rate, 1 animal

was excluded because it was duplicated and 82 animals were excluded because of high relatedness with other animals. After data editing, a total of 47,406 autosomal SNPs, 35 goats in Group A and 740 goats in Group B were retained. The MDS plot, produced on the imputed dataset (Figure S1), showed a clear separation between breeds raised in Italy and Pakistan, with no overlap between the two major clusters.

3.1. *Run of homozygosity*

Using the selected thresholds previously mentioned, six ROH with normalized H scores ≥ 3.4 were considered regions of high homozygosity in the group A (Figure 2, Supplementary Figures S2-5). Two regions were located on chromosome 5 with lengths of 308.49 Kb and 1.75 Mb respectively (16,323,819-16,632,308 bp and 17,885,772-19,634,050 bp). Two other regions were located on chromosome 6, with lengths of 1.31Mb and 230.0 Kb (13,341,998-14,648,705 bp and 69,566,293-69,796,504 bp). One region was located on chromosome 13, with length of 308.19 Kb (77,687,773-77,995,965) and on chromosome 24, with length 706.19Kb (46,684,650-47,390,842 bp). The region with the highest H score was the 1.7 Mb region on chromosome 5, with the max H value of 0.71 reached within the sub region between 18,127,496 and 18,846,626 bp. The remaining regions showed lower H values that ranged from 0.46 to 0.5. Among these regions, three were also above the dROH analysis threshold that compared ROH of group A with all ROH in each breed of group B (Figure 2): the 1.7Mb region on chromosome 5, the 230.0 Kb region chromosome 6, and the 308.19 Kb region on chromosome 13. The region on chromosome 5 contained 9 genes, the region on chromosome 6 contained 1 gene and the region on chromosome 13 contained 4 genes. The full list of the 14 genes included in these 3 regions is reported in Table S2. Among these genes, a marker on chromosome 5, which had the max H value of 0.71 (snp14289-scaffold157-1998233, 18,127,496 bp), was located within the *KITLG* gene (from 18,044,632 to 18,151,924 bp).

3.2. *Haplotype analysis*

The XP-EHH analysis was performed to discover regions of the genome that were under strong directional selection. Normalized XP-EHH analysis identified a total of 18 regions of the genome with more than 1 consecutive SNP that were under strong directional selection (Figure 2, Supplementary Table S3). Among these, the longest region was located on chromosome 5, from 18,127,496 to 19,478,621 bp. This region overlaps the region identified by ROH-based analyses and contained the *KITLG* gene. These results indicate that this region (and gene) was under strong selection. In addition, no other regions were in common with the ROH identified region, which, highlights the importance of this region.

4. Discussion

Runs of homozygosity analyses identified several regions of the genome that might be associated with roan coat color. The region with the highest H score for Group A was on chromosome 5, in a 1.7 Mb region from 17,448,053 to 19,198,567 bp, which included a genetic marker within the *KITLG* gene sequence. This region was also identified by standardized H score analyses, which compared the ROH in the two populations pooled together with all other ROH in all other populations. Furthermore, this same region was identified by XP-EHH. This approach also identified several other regions that were not found using ROH, which indicates that those signals were either non-specific or the result of different selection forces. Finally, even after removing animals from Italy and the most divergent from Pakistan based on the MDS plot, the results still confirm this region.

An additional clue of the potential involvement of this gene in this pigmentation pattern is given by the minor allele frequency of the marker within the *KITLG* gene sequence (snp14289-scaffold157-1998233; 18,127,496 bp). This marker was completely fixed in the Group A (MAF = 0), whereas it was variable in Group B individuals (MAF = 0.294).

The *KIT* ligand (*KITLG*) gene is known as Mast Cell Grow Factor (*MGF*) or Stem Cell Factor (*SCF*). and it is involved in and can affect many biological processes, such as hematopoiesis, gametogenesis and melanogenesis. The *KITLG* gene affects pigmentation in both human and mice (Guenther et al., 2014). A missense mutation in the 7th exon of this gene was associated with the roan phenotype in cattle (Seitz et al., 1999). The genetic basis of the roan pattern has also been studied in the horse and pig, but no mutation in the *KITLG* gene has been associated with the roan phenotype, whereas the *KIT* gene has been found as the major candidate for these species (Fontanesi et al., 2010; Cho et al., 2011; Marklund et al., 1999). Cattle and goats belong to the same *Bovidae* family, have the same number of chromosomes and few rearrangements at the genomic level (Schibler et al., 2009; Cribiu et al., 2001; Dong et al., 2013). Therefore, it is not unusual to find similarities both at the genetic and biological level. Polymorphisms in the *KITLG* gene have already been associated with litter size in goat (An et al., 2012, 2015). No phenotypic information about litter size was available in our dataset, and therefore this possible association could not be tested with this present dataset.

5. Conclusions

The results of this study indicate that a region that includes *KITLG* gene has undergone a loss of variability in two caprine populations that exhibit the roan pattern. Particularly, a SNP located within the gene is monomorphic in the roan group compared to many other breeds. These results indicate that *KITLG* gene is the likely a strong gene candidate for this roan pattern phenotype in goats and should be investigated further.

6. Funding

This work was supported by State of Iowa funds, Iowa State University, the Ensminger International Animal Agriculture fund; the Italian Ministry of Agriculture (grant INNOVAGEN);

the European Union's Seventh Framework Programme (FP7/2010-2014) "NEXTGEN", and by the U.S. Agency for International Development through the Pakistan – U.S. Science & Technology Cooperation Program. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the U.S. Agency for International Development.

7. Data Availability

We have deposited the primary genotyping data on Dryad.

8. Competing interests

The authors declare that they have no competing interests.

9. References

An, X.P., J.X. Hou, T.Y. Gao, Y.N. Lei, Y.X. Song, J.G. Wang, and B.Y. Cao. 2015. Association analysis between variants in KITLG gene and litter size in goats. *Gene*. 558:126–130. doi:10.1016/j.gene.2014.12.058.

An, X.P., J.X. Hou, G. Li, Y.X. Song, J.G. Wang, Q.J. Chen, Y.H. Cui, Y.F. Wang, and B.Y. Cao. 2012. Polymorphism identification in the goat KITLG gene and association analysis with litter size. *Anim. Genet.* 43:104–107. doi:10.1111/j.1365-2052.2011.02219.x.

Bertolini, F., B. Gandolfi, E.S. Kim, B. Haase, L.A. Lyons, and M.F. Rothschild. 2016. Evidence of selection signatures that shape the Persian cat breed. *Mamm. Genome*. 27:144–155. doi:10.1007/s00335-016-9623-1.

Bickhart, D.M., B.D. Rosen, S. Koren, B.L. Sayre, A.R. Hastie, S. Chan, J. Lee, E.T. Lam, I. Liachko, S.T. Sullivan, J.N. Burton, H.J. Huson, J.C. Nystrom, C.M. Kelley, J.L. Hutchison, Y. Zhou, J. Sun, A. Crisà, F. Abel Ponce de León, J.C. Schwartz, J.A. Hammond, G.C. Waldbieser, S.G. Schroeder, G.E. Liu, M.J. Dunham, J. Shendure, T.S. Sonstegard, A.M.

Phillippy, C.P. Van Tassell, and T.P. L Smith. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Publ. Gr.* 49:4–3. doi:10.1038/ng.3802.

Brooks, S.A., and E. Bailey. 2005. Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mamm. Genome*. 16:893–902. doi:10.1007/s00335-005-2472-y.

Browning, B.L. 2011. Beagle 3.3.2. 1–30.

Chang, C.C., C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 4:7. doi:10.1186/s13742-015-0047-8.

Cho, I.-C., T. Zhong, B.-Y. Seo, E.-J. Jung, C.-K. Yoo, J.-H. Kim, J.-B. Lee, H.-T. Lim, B.-W. Kim, J.-H. Lee, M.-S. Ko, and J.-T. Jeon. 2011. Whole-genome association study for the roan coat color in an intercrossed pig population between Landrace and Korean native pig. *Genes Genomics*. 33:17–23. doi:10.1007/s13258-010-0108-4.

Cribiu, E.P., D. Di Berardino, G.P. Di Meo, A. Eggen, D.S. Gallagher, I. Gustavsson, H. Hayes, L. Iannuzzi, C.P. Popescu, J. Rubes, S. Schmutz, G. Stranzinger, A. Vaiman, and J. Womack. 2001. International System for Chromosome Nomenclature of Domestic Bovids (ISCNDB 2000). *Cytogenet. Cell Genet.* 92:283–99. doi:56917.

Dong, Y., M. Xie, Y. Jiang, N. Xiao, X. Du, W. Zhang, G. Tosser-Klopp, J. Wang, S. Yang, J. Liang, W. Chen, J. Chen, P. Zeng, Y. Hou, C. Bian, S. Pan, Y. Li, X. Liu, W. Wang, B. Servin, B. Sayre, B. Zhu, D. Sweeney, R. Moore, W. Nie, Y. Shen, R. Zhao, G. Zhang, J. Li, T. Faraut, J. Womack, Y. Zhang, J. Kijas, N. Cockett, X. Xu, S. Zhao, J. Wang, and W. Wang. 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat. Biotechnol.* 31:135–141. doi:10.1038/nbt.2478.

Dürig, N., R. Jude, H. Holl, S.A. Brooks, C. Lafayette, V. Jagannathan, and T. Leeb. 2017. Whole genome sequencing reveals a novel deletion variant in the *KIT* gene in horses with white spotted coat colour phenotypes. *Anim. Genet.* 48:483–485. doi:10.1111/age.12556.

Fontanesi, L., F. Beretti, V. Riggio, S. Dall'Olio, E.G. González, R. Finocchiaro, R. Davoli, V. Russo, and B. Portolano. 2009a. Missense and nonsense mutations in melanocortin 1 receptor (MC1R) gene of different goat breeds: association with red and black coat colour phenotypes but with unexpected evidences. *BMC Genet.* 10:47. doi:10.1186/1471-2156-10-47.

Fontanesi, L., F. Beretti, V. Riggio, E. Gómez González, S. Dall'Olio, R. Davoli, V. Russo, and B. Portolano. 2009b. Copy Number Variation and Missense Mutations of the Agouti Signaling Protein (*ASIP)* Gene in Goat Breeds with Different Coat Colors. *Cytogenet. Genome Res.* 126:333–347. doi:10.1159/000268089.

Fontanesi, L., E. D'Alessandro, E. Scotti, L. Liotta, A. Crovetti, V. Chiofalo, and V. Russo. 2010. Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. *Anim. Genet.* 41:478–492. doi:10.1111/j.1365-2052.2010.02054.x.

Guenther, C.A., B. Tasic, L. Luo, M.A. Bedell, and D.M. Kingsley. 2014. A molecular basis for classic blond hair color in Europeans. *Nat. Genet.* 46:748–52. doi:10.1038/ng.2991.

Li, W., A. Sartelet, N. Tamma, W. Coppieters, M. Georges, and C. Charlier. 2016. Reverse genetic screen for loss-of-function mutations uncovers a frameshifting deletion in the melanophilin gene accountable for a distinctive coat color in Belgian Blue cattle. *Anim. Genet.* 47:110–113. doi:10.1111/age.12383.

Marklund, S., M. Moller, K. Sandberg, and L. Andersson. 1999. Close association between sequence polymorphism in the KIT gene and the roan coat color in horses. *Mamm. Genome.* 10:283–288. doi:10.1007/s003359900987.

Nicoloso, L., L. Bomba, L. Colli, R. Negrini, M. Milanesi, R. Mazza, T. Sechi, S. Frattini, A. Talenti, B. Coizet, S. Chessa, D. Marletta, M. D'Andrea, S. Bordonaro, G. Ptak, A. Carta, G. Pagnacco, A. Valentini, F. Pilla, P. Ajmone-Marsan, and P. Crepaldi. 2015. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet. Sel. Evol.* 47. doi:10.1186/s12711-015-0140-6.

Nicoloso, L., R. Negrini, P. Ajmone-Marsan, and P. Crepaldi. 2012. On the way to functional agro biodiversity: coat colour gene variability in goats. *Animal*. 6:41–9. doi:10.1017/S175173111100139X.

Quinlan, A.R., and I.M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26:841–2. doi:10.1093/bioinformatics/btq033.

Sabeti, P.C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E.H. Byrne, S.A. Mccarroll, S.F. Schaffner, E.S. Lander, and T.I. Hapmap. 2007. Genome-wide detection and characterization of positive selection in human populations. *October*. 449:913–918. doi:10.1038/nature06250.Genome-wide.

Schibler, L., G.P. Di Meo, E.P. Cribiu, and L. Iannuzzi. 2009. Molecular cytogenetics and comparative mapping in goats (Capra hircus, 2n = 60). *Cytogenet. Genome Res.* 126:77–85. doi:10.1159/000245908.

Seitz, J.J., S.M. Schmutz, T.D. Thue, and F.C. Buchanan. 1999. A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mamm. Genome*. 10:710–712. doi:10.1007/s003359901076.

Szpiech, Z.A., and R.D. Hernandez. 2014. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* 31:2824–2827. doi:10.1093/molbev/msu211.

Tosser-Klopp, G., P. Bardou, O. Bouchez, C. Cabau, R. Crooijmans, Y. Dong, C. Donnadieu-Tonon, A. Eggen, H.C.M. Heuven, S. Jamli, A.J. Jiken, C. Klopp, C.T. Lawley, J. McEwan, P. Martin, C.R. Moreno, P. Mulsant, I. Nabihoudine, E. Pailhoux, I. Palhière, R. Rupp, J. Sarry, B.L. Sayre, A. Tircazes, Jun Wang, W. Wang, W. Zhang, and and the I.G.G. Consortium. 2014. Design and Characterization of a 52K SNP Chip for Goats. *PLoS One*. 9:e86227.

10. Figure Captions

Figure 1. Pictures of a) Barri and b) Beetal Muki Cheni goat breeds. The distribution of roan ranged from almost colored to almost white with only few spots.

Figure 2. Circos plot of the ROH H scores and dH score (outer circles, grey and blue tracks, respectively), XP-EHH (inner circle, green track); for each analysis the top 0.2% SNPs are marked in red. See online version for full colors.

Figure 1a

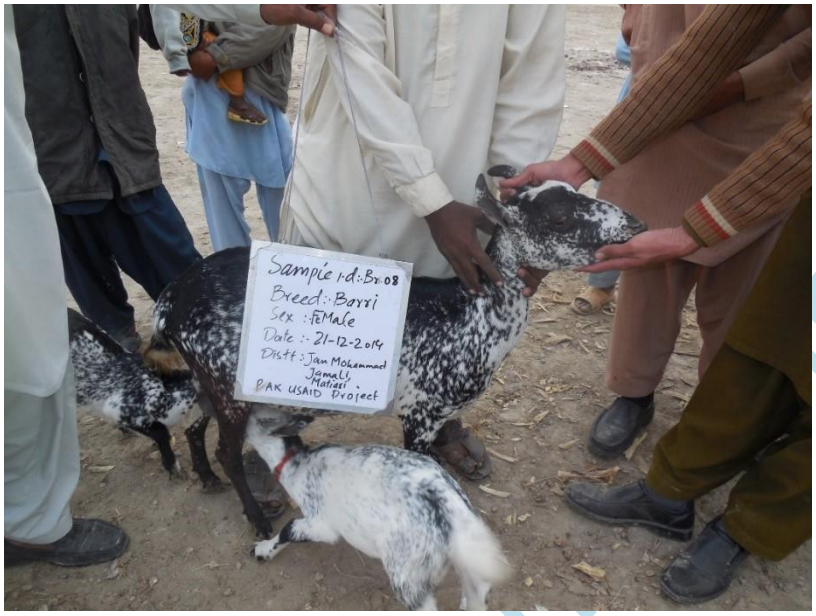

Figure 1b

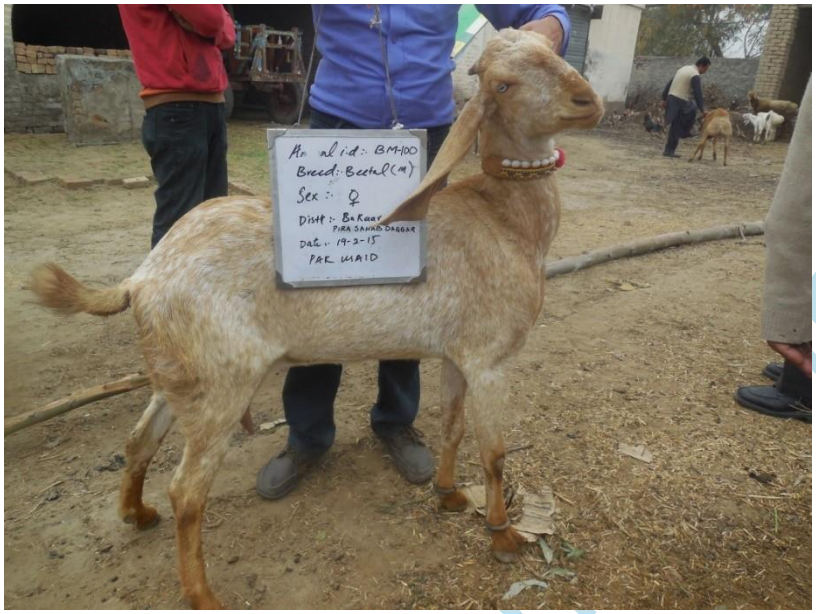

Figure 2