Linkage disquilibrium analysis on X-chromosome of indigenous breeds of cattle (Bos indicus) adapted to extreme agro-climatic conditions of tropical and temperate ecosystem

[1,2]Jigyasha Tiwari, [2]Sanjeev Singh[*],[2] Indrajit Ganguly and [2]S. P. Dixit

[1]*ICAR-National Dairy Research Institute, Karnal-132001 (Haryana) India*

[2]*ICAR-National Bureau of Animal Genetic Resources, Karnal-132001 (Haryana) India*

ABSTRACT

The present study was undertaken to analyze the breed wise Linkage disequilibrium (LD) decay in indigenous cattle breeds belonging to contrasting landscape (hot semi-arid /arid viz. Hallikar, Kankrej versus cold high altitude viz. Siri and Ladakhi) so as to ascertain the effective number of SNPs required for the preparation of X chromosome specific SNP panel. A total of 46 animals belonging to four different indigenous breeds of cattle were genotyped by using Illumina 777k Bovine HD chip. The LD decay analysis was carried out by using PLINK 1.9 ver. software. The LD between SNP pairs (r^2) varied from 0.14 to 0.21 among the breeds with an average of 0.18 for all the breeds. The average inter marker distance on X chromosome in tropical cattle (hot semi-arid /arid) varied from 60 - 100 kb; whereas for temperate cattle it was about 200kb at $r^2<0.2$. Therefore, the effective number of SNPs required for the preparation of X chromosome specific SNP panel was estimated to be about 2,638 for tropical breeds and 791 for temperate breeds. The effective number of SNPs identified in the present study might be helpful in the development of LD maps on X-chromosomes in indigenous breeds of cattle for association studies.

Keywords: Linkage disequilibrium decay, Indigenous Cattle, X- Chromosome, SNP panel

***Corresponding Author**: sssanjeev197@gmail.com

INTRODUCTION

Linkage disequilibrium is the non-random association between alleles at different loci which helps in identifying the pattern of association between alleles. It is estimated by measuring the correlations (r^2) between adjacent SNPs and polymorphism in DNA sequence. The subsets of SNPs that have very high or complete linkage disequilibrium (LD), *i.e.* r^2 value of 1 or near to 1, causes substantial redundancy in loci (Larmer et al. 2014). The amount of such redundancy is expected to be especially large for genomic data from populations with a less effective population size suggesting they have higher LD rates. Moreover, the magnitude of LD and the distance at which it decayed helps in ascertaining the effective number of SNPs required for performing the association analysis on X chromosomes. Several studies had reported the presence of selection footprints on the X chromosome in various livestock species and observed genes relevant to meat quality, reproduction and immune system in potential selection regions indicating the economic significance of this chromosome in livestock production system (Zhu et al. 2019; Ma et al. 2014, Stella et al. 2010). Further, Arishima and coworkers (2017) have identified the risk allele associated with recurrent infertility in repeat breeding Heifers of Japanese black cattle in the upstream of FOXP3 gene on the X chromosome, which causes severe economic losses, using genome wide association studies. However, the LD based genomic studies on X chromosomes in Indigenous cattle breeds are lacking. Therefore, the present study was undertaken with an aim to investigate the LD decay in Indian cattle breeds so that effective number of SNPs required for the genomic selection and association studies can be ascertained.

MATERIALS AND METHODS

Blood collection, DNA isolation and genotyping of samples:

The details of the sample collection and methodology of DNA isolation was described in our earlier published report (Bhardwaj et al. 2021). In brief; HiPurA™ SPP Blood DNA isolation kit (Himedia, Lab. Pvt. Ltd.) was used for the extraction of genomic DNA from whole blood following the manufacturer's instructions.The concentrations of DNA samples were evaluated using a Quantus™ fluorometer and the quality of the samples was assessed in a 1% agarose gel electrophoresis (Promega Coorporation, USA). A total of 46 high-quality DNA samples from Siri (9), Ladakhi (11) Kankrej (14) and Hallikar (12) were genotyped using the BovineHD BeadChip at Agri Genome Labs Pvt. Ltd. in Kerala, India (Illumina, Inc. San Diego, CA, USA).

Quality filtration of genotyped data: Quality control checking on the genotyped data was performed by using

PLINK v 1.9 (Purcell et al. 2007; Chang et al. 2015). PLINK –geno 0.1 –mind 0.1 –maf 0.05 –hwe 0.001 commands were used for the data pruning so as to remove poorly genotyped individuals (<90% of SNPs), loci genotyped in <90% of individuals, rare alleles (MAF < 0.05) and loci not in HW Equilibrium (P < 0.001) (Turner et al. 2011). Only the SNPs located on X chromosomes were used for the further analysis using –keep -chrX command. This retained a total of 38,830 SNPs for further analysis.

Linkage disequilibrium (LD) pattern: Correlation coefficient (r^2) for every pair of SNPs was calculated using PLINK v 1.9 with a window size limit of 500 kb between pair-wise SNPs for calculating the LD in all the four indigenous cattle breeds. Breed wise LD decay ($r^2<0.2$) was visualized for identifying the effective number of SNPs required for the preparation of X chromosome specific SNP panel, ($r^2<0.2$) with increasing distance between SNPs. R script was used for plotting the corresponding graphs of LD for each breed (Figure1). For ascertaining the effective number of SNPs in X chromosome specific SNP panel, the cattle X chromosome size (about 158.337 Mb) was divided by the average inter-marker distance at $r^2<0.2$ for each breed.

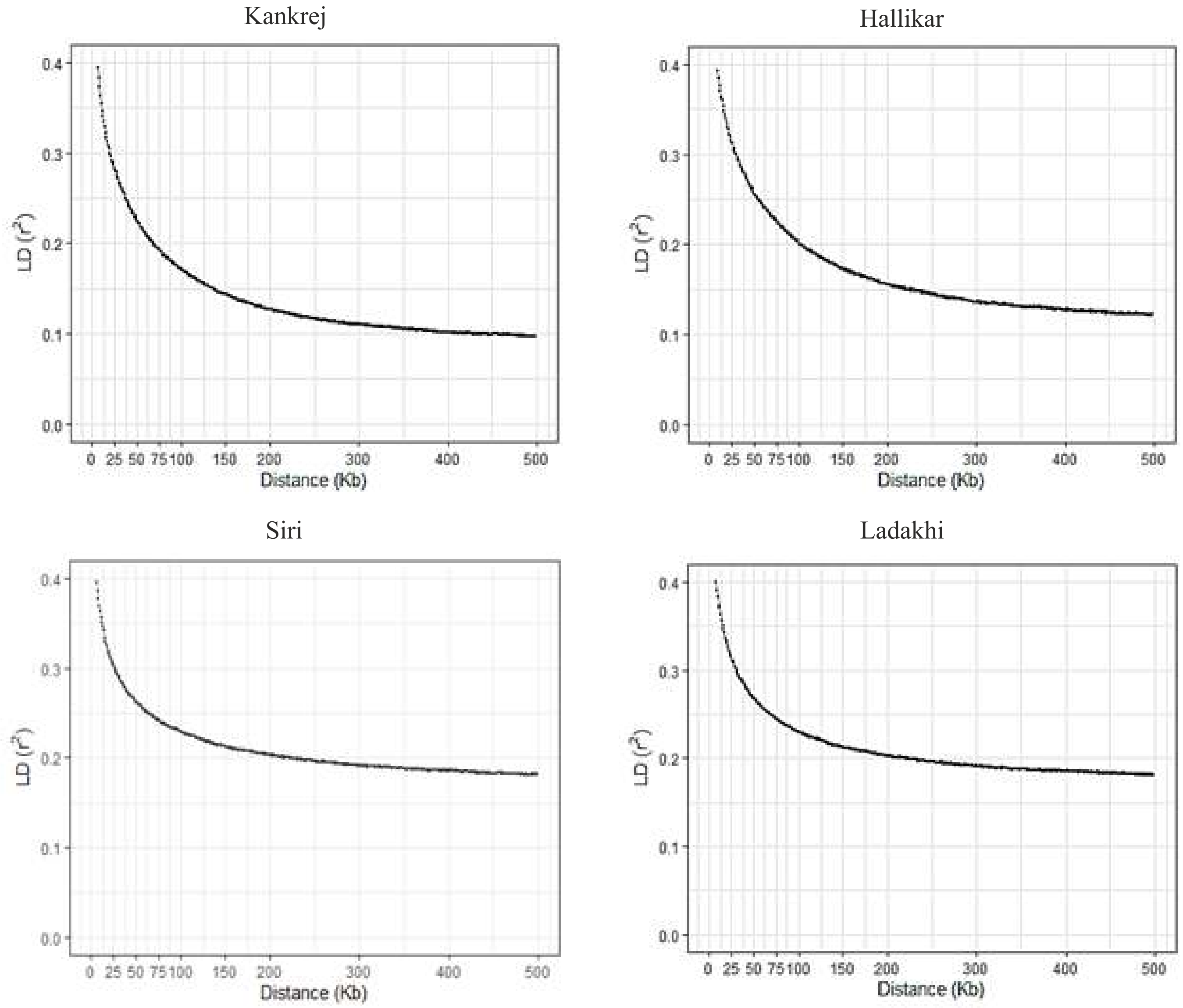

Figure 1: Breed wise LD Decay graph of four cattle breeds (Kankrej, Hallikar, Siri and Ladakhi).

RESULTS AND DISCUSSION

Most of the LD studies on indigenous breeds were carried out on the autosomes. This is the first reported study of the distribution of LD (r^2) on X chromosome in our indigenous cattle breeds. Mean LD (r^2) between adjacent SNPs ranged from 0.14 in tropical (Kankraj) breed to 0.21 in temperate breeds with an overall mean of 0.18 among the four indigenous cattle breeds. The LD decay ($r^2<0.2$) ranged from 60 kb (Kankrej) to 100 kb (Hallikar) inter-marker distance in hot semi-arid/arid adapted breeds, and at about 200 kb in cold adapted breeds (Ladakhi and Siri).

Therefore, the effective number of SNPs required for the preparation of X chromosome specific SNP panel was estimated to be about 2,638 for tropical breeds and 791 for temperate breeds.

Mustafa et al. (2018) also identified the overall mean value of 0.18 LD (r^2) across all the chromosomes in Sahiwal cattle. Lower estimates of LD were also reported in Nellore ($r^2=0.17$) (Espigolan et al. 2013) and in Bradford cattle ($r^2=0.18$) (Biegelmeyer et al. 2016). However, higher estimates of LD ($r^2 = 0.33$-0.35) were reported in Sahiwal, Tharparkar and Gir cattle (Das et al. 2017). The results obtained in present study are also in consonance with the LD decay estimated by Makina et al. (2015) in Afrikaner cattle in which LD ($r^2 \geq 0.2$) extended up to 100 kb inter-marker distance and with Dash et al. (2017) in Sahiwal and Tharparkar cattle breeds (140 Kb at $r^2 \geq 0.2$). Higher LD values obtained in cold adapted breeds (0.21) in comparison to hot adapted breeds (0.14 - 0.17) might be due to the lower effective population size in these breeds as compared to the breeds of the tropical environment.

Finally, it is concluded that the effective number of SNPs identified in the present study might be helpful in genome wide selection and association studies on X-chromosomes of Indigenous cattle breeds.

ACKNOWLEDGEMENTS

The Authors are thankful to the Director, ICAR-NBAGR, Karnal, Incharges of various livestock farms and Veterinary Officers of State Animal Husbandry departments for providing the necessary facilities and needful help in blood sample collection during the research work. The authors also acknowledge the funding and support received from the ICAR-National Bureau of Animal Genetic Resources, Karnal for the institute project.

REFERENCES

Arishima T, Sasaki S, Isobe T, Ikebata Y, Shimbara S, Ikeda S, Kawashima K, Suzuki Y, Watanabe M, Sugano S, Mizoshita K and Sugimoto Y. 2017. Maternal variant in the upstream of FOXP3 gene on the X chromosome is associated with recurrent infertility in Japanese Black cattle. *BMC genetics*. 18(1): 103.

Bhardwaj S, Singh S, Ganguly I, Bhatia AK, Bharti VK and Dixit SP. 2021. Genome-wide diversity analysis for signatures of selection of *Bos indicus* adaptability under extreme agro-climatic conditions of temperate and tropical ecosystems. *Animal Gene*. 20: 200115

Biegelmeyer P, Gulias-Gomes CC, Caetano AR, Steibel JP and Cardoso, FF. 2016. Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. *BMC Genetics*. 17:32.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM and Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*.4: 7.

Dash S, Singh A, Bhatia AK, Jayakumar S, Sharma A, Singh S, Ganguly I and Dixit SP. 2017. Evaluation of Bovine High-Density SNP Genotyping Array in Indigenous Dairy Cattle Breeds. *Anim. Biotechnolo*. 29(2): 129–135.

Espigolan R, Fernando B, Arione AB, Fabio RPS, Daniel GMG, Rafael LT, Diercles FC, Henrique NO, Humberto T and Mehdi S. 2013. Study of whole genome linkage disequilibrium in Nellore cattle.*BMC Genomics.*14:305

Larmer SG, Sargolzaei M and Schenkel FS. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. *J. Dairy Sci.* 97(5): 3128–3141

Ma Y, Zhang H, Zhang Q and Ding X. 2014. Identification of Selection Footprints on the X Chromosome in Pig. *PLoS ONE*. 9:e94911.

Makina SO, Taylor JF, van Marle-Koster E, Muchadeyi FC, Makgahlela ML, MacNeil MD and Maiwashe A. 2015. Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. *Front. Genet*. 6: 337

Mustafa H, Ahmad N, Heather HJ, Eui-soo K, Khan WA, Ajmal A, Javed K, Pasha TN, Ali A, Kim JJ and Sonstegard TS.2018. Whole genome study of linkage disequilibrium in Sahiwal cattle. *S. Afr. J. Anim. Sci.*48:353-360.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker

PIW, Daly MJ and Sham PC. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.*81(3):559–575

Stella A, Ajmone-Marsan P, Lazzari B and Boettcher P. 2010. Identification of selection signatures in cattle breeds selected for dairy production. *Genetics*. 185(4): 1451–1461

Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, Andrade M, Doheny KF, Haines JL, Hayes G, Jarvik G, Jiang L, Kullo IJ, Li R, Ling H, Manolio TA, Matsumoto M, McCarty CA, McDavid, AN, Mirel DB, Paschall JE, Pugh EW, Rasmussen LV, Wilke RA, Zuvich RL and Ritchie MD. 2011. Quality control procedures for genome-wide association studies. *Curr. Protocols Human Genet*. 68.

Zhu C, Li M, Qin S, Zhao F and Fang, S. 2019. Detection of copy number variation and selection signatures on the X chromosome in Chinese indigenous sheep with different types of tail. *Asian-australas. J. Anim. Sci.*33(9): 1378–1386