Journal of Animal & Plant Sciences, 34(3): 2024, Page: 809-816
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2024.3.0767
Short Communication

GENETIC DIFFERENTIATION AMONG RED COLORED BOS INDICUS SUBCONTINENT ORIGIN BREEDS USING SNP CHIP DATA

M. Moaeen-ud-Din*[1Ψ], N. Abbas[1Ψ], M. S. Khan[2], A. Yousaf[3], R. D. Muner[1] and S. Mehreen[1]

[1]Laboratory of Functional Genomics, Department of Animal Breeding & Genetics, Faculty of Veterinary & Animal Sciences, PMAS Arid Agriculture University, Rawalpindi, 46300, Pakistan
[2]Cholistan University of Veterinary & Animal Sciences, Bahawalpur, Pakistan
[3]Department of Clinical Studies, Faculty of Veterinary & Animal Sciences, PMAS Arid Agriculture University, Rawalpindi, 46300, Pakistan
*Corresponding author's email: drmoinawan@gmail.com
[Ψ]These authors contributed equally

ABSTRACT

In Pakistan, Sahiwal and Cholistani cattle are considered as dairy breeds with traits of disease resistance, heat tolerance, and adaptability to high temperature. These breeds are considered two different breeds although Sahiwal is considered an off shoot of Cholistani cattle. However, these breeds have never been characterized genotypically using SNP data. Moreover, information on genotypic differentiation is unavailable for other subcontinent red colored breeds i.e., Gir, Brahman and Red Sindhi. Thus, present study was meant to investigate distinctive genetic architect if any of these red color breeds of same region of the world. Genotype data of 50K SNP Bead chip was retrieved online from Dryad database. Quality control measures were performed in Plink 1.9 software. Admixture, Principal component analysis and Multidimensional scaling were used for breeds distinction while genetic diversity was estimated using linkage disequilibrium. The genetic structure of these breeds clearly differentiated them from each other performed by admixture analysis from K =2 to K =5. Similarly, all breeds principally differentially scattered in PCA and MDS results. The study revealed different genetic structure for Cholistani and Sahiwal indicating as distinct breeds while Brahman had mixed genetic architect of other *Bos indicus* subcontinent breeds. The findings of the current study stress for separate genetic improvement program of Sahiwal and Cholistani cattle.

Key words: Cholistani, Sahiwal, SNPs chip, Admixture, Pakistan.

INTRODUCTION

Pakistani cattle breeds are of *Bos indicus* origin. The cattle population in Pakistan is currently 51.5 million heads (GoP, 2021). Cholistani cattle are inhabitants of the desert areas of Cholistan, Punjab, Pakistan. Cholistan is the driest and hottest zone of Pakistan. Being the inhabitant of this area, Cholistani cattle breed can tolerate high temperatures (Farooq *et al*., 2010). Cholistani breed is a dairy cattle breed found in a range of coat colors i.e., white with brown or black spots, pure white, pure grey and pure dark brown (similar to the other red colored breeds of the region) with long ears and the small horns apart from the notion that it considered as parent breed for development of Sahiwal breed. Sahiwal cattle breed is the main local dairy cattle breed of Pakistan belonging to *Bos indicus.* Sahiwal cattle basically originate from the Sahiwal district of Punjab province in Pakistan (McNamara and McSweeney, 2022) and is considered as an offshoot of Cholistani cattle (Farooq *et al*., 2010). It has characteristics of heat tolerance and can adapt under changing environmental conditions. This breed is famous for its high milk production, disease resistance, tick resistance, and its ability to adapt to high temperature environment. Sahiwal cattle have dark brown coat color with a well-developed dewlap and udder (McNamara and McSweeney, 2022).

Phenotypically Sahiwal and Cholistani cattle breeds are characterized and considered to be two different breeds except common dark brown coat color (Sahiwal has dark brown coat while animals of same color are also found in Cholistanti). Moreover, Sahiwal cattle has been developed as dairy breed of tropics during last 100 years in the central-south Punjab (Farooq *et al*., 2010). However, there is very limited information available about their genotypic differentiation whether these red colored breeds i.e., Sahiwal, Cholistani, Red Sindhi, Gir and Brahman are one and same or different.

The study of genetic differences at the molecular level has developed rapidly since the 20th century. Thus, information on genetic variations based on molecular data and the use of different available technologies have made it possible to find the genetic structure among populations, breeds, and species (Elsik *et al.*, 2009). In the 1970s, the study of the first genetic marker protein polymorphism was used in livestock species and in 1980s, it was replaced by restriction enzyme analysis. For genetic data analysis, the polymerase chain reaction (PCR) based on the genotyping data has been provided rapid testing, however, Single Nucleotide Polymorphism (SNP) is the common form of polymorphism and this genetic marker contains more information about the genetic variations among populations. UMD 3.1 and Btu_4.2 are different Bovine genome assemblies available on the NCBI website that have helped to develop the Illumina Bovine HD SNP and BeadChip genotyping array. The Bovine HD Chip has the superior power to identify the genetic structure of many cattle breeds. The main objective of the Bovine HD SNP Chip is the identification of genetic differentiations, linkage disequilibrium, minor allele frequency, breed characterizations, and genetic evaluation studies (Elsik *et al.*, 2009). Finally, various techniques have provided evidence for the usefulness of SNP markers in studying genetic variations and population structures for numerous cattle breeds (Behl *et al.*, 2016, Kamiti *et al.*, 2016, Kumar *et al.*, 2016).

The earliest breed characterization of Pakistani livestock was performed by using morphometric data and their correlation with reproductive and performance attributes. There has been infrequent research on the genetic differentiation of Pakistani cow breeds using mitochondrial and low-density microsatellite markers (Rehman and Khan, 2009, Hussain *et al.*, 2016). Low-density markers have also been used to characterize the features of Pakistani cattle breed adaptation and tick resistance however, among these breeds, there is little knowledge of the full genome or significant functional genomic region (Rehman and Khan, 2009). However, no systematic genome-wide studies have been done to investigate genetic diversity among Pakistani native breeds using genomic technologies (Mustafa *et al.*, 2014). Evaluation of genetic diversity may help to create local reference population of cattle for the creation of future synthetic breeds (Edea *et al.*, 2015).

The availability of SNP chips is an important approach for the documentation of dairy cattle genotypes and for the measurement of their degree of divergence. Bovine SNP chips have been used for estimation of population structure, genetic diversity, and genetic variations of dairy cattle (Chagunda *et al.*, 2018). Many versions of the Bovine SNP chip (High, medium, low density) are available for the measurement of dairy cattle genotypes (Dassonneville *et al.*, 2012). In Pakistan, there is limited use of Bovine SNP chips for dairy cattle breeds differentiation. Therefore, this study was designed to investigate if: i. Sahiwal and Cholistani cattle are genotypically two distinct breeds or not by using Bovine 50K SNP chips despite of common origin (Sahiwal is off shoot of Cholistani cattle) and having common brown color between two breeds and ii. if other red color breeds i.e., Gir, Brahman and Red Sindhi are genetically distinct or not. This new knowledge may help to guide future breeding policies for the conservation and improvement of our local genetic resources of Sahiwal and Cholistani cattle in particular.

MATERIALS AND METHODS

Genotype data were obtained from Dryad repository online resources (https://www.datadryad.org/) deposited by Division of Animal Sciences, University of Missouri, Columbia, Missouri, United States of America (Decker *et al.*, 2014). The data were originally generated by genotyping using Illumina Bovine 50K SNP Chip by Illumina, San Diego, Calf, USA and made available by Decker *et al.*, (2014) on https://www.datadryad.org/. Genotypic data used in current study that was retrieved from aforementioned source included total of 78 animals from five different breeds viz. Sahiwal (n = 19), Cholistani (n = 11), Red Sindhi (n = 10), Gir (n = 20) and Brahman (n = 20). Quality Control tests were performed to filter the genotype data by using the Plink 1.9 available on http://pngu.mgh.harvard.edu/purcell/plink/ (Purcell *et al.*, 2007) with a genotyping rate of 90%, a minor allele frequency (MAF) of 5% and a Hardy Weinberg Equilibrium (HWE) threshold level of 0.001 yielding 43043 SNPs for final analysis.

Clustering of aforementioned breeds was done with Admixture Analysis (AA) that was confirmed by Principal Component Analysis (PCA) and Multidimensional Scaling (MDS). AA was performed with Admixture 1.3 (Alexander *et al.*, 2009) for 78 animals genotyped using Illumina Bovine 50K SNP Chip with 43043 SNPs included in final analysis. The admixture analysis was carried out to predict the best fitted K values (number of populations) for the dataset (Alexander *et al.*, 2009). K values (the parameter explaining subpopulation number and input dataset provided the total populations) and a cross-validation was carried out for K = 2 to K = 5 populations. The Q values retrieved from breeds were plotted by representing the ancestry population of each individual animal. PCA was performed in Plink 1.9 software to identify the breeds (Purcell *et al.*, 2007). Eigenvectors produced for each individual of respective breeds were plotted using Package QQman, and XQuartz-2.7.11 in R program (Sharma *et al.*, 2016). MDS analysis, another statistical approach was used for population stratification to affirm the PCA results using Plink 1.9 software (Purcell *et al.*, 2007). MDS analysis results were visualized as two-dimensional graphical clustering in the R program (Sharma *et al*., 2016).

Pattern of Linkage disequilibrium (LD) – a measure of population breeding history and evolution; was measured by the square of correlation coefficient (r^2) of allelic frequencies at a pair of loci. Pair-wise (r^2) between adjacent SNPs was calculated for each autosome and the genome-wide LD over all autosomes was estimated in each breed using of the Bovine 50K SNPs chip dataset. The genotypic dataset of selected individuals was performed for the estimating of LD using Plink software. Breed-wise LD (r^2) values were calculated for each pair of 1,000 consecutive SNPs at most 1 Mb apart using a sliding window (Kim *et al*., 2018).

RESULTS AND DISCUSSION

Application of quality control tests yielded 43043 SNPs for further analyses. The SNPs removed as a result of quality control tests included: 504 SNPs due to missing genotype data, 324 SNPs due to the Hardy-Weinberg test and 19382 due to minor allele frequency.

Conservation of the cattle breeds and information about population structure and genetic diversity is important among the cattle breeds for genetic improvement as well as judicious use of genetic resources. The current study explored the population structure of five cattle breeds to differentiate among these breeds.

The admixture analysis based on 43043 SNPs for five cattle breeds viz. Cholistani, Sahiwal, Red Sindhi, Gir, and Brahman for different number of population (K=2 to 5) differentiated five cattle breeds. The population structures were clearly distinguished into pre-defined K values and some indicated shared ancestry of paternal lineages (Figure 1). The cross validation concluded five distinct populations for analysis viz. K=5. Moreover, as explained in Figure 1; we did observe shared ancestry between Brahman and Red Sindhi at K = 2 with lowest cross-validation where other breeds were independent but clustered as one breed i.e., Sahiwal, Gir and Cholistani. At K = 3, shared ancestry between Cholistani and Red Sindhi is evident and some mixing with Sahiwal and Brahman breed is observed. At K = 4 and K = 5; Brahman and Red Sindhi showed admixed background, however, Sahiwal, Cholistani and Gir were quite distinct. In summary, Cholistani and Sahiwal breeds were clearly separated from the other three cattle breeds at K = 4 and K = 5. Gir cattle was also differentiated from other breeds while admixture portion is represented in Brahman and Red Sindhi. Furthermore, in previous study, admixture analysis performed among Iraqi cattle and four reference breeds (African taurine N'Dama Guinea, Holstein Friesian, African crossbreed Sheko and Asian Zebu Nellore). The selected Iraqi breeds Rustaqi and Jenoubi represented shared the ancestry with four breeds higher in Jenoubi than in Rustaqi which had mixed background with Holstein-Friesian (Alshawi *et al*., 2019). This suggested a probable event of crossbreeding in recent past which also seems to be case in the current study where Brahman samples taken were from Brazil which probably mixed with other *Bos indicus* cattle i.e., Red Sindhi which was imported to Brazil at same time during British era from the subcontinent. Thus, probably in Brazil these breeds mixed somehow through intentionally or unintentionally crossbreeding event which indicated historical gene flow (Bhuiyan *et al*., 2021). Moreover, the results of current study are similar to previous in term of distinctness of Sahiwal studied with other subcontinent breeds and African breeds (Dixit *et al*., 2020, Gebrehiwot *et al*., 2020, Dixit *et al*., 2021) and Gir cattle when studied with other subcontinent breeds (Dixit *et al*., 2020, Dixit *et al*., 2021) using genotype data.

PCA and MDS refer to series of statistical methods which reduce the complexity of data, enhance data visualization by exploring relational structure on the basis of similarity among the groups (Lewis *et al*., 2011, Hout *et al*., 2013). Consistent results were also obtained from MDS and PCA for five, three and two breeds populations in the current study (Figures 2 and 3). Among these five breeds Brahman and Gir are clearly separated from others three breeds. The other three breed's populations were observed clear divergence between Cholistani and Sahiwal breeds. The mixing of some individuals was observed between Cholistani and Red Sindhi breeds population which might be due to gene flow as result of crossbreeding. Plotting for V2:3 clearly differentiated 5 breeds although not as well distanced (Figure 2). In conclusion, based on genotype data, our results clearly demonstrated that five cattle breeds populations were distinct due to their independent clusters given by PCA and MDS plots. Furthermore, Brahman had relatively wide spread as compared rest of the breeds probably due to mixed origin as revealed in Admixture results (Figure 1), thus, resulting into a comparatively diversified population. However, MDS revealed a weak differentiation among Pakistani origin breeds viz. Red Sindi vs Cholistani and Sahiwal vs Cholistani that was probably due to a weak differentiation among these breeds as previously observed in case of Bangladeshi cattle population (Bhuiyan *et al*., 2021). Moreover, Sahiwal and Gir were also distinctly plotted in previous studies when compared to other subcontinent and African breeds (Dixit *et al*., 2020, Gebrehiwot *et al*., 2020, Dixit *et al*., 2021).

The results of LD decay corresponding with the distance between SNPs, the r^2 level starts at 0.43, 0.40, 0.39, 0.37, 0.35 for Cholistani, Sahiwal, Red Sindhi, Gir, and Brahman respectively (Figure 4) when using the 250 Kb bin of SNPs. The decay of LD was more rapid in Brahman, and Gir than other three breeds. After these declines, the decay of each breed was less at a distance of 1000 Kb. These cattle breeds were selected for displayed the extremes of LD level and decay as a function of increasing the physical distance. Extremes LD level (r^2=0.43) was observed in Cholistani cattle persistently higher values compared with others breed. Moreover, the lowest was observed in Brahman cattle (r^2= 0.34). It may be noticed that the decay of LD sharply declines towards the end of marker distance. In previous studies the results of LD values across the genomic distance was observed higher displayed in Sahiwal cattle as compared to other breeds viz. Non descriptive Desi (DES), Deshi (DES), North Bengal Gery (NBG), Pabna (PC), Munshiganj (MC), and Red Chittagong (RCC) while RCC had the lowest value (Bhuiyan *et al*., 2021). The linkage disequilibrium (LD) was observed in sheep and can be compared to other domestic species. Border Leicester which had the highest LD and lower values of r^2 with comparable distance than present in dogs and dairy cattle breeds while the Merino display a lower extend of LD than other species (Kijas *et al*., 2014).

In conclusion, the current study revealed a distinction between Sahiwal and Cholistani cattle as well as other red colored humped cattle of the subcontinent origin.

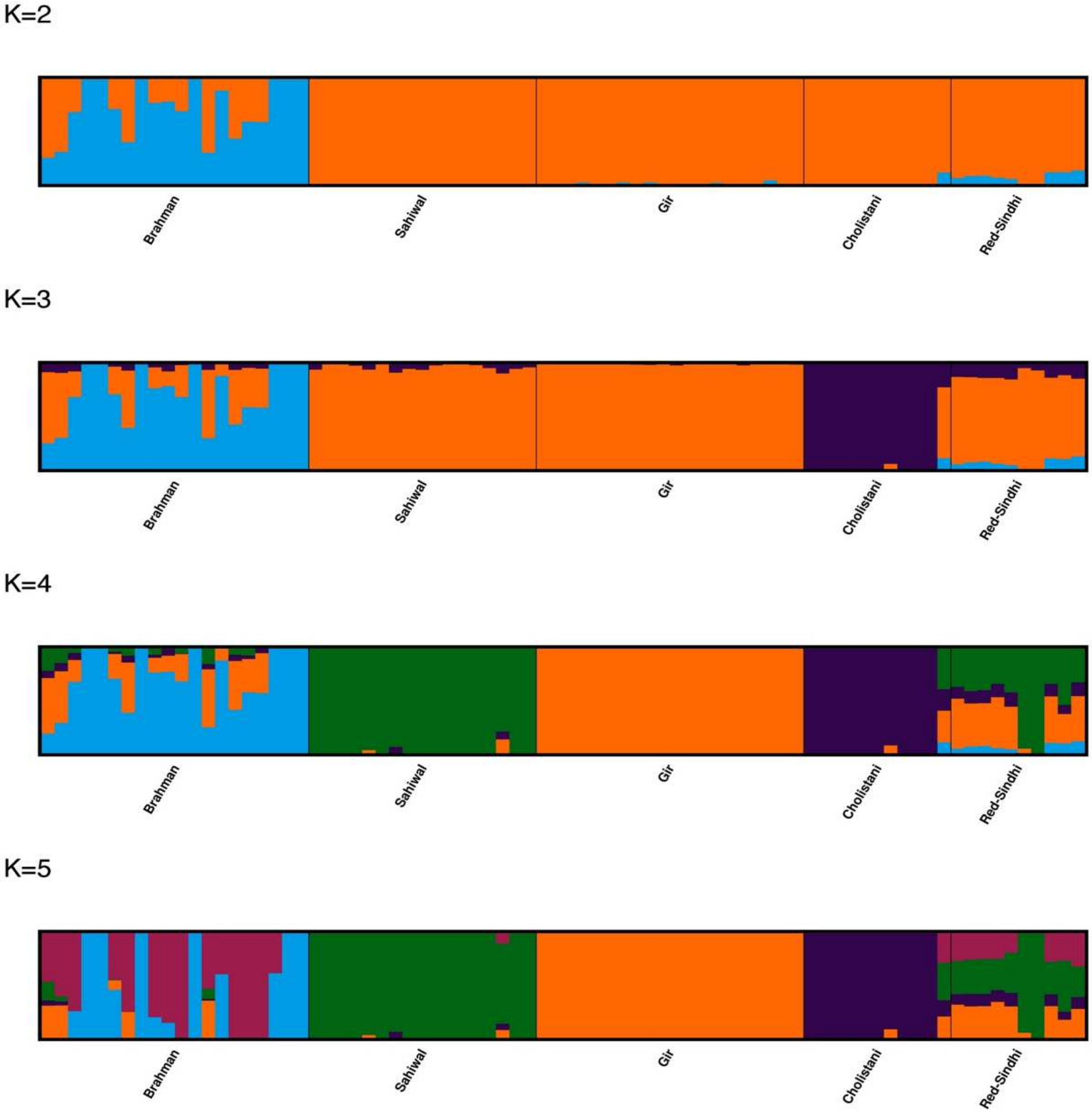

Figure 1. Genetic structure of five cattle breeds viz. Brahman, Cholistani, Gir, Red Sindhi, and Sahiwal at K = 2-5 levels

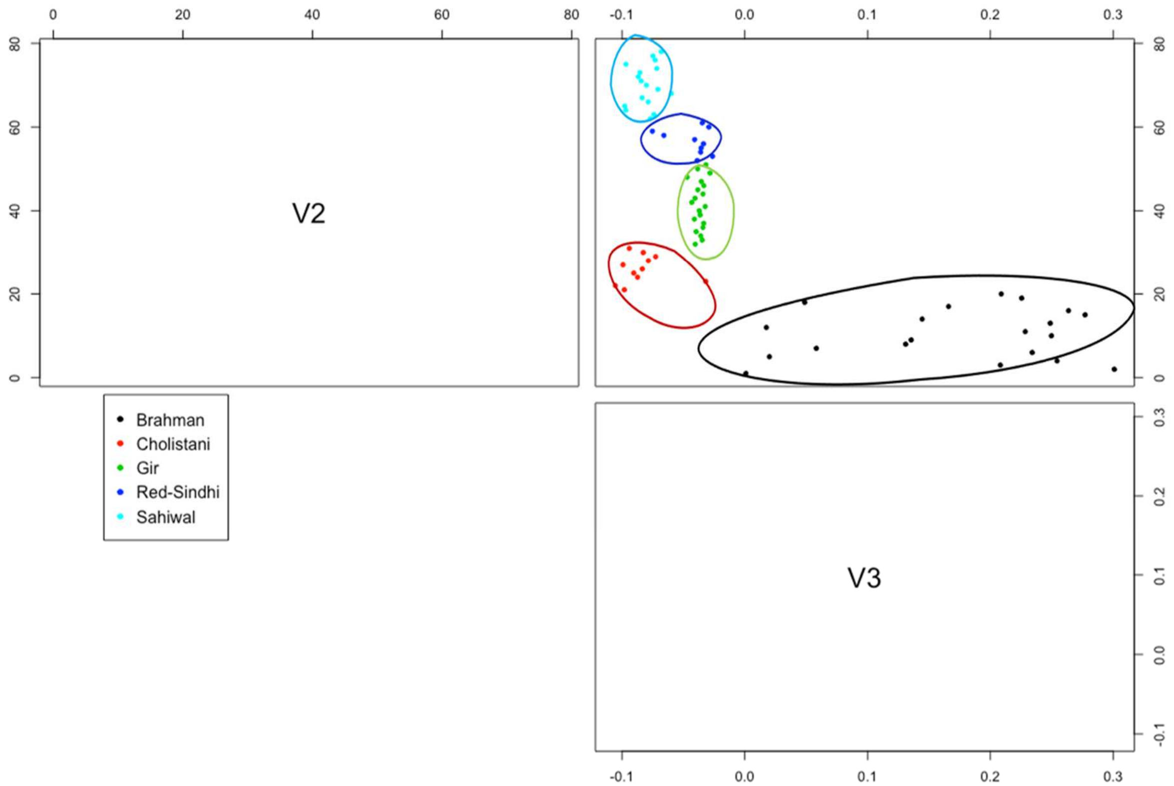

Figure 2. PCA plot of V3:V4 for five *Bos Indicus* of subcontinent origin breeds viz. Brahman, Cholistani, Gir, Red Sindhi and Sahiwal

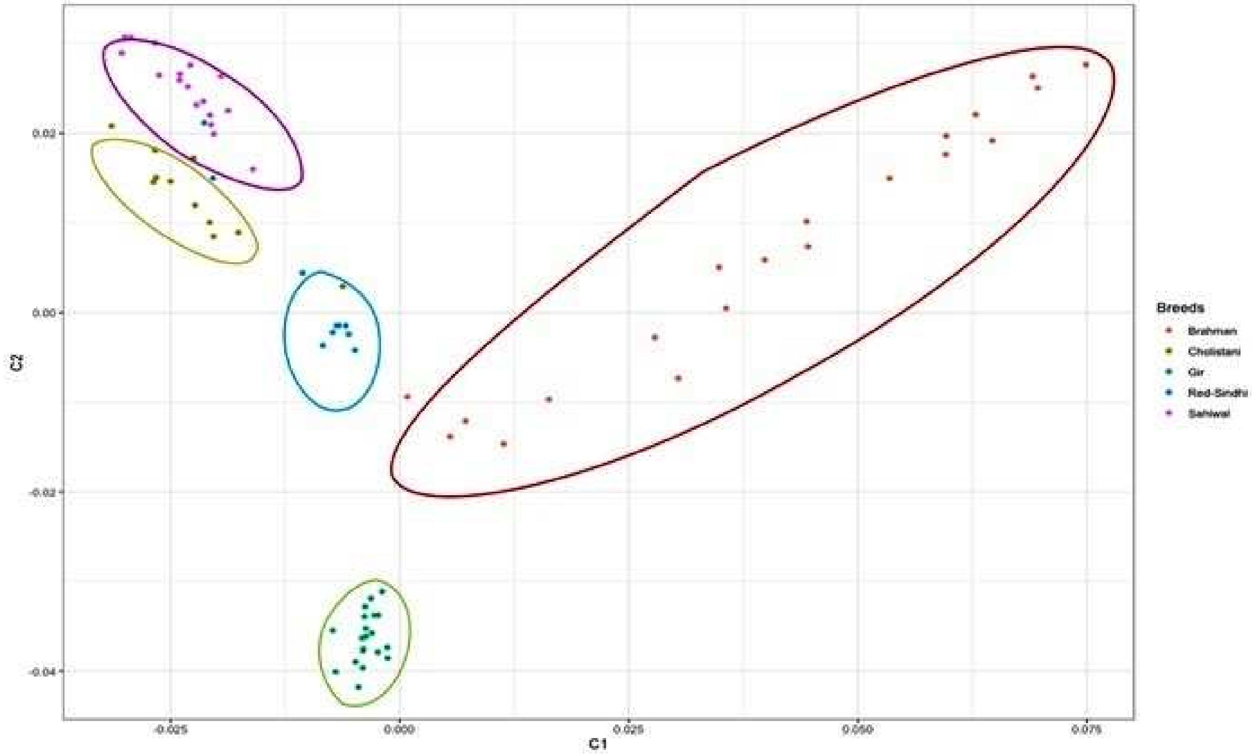

Figure 3. MDS plot for five *Bos Indicus* of subcontinent origin breeds viz. Brahman, Cholistani, Gir, Red Sindhi and Sahiwal

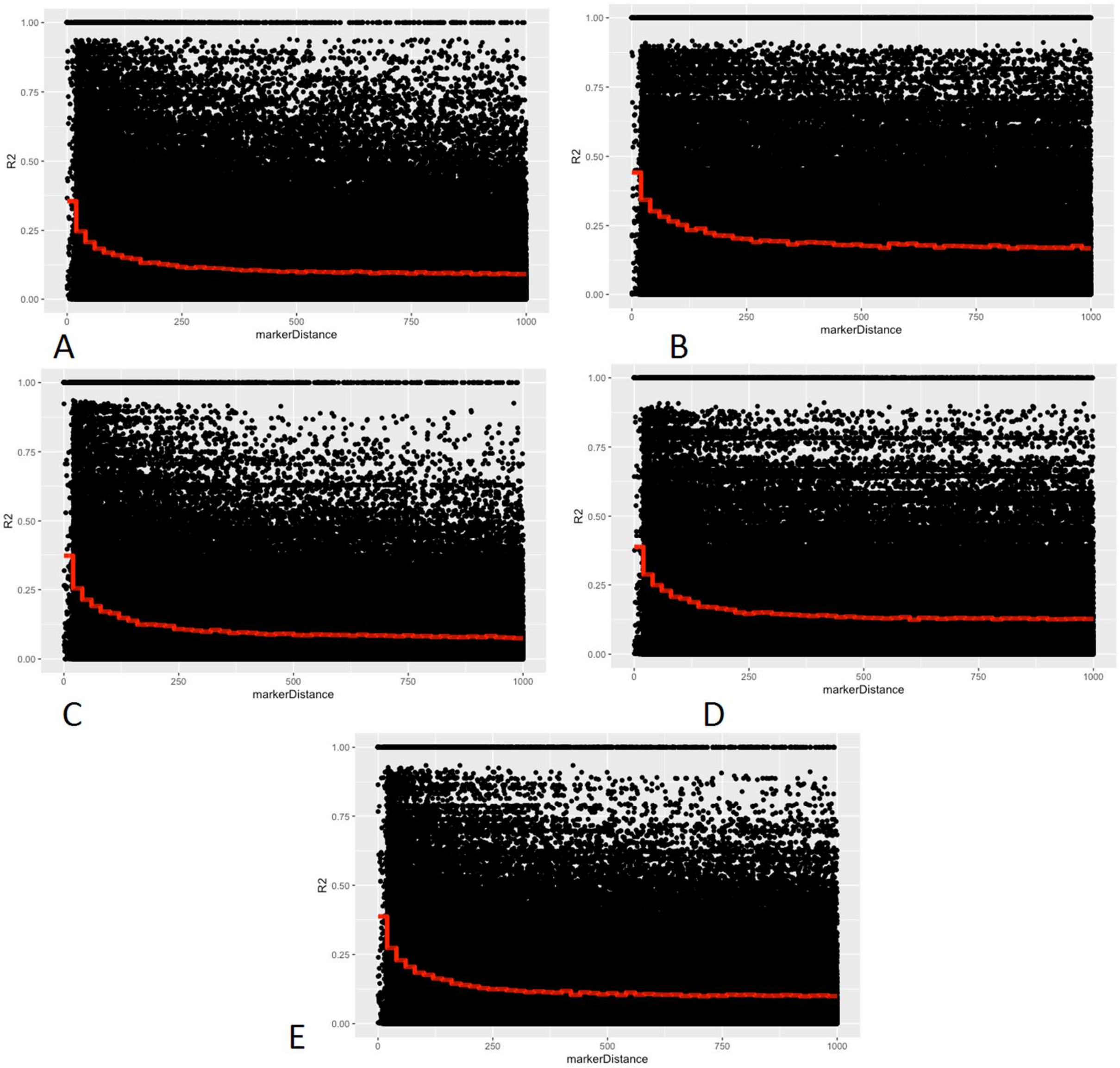

Figure 4. Linkage disequilibrium in, A: Brahman, B: Cholistani, C: Gir, D: Red Sindhi and E: Sahiwal cattle breeds

Acknowledgment/Disclaimers/Conflict of interest: The authors acknowledge the data acquisition from a public repository Dryad Digital Repository: http://dx.doi.org/10.5061/Dryad.55f3s. moreover, authors declare no conflict of interest.

REFERENCES

Alexander, D. H., J. Novembre, and K. Lange (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19(9):1655-1664. doi: 10.1101/gr.094052.109

Alshawi, A., A. Essa, S. Al-Bayatti, and O. Hanotte (2019). Genome Analysis Reveals Genetic Admixture and Signature of Selection for Productivity and Environmental Traits in Iraqi Cattle. Front. Genet. 10:609. DOI: 10.3389/fgene.2019.00609

Behl, J. D., P. Mishra, N. K. Verma, S. K. Niranjan, P. S. Dangi, R. Sharma, and R. Behl (2016). Nucleotide polymorphisms in the bovine lymphotoxin A gene and their distribution among Bos indicus zebu cattle breeds. Gene. 579(1):82-94. DOI: 10.1016/j.gene.2015.12.049

Bhuiyan, M. S. A., S. H. Lee, S. M. J. Hossain, G. K. Deb, M. F. Afroz, S. H. Lee, and A. K. F. H. Bhuiyan (2021). Unraveling the Genetic Diversity and Population Structure of Bangladeshi Indigenous Cattle Populations

Using 50K SNP Markers. Animals (Basel). 11(8):2381. DOI: 10.3390/ani11082381

Chagunda, M. G. G., F. D. N. Mujibi, T. Dusingizimana, O. Kamana, E. Cheruiyot, and O. A. Mwai (2018). Use of High Density Single Nucleotide Polymorphism (SNP) Arrays to Assess Genetic Diversity and Population Structure of Dairy Cattle in Smallholder Dairy Systems: The Case of Girinka Programme in Rwanda. Front. Genet. 9:438. DOI: 10.3389/fgene.2018.00438

Dassonneville, R., S. Fritz, V. Ducrocq, and D. Boichard (2012). Imputation performances of 3 low-density marker panels in beef and dairy cattle. J. of Dairy Sci. 95(7):4136-4140. DOI: 10.3168/jds.2011-5133

Decker, J. E., S. D. McKay, M. M. Rolf, J. Kim, A. Molina Alcala, T. S. Sonstegard, O. Hanotte, A. Gotherstrom, C. M. Seabury, L. Praharani, M. E. Babar, L. Correia de Almeida Regitano, M. A. Yildiz, M. P. Heaton, W. S. Liu, C. Z. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel, and J. F. Taylor (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genetics. 10(3):e1004254. DOI: 10.1371/journal.pgen.1004254

Dixit, S. P., A. K. Bhatia, I. Ganguly, S. Singh, S. Dash, A. Sharma, N. Anandkumar, A. K. Dang, and S. Jayakumar (2021). Genome analyses revealed genetic admixture and selection signatures in Bos indicus. Sci. Rep. 11(1):21924. DOI: 10.1038/s41598-021-01144-2

Dixit, S. P., S. Singh, I. Ganguly, A. K. Bhatia, A. Sharma, N. A. Kumar, A. K. Dang, and S. Jayakumar (2020). Genome-Wide Runs of Homozygosity Revealed Selection Signatures in Bos indicus. Front. Genet. 11:92. DOI: 10.3389/fgene.2020.00092

Edea, Z., M. S. Bhuiyan, T. Dessie, M. F. Rothschild, H. Dadi, and K. S. Kim (2015). Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. Animal. 9(2):218-226. DOI: 10.1017/S1751731114002560

Elsik, C. G., R. L. Tellam, K. C. Worley, R. A. Gibbs, B. G. S. A. Consortium (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 324(5926):522-528. DOI: 10.1126/science.1169588

Farooq, U., F. Sher, F. Asim, and M. A. Khan (2010). Cholistan and Cholistani Breed of Cattle. Pakistan Vet. J. 30(2):126-130.

Gebrehiwot, N. Z., E. M. Strucken, H. Aliloo, K. Marshall, and J. P. Gibson (2020). The patterns of admixture, divergence, and ancestry of African cattle populations determined from genome-wide SNP data. BMC Genomics. 21(1):869. DOI: 10.1186/s12864-020-07270-x

GoP. (2021). Pakistan Economic Survey. Economic Adviser’s Wing, Finance Division, Government of Pakistan, Islamabad

Hout, M. C.Papesh, and M. H.Goldinger (2013). Multidimensional scaling. Wiley Interdiscip Rev. Cogn. Sci. 4(1):93-103. DOI: 10.4135/9781412985130

Hussain, T., M. E. Babar, S. O. Peters, A. Wajid, A. Ali, A. Azam, Z. Ahmad, M. Wasim, A. Ali, and A. Kizilkaya (2016). Microsatellite Markers Based Genetic Evaluation of Pakistani Cattle Breeds. Pakistan J. Zool. 48(6):1633-1641

Kamiti, D., E. Ilatsia, R. Bett, and A. Kahi. 2016. Population structure and demographic trends of the registered Sahiwal cattle in Kenya. Trop. Anim. Health and Prod. 48(5):1029-1036. DOI: 10.1007/s11250-016-1055-0

Kijas, J. W., L. Porto-Neto, S. Dominik, A. Reverter, R. Bunch, R. McCulloch, B. J. Hayes, R. Brauning, J. McEwan, and C. International Sheep Genomics (2014). Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Anim. Genet. 45(5):754-757. DOI: 10.1111/age.12197

Kim, S., H. S. Cheong, H. D. Shin, S.-S. Lee, H.-J. Roh, D.-Y. Jeon, and C.-y. Cho (2018). Genetic diversity and divergence among Korean cattle breeds assessed using a BovineHD single-nucleotide polymorphism chip. Asian-Australasian J. Anim. Sci. 31(11):1691-1699. DOI: 10.5713/ajas.17.0419

Kumar, A., S. A. Waiz, T. Sridhar Goud, R. K. Tonk, A. Grewal, S. V. Singh, B. R. Yadav, and R. C. Upadhyay (2016). Assessment of adaptability of zebu cattle (Bos indicus) breeds in two different climatic conditions: using cytogenetic techniques on genome integrity. Int. J. Biometeorol. 60(6):873-882. DOI: 10.1007/s00484-015-1080-0

Lewis, J.Abas, Z.Dadousis, C. Lykidis, D. Paschou, and P.Drineas (2011). Tracing cattle breeds with principal components analysis ancestry informative SNPs. PLoS One. 6(4):e18007. DOI: 10.1371/journal.pone.0018007

McNamara J. P. and P. L. H. McSweeney (2022). Encyclopedia of dairy sciences (Third). Elsevier Academic Press.

Mustafa, H., H. J. Heather, K. EuiSoo, A. Ahmad, A. Ali, W. A. Khan, T. N. Pasha, M. Z. Farooq, K. Javed, A. Ajmal, and T. S. Sonstegard (2014). Comparative analysis of genome wide difference in Red Sindhi and Holstein cattle breeds using dense SNP marker. Int. J. Adv. Res. 2(4):300-304.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, and M. J. Daly (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81(3):559-575. DOI: 10.1086/519795

Rehman, M. S. and M. S. Khan (2009). Genetic diversity of Hariana and Hissar cattle from Pakistan using microsatellite analysis. Pakistan Vet. J. 29(2):67-71.

Sharma, A., S. H. Lee, D. Lim, H. H. Chai, B. H. Choi, and Y. Cho (2016). A genome-wide assessment of genetic diversity and population structure of Korean native cattle breeds. BMC Genet. 17(1):139. DOI: 10.1186/s12863-016-0444-8.