Genetic insights into crossbred dairy cattle of Pakistan: exploring allele frequency, linkage disequilibrium, and effective population size at a genome-wide scale

Nisa, Fakhar Un; Kaul, Haiba; Asif, Muhammad; Amin, Imran; Mrode, Raphael; Mansoor, Shahid; Mukhtar, Zahid

Published in:
Mammalian Genome

DOI:
10.1007/s00335-023-10019-y

Print publication: 01/12/2023

Document Version
Peer reviewed version

Genetic Insights into Crossbred Dairy Cattle of Pakistan: Exploring Allele Frequency, Linkage Disequilibrium, and Effective Population Size at a Genome-Wide Scale

Fakhar un Nisa [1,2,3], Haiba Kaul [3], Muhammad Asif [1,2], Imran Amin [1,2], Raphael Mrode [4,5], Shahid Mansoor[1,2,6] Zahid Mukhtar [1,2,*]

Affiliations:

1 Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering College (NIBGE-C), Faisalabad - Pakistan

2 Pakistan Institute of Engineering & Applied Sciences (PIEAS), Nilore, Islamabad, Pakistan.

3 Department of Animal Breeding and Genetics, University of Veterinary and Animal Sciences, Lahore, Pakistan.

4 Animal Biosciences, International Livestock Research Institute, Nairobi, Kenya.

5 Animal and Veterinary Sciences, Scotland's Rural College, Edinburgh, United Kingdom.

6 International Centre for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan.

***Corresponding Author:**

Zahid Mukhtar

E-mail: zahidmukhtar@yahoo.com

Email Addresses:

FU Nisa: fakhar.nisa@uvas.edu.pk / dr.fakharunnisa@gmail.com

H Kaul: haiba.kaul@uvas.edu.pk

M Asif: asif1.nibge@gmail.com

I Amin: imranamin1@yahoo.com

R Mrode: raphael.mrode@sruc.ac.uk

S Mansoor: shahidmansoor7@gmail.com

Z Mukhtar: zahidmukhtar@yahoo.com

Abstract

Linkage disequilibrium (LD) affects genomic studies accuracy. High-density genotyping platforms identify SNPs across animal genomes, increasing LD evaluation resolution for accurate analysis. This study aimed to evaluate the decay and magnitude of LD in a cohort of 81 crossbred dairy cattle using the GGP_HDv3_Ci Bead Chip. After quality control, 116,710 Single Nucleotide Polymorphisms (SNPs) across 2520.241 Mb of autosomes were retained. LD extent was assessed between autosomal SNPs within a 10 Mb range using the r^2 statistics. LD value declined as inter-marker distance increased. The average r^2 value was 0.24 for SNP pairs <10 kb apart, decreasing to 0.13 for 50-100 kb distances. Minor allele frequency (MAF) and sample size significantly impact LD. Lower MAF thresholds result in smaller r^2 values, while higher thresholds show increased r^2 values. Additionally, smaller sample sizes exhibit higher average r^2 values, especially for larger physical distance intervals (>50 kb) between SNP pairs. Effective population size and inbreeding coefficient were 150 and 0.028 for the present generation, indicating a decrease in genetic diversity over time. These findings imply that the utilization of high-density SNP panels and customized/breed-specific SNP panels represent a highly favorable approach for conducting Genome-Wide Association Studies (GWAS) and implementing Genomic Selection (GS) in the Bos indicus cattle breeds, whose genomes are still largely unexplored. Furthermore, it is imperative to devise a meticulous breeding strategy tailored to each herd, aiming to enhance desired traits while simultaneously preserving genetic diversity.

Keywords

Linkage Disequilibrium, Minor allele frequency, Effective population size, Crossbred dairy cattle, SNP Chip, Pakistan

Introduction

When alleles are sufficiently close together on a chromosome, they tend to be inherited together through linked inheritance rather than being passed down independently. This means that the offspring receives blocks of alleles or haplotypes from each parent, rather than individual alleles (O'Brien et al., 2014, Ardlie et al., 2002). Linkage disequilibrium (LD) arises due to the interconnection of alleles on a chromosome, which results in a certain degree of correlation among them. This phenomenon can be observed not only for nucleotides in the genome but also for different types of genetic markers, including single nucleotide polymorphisms (SNPs) (O'Brien et al., 2014). In essence, LD between molecular markers represents the extent to which the genotypes of two SNPs are correlated (Porto-Neto et al., 2014). While physical proximity plays a significant role in determining LD, it is important to acknowledge that other factors such as evolutionary processes and historical events can also affect the correlation between molecular markers. These factors encompass inbreeding, selection, population stratification, genetic drift, genetic bottleneck, effective population size, mutation, recombination rate, and migration (Karimi et al., 2015, Ardlie et al., 2002, Reich et al., 2001).

Comprehending the concept of LD holds paramount importance in the mapping of genes and their relevance to genomic studies. Through the analysis of LD between SNPs, researchers can obtain valuable insights into the diversity across various breeds, estimate recombination event frequencies, investigate fluctuations in effective population size across generations, and identify genomic regions suitable for enhancing economically important traits. Such

understanding contributes significantly to the advancement of genetic research and the targeted improvement of desired traits(O'Brien et al., 2014, Espigolan et al., 2013, McKay et al., 2007).

Effective population size is an important genetic parameter that estimates the effect of genetic drift in a population (Crow and Kimura, 1970). It is one of the best quantitative indicators of genetic diversity, (Makanjuola et al., 2020) that determining help determine the number of independent chromosome segments which is required for genomic predictions (Mrode et al., 2019). Estimating the effective population size is valuable not only from an evolutionary standpoint but also for enhancing models used in the mapping of genes related to quantitative traits (Li and Kim, 2015).

The decreasing cost of conducting high-throughput genotyping assays has opened up new possibilities for conducting large-scale genomic studies. The effectiveness and precision of these genomic investigations are greatly influenced by the extent and structure of LD observed between SNPs throughout the genome. As LD patterns determine the level of correlation between markers, understanding LD is crucial for maximizing the utility and accuracy of GWAS and GS approaches (Goddard and Hayes, 2012). Investigations have been carried out to examine the LD between markers within the genomes of diverse taurine and indicine cattle breeds. (Espigolan et al., 2013, Makina and Taylor, 2015). The findings of their study showed that moderate LD with a value of $r^2 = 0.20$ was observable in distances of less than 100 kilobases. This suggests that a set of 50,000 SNPs is sufficient to capture most of the LD information needed for conducting GWAS in taurine breeds (McKay et al., 2007, Espigolan et al., 2013). In contrast, the researchers observed a lower extent of LD ($r^2 = 0.20$ to 0.34) within indicine cattle at distances less than 30 kilobases. This suggests that a higher-density SNP chip would be necessary to capture the LD information required for conducting genomic studies in these cattle (Makina and Taylor, 2015). The diminished LD extent observed in indicine cattle

breeds can be attributed to a potential ascertainment bias inherent in the SNP chips utilized for genotyping.

The practice of crossbreeding in dairy cows has been identified as a highly effective approach for enhancing livestock productivity, reproductive efficiency, and sustainability (Leroy et al., 2016, Mbole-Kariuki et al., 2014, Bebe et al., 2003). In a single lactation, the crossbred offspring of Sahiwal(Bos indicus) and HF(Bos taurus) cows have the ability to produce approximately 4,000 liters of milk containing 4% fat (Kumar et al., 2018). Crossbreeding Sahiwal and HF cows result in offsprings that benefit from hybrid vigor, which leads to improved health traits and increased productivity. This is due to the combination of the high milk yield of HF cows and the adaptability and heat tolerance of Sahiwal cows, resulting in offsprings that are well-suited to local conditions and have improved milk yield.

In a recent study, admixture patterns and signature of selection for the same samples were studied (Unpublished data). However, the investigation of genome-wide LD and its pattern using this specific chip remains unexplored. The prediction accuracy from genomic selection (GS) is affected by marker density, minor allele frequency (MAF), and genetic architecture of the target trait (Zhang et al., 2019). Also, the accuracy of the genomic prediction depends on the amount of genetic variation explained by the markers resulting from the LD between the marker and QTL (Goddard, 2009). Therefore, this study examines the distribution of allelic frequencies, determines the level of LD (measured using r^2), and estimates the effective population size in the population of crossbred dairy cattle, which have a major impact on the accuracy of GS in this admixed population. This study will provide valuable insights for estimating marker density in genomic studies of crossbred dairy cattle.

Materials and methods

Ethics statement:

To ensure the ethical and humane treatment of animals, the study described in this research paper was approved by the institutional review committee. During blood collection, a professional veterinarian was there to ensure minimal distress and harm to the animals. Before collecting any samples, the researchers met with the owners of the farm where the animals were housed to explain the purpose of the study and obtain informed consent verbally.

Animal sample and genotype quality control

The sample size for this study consisted of 81 crossbred cattle from the Military Farm located in Renala Khurd near Okara, Punjab. These animals were selected based on having varying percentages of HF and Sahiwal genetics from different lactations. Due to the nature of crossbreeding, the breed composition varies from individual to individual due to variations in Sahiwal and HF inheritance. For instance, some crossbreds have an approx. 50% inheritance from both HF and Sahiwal, while others have 31/32 parts of HF inheritance the remaining being Sahiwal. Blood sampling was carried out in different visits to cattle farms during 2021 and 2022.

DNA was extracted from the blood samples using the FavorPrep^{TM} Blood Genomic DNA Extraction Mini Kit, following the manufacturer's guidelines. The quality and quantity of the DNA were evaluated using different methods, including a NanoDrop spectrophotometer, agarose gel electrophoresis, and a Qubit spectrophotometer. The extracted DNA was genotyped using the GGP_HDv3_C (GeneSeek® Genomic Profiler™) and commercially available services at GeneSeek (Neogen Corporation, Lincoln, NE, United States). The genotypes were identified and analyzed using the Genome Studio software from Illumina, Inc. The analysis was based on the bovine genome assembly, ARS-UCD1.2.

Quality control (QC):

After the genotyping process, the initial raw data comprised 139,376 SNPs for the crossbred individuals. Quality control measures were applied using the PLINK v1.9 software, (Slifer, 2018) which involved removing SNPs that had a call rate of less than 95%, minor allele frequency (MAF) of less than 0.02, and a Hardy-Weinberg equilibrium (HWE) of less than $10E^{-05}$. For subsequent analysis, only autosomal SNPs were considered.

Marker Statistics:

The R software was employed to estimate multiple characteristics of the autosomes. These included the length of each chromosome in megabases (Mb), the count of markers on each autosome, the longest and shortest intervals between SNPs, and the average interval between SNPs across all autosomes (Team, 2020).

Minor allele frequency (MAF):

To calculate the MAF of autosomal SNPs, default settings in PLINK v1.9 software were used with the command "--file data --freq" (Slifer, 2018). The distribution of allele frequencies across various chromosomes was analyzed using the R software. Additionally, a plot was generated to visualize the proportion of SNPs falling within different frequency categories, namely 0.02-0.10, 0.10-0.20, 0.20-0.30, 0.30-0.40, and 0.40-0.50 (Team, 2020).

Inbreeding Coefficient (F) and Effective Population Size (Ne):

To estimate F, the expected and observed homozygote differences were used with PLINK v1.9 software (Slifer, 2018) using the formula $F_i = (O_i - E_i) / L_i - E_i$. In this equation, F_i is the estimated inbreeding coefficient of the i^{th} animal, O_i represents the number of observed homozygous loci, Ei represents the count of expected homozygous loci, while Li represents the count of genotyped autosomal loci. The calculation was performed using PLINK v1.9 software (Slifer, 2018).

To estimate the effective population size (Ne), the SNeP tool was utilized, leveraging the relationship between Ne, linkage disequilibrium (represented by r^2), and recombination rate (c) (Barbato et al., 2015). This is given by:

$N_T(t) = 1/\{4f(C_t)\} * [1/\{E(r^2_{adj}) C_t\} - \alpha]$ (Corbin et al., 2012)

In the provided equation: $N_T(t)$ represents the estimated effective population size t generations ago in the past, C_t denotes the recombination rate t generations ago in the past, r_{2adj} signifies the adjusted LD estimation, accounting for sampling bias, f is mapping function and α represents a constant value.

Linkage Disequilibrium (LD):

The assessment of LD was carried out using the square of the correlation coefficient between two loci, represented as r^2. This metric is regarded as robust and unaffected by fluctuations in allele frequency and population size (Zhao et al., 2007). The estimation of r^2 plays a crucial role in determining the number of loci needed for conducting GWAS and quantitative trait loci (QTL) mapping. This measure helps to assess the extent of LD between markers and assists in designing the appropriate sample size and marker density for such studies (Makina and Taylor, 2015). The equation for estimating LD using the r^2 value is expressed as follows:

$D = (pAiAi)(pBjBj) - (pAiBj)(pBjAi)$

$r2 = D2/ pAi(1-pAi) pBj (1-pBj)$

In this context, the frequency of the i^{th} allele at locus A is denoted as p_{Ai}, while the frequency of the j^{th} allele at locus B is represented as p_{Bj}. Additionally, the frequency of the haplotype A_iB_j in the population is denoted as pA_iB_j.

The MapThin v1.11 was used to thin the map files, selecting 20 SNPs per 10^6 bp positions to minimize false positive results and increase the efficiency of the analysis (Howey and Cordell,

2011). The PLINK v1.9 software was used with the default command "--ld-snp-list mysnplist --ld-window-kb 186000 --ld-window 99999 --ld-window r^2 0" to estimate the r^2 between all pairs of SNPs on autosomes as the length of the longest chromosomes (chr1) is around 186kb (Slifer, 2018).

To verify the occurrence of free recombination at a physical distance greater than 10 Mb, two types of analyses were conducted on our dataset: one without considering a 10 Mb window, and the other with considering a 10 Mb window (Zhao et al., 2014).

1. To analyze the decay of LD value, the genomic regions were divided into eight categories based on a range of 20Mb each, namely 0-20 Mb, 20-40 Mb, 40-60 Mb, 60-80 Mb, 80-100 Mb, 100-120 Mb, 120-140 Mb, and 140-160 Mb. The LD value was then calculated for all possible regions within each category.
2. By taking into account a maximum distance of 10 Mb between SNP pairs, the LD decay was calculated for all possible SNP pairs across the autosomes. The trend in Linkage Disequilibrium (LD) decay for crossbred individuals was plotted across the entire first 10 megabases (MB) of the genome. This analysis likely provides insights into how LD changes over increasing physical distances within this specific genomic region for crossbred animals.

The calculated LD decay was then categorized into eight intervals based on distance ranges. These intervals included: 0-10 kb, 10-25 kb, 25-50 kb, 50-100 kb, 100-500 kb, 0.5-1 Mb, 1-5 Mb, and 5-10 Mb. This categorization allows for a comprehensive assessment of LD decay patterns across varying distances, providing valuable insights into the dynamics of LD in the autosomal genome and plotted against distance range.

Minor Allele Frequency and Sample Size Impact:

To assess the impact of minor allelic frequency (MAF) and sample size on LD, the analysis was extended. For a physical distance of 10 Mb, LD was computed using four distinct MAF thresholds (0.05, 0.10, 0.15, and 0.2). Furthermore, seven random subsets of the population were selected with different sample sizes (N=10, 20, 30, 40, 50, 60, and 70) to investigate the impact of sample size on r2-based LD. The extent of LD was assessed for each subset, and the impact of sample size and MAF on LD (r^2) was also depicted through plotting.

RESULTS

Quality control:

Table 1 summarizes the quality control results for different MAF thresholds. For example, for a 0.02 MAF threshold, 1804 SNPs were removed due to an MAF less than 0.02; 216 SNPs were removed based on Hardy-Weinberg Equilibrium (HWE); and 1111 SNPs were excluded due to a call rate threshold criterion. Therefore, a total of 116,710 autosomal SNPs with a genotypic rate of 0.99 were available for downstream analysis. These steps were repeated for different MAF values. All the remaining parameters were the same therefore the effect of different maf values on the final number of SNPs left for downstream analysis is depicted in Figure 1.

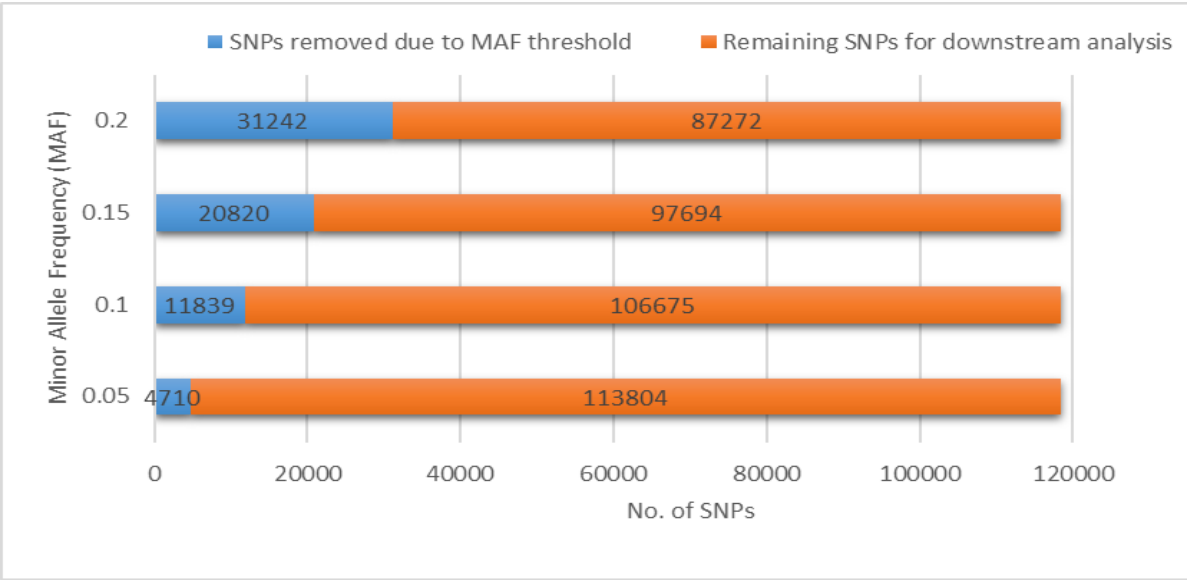


Figure 1 Effect of different MAF thresholds on the total no. of SNPs left for downstream analysis

Marker statistics:

The quality control process resulted in a total of 2520.241 Mb of retained SNPs across the genome of crossbred dairy cattle, with an average chromosome length of 86.90 Mb. The longest chromosome was BTA1, with a length of 158.8551 Mb, while the shortest was BTA25, with a length of 42.85 Mb. The number of SNPs on each chromosome exhibited a proportional relationship with the length of the respective chromosome. Notably, the highest number of SNPs was observed on BTA1 (7078), while the lowest number was recorded on BTA25 (1945). On average, the distance between adjacent SNPs was approximately 21.70 kb. The longest distance between SNPs was observed on BTA5 (612 kb), and the longest distance between SNPs on the same chromosome was found on BTA5 (3882 kb). Conversely, the mean shortest distance between SNPs was 0.16 kb, with the shortest distance occurring on BTA18 (0.002 kb). Descriptive statistics for each autosome's SNP markers are provided in Table 2.

240 *Table 1 Snapshot of the SNP markers studied and their minor allele frequency (MAF) across*

241 *the autosomal chromosomes (BTA)*

Chromosome no	Length (Mb)	Number of SNPs	Longest SNP Interval (kb)	Shortest SNP Interval (kb)	Average SNP Interval ± SD (kb)
BTA1	158.8551	7078	561.588	0.066	22.444±17.782
BTA2	136.7696	6284	206	0.075	21.759±15.989
BTA3	123.149	5659	1745.571	0.007	21.759±28.315
BTA4	120.6419	5379	211.953	0.163	22.429±16.104
BTA5	125.0587	5899	3882.807	0.007	21.198±53.387
BTA6	122.5097	6562	1919.865	0.005	18.667±32.844
BTA7	112.6101	5585	1146.872	0.003	20.165±23.583
BTA8	113.6262	5004	446.389	0.084	22.707±19.746
BTA9	105.689	4785	135.389	0.247	22.089±15.86
BTA10	104.2536	4703	395.732	0.274	22.167±18.657
BTA11	107.283	4814	145.246	0.178	22.282±16.311
BTA12	91.13102	4043	895.952	0.102	22.539±31.517
BTA13	84.22998	3733	373.764	0.038	22.561±19.002
BTA14	84.62824	4612	525.259	0.211	18.225±20.71
BTA15	85.27231	3835	226.131	0.092	22.222±16.271
BTA16	81.68807	3636	772.05	0.568	22.463±21.368
BTA17	75.13293	3338	509.683	0.193	22.508±20.518
BTA18	65.9992	2960	248.273	0.002	22.289±18.395

BTA19	64.04478	2844	180.7	0.239	22.465±15.837
BTA20	71.97177	3688	192.949	0.042	19.5±16.124
BTA21	71.5735	3192	524.177	0.047	22.41±19.349
BTA22	61.3782	2762	355.005	0.246	22.218±21.218
BTA23	52.46563	2390	358.949	0.129	21.955±17.689
BTA24	62.6437	3236	152.828	0.011	19.084±15.509
BTA25	42.85112	1945	136.474	0.185	22.018±14.725
BTA26	51.68014	2358	394.35	0.281	21.87±17.185
BTA27	45.38817	2037	597.764	0.592	22.284±23.675
BTA28	46.22406	2045	128.661	0.036	22.612±15.776
BTA29	51.49248	2277	394.565	0.613	22.576±18.563

Minor Allele Frequency (MAF):

The mean MAF observed across all autosomes was recorded as 0.30. Figure 2 provides a visual representation of the distribution of MAF on all autosomes.

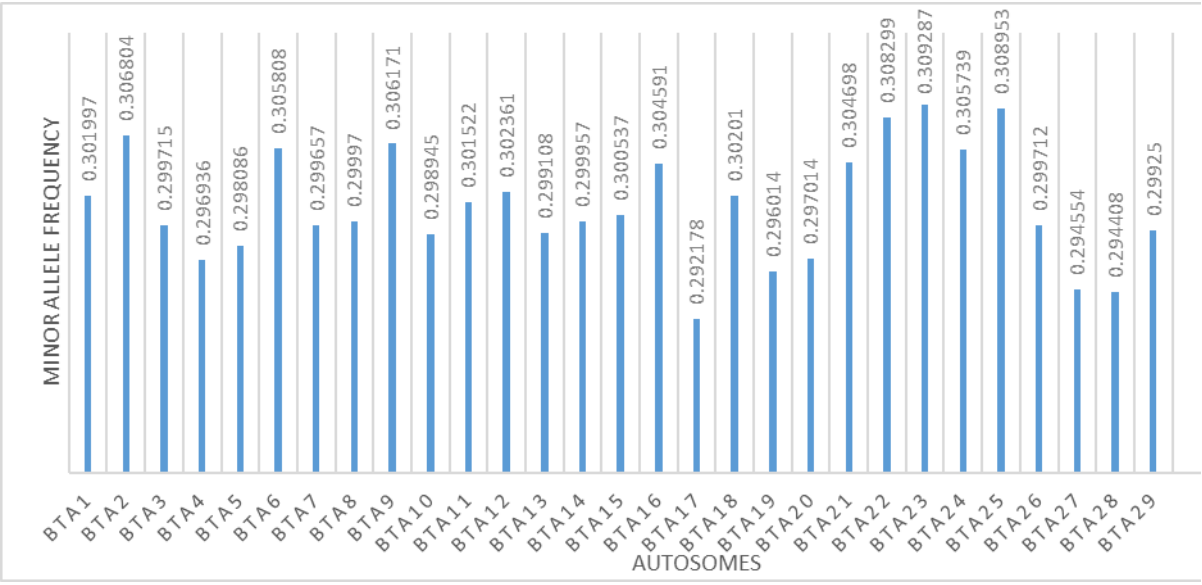


Figure 2 Minor Allele Frequency (MAF) in all autosomes

Similarly, the distribution of MAF indicated that a significant percentage of SNPs exhibited elevated MAF values (Figure 3). Specifically, around 54% of the SNPs were categorized in the last two MAF groups (MAF ≥ 0.3), while a lower percentage of SNPs fell into the initial categories. On average, around 25.20% of the SNPs displayed an MAF value lower than 0.2. This distribution highlights the predominance of SNPs with higher MAF values in the analyzed dataset.

It is noteworthy that all autosomes exhibited a similar trend, with a greater percentage of SNPs falling into the last two groups (MAF ≥ 0.3). However, BTA25, BTA6, BTA22, BTA9, and BTA23 had a higher percentage of SNPs in the last category (MAF ≥ 0.4). Among the chromosomes, BTA19 (10.04%), BTA17 (9.67%), BTA4 (9.61%), and BTA14 (9.54%) had a higher percentage of SNPs with MAF values between 0.02 and 0.1.

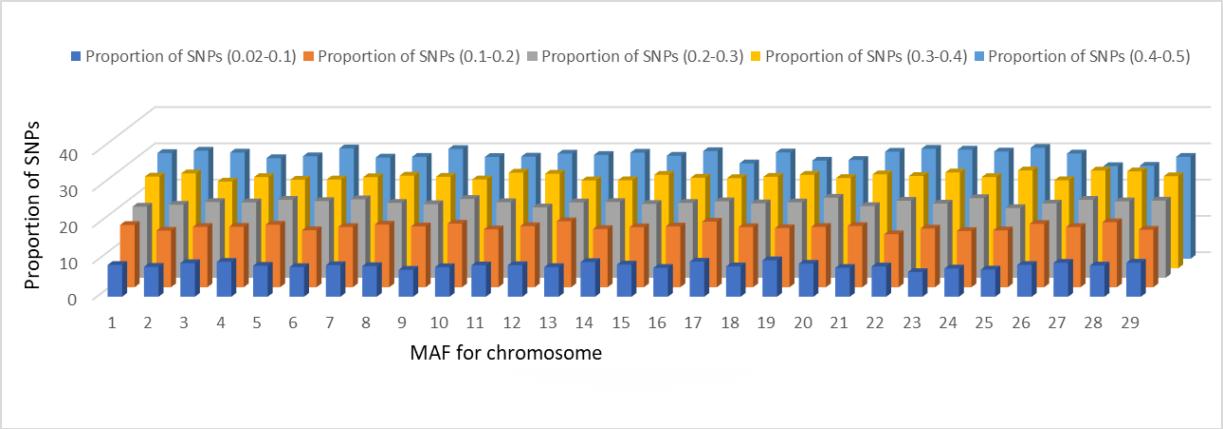


Figure 3 Proportion of SNPs categorized by minor allele frequencies (MAF) across autosomal chromosomes

Inbreeding Coefficient (F) and Effective Population Size (Ne):

The average value of F was estimated to be 0.028, indicating that the risk of negative impacts due to inbreeding depression can be considered insignificant at this level of inbreeding.

The Ne of the crossbred dairy cattle was estimated throughout the past 1000 generations based on the average r^2 values, as presented in Table 2. The results showed a declining trend in Ne, which decreased from 2775 (995 generations ago) to 150 (13 generations ago). This suggests that the crossbred dairy cattle population has experienced a decrease in genetic diversity over time.

Table 2 Effective Population Size (Ne) across generations determined through linkage disequilibrium (r^2)

Generations Ago	Ne	Distance	r^2	r^2 (SD)
13	150	3748832	0.042443	0.058646
15	165	3272844	0.044289	0.061595
17	183	2843931	0.045914	0.064077
20	203	2459746	0.047784	0.066845
23	226	2116420	0.049734	0.070699
27	252	1811032	0.05185	0.074908
32	286	1541094	0.053603	0.077165
38	328	1303393	0.055299	0.080254
45	377	1095203	0.057112	0.082466
54	435	914086	0.059203	0.084556
65	508	757766	0.061023	0.088151
80	598	623567	0.062809	0.090512
98	705	509487	0.065115	0.093385
120	827	413344	0.068169	0.097614
150	969	333170	0.071866	0.102766
187	1130	267075	0.076524	0.11055

234	1306	213281	0.082388	0.117489
293	1489	170160	0.089823	0.126196
366	1666	136285	0.09919	0.138301
454	1842	110080	0.109739	0.150293
553	1993	90367	0.121873	0.162818
658	2185	75961	0.130922	0.171969
759	2349	65842	0.139131	0.178594
847	2461	59027	0.146845	0.185381
914	2542	54693	0.152421	0.19237
958	2567	52141	0.157367	0.196599
983	2694	50818	0.154391	0.189964
995	2775	50240	0.152043	0.192054

Extent of LD across the genome

Without considering the window, we obtained a total of 47,054,338 possible pairs in the whole dataset, with a mean r^2 value of 0.020025 (Table 4). While a total of 9,115,588 combination pairs across the autosomes were analyzed to estimate LD for SNP pairs with a physical distance of ≤10 Mb. The mean r^2 value for markers at a 10 Mb distance was determined to be 0.128. Table 4 provides the mean LD (r^2) values for different intervals of physical distance.

Table 3 Statistical summary of linkage disequilibrium (r2) over the entire genome and up to 10 MB SNP distance

Entire genome			Upto 10 Mb distance between SNP pairs		
Distance (Mb)	No of SNP pairs	Mean r^2	Distance	No of SNP pairs	Mean r^2

0-20	17129667	0.0358	0-10 kb	171	0.243576
20-40	12674399	0.0212	10-25 kb	1970	0.242189
40-60	8343391	0.0182	25-50 kb	22927	0.182909
60-80	4946708	0.0173	50-100 kb	48204	0.141243
80-100	2610172	0.0169	100-500 kb	387883	0.076373
100-120	1004252	0.0167	0.5-1 Mb	481614	0.057482
120-140	279260	0.0165	1-5 Mb	3745729	0.0451343
140-160	66489	0.0176	5-10 Mb	4427090	0.035091

Table 4 shows that considering a 10 Mb window is important because there is an inverse relationship between LD (r^2) and the distance between SNP pairs. The mean r^2 values are higher when the SNP pair distance is smaller, i.e., between 0-10 kb, and decrease from 0-10 kb to 5-10 Mb. Therefore, it is suggested that SNP pairs within a distance of 10 Mb should be explored further.

The level of Linkage Disequilibrium (LD) decay, about the distance between pairs of Single Nucleotide Polymorphisms (SNPs), is depicted in Figure 4 for all autosomes. Notably, higher levels of LD were predominantly observed at shorter distances between SNP pairs, highlighting the rapid decay of LD as the physical distance between SNPs increases.

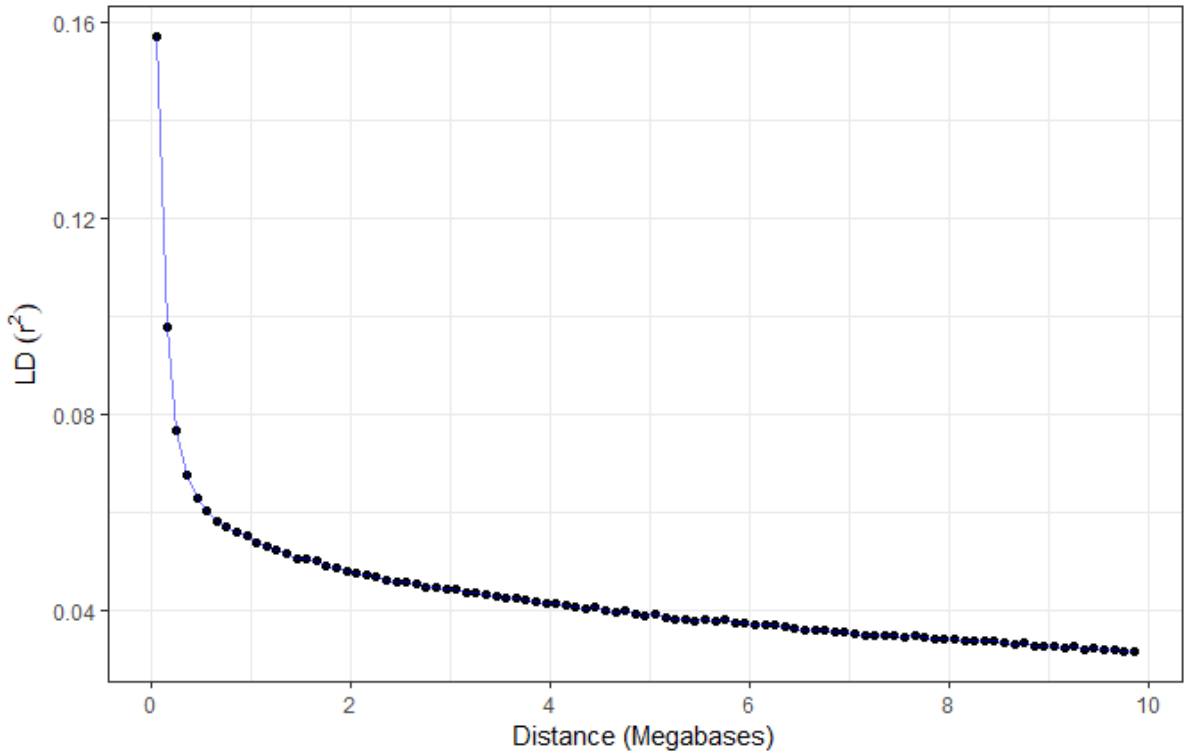


Figure 4 Average linkage disequilibrium (LD) decay in relation to the SNP pair distance

The level of LD measured by r^2 varied across each chromosome, and was dependent on the physical distance between genetic markers. To explore the relationship, the mean r^2 was computed for various physical distance intervals of markers across each chromosome. Chromosomes BTA22, BTA19, BTA18, and BTA7 exhibited higher levels of LD. When considering markers separated by <10 kb, the average r^2 was found to be 0.2332, which decreased to 0.1792 for markers with distances between 25-50 kb. The average r^2 continued to decline with increasing distance, reaching a final value of 0.0344 for the 5-10 Mb category. These results indicate that the mean r^2 values decrease as the physical distance between markers increases, demonstrating a decline in LD with increasing genetic distance (Figure 4).

The average r^2 values showed a significant difference across various autosomes, especially for SNP distances less than 10kb. On the other hand, for SNP distances greater than 100kb, lower mean r^2 values with relatively little variation across different autosomes were observed (Fig 6).

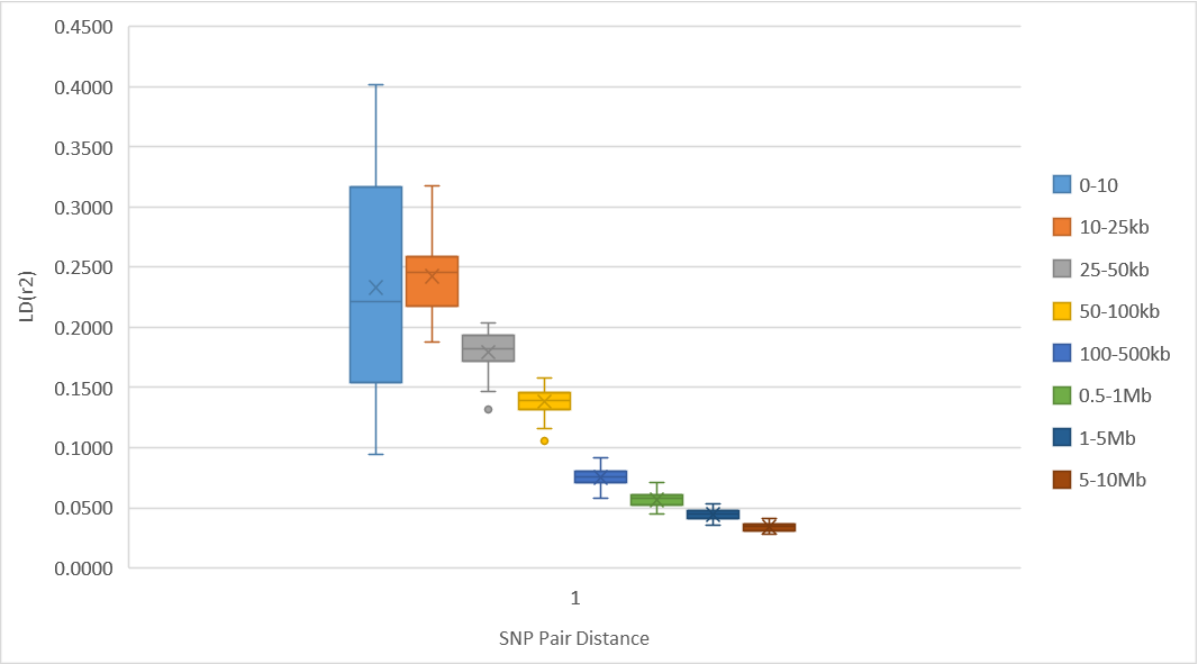


Figure 5 Box Plot of mean r2 and SNP Pair distance up to 10 Mb for all 29 autosomes

Minor Allele Frequency (MAF) and Linkage Disequilibrium (LD) estimates

The impact of MAF on the magnitude of LD was examined by employing four threshold levels: 0.05, 0.10, 0.15, and 0.20. This analysis focused on SNP pairs with 10 Mb physical distances (Fig 3). The findings revealed a notable influence of the MAF threshold on the average r^2, especially in the case of shorter distances between SNPs. A decrease in the r^2 value between SNP pairs was observed when the MAF threshold was set to a lower value (0.05), while a substantial increase in the r^2 value was noted at higher thresholds of MAF. The mean r^2 values ranged from 0.03 to 0.25 for MAF > 0.05, 0.03 to 0.27 for MAF > 0.10, 0.03 to 0.32 for MAF > 0.15, and 0.04 to 0.34 for MAF > 0.20. These results indicated that the MAF threshold has a considerable effect on the LD extent between SNPs, with higher MAF thresholds resulting in stronger LD.

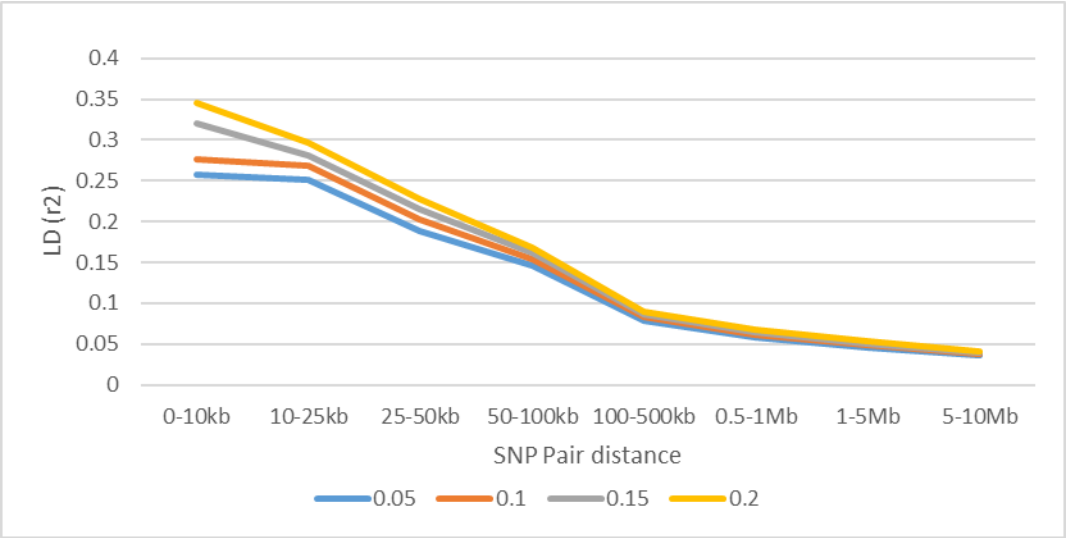
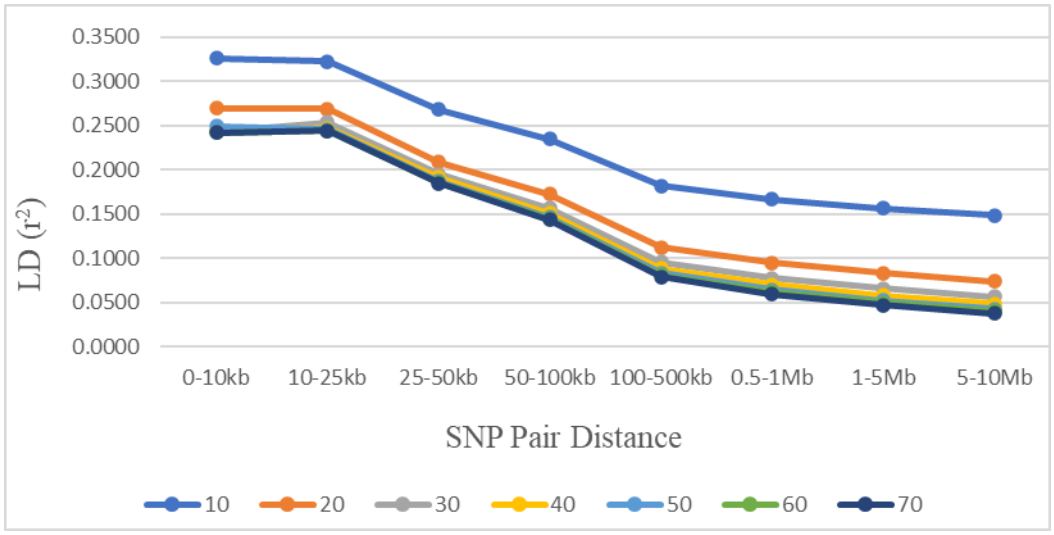


Figure 6 Effect of Minor Allele Frequency (MAF) on Linkage Disequilibrium Extent

Sample size and LD estimates

To study the effect of sample size, random samples of sizes 10, 20, 30, 40, 50, 60, and 70 were selected from the total population for analysis. A noteworthy finding of this study is that the average r^2 increased for smaller sample sizes, particularly when the physical distance intervals between SNP pairs exceeded 50 kb (Figure 4). These results suggest that a larger sample size (at least 40 animals) is needed for an accurate estimation of r^2.



331

332 *Figure 7 Effect of different sample size on mean r^2 estimates*

333 **Discussion**

334 Following rigorous quality control measures, a final set of 116,710 autosomal SNPs were
335 retained for analysis. These SNPs spanned a genomic region of 2520.241 Mb in crossbred dairy
336 cattle. The average MAF was found to be 0.30, aligning with previously reported MAF values
337 observed in diverse cattle breeds (Makina and Taylor, 2015). These findings align with previous
338 studies on other taurine breeds of cattle (O'Brien et al., 2014, Matukumalli et al., 2009, McKay
339 et al., 2007). However, the average MAF values observed in this study were notably higher
340 compared to indicine breeds of cattle, which typically exhibit MAF values ranging from 0.19-
341 to 0.20 (O'Brien et al., 2014, Espigolan et al., 2013, Silva et al., 2010). In contrast to taurine
342 cattle, indicine breeds typically display a distinct pattern in MAF levels, characterized by a
343 greater representation of alleles with lower frequencies (< 0.2)(O'Brien et al., 2014, Gibbs et
344 al., 2009, Villa-Angulo et al., 2009). The variation mentioned above could be ascribed to the
345 increased genetic diversity identified in indicine breeds (Murray et al., 2010, Gibbs et al.,
346 2009). As there are few such reports on crossbred cattle, especially with Bos Indicus and Bos
347 taurus crossbreds, this may serve as a positive contribution and important reference for the
348 future studies of native breeds. Furthermore, the commercially available SNP panel used in this

study predominantly utilized sequence data from Bos taurus breeds. Thus, it may lead to the ascertainment bias leading to a greater proportion of SNPs with low MAF in indicine breeds of cattle.

The distribution of MAF has a direct impact on the extent of LD, as a lower MAF can result in a greater difference in allelic pair frequencies, leading to an underestimation of LD (Wray, 2005). To examine the effect of MAF on LD, four different MAF thresholds were selected. The results indicated that higher thresholds of MAF (>0.20) were associated with higher average LD (r^2) between SNPs, particularly at shorter distances(O'Brien et al., 2014, Sargolzaei et al., 2008, Uimari et al., 2005). At a lower MAF threshold (e.g., 0.05), there might be the inclusion of rare variants in the analysis. Rare variants may behave differently in terms of LD leading to the observed decrease in r2 as they are less likely to be in strong linkage with other variants. Although, it has been established till now that the r2 method is less affected by the sample size it may also be a contributing factor to this finding indicating that SNP chips with a higher density of SNPs and studies on larger populations may be preferable for genomic studies in native cattle breeds of Pakistan.

The estimation of LD between SNP pairs was conducted using the correlation (r^2) method, which is known to be less affected by MAF (Ardlie et al., 2002) and small sample size (Zhao et al., 2014). To assess the decay of LD, the physical distance between markers was divided into distinct intervals. The findings demonstrated a swift decrease in r^2 beyond a threshold of 100 kb. Furthermore, r^2 declined from 0.24 to 0.17 when considering marker distances of 10 kb and 50 kb, respectively. For inter-marker distances of up to 25 kb, the average r^2 value was 0.24, which is comparatively lower than previous LD estimates, documented for taurine breeds such as Angus (0.46) and Hereford (0.49), as well as indicine breeds like Brahman (0.25) and Nellore (0.27) cattle(Porto-Neto et al., 2014, Espigolan et al., 2013, Lu et al., 2012). The results

showed higher mean r^2 values for BTA22, BTA19, BTA18, and BTA7, while lower mean r^2 values were observed for BTA27, BTA21, BTA9, and BTA26.

LD (r^2) values that surpass 0.3 are deemed valuable for dependable association studies and precise genomic predictions (Ardlie et al., 2002, Meuwissen et al., 2001, Kruglyak, 1999). In this study, regions up to 10 kb on BTA5, BTA7, BTA13, BTA18, BTA22, BTA26, and BTA27 exhibited r^2 values larger than 0.3. On the other hand, BTA19 showed a slower decay in LD, achieving the same level of r^2 up to a distance of 25 kb. These findings deviate from the average LD levels documented in other taurine breeds such as Angus, Holstein, Brown Swiss, and Fleckvieh reach an average r^2 value of 0.3 at distances ranging from 40 to 50 kb (O'Brien et al., 2014). These results align with the findings observed in indicine breeds such as Gyr and Nelore, which exhibit a more rapid decline in LD, reaching a similar r^2 value at distances of approximately 20 kb. Similar outcomes have been observed in other taurine breeds, where r^2 values remained comparable for distances equal to or less than 30 kb (Larmer et al., 2014, Bolormaa et al., 2011). Nevertheless, no correlation was identified between chromosomal size and r^2 estimates (Bohmanova et al., 2010).

The LD decay analysis within a range of up to 10 megabases (MB), employing 10-kilobase (kb) windows, is presented in Figure 4 for crossbred individuals. This analysis revealed a pronounced decline in LD at shorter distances between pairs of SNPs. This behavior is likely attributed to the relatively low number of SNP pair comparisons available at these close distances.

These findings suggest that, when utilizing the set of SNPs contained in the GGPHDv3-C chip, there may not be consistent LD levels expected for genomic distances less than 10MB. Importantly, these results align with those reported by (O'Brien et al., 2014)in their LD analysis conducted across different taurine and indicine breeds.

To examine the influence of sample size on the extent of LD, various sample sizes were employed in the computation of r^2 values as it is previously reported that a small sample size may lead to overestimation of LD (Yan et al., 2009, Khatkar et al., 2008). In the present investigation, a sample size of 40 cattle did not influence r^2, which aligns with previous findings (Bohmanova et al., 2010, Singh et al., 2021). However, various other studies have reported different threshold limits for sample size. For example, Zhu et al. (2013) suggested a minimum sample size of 100. In the case of Holstein cattle a minimum sample size of 400 is necessary for reliable LD decay analysis (Khatkar et al., 2008). Human studies have indicated even higher sample sizes (Chen et al., 2006). The minimum threshold for sample sizes appears to be around 75, as r^2 accuracy is significantly compromised below this value (Khatkar et al., 2008). Similarly, another study suggested a minimum sample size of 55 (Bohmanova et al., 2010).

Previous studies have consistently emphasized the significance of employing a larger number of SNPs to adequately cover the genome in genomic evaluations, especially when analyzing data from crossbreds and indicine breeds (Makina and Taylor, 2015, Espigolan et al., 2013). The findings from our study support this notion, that a higher density SNP array provides more information and enhances the reliability of GWAS and GS in crossbred dairy cattle populations. These results are consistent with other studies that have also emphasized the benefits of utilizing a higher SNP density for such analyses (Singh et al., 2021).

To gain a better understanding of the population diversity and structure, we estimated F and Ne. Our study revealed an inbreeding coefficient of 0.028. This finding is comparable to previously reported inbreeding coefficients of 3%, 4%, and 6% in the Vrindavani crossbred cattle population of India (Chhotaray et al., 2021, Singh et al., 2021, Elavarasan et al., 2023). The Karan Fries crossbred cattle of Karnal exhibited an inbreeding coefficient of 3.68% (Mumtaz et al., 2021). In the case of Sahiwal cattle, one of the 11 breeds studied by Bang et al. (2022), the inbreeding coefficient was 0.9%. In Tharparker cattle, different methods were

422 employed to estimate genomic inbreeding coefficients, resulting in values of 0.0589 (FROH),
423 0.0215 (FHOM), 0.0532 (FGRM), and 0.0160 (FUNI) (Saravanan et al., 2022). Another study
424 was conducted on Holstein, Montebeliarde, and Normande breeds, an inbreeding coefficient of
425 4.5-5% (Dezetter et al., 2015). For pure African taurine (Baoulé) and its crossbreeds with
426 indicine Zebu cattle, genomic inbreeding coefficients ranged from 0 to 4% (Ouédraogo et al.,
427 2021). Among nine breeds, the mean genomic inbreeding estimates were highest for Jersey
428 (0.173) and lowest for Hereford (0.051) (Kelleher et al., 2017). Lower inbreeding was observed
429 in six Columbian cattle breeds, ranging from 0.5 to 4.5% (Martinez et al., 2023). In a study of
430 171 cattle groups conducted by Tian et al. (2023), the average inbreeding coefficient ranged
431 from 0.22 to 0.05.

432 In our study, we observed a decrease in effective population size (Ne) over multiple generations
433 in crossbred dairy cattle. The estimated Ne in our population was 150, and a decreasing trend
434 in Ne was observed specifically 13 generations ago. When the effective population size (Ne)
435 decreases, the genetic diversity available for selection in genomic breeding is constrained. This
436 reduction in diversity limits the number of allelic variants that can be considered. Additionally,
437 increased inbreeding resulting from a smaller Ne compromises fitness and undermines the
438 accuracy of predictions due to the correlation of genetic variants. Moreover, the limited genetic
439 contributions and heightened genetic drift further hinder the effectiveness of GS. When
440 comparing our findings to previous studies, a range of 33-153 Ne was observed in different
441 studies on dairy cattle (Doekes et al., 2018, Rodríguez-Ramilo et al., 2015, Stachowicz et al.,
442 2011). Effective population sizes for European taurine breeds ranged from 98 to 152, with
443 Brown Swiss exhibiting the lowest value (98), and Limousine and Piedmontese showing the
444 highest values (138 and 144, respectively). African taurine breeds recorded a range of 120 to
445 175 for Ne. Among indicine cattle breeds, Gir had the highest Ne estimate (180), while
446 Tharparker had the lowest (63) (Barbato et al., 2020). In the context of buffalo breeds, both

purebred and crossbred populations demonstrated a decreasing trend in recent Ne, with estimated values closer to 387 and 113, respectively, 13 generations ago. This suggests that these animals have undergone strong selection or genetic drift, resulting in a decline in population size (Deng et al., 2019).

Conclusion

This study aimed to assess the level of LD between markers in crossbred dairy cattle from Pakistan using the GGP_HDv3_C (GeneSeek® Genomic Profiler™) SNP panel. The average estimated value of r^2 was 0.24, which was lower compared to both indicine and taurine breeds. This suggests that a denser SNP panel is necessary to obtain more precise and accurate results in whole genome association studies for crossbred dairy cattle.

Additionally, the study observed a declining trend in the estimates of effective population size (Ne) in the population. This indicates the need for a well-designed breeding plan that can maintain a sufficiently large Ne to mitigate the negative effects of genetic drift and inbreeding.

Conducting studies on a larger population using a high-density array of SNPs would provide more comprehensive and reliable information regarding the extent of LD and the effective population size in crossbred dairy cattle of Pakistan.

List of Abbreviations:

MAF: Minor Allele Frequency

LD: Linkage Disequilibrium

Ne: Effective Population Size

SNP: Single Nucleotide Polymorphism

GWAS: Genome-Wide Association Study

r2: Correlation coefficient

Declarations:

Ethics statement:

To ensure the ethical and humane treatment of animals, the study described in this research paper was approved by the Research Ethics Committee of the National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan on 10-06-2020. During blood collection, a professional veterinarian was there to ensure minimal distress and harm to the animals. Before collecting any samples, the researchers met with the owners of the farm where the animals were housed to explain the purpose of the study and obtain informed consent verbally.

Availability of data and materials:

All the necessary files are provided with the paper.

Competing interests:

The authors declare that they have no competing interests

Funding:

This study is funded by the PAKISTAN AGRICULTURAL RESEARCH COUNCIL, AGRICULTURAL LINKAGES PROGRAMME (ALP) with Project Identification No. AS 016 titled “Development and application of genomic selection in foreign and local cattle breeds for improvement in dairy-related traits”.

Author Contributions

Conceptualization: Fakhar un Nisa Raphael Mrode

Data curation: Fakhar un Nisa, Muhammad Asif.

Formal analysis: Fakhar un Nisa, Haiba Kaul.

Investigation: Fakhar un Nisa.

Methodology: Fakhar un Nisa.

Project administration: Shahid Mansoor, Zahid Mukhtar, Imran Amin.

Resources: Shahid Mansoor, Imran Amin.

Software: Fakhar un Nisa.

Supervision: Raphael Mrode, Zahid Mukhtar, Shahid Mansoor.

Visualization: Fakhar un Nisa, Haiba Kaul.

Writing – original draft: Fakhar un Nisa.

Writing – review & editing: Raphael Mrode, Haiba Kaul, Zahid Mukhtar, Muhammad Asif, Shahid Mansoor, Imran Amin.

Acknowledgements:

We acknowledge the Higher Education Commission, Pakistan for providing funding to the first author for her PhD Studies.

References

ARDLIE, K. G., KRUGLYAK, L. & SEIELSTAD, M. 2002. Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics,* 3**,** 299-309.

BANG, N. N., HAYES, B. J., LYONS, R. E., RANDHAWA, I. A., GAUGHAN, J. B. & MCNEILL, D. M. 2022. Genomic diversity and breed composition of Vietnamese smallholder dairy cows. *Journal of Animal Breeding and Genetics,* 139**,** 145-160.

BARBATO, M., HAILER, F., UPADHYAY, M., DEL CORVO, M., COLLI, L., NEGRINI, R., KIM, E.-S., CROOIJMANS, R. P., SONSTEGARD, T. & AJMONE-MARSAN, P. 2020. Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. *Scientific reports,* 10**,** 1-11.

BARBATO, M., OROZCO-TERWENGEL, P., TAPIO, M. & BRUFORD, M. W. 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in genetics,* 6**,** 109.

BEBE, B. O., UDO, H. M., ROWLANDS, G. J. & THORPE, W. 2003. Smallholder dairy systems in the Kenya highlands: breed preferences and breeding practices. *Livestock Production Science,* 82**,** 117-127.

BOHMANOVA, J., SARGOLZAEI, M. & SCHENKEL, F. S. 2010. Characteristics of linkage disequilibrium in North American Holsteins. *BMC genomics,* 11**,** 1-11.

BOLORMAA, S., HAYES, B., SAVIN, K., HAWKEN, R., BARENDSE, W., ARTHUR, P., HERD, R. & GODDARD, M. 2011. Genome-wide association studies for feedlot and growth traits in cattle. *Journal of animal science,* 89**,** 1684-1697.

CHEN, Y., LIN, C.-H. & SABATTI, C. 2006. Volume measures for linkage disequilibrium. *BMC genetics,* 7**,** 1-8.

CHHOTARAY, S., PANIGRAHI, M., PAL, D., AHMAD, S. F., BHANUPRAKASH, V., KUMAR, H., PARIDA, S., BHUSHAN, B., GAUR, G. & MISHRA, B. 2021. Genome-wide estimation of inbreeding coefficient, effective population size and haplotype blocks in Vrindavani crossbred cattle strain of India. *Biological Rhythm Research,* 52**,** 666-679.

CORBIN, L. J., LIU, A., BISHOP, S. & WOOLLIAMS, J. 2012. Estimation of historical effective population size using linkage disequilibria with marker data. *Journal of Animal Breeding and Genetics,* 129**,** 257-270.

CROW, J. & KIMURA, M. 1970. An Introduction to Population Genetics Theory. New York: Harper & Row.

DENG, T., LIANG, A., LIU, J., HUA, G., YE, T., LIU, S., CAMPANILE, G., PLASTOW, G., ZHANG, C. & WANG, Z. 2019. Genome-wide SNP data revealed the extent of linkage disequilibrium, persistence of phase and effective population size in purebred and crossbred buffalo populations. *Frontiers in genetics,* 9**,** 688.

DEZETTER, C., LECLERC, H., MATTALIA, S., BARBAT, A., BOICHARD, D. & DUCROCQ, V. 2015. Inbreeding and crossbreeding parameters for production and fertility traits in Holstein, Montbéliarde, and Normande cows. *Journal of Dairy Science,* 98**,** 4904-4913.

DOEKES, H. P., VEERKAMP, R. F., BIJMA, P., HIEMSTRA, S. J. & WINDIG, J. J. 2018. Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. *Genetics Selection Evolution,* 50**,** 1-16.

ELAVARASAN, K., KUMAR, S., AGARWAL, S., VANI, A., SHARMA, R., KUMAR, S., CHAUHAN, A., SAHOO, N. R., VERMA, M. R. & GAUR, G. K. 2023. Estimation of microsatellite-based autozygosity and its correlation with pedigree inbreeding coefficient in crossbred cattle. *Animal Biotechnology***,** 1-14.

ESPIGOLAN, R., BALDI, F., BOLIGON, A. A., SOUZA, F. R., GORDO, D. G., TONUSSI, R. L., CARDOSO, D. F., OLIVEIRA, H. N., TONHATI, H. & SARGOLZAEI, M. 2013. Study of whole genome linkage disequilibrium in Nellore cattle. *BMC genomics,* 14**,** 1-8.

GIBBS, R. A., TAYLOR, J. F., VAN TASSELL, C. P., BARENDSE, W., EVERSOLE, K. A., GILL, C. A., GREEN, R. D., HAMERNIK, D. L. & KAPPES, S. M. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science,* 324**,** 528-532.

GODDARD, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica,* 136**,** 245-257.

GODDARD, M. & HAYES, B. 2012. Genome-wide association studies and linkage disequilibrium in cattle. *Bovine genomics***,** 192-210.

HOWEY, R. & CORDELL, H. 2011. MapThin. *http://www.staff.ncl.ac.uk/richard.howey/mapthin/*.

KARIMI, K., ESMAILIZADEH KOSHKOIYEH, A. & GONDRO, C. 2015. Comparison of linkage disequilibrium levels in Iranian indigenous cattle using whole genome SNPs data. *Journal of animal science and technology,* 57**,** 1-10.

KELLEHER, M., BERRY, D., KEARNEY, J., MCPARLAND, S., BUCKLEY, F. & PURFIELD, D. 2017. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. *Animal,* 11**,** 15-23.

KHATKAR, M. S., NICHOLAS, F. W., COLLINS, A. R., ZENGER, K. R., CAVANAGH, J. A., BARRIS, W., SCHNABEL, R. D., TAYLOR, J. F. & RAADSMA, H. W. 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC genomics,* 9**,** 1-18.

KRUGLYAK, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature genetics,* 22**,** 139-144.

KUMAR, S., ALEX, R., GAUR, G., MUKHERJEE, S., MANDAL, D., SINGH, U., TYAGI, S., KUMAR, A., DAS, A. & DEB, R. 2018. Evolution of Frieswal cattle: A crossbred dairy animal of India. *Indian J Anim Sci,* 88**,** 265-275.

LARMER, S., SARGOLZAEI, M. & SCHENKEL, F. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. *Journal of dairy science,* 97**,** 3128-3141.

LEROY, G., BAUMUNG, R., BOETTCHER, P., SCHERF, B. & HOFFMANN, I. 2016. Sustainability of crossbreeding in developing countries; definitely not like crossing a meadow…. *Animal,* 10**,** 262-273.

LI, Y. & KIM, J.-J. 2015. Effective Population Size and Signatures of Selection Using Bovine 50K SNP Chips in Korean Native Cattle (Hanwoo). *Evolutionary Bioinformatics,* 11**,** EBO.S24359.

LU, D., SARGOLZAEI, M., KELLY, M., LI, C., VANDER VOORT, G., WANG, Z., PLASTOW, G., MOORE, S. & MILLER, S. P. 2012. Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Frontiers in Genetics,* 3**,** 152.

MAKANJUOLA, B. O., MIGLIOR, F., ABDALLA, E. A., MALTECCA, C., SCHENKEL, F. S. & BAES, C. F. 2020. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *Journal of dairy science,* 103**,** 5183-5199.

MAKINA, S. & TAYLOR, J. 2015. van Marle-Kö ster E, Muchadeyi FC, Makgahlela ML, MacNeil MD, et al. Extent of linkage disequilibrium and effective population size in four South African Sanga cattle breeds. *Frontiers in genetics,* 6**,** 337.

MARTINEZ, R., BEJARANO, D., RAMÍREZ, J., OCAMPO, R., POLANCO, N., PEREZ, J. E., ONOFRE, H. G. & ROCHA, J. F. 2023. Genomic variability and population structure of six Colombian cattle breeds. *Tropical Animal Health and Production,* 55**,** 1-8.

MATUKUMALLI, L. K., LAWLEY, C. T., SCHNABEL, R. D., TAYLOR, J. F., ALLAN, M. F., HEATON, M. P., O'CONNELL, J., MOORE, S. S., SMITH, T. P. & SONSTEGARD, T. S. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PloS one,* 4**,** e5350.

MBOLE-KARIUKI, M. N., SONSTEGARD, T., ORTH, A., THUMBI, S., BRONSVOORT, B. D. C., KIARA, H., TOYE, P., CONRADIE, I., JENNINGS, A. & COETZER, K. 2014. Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity,* 113**,** 297-305.

MCKAY, S. D., SCHNABEL, R. D., MURDOCH, B. M., MATUKUMALLI, L. K., AERTS, J., COPPIETERS, W., CREWS, D., NETO, E. D., GILL, C. A. & GAO, C. 2007. Whole genome linkage disequilibrium maps in cattle. *BMC genetics,* 8**,** 1-12.

MEUWISSEN, T. H., HAYES, B. J. & GODDARD, M. 2001. Prediction of total genetic value using genome-wide dense marker maps. *genetics,* 157**,** 1819-1829.

MRODE, R., OJANGO, J. M. K., OKEYO, A. & MWACHARO, J. M. 2019. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. *Frontiers in genetics,* 9**,** 694.

MUMTAZ, S., MUKHERJEE, A., PATHAK, P. & PARVEEN, K. 2021. Effects of inbreeding on performance traits in Karan Fries crossbred cattle.

MURRAY, C., HUERTA-SANCHEZ, E., CASEY, F. & BRADLEY, D. G. 2010. Cattle demographic history modelled from autosomal sequence variation. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 365**,** 2531-2539.

O'BRIEN, A. M. P., MÉSZÁROS, G., UTSUNOMIYA, Y. T., SONSTEGARD, T. S., GARCIA, J. F., VAN TASSELL, C. P., CARVALHEIRO, R., DA SILVA, M. V. & SÖLKNER, J. 2014. Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livestock Science,* 166**,** 121-132.

OUÉDRAOGO, D., OUÉDRAOGO-KONÉ, S., YOUGBARÉ, B., SOUDRÉ, A., ZOMA-TRAORÉ, B., MÉSZÁROS, G., KHAYATZADEH, N., TRAORÉ, A., SANOU, M. & MWAI, O. A. 2021. Population structure, inbreeding and admixture in local cattle populations managed by community-based breeding programs in Burkina Faso. *Journal of Animal Breeding and Genetics,* 138**,** 379-388.

PORTO-NETO, L. R., KIJAS, J. W. & REVERTER, A. 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genetics Selection Evolution,* 46**,** 1-5.

REICH, D. E., CARGILL, M., BOLK, S., IRELAND, J., SABETI, P. C., RICHTER, D. J., LAVERY, T., KOUYOUMJIAN, R., FARHADIAN, S. F. & WARD, R. 2001. Linkage disequilibrium in the human genome. *Nature,* 411**,** 199-204.

RODRÍGUEZ-RAMILO, S. T., FERNÁNDEZ, J., TORO, M. A., HERNÁNDEZ, D. & VILLANUEVA, B. 2015. Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. *PLoS One,* 10**,** e0124157.

SARAVANAN, K., PANIGRAHI, M., KUMAR, H., PARIDA, S., BHUSHAN, B., GAUR, G., KUMAR, P., DUTT, T., MISHRA, B. & SINGH, R. 2022. Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. *Animal biotechnology,* 33**,** 297-311.

SARGOLZAEI, M., SCHENKEL, F., JANSEN, G. & SCHAEFFER, L. 2008. Extent of linkage disequilibrium in Holstein cattle in North America. *Journal of Dairy Science,* 91**,** 2106-2117.

SILVA, C., NEVES, H., QUEIROZ, S., SENA, J. & PIMENTEL, E. Extent of linkage disequilibrium in Brazilian Gyr dairy cattle based on genotypes of AI sires for dense SNP markers. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig, 2010. 1-29.

SINGH, A., KUMAR, A., MEHROTRA, A., PANDEY, A. K., MISHRA, B. & DUTT, T. 2021. Estimation of linkage disequilibrium levels and allele frequency distribution in crossbred Vrindavani cattle using 50K SNP data. *Plos one,* 16**,** e0259572.

SLIFER, S. H. 2018. PLINK: key functions for data analysis. *Current protocols in human genetics,* 97**,** e59.

STACHOWICZ, K., SARGOLZAEI, M., MIGLIOR, F. & SCHENKEL, F. 2011. Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. *Journal of dairy science,* 94**,** 5160-5175.

TEAM, R. C. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

TIAN, R., ASADOLLAHPOUR NANAIE, H., WANG, X., DALAI, B., ZHAO, M., WANG, F., LI, H., YANG, D., ZHANG, H. & LI, Y. 2023. Genomic adaptation to extreme climate conditions in beef cattle as a consequence of cross-breeding program. *BMC genomics,* 24**,** 1-10.

UIMARI, P., KONTKANEN, O., VISSCHER, P. M., PIRSKANEN, M., FUENTES, R. & SALONEN, J. T. 2005. Genome-wide linkage disequilibrium from 100,000 SNPs in the East Finland founder population. *Twin Research and Human Genetics,* 8**,** 185-197.

VILLA-ANGULO, R., MATUKUMALLI, L. K., GILL, C. A., CHOI, J., VAN TASSELL, C. P. & GREFENSTETTE, J. J. 2009. High-resolution haplotype block structure in the cattle genome. *BMC genetics,* 10**,** 1-13.

WRAY, N. R. 2005. Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Research and Human Genetics,* 8**,** 87-94.

YAN, J., SHAH, T., WARBURTON, M. L., BUCKLER, E. S., MCMULLEN, M. D. & CROUCH, J. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PloS one,* 4**,** e8451.

ZHANG, H., YIN, L., WANG, M., YUAN, X. & LIU, X. 2019. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Frontiers in genetics,* 10**,** 189.

ZHAO, F., WANG, G., ZENG, T., WEI, C., ZHANG, L., WANG, H., ZHANG, S., LIU, R., LIU, Z. & DU, L. 2014. Estimations of genomic linkage disequilibrium and effective population sizes in three sheep populations. *Livestock Science,* 170**,** 22-29.

ZHAO, H., NETTLETON, D. & DEKKERS, J. C. 2007. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genetics Research,* 89**,** 1-6.

ZHU, M., ZHU, B., WANG, Y., WU, Y., XU, L., GUO, L., YUAN, Z., ZHANG, L., GAO, X. & GAO, H. 2013. Linkage disequilibrium estimation of Chinese beef Simmental cattle using high-density SNP panels. *Asian-Australasian Journal of Animal Sciences,* 26**,** 772-779.