Genome Wide Identification of Novel SNPs and Diversity Analysis in Tharparkar Cattle

Thesis

Submitted to the
DEEMED UNIVERSITY
Indian Veterinary Research Institute
Izatnagar - 243 122 (U.P.), India

Dr. Anal Bose
Roll No. M-6668

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Veterinary Science
(Animal Genetics and Breeding)

2025

`

Dedicated to....

My Beloved Mother

&

Respected Guide

भा.कृ.अनु.प.–भारतीय पशु चिकित्सा अनुसंधान संस्थान
(सम विश्वविद्यालय)
इज्जतनगर –243122, (उ.प्र.), भारत

DIVISION OF ANIMAL GENETICS
ICAR-INDIAN VETERINARY RESEARCH INSTITUTE
(Deemed University)
IZATNAGAR - 243 122, U.P., INDIA

Dr. Manjit Panigrahi
M.V.Sc., Ph.D.
Senior Scientist

Dated: 25/08/ **2025**

Certificate

This is to be certified that the research work embodied in this thesis entitled **"Genome Wide Identification of Novel SNPs and Diversity Analysis in Tharparkar Cattle"** *submitted by* **Dr. Anal Bose, Roll No. M-6668,** *for the award of* **Master of Veterinary Science** *Degree in* **Animal Genetics and Breeding** *at ICAR-Indian Veterinary Research Institute, Izatnagar, is the original work carried out by the candidate himself under my supervision and guidance.*

It is further certified that **Dr. Anal Bose, Roll No. M-6668,** *has completed 4 semesters in the Institute and has put in more than 6 months attendance under me from the date of registration for the Master of Veterinary Science Degree in this Deemed University, as required under the relevant ordinance.*

M. Panigrahi

(MANJIT PANIGRAHI)
Chairman
Advisory Committee

Certificate

We the undersigned members of Advisory Committee of **Dr. Anal Bose,** *Roll No. M-6668, a candidate for the degree of* **Master of Veterinary Science** *with the major discipline* **Animal Genetics and Breeding,** *agree that the thesis entitled* **"Genome Wide Identification of Novel SNPs and Diversity Analysis in Tharparkar Cattle"** *may be submitted in partial fulfillment of the requirement for the degree.*

We have gone through the contents of the thesis and are fully satisfied with the work carried out by the candidate, which is being presented for the award of **Master of Veterinary Science** *Degree of this Institute.*

It is further certified that the candidate has completed all the prescribed requirements governing the award of **Master of Veterinary Science** *Degree of the Deemed University, ICAR-Indian Veterinary Research Institute, Izatnagar.*

Signature
Name Dr. D. Kumar
External Examiner

Date : 11/09/2025

(Manjit Panigrahi)
Chairman
Advisory Committee

Date : 11/09/2025

MEMBERS OF STUDENT'S ADVISORY COMMITTEE

Dr. Pushpendra Kumar, Principal Scientist
Division of Animal Genetics, ICAR-IVRI, Izatnagar

Dr. G.V.P.P.S. Ravi Kuamr, Principal Scientist & Head
Division of Animal Biotechnology, ICAR-NBAGR, Karnal

Dr. Ajay Kumar, Principal Scientist
Division of Biochemistry, ICAR-IVRI, Izatnagar

ACKNOWLEDGEMENTS

Life is filled with countless beautiful moments that deserve to be cherished and remembered and this thesis stands as a reflection of many such precious experiences that I gathered during the course of my work.

First and foremost, I bow in deep gratitude to the **Almighty** for His infinite blessings, guidance and opportunities that enabled me to successfully complete this thesis.

I am profoundly indebted to my family for their unwavering love and support. My heartfelt gratitude goes to my **mother, Mrs. Soma Bose**, my **father, Mr. Pradip Kumar Bose** and my **brother, Mr. Sudipto Bose**, whose constant encouragement, care and faith in me have been my greatest strength throughout this journey.

With the deepest sense of respect, I express my sincere appreciation to my revered Chairman and guide, **Dr. Manjit Panigrahi**, Senior Scientist, Animal Genetics Division, **ICAR-IVRI, Izatnagar**. His invaluable guidance, constant support and fatherly mentorship have been instrumental in shaping this research. His expertise, patience and encouragement not only enhanced the scientific merit of this thesis but also imparted life lessons that I will always treasure.

I am also profoundly thankful to **Dr. Subodh Kumar**, Head and Principal Scientist, Division of Animal Genetics, ICAR-IVRI, **Dr. A.K. Pandey**, Principal Scientist and **Dr. Anuj Chauhan,** Principal Scientist, ICAR-IVRI, for their continuous motivation, constructive advice, and kind support during my M.V.Sc. programme.

I am sincerely grateful to my advisory committee members - **Dr. Pushpendra Kumar**, Principal Scientist (AG); **Dr. Ravi Kumar Gandham V.P.P.S.,** Principal Scientist and Head (BTY) ICAR-NBAGR and **Dr. Ajay Kumar**, Senior Scientist (BCT), **ICAR-IVRI** - for their constructive suggestions, kind support and valuable guidance throughout the course of my work.

I extend my heartfelt thanks to **Dr. Sita Ram Gupta**, Officer In Charge, Livestock Research Station (Beechwal) and **Dr. Vijay Bishnoi**, Officer In Charge, Livestock Research

Station (Bikaner), **RAJUVAS Rajasthan**, for granting permission to collect blood samples and for their immense cooperation during sample collection, which was vital for this research.

I gratefully acknowledge the esteemed Director, **Dr. T. Dutt** and the Joint Director (Research), **Dr. S.K. Singh**, **ICAR-IVRI**, for providing essential facilities and infrastructure. I am also thankful to all the respected faculty members of the **Animal Genetics Division, ICAR-IVRI**, for their excellent teaching and academic guidance during my coursework, which laid the foundation for this research.

My heartfelt gratitude goes to my lab seniors, **Dr. Divya Rajawat** and **Dr. Sonali Sonejita Nayak**, for patiently teaching me research fundamentals and guiding me through experiments. I also thank my divisional seniors, juniors and the technical as well as non-technical staff of the **Division of Animal Genetics** for their constant cooperation and support.

I am especially grateful to **Dr. Anurodh Sharma**, **Dr. Suryakant Verma**, **Mr. Vijendra** and **Ms. Nooreen** for their continuous support and kindness, which made my laboratory work and stay more pleasant. I also extend my warm thanks to my seniors - **Dr. Roshni**, **Dr. Nitish Gaitri**, **Dr. Amritanshu** and **Dr. Sarath** - for their encouragement and selfless assistance.

To my dear friends and batchmates - **Dr. Nishu**, **Dr. Ajay**, **Dr. Shraddha**, **Dr. Shriram**, **Dr. Ankur**, **Dr. Shivani**, **Dr. Shreyasi**, **Dr. Ansiya**, **Dr. Priyanka**, **Dr. Mir Mehroz and Dr. Kaivalya** along with my juniors - I owe my deepest gratitude for your friendship, encouragement and support. Your presence turned this academic journey into an unforgettable experience.

My heartfelt thanks to my undergraduate best friend at IVRI, **Dr. Pallavi**, without whose unwavering support my degree and stay at IVRI would not have been as seamless.

A special note of thanks goes to my friends from Ranchi - **Dr. Satyam**, **Dr. Nitu** and **Dr. Dipshikha** - who stood as pillars of strength throughout my stay at IVRI. Your constant encouragement, both academically and personally, has been invaluable and I remain truly indebted to you.

I am also thankful to **Mr. Dharmendra (Chachu)**, **Kuldeep** and **Amaan** for their kind help in preparing and arranging this manuscript. My heartfelt appreciation goes to **U.P** mess staff for ensuring my daily needs with utmost care and dedication.

I sincerely acknowledge my undergraduate professors from Ranchi Veterinary College *for their mentorship, which built the academic foundation that strengthened my pursuit of higher studies.*

I gratefully acknowledge the financial support from the National Agricultural Science Fund (NASF), ICAR*, which enabled me to pursue this research, as well as the* Institutional Fellowship*, which supported me during my M.V.Sc. program.*

Lastly, I extend my gratitude to all those who, in one way or another, contributed to this journey. Though every name cannot be mentioned here, none are forgotten.

Anal Bose

(Anal Bose)

Date: 25/08/2025

Place: *ICAR-IVRI, Izatnagar*

ABBREVIATIONS

%	:	Percentage
<	:	less than
=	:	equal to
AE buffer	:	DNA elution buffer (Qiagen kit specific)
AL buffer	:	Lysis buffer (Qiagen kit specific)
AMOVA	:	Analysis of Molecular Variance
AMPK	:	AMP-activated Protein Kinase
ARS-UCD 2.0	:	Bos taurus genome reference assembly version
AW1, AW2 buffer	:	Wash buffers (Qiagen kit specific)
BAHS	:	Basic Animal Husbandry Statistics
BAM	:	Binary Alignment Map
BED	:	Browser Extensible Data format
bp	:	base pair
BWT	:	Burrows–Wheeler Transform
CCBF	:	Central Cattle Breeding Farm
CNV(s)	:	Copy Number Variation(s)
CRoPS	:	Complexity Reduction of Polymorphic Sequences
DAHD	:	Department of Animal Husbandry and Dairying
DAVID	:	Database for Annotation, Visualization and Integrated Discovery
δ (Delta)	:	Allele frequency difference statistic
ddRAD / ddRADseq	:	Double-Digest Restriction Site-Associated DNA Sequencing
D	:	Average Pairwise Genetic Distance
DNA	:	Deoxyribonucleic Acid
ENA	:	European Nucleotide Archive
FASTQ	:	File format for storing sequence data with quality scores
FIFS	:	Frequent Item Feature Selection
F_{PED}	:	Pedigree-based Inbreeding Coefficient
FSNP	:	SNP-based Inbreeding Coefficient
F	:	Inbreeding coefficient (general)
FIS	:	Inbreeding coefficient within subpopulations (Fixation Index)
F_{ROH}	:	Inbreeding coefficient based on Runs of Homozygosity
F_{ST}	:	Fixation index (Wright's F-statistic; population differentiation)

F_{UNI}	:	Inbreeding coefficient based on correlation between uniting gametes
GB	:	Gigabyte
GBS	:	Genotyping by Sequencing
GDP	:	Gross Domestic Product
GO	:	Gene Ontology
GR	:	Genotyping Rate
GVA	:	Gross Value Added
He	:	Expected heterozygosity
Ho	:	Observed heterozygosity
HSP	:	Heat Shock Protein
HWE	:	Hardy–Weinberg Equilibrium
IAEC	:	Institutional Animal Ethics Committee
IBD	:	Identical by Descent
IBS	:	Identical by State
iNOS	:	Inducible Nitric Oxide Synthase
In	:	Informativeness for Assignment
LD	:	Linkage Disequilibrium
LRS	:	Livestock Research Station
mL	:	millilitre
MAF	:	Minor Allele Frequency
MAF-LD	:	Minor Allele Frequency – Linkage Disequilibrium method
MCMC	:	Markov Chain Monte Carlo
MoSPI	:	Ministry of Statistics and Programme Implementation
mtDNA	:	Mitochondrial DNA
NanoDrop	:	Spectrophotometer for nucleic acid quantification
NBAGR	:	National Bureau of Animal Genetic Resources
NCBI	:	National Center for Biotechnology Information
NDDB	:	National Dairy Development Board
Ne	:	Effective population size
NGS	:	Next-Generation Sequencing
NPBBDD	:	National Programme for Bovine Breeding and Dairy Development
π (Pi)	:	Nucleotide diversity
PCA	:	Principal Component Analysis
QTL	:	Quantitative Trait Locus

QTLdb	:	Quantitative Trait Loci Database
RAD-seq	:	Restriction Site-Associated DNA Sequencing
RAJUVAS	:	Rajasthan University of Veterinary and Animal Sciences
RD	:	Read Depth
R1, R2	:	Forward and reverse paired-end reads
ROH	:	Runs of Homozygosity
rpm	:	revolutions per minute
SAM	:	Sequence Alignment Map
SNP(s)	:	Single Nucleotide Polymorphism(s)
SRA	:	Sequence Read Archive
SSR	:	Simple Sequence Repeat (Microsatellite)
TRES	:	Toolbox for Ranking and Evaluation of SNPs
Ts/Tv	:	Transition-to-Transversion ratio
VCF	:	Variant Call Format
VEP	:	Variant Effect Predictor
WGS	:	Whole Genome Sequencing
µg	:	microgram
µL / µl	:	microlitre
ng/µL	:	nanogram per microlitre

LIST OF TABLES

Table No.	Title	On/After Page No.
Table 1	Quality parameters of genomic DNA isolated from 48 samples, showing concentration, purity (A260/280), eluted volume, and total yield	22
Table 2	Raw sequencing data statistics of paired-end FASTQ files generated from ddRAD sequencing of 48 Tharparkar cattle	34
Table 3	Summary of clean paired-end FASTQ reads of 48 Tharparkar cattle after adapter trimming and quality filtering	37
Table 4	Summary of observed and expected heterozygosity values across autosomes	41
Table 5	Details of Number of Animals per breed in reference and validation dataset	43
Table 6	Details of SNPs selected through pre-selection methods and retained after the first MAF–LD filtering step, along with the final set of breed-specific SNPs.	44
Table 7	Genes overlapping with genomic regions harboring Tharparkar-specific SNPs, with chromosome position, SNP ID, gene symbol, and gene name	47
Table 8	Summary of QTLs overlapping with breed-specific Tharparkar SNPs and their associated genes	48

LIST OF FIGURES

Figure No.	Title	On/After Page No.
Fig. 1:	Gel band image of isolated genomic DNA of 48 Tharparkar cattle	22
Fig. 2a:	Classification of variants based on type	41
Fig. 2b:	Single Nucleotide Polymorphism effect distribution	41
Fig. 2c.	Number of transitions and transversions in annotated variants	41
Fig. 2d:	Number of effects by impact	41
Fig. 2e:	Number of effects by functional class	41
Fig. 3:	Diagrammatic representation of Chromosome wise number of variants	41
Fig. 4:	Boxplot depicting observed (O.HET) and expected (E.HET) heterozygosity values	41
Fig. 5:	Diagrammatic representation of chromosome wise observed (Ho) and expected (He) heterozygosity	41
Fig. 6:	Pie chart showing the distribution of Minor Allele Frequency (MAF) across different class intervals	41
Fig. 7a:	Class wise distribution of ROH count	43
Fig. 7b:	F_{ROH} estimates according to various ROH class lengths	43
Fig. 8:	Chromosome wise F_{ROH} values in Tharparkar cattle	43
Fig. 9:	Violin plot depicting genome-wide F_{ROH} distribution. The central red line in (a) signifies the median value of the population's F_{ROH} while the upper and lower boundaries of the central black box denote the upper and lower quartiles of the F_{ROH} distribution within the population. In (b) the central red dot signifies the mean value of the population's F_{ROH}. The width of the violin plot characterizes the probability density distribution of the population's F_{ROH}	43
Fig. 10:	Bar plot of F_{ROH} for each individual in the population	43

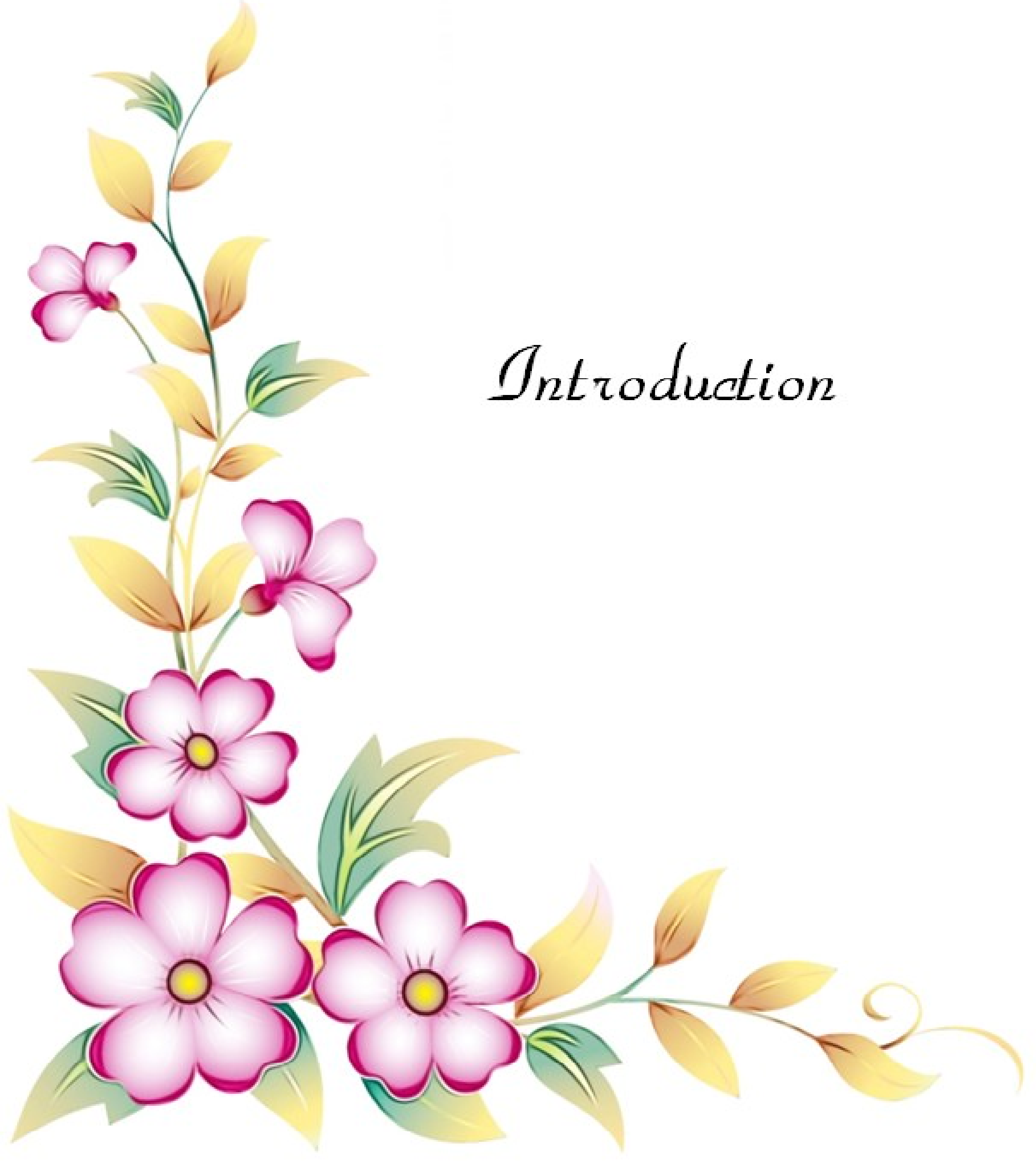
Figure No.	Title	On/After Page No.
Fig. 11:	Manhattan plot showing the percentage of SNPs included in Runs of Homozygosity (ROH) across autosomal chromosomes in Tharparkar cattle. Each dot represents a SNP, with alternating colors (red and blue) distinguishing adjacent chromosomes. The y-axis represents the proportion of SNPs occurring within ROH, while the x-axis corresponds to the chromosome number.	43
Fig. 12:	Boxplot depicting genomic inbreeding coefficients	43
Fig. 13:	Declining trend of effective population size (Ne) over the past 150 generations	43
Fig. 14:	Principal Component Analysis (PCA) plot showing the genetic variation among individuals, confirming the presence of five distinct breed clusters.	45
Fig. 15a:	SNPs selected by TRES for distinguishing Tharparkar breed from other breeds using three pre-selection methods - Delta, pairwise Wright's F_{ST}, and Informativeness for Assignment - with common SNPs identified by intersecting the results across methods through VENNY 2.1.0.	45
Fig. 15b:	Rathi breed versus other breeds	45
Fig. 15c:	Red Sindhi breed versus other breeds	45
Fig. 15d:	Gir breed versus other breeds	45
Fig. 15e:	Sahiwal breed versus other breeds	45
Fig. 16a:	Open bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Reference dataset	47
Fig. 16b:	Close bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Reference dataset	47
Fig. 17a:	Open bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Validation dataset	47
Fig. 17b:	Close bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Validation dataset	47

Figure No.	Title	On/After Page No.
Fig. 18:	F_{ST} plot generated using STRUCTURE v2.3.4, illustrating the differentiation of Tharparkar breed from the other breeds in reference dataset	47
Fig. 19:	F_{ST} plot generated using STRUCTURE v2.3.4, illustrating the differentiation of Tharparkar breed from the other breeds in validation dataset.	47
Fig. 20:	Summary of functional annotation of 77 breed-specific SNPs in Tharparkar cattle obtained using the Variant Effect Predictor (VEP).	47

CONTENTS

Sl. No.	CHAPTER	PAGE NO.
1.	INTRODUCTION	01-05
2.	REVIEW OF LITERATURE	06-19
3.	MATERIALS AND METHODS	20-32
4.	RESULTS	33-49
5.	DISCUSSION	50-56
6.	SUMMARY AND CONCLUSIONS	57-60
7.	MINI ABSTRACT	61
8.	HINDI ABSTRACT	62
9.	REFERENCES	63-80

Introduction



1 INTRODUCTION

The livestock sector is a cornerstone of India's rural economy, significantly contributing to the nation's Gross Domestic Product (GDP). It serves as an essential source of income for households engaged in agriculture as well as landless communities, with approximately 20.5 million people depending on this sector for their livelihoods. According to the provisional estimates of the Gross Value Added (GVA) by the Ministry of Statistics and Programme Implementation (MoSPI) released on May 31, 2024, the livestock sector contributed 13,55,460 crores at current prices during the financial year 2022-23. This accounts for 30.23% of the Agricultural and Allied Sector GVA and 5.50% of the total GVA.

Among all the agricultural and allied sector activities, dairy stands out as one of the largest agricultural commodities, contributing 5% to the national economy and directly supporting the livelihoods of over 8.5 crore farmers. India holds a prominent position globally as the leading producer of milk, contributing 25% of the world's total milk production. During 2023-24, India's milk production reached an estimated 239.30 million tonnes, reflecting a remarkable growth of 51.05% over the past decade. Notably, milk production from indigenous and non-descript cattle showed a substantial increase of 44.76% during 2023-24, surpassing the 8% growth observed in exotic and crossbred cattle (BAHS, 2024).

India is home to the largest livestock population globally, with its cattle comprising 19.5% of the world's total. Indigenous and non-descript cows contributed 11.36% and 10.11%, respectively, to the total milk produced in the country in 2023-24 (BAHS, 2024). At present, India boasts 53 recognized indigenous cattle breeds, categorized based on their utility as dairy,

draught, or dual-purpose breeds (NBAGR). The growing preference for indigenous cattle over exotic and crossbred varieties stems from their numerous advantages, including disease resistance, adaptability to extreme weather conditions and ability to thrive on low-quality roughages and grasses. These figures highlight the vast potential of India's livestock sector and the increasing focus on conserving and promoting indigenous breeds. This transition aligns with the objectives of sustainable livestock production while emphasizing the need to preserve the rich genetic diversity of Indian cattle.

The evolution of modern cattle has played an important role in shaping their genetic traits, physical characteristics and ability to adapt to diverse environments. This evolutionary journey has been influenced by natural selection, environmental pressures, human intervention and agricultural needs over thousands of years. Present-day cattle trace their ancestry to two wild aurochs species: *Bos primigenius primigenius* (the European aurochs) and *Bos primigenius namadicus* (the Indian aurochs) (Ajmone-Marsan *et al*., 2010). Historically, aurochs were widespread across Eurasia and northern Africa but became extinct in 1627 AD, with the last recorded sighting in Poland (Götherström *et al*., 2005; Zeuner, 1963). Approximately 10,000 years ago, *Bos primigenius* gave rise to taurine cattle through domestication in the Fertile Crescent. A separate domestication event, occurring about 1,500 years later in the Indus Valley, led to the development of indicine or zebu cattle, which belong to a distinct lineage from taurine cattle (Bruford *et al*., 2003; Loftus *et al*., 1994). Zebu (*Bos indicus*) cattle have been shaped over centuries by traditional Indian animal husbandry practices with minimal scientific intervention. As a result, they are highly adapted to challenging local environments, demonstrating resistance to tropical diseases and parasites and thriving on low-quality roughages and grasses (Sharma *et al*., 2015).

Tharparkar, also known as White Sindhi, Grey Sindhi, or Thari, is a dual-purpose indigenous breed of India that is renowned for its adaptability to harsh climatic conditions and its valuable contribution to both milk and draught power. It is recognized as a significant dairy breed native to the western arid regions of India, including western Rajasthan and the Rann of Kutch in Gujarat. The breed traces its origins to the Tharparkar district in the Sind region of Pakistan. The total population of Tharparkar cattle in India stands at 5,82,257, accounting for

only 0.4% of the country's indigenous cattle population. Among these, 1,51,056 are purebred, while 4,31,201 are graded, exhibiting more than 50% phenotypic resemblance but not considered pure (Breed Survey 2019, DAHD). Comparing this with data from the previous Breed Survey (2013, DAHD), the Tharparkar population was reported at 7,32,473, including 1,97,291 purebred and 5,35,182 graded cattle, indicating a possible declining trend in the breed's population. There was also a 6% decline in the total indigenous cattle population (both descript and non-descript) compared to the previous census (DAHD report, 2022-23). This decline can be attributed to various factors, including the implementation of crossbreeding programs aimed at increasing productivity, which led to faster genetic degradation and breed dilution, particularly due to unplanned breeding and the introduction of exotic germplasm. The situation was further exacerbated by indiscriminate crossbreeding, modern purpose-driven farming practices, climate change and other challenges (Srivastava, 2019). Additionally, the productive performance of Tharparkar cattle is declining due to the lack of efficient selection strategies, which underscores the urgent need to conserve and enhance the productivity of indigenous cattle breeds. Genetic characterization is crucial for identifying genetic variability, which is fundamental to effective breed improvement and conservation efforts. Natural and artificial selection pressures across the genome enhance disease resistance and adaptive potential within populations (Kristensen *et al*., 2015; Berry, 2018). Preserving adequate genetic variation is vital for maintaining a population's ability to adapt to environmental challenges, as its absence can limit selection response and genetic progress. Consequently, profiling genetic diversity and conducting genome-wide scans to identify regions under selection are essential for designing effective conservation programs and breed improvement strategies (Panigrahi *et al*., 2023).

The advent of high-throughput next-generation sequencing (NGS) techniques and the development of molecular markers have significantly broadened the scope of livestock genomics studies. Among these, Single Nucleotide Polymorphisms (SNPs) have gained immense popularity due to their simplicity, bi-allelic nature, codominance, genetic stability and widespread distribution throughout the genome (Matukumalli *et al*., 2009). Genomic tools like the Bovine SNP50 BeadChip v3 and the BovineHD (777K) SNP chip have been employed in various genomic applications, including admixture analysis, breed-specific SNP panel development (Kumar *et*

al., 2019; Kumar *et al*., 2021a; Kumar *et al*., 2021b), detection of copy number variations (CNVs) (Kumar *et al*., 2023) and the identification of selection signatures (Saravanan *et al*., 2020a). These tools have also facilitated the analysis of genetic diversity parameters and haplotype block structures (Saravanan *et al*., 2022; Chhotaray *et al*., 2021).

However, SNP array-based technologies come with certain limitations, such as ascertainment bias and reduced capacity to detect novel and rare variants (DeDonato *et al*., 2013). Moreover, bovine SNP chips are predominantly constructed using SNPs identified in taurine breeds, limiting their ability to accurately represent diversity parameters and identify novel rare variants in the genomic data of indigenous cattle breeds (DeDonato *et al*., 2013). Advances in sequencing technologies, such as Whole Genome Sequencing (WGS) and Genotyping by Sequencing (GBS), have addressed these challenges by minimizing ascertainment bias and enabling the detection of novel and rare SNPs (Peterson *et al*., 2012). These methods provide more comprehensive and in-depth genotyping, enhancing the analysis of multiple individuals with greater efficiency and precision.

Reduced representation sequencing, commonly known as RADseq, is a cost-efficient and effective technique that uses restriction enzymes to focus on a specific portion of the genome. This method provides higher coverage depth for each locus (Davey *et al*., 2013). RADseq loci are generally conserved within populations and can be located in both coding and non-coding regions of the genome. Variations of this approach include single-enzyme methods, such as the original RADseq and Genotyping by Sequencing, as well as dual-enzyme strategies like double-digest Restriction site-Associated DNA sequencing (ddRAD) and Complexity Reduction of Polymorphic Sequences (CRoPS) (Andrews *et al*., 2016).

The study of genetic diversity and the identification of novel breed-specific SNPs are fundamental to the effective conservation and management of indigenous cattle breeds. These breeds harbour unique genetic traits that have evolved over centuries through natural selection, isolation and traditional breeding practices, making them well-suited to specific environments. Traits such as resistance to local diseases, tolerance to harsh climatic conditions and efficiency in utilizing limited resources are vital for sustainable agricultural systems. However, the genetic

diversity of these breeds is increasingly threatened by crossbreeding with high-yielding exotic breeds and the subsequent replacement of local populations. This highlights the need for detailed genetic characterization and strategic conservation planning (Groeneveld *et al*., 2010).

The identification of breed-specific Single Nucleotide Polymorphisms (SNPs) using the ddRAD sequencing approach complements these efforts by offering a cost-effective and efficient tool for advancing genomic studies in livestock. This method enables the discovery of unique SNPs that define the genetic identity of indigenous breeds, addressing challenges such as ascertainment bias and the detection of rare variants often missed in traditional SNP arrays. The development of breed-specific SNP panels through ddRAD sequencing facilitates precise genetic characterization, differentiation and parentage verification, while also supporting tailored breeding strategies. These advances enhance the productivity and resilience of indigenous breeds, ensuring their conservation and sustainable use for future generations (Kumar *et al*., 2021b). Collectively, these efforts safeguard the genetic diversity of indigenous breeds, ensuring their ability to address emerging challenges such as climate change, disease outbreaks and evolving agricultural demands.

OBJECTIVES:

a) To estimate the genetic diversity parameters in Tharparkar cattle by using ddRAD sequencing data.

b) To identify novel breed-specific SNPs using ddRAD sequencing data and annotation of those SNPs.

Review of Literature

2 REVIEW OF LITERATURE

2.1 Cattle Domestication: From Aurochs to Modern Breeds

The domestication of animals has played a pivotal role in shaping human demographic and cultural development, with cattle being the first livestock species domesticated for agricultural purposes. Archaeological and biological findings suggest that *Bos taurus* and *Bos indicus* were independently domesticated from the ancestral species *Bos primigenius* (wild aurochs) around 8,000 to 10,000 years ago in the Middle East and Indian subcontinent (Bollongino *et al*., 2012). These ancestral populations migrated across the globe, adapting to various geographical and ecological conditions, including different temperatures, humidity levels, climates and feed availability. Evolutionary forces and demographic processes, such as genetic drift, non-random mating, selection, mutation and migration, contributed to the development of a diverse range of cattle breeds with distinct phenotypes (Groenveld *et al*., 2010).

After domestication, cattle breeds were selectively bred for desirable traits, particularly for milk and meat production, as well as draught power. Artificial selection is a process in animal breeding that has led to the evolution of breeds through human intervention. Darwin (1868) classified two types of artificial selection: methodical and unconscious (Gregory, 2009). Unconscious selection was more common in ancient times, occurring without any deliberate effort to modify a breed. Over time, breeders adopted methodical or deliberate selection, aimed at achieving specific objectives. Both selection and crossbreeding have played significant roles in the development of composite cattle breeds.

2.2 Indigenous cattle resources and their depletion

India is one of the world's 12 mega-biodiversity countries and ranks as the third richest biodiversity nation in Asia (Dash *et al*., 2022). It is home to a vast array of cattle genetic resources, with 53 registered indigenous cattle breeds (NBAGR, 2024) and many more yet to be explored. These breeds exhibit significant diversity in phenotypic traits, utility patterns and adaptability, shaped by varying agro-climatic conditions and production systems across the country (Srivastava *et al*., 2019). Indigenous cattle represent 22% of India's total cattle population of 192.5 million (DAHD report, 2022-23; 20th All India Livestock Census, 2019), yet they are experiencing a concerning decline, with a 6% decrease noted in the 20th livestock census. These indigenous breeds have evolved to thrive in harsh climatic conditions with minimal management inputs, efficiently converting low-quality feeds into valuable products while being more resilient to tropical diseases. They play an important role in the country's agricultural economy. However, due to unregulated breeding practices and the introduction of exotic germplasm through crossbreeding, many of these breeds are facing rapid genetic degradation and dilution (Groeneveld *et al*., 2010). Consequently, several indigenous breeds are becoming endangered and the unique germplasm that provides heat tolerance and disease resistance is depleting. Breed dilution, characterized by the loss of purebred animals that conform to specific traits, is a growing concern, especially in developing nations. The conservation of genetic diversity is critical for the long-term survival of these species, particularly in the face of changing environmental conditions (Tesfa *et al*., 2023).

2.3 Tharparkar Cattle

Tharparkar is a prominent indigenous cattle breed of India, primarily distributed in the western arid regions, including western Rajasthan and the Rann of Kutch in Gujarat. It originated from the Tharparkar district in the southeast Sindh region of Pakistan. Known by various names such as Thari, Kutchi, Grey Sindhi, or White Sindhi (Mason, 1996), the breed is highly valued for its dual-purpose utility, excelling in both milk production and draught power. Milk yields range from approximately 1,135 kg per lactation on arid pastures to 1,980 kg in village settings (Choudhary *et al*., 2018). The Tharparkar cattle are of medium-sized, deep-bodied

and robust, with straight limbs, sturdy feet and a predominantly white or light gray coat. Males often exhibit darker gray shading on the fore and hindquarters, with a distinctive light gray stripe along the backbone. Characterized by a medium-sized head with a broad forehead, full bright eyes, lyre-shaped horns and a moderately developed hump in males, the breed also features a medium-sized dewlap and a prominent navel flap in females.

Renowned for its exceptional adaptability to extreme desert climates, the Tharparkar thrives in arid conditions with minimal resources. It is heat-tolerant, disease-resistant and capable of efficiently converting low-quality fodder into milk, with yields ranging from 913 to 2,147 kg per lactation. This resilience to tropical diseases and environmental stress makes it a cornerstone for sustainable dairy farming. Additionally, its genetic traits, such as thermotolerance and disease resistance, make it invaluable for crossbreeding programs aimed at enhancing productivity while retaining adaptability.

Extensive genetic and physiological research has been conducted to establish the thermo-tolerance abilities of Tharparkar cattle. Thermo-tolerance is assessed by the expression or activation of specific biological markers, including heat shock transcription factors (HSF), heat shock proteins (HSP70, HSP90 and HSP27), the slick hair gene (Dikmen *et al*., 2008), ATP1A1 (Liu *et al*., 2011), ATP1B2 (Wang *et al*., 2011) and adenosine 5'-monophosphate (AMP)-activated protein kinase (AMPK) (Min *et al*., 2017). The ATP1A1 gene, encoding the alpha-1 chain of Na+/K+-ATPase, plays a critical role in mitigating heat stress in Tharparkar cattle (Kashyap *et al*., 2014). Studies have shown that genes from the HSP70 protein family are significantly upregulated in Tharparkar cattle during the summer, enhancing their ability to adapt to extreme environments (Kumar *et al*., 2015). Bharati *et al*. (2017) observed a biphasic expression pattern of HSP70 in these cattle, which likely provides an additional layer of protection under prolonged heat stress. Furthermore, research has revealed higher basal levels of HSP70, along with reduced levels of inducible MMPs and iNOS, indicating that Tharparkar cattle are more thermo-tolerant than Karan Fries and Murrah buffaloes (Choudhary *et al*., 2018). Bhat *et al*. (2016) reported that the AA genotype, associated with HSP70 polymorphism, demonstrates superior thermotolerance in the Tharparkar cattle. Several studies

also reported a reduced incidence of mastitis in Tharparkar compared to other breeds (Jingar *et al*., 2014; Alhussien *et al*., 2016).

The population of Tharparkar cattle in India is facing a potential decline. As per the Breed Survey 2019 (DAHD), the total population stands at 5,82,257, representing only 0.4% of the country's indigenous cattle population. Of these, 1,51,056 are purebred, while 4,31,201 are graded cattle that exhibit more than 50% phenotypic resemblance but are not considered purebred. Comparatively, the Breed Survey 2013 (DAHD) reported a population of 7,32,473, including 1,97,291 purebred and 5,35,182 graded cattle, indicating a decreasing trend in the breed's numbers.

The Government of India, along with various organizations, is actively promoting the conservation and development of indigenous cattle breeds, including Tharparkar. Initiatives like the Rashtriya Gokul Mission, under the National Programme for Bovine Breeding and Dairy Development (NPBBDD), focus on breed improvement and conservation. National Dairy Development Board (NDDB) initiated the National Dairy Plan-I, a World Bank-assisted project implemented in 18 major dairy states, which supports the development of six indigenous breeds of cattle, mainly Gir, Kankrej, Tharparkar, Sahiwal, Rathi and Hariana. Facilities like the Central Frozen Semen Production & Training Institute, Central Cattle Breeding Farms (CCBFs) and Government Livestock Farms (GLFs) are dedicated to maintaining purebred indigenous animals. Additionally, Rajasthan Gau Sewa Sangh Gaushala in Jaipur houses over 100 pure Tharparkar cattle (DAHDF, 2019). Preserving and documenting India's diverse cattle genetic resources is vital for their conservation, genetic improvement and sustainable utilization (Saravanan *et al*., 2020a).

2.4 Genetic Diversity and Its Importance

The estimation of genetic diversity is fundamental for understanding the adaptability and evolutionary potential of animal populations. Genetic diversity within a population is crucial for responding to environmental challenges such as climate change, emerging diseases and resource variability (Hoffmann and Willi, 2009). High genetic variation enables populations to adapt effectively, while low variation increases vulnerability and limits long-term survival (Amos

and Harwood, 1998). Key factors that influence the genetic diversity include mutation, migration, selection, genetic drift and mating patterns, which collectively shape the genetic makeup of the population.

In livestock, genetic diversity underpins sustained genetic improvement and adaptation to evolving breeding goals. It serves as the raw material for selection, enabling breeders to develop improved breeds with desirable traits (Saravanan *et al.*, 2020a). Additionally, diversity is vital for harnessing heterosis and minimizing the transmission of unfavorable alleles to future generations (Notter, 1999). Declining genetic diversity due to inbreeding or artificial selection can increase the risk of genetic erosion, reducing resilience to environmental pressures (Visser *et al.*, 2016). Consequently, estimating and maintaining genetic diversity is essential for developing effective livestock improvement programs and ensuring long-term sustainability in animal breeding.

2.4.1 Trends in Methods of Genetic Diversity Analysis

The field of genetic diversity analysis has evolved significantly over time, transitioning through various methodologies to accommodate the need for more precise and efficient tools. Early studies in the 1970s relied on blood groups and allozyme markers to analyze domestic animal genetic diversity. Although these markers facilitated initial insights, their limited polymorphism restricted their utility (Baker and Manwell, 1980). The advent of the polymerase chain reaction and whole-genome sequencing in the 1980s revolutionized genetic diversity studies. Techniques such as restriction fragment length polymorphism, random amplified polymorphic DNA and amplified fragment length polymorphism emerged. While these methods offered greater reliability and selective neutrality, they were labour-intensive, time-consuming and displayed limited polymorphism compared to newer approaches (Yang *et al.*, 2013).

The introduction of microsatellites or simple sequence repeats (SSRs) marked a significant advancement. Microsatellites, known for their high polymorphism, reproducibility and ability to provide detailed genetic information, were widely used for over two decades. However, their application became increasingly limited due to high costs, labour-intensive processes and challenges in designing markers (Sheriff and Alemayehu, 2018; Yang *et al.*, 2013).

The development of advanced markers, such as mitochondrial DNA (mtDNA), copy number variations (CNVs) and single-nucleotide polymorphisms (SNPs), further revolutionized genetic diversity studies. mtDNA became instrumental in evolutionary and domestication studies, while CNVs were applied to analyze structural genomic variations. CNVs, alongside SNPs, have jointly contributed to understanding genetic variation in livestock species (Bhanuprakash *et al*., 2018; Pierce *et al*., 2018).

SNPs have emerged as the most reliable marker for genetic diversity studies due to their abundance, genetic stability and cost-effectiveness. SNP genotyping allows high-throughput automated analyses and has facilitated genome-wide studies, including linkage disequilibrium analysis, genome-wide association studies and QTL identification (Matukumalli *et al*., 2009; Koopaee and Koshkoiyeh, 2014). Technologies like the Illumina BovineSNP50 BeadChip, first introduced in 2008, have enabled the identification of thousands of SNPs, transforming population genetics research (Matukumalli *et al*., 2009).

The genomic revolution has been further accelerated by reduced representation genome sequencing methods, such as ddRADseq and genotyping by sequencing (GBS), which eliminate the need for whole-genome sequencing. These approaches provide cost-effective and efficient tools for assessing genetic diversity and identifying genomic regions of interest (Peterson *et al*., 2012; Van Tassell *et al*., 2008).

2.4.2 Parameters of genetic diversity estimation

Genetic diversity can be quantified at two levels: within populations (intrapopulation) and between populations (interpopulation). Intrapopulation diversity is measured using parameters such as effective population size (Ne), linkage disequilibrium (LD), nucleotide diversity (ð), genomic inbreeding coefficients, allelic richness, minor allele frequency (MAF), average expected heterozygosity and the proportion of polymorphic loci. Metrics like the extent of inbreeding (FIS) and the average genetic distance (D) are also utilized to assess genetic relationships among individuals in a population.

For interpopulation diversity, metrics include Analysis of Molecular Variance (AMOVA), Wright's F-statistics (e.g., the F_{ST} index) and measures of genetic differentiation across single

or multiple loci. These tools collectively provide insights into the genetic variation and relationships both within and among populations.

2.4.2.1 Heterozygosity, MAF and Average Pairwise Genetic Distance (D)

Heterozygosity, the occurrence of different alleles at a genetic locus within an individual, is a key indicator of genetic variability in populations (Brito *et al*., 2017). It can be assessed in two ways: observed heterozygosity (Ho), which represents the actual proportion of heterozygotes at a locus and expected heterozygosity (He), which is derived from allele frequencies under Hardy-Weinberg equilibrium. The comparison between observed and expected heterozygosity allows for the identification of population substructures and deviations from equilibrium, which may signal evolutionary processes like selection or genetic drift (Chagunda *et al*., 2018). Populations with high heterozygosity levels are considered genetically diverse, leading to greater adaptability, while populations with low heterozygosity may face reduced fitness and increased risk of inbreeding (Brito *et al*., 2017). Recent studies on Indian cattle highlight this variation in heterozygosity. Using ddRAD sequencing, Masharing *et al*. (2023) reported Ho values ranging from 0.464 to 0.551 and He values from 0.448 to 0.535 across six indigenous breeds-Tharparkar, Rathi, Red Sindhi, Sahiwal, Gir and Kankrej. In Tharparkar cattle, analysis with the Illumina Bovine SNP50 array revealed mean Ho and He values of 0.339 ± 0.156 and 0.325 ± 0.129, respectively (Saravanan *et al*., 2020). Similarly, a study on Vrindavani crossbred cattle using SNP50K data estimated average homozygosity of 0.67 ± 0.02 (observed) and 0.68 ± 0.001 (expected), indicating comparatively lower heterozygosity (Chhotaray *et al*., 2019).

The Minor Allele Frequency (MAF) is a fundamental metric in population genetics, used to identify the frequency of the least common allele within a population (Dash *et al*., 2018). MAF offers insight into genetic diversity, helping distinguish between the common and rare alleles. High MAF suggests genetic stability or lack of population subdivision, whereas low MAF can indicate genetic bottlenecks, mutations, or geographic isolation (Marth *et al*., 2004). Filtering SNPs based on MAF ensures reliable genetic analyses, as variants with very low frequencies often lack statistical power (Linck & Battey, 2019). Estimating MAF can be

done using tools like PLINK (Purcell *et al.*, 2007), which allows the calculation of MAF for each autosome and provides insights into population structure. In Indian zebu cattle, the minor allele frequency (MAF) has been assessed using different genotyping platforms, yielding varying estimates. Analysis with the Illumina Bovine HD SNP chip reported average MAF values of 0.248 ± 0.006 in Sahiwal, 0.241 ± 0.007 in Tharparkar and 0.242 ± 0.009 in Gir cattle (Dash *et al.*, 2018). In contrast, comparatively lower MAFs of 0.193, 0.197 and 0.191 were observed in the same breeds when genotyped with the Illumina Bovine SNP50K BeadChip (Ahmad *et al.*, 2018). More recently, Rajawat *et al.* (2024) employed a ddRAD sequencing approach in Sahiwal cattle and estimated an average MAF of 0.221 ± 0.119.

Average pairwise genetic distance (D) quantifies the genetic differences between individuals in a population. This parameter is calculated by determining the proportion of alleles shared between pairs of individuals. Larger genetic distances indicate greater genetic differentiation, suggesting isolated or divergent populations. The calculation of D is facilitated by software like PLINK (Purcell *et al.*, 2007) and it is particularly useful for examining evolutionary patterns, genetic relationships and the overall genetic structure within populations.

2.4.2.2 Effective Population Size (N_e)

The concept of Effective Population Size refers to the number of breeding individuals in an idealized population that would exhibit the same level of genetic drift and inbreeding as observed in the actual population (Falconer & Mackay, 1996). Ne is crucial as it influences the rate at which genetic variation is lost, the fixation of harmful alleles and the occurrence of inbreeding (Wright, 1969). A smaller Ne leads to a faster loss of genetic diversity, which can reduce the adaptability of a population and increase the risk of inbreeding depression (Gasca-Pineda *et al.*, 2013).

The estimation of effective population size can be conducted using three main methodological approaches: demographic, pedigree-based and marker-based methods (Flury *et al.*, 2010). In the absence of pedigree information, linkage disequilibrium (LD)-based methods are commonly employed to estimate Ne (Sved, 1971; Hill, 1981). The LD method relies on SNP marker data to estimate Ne by assessing the extent of allele frequency correlations in a population (Barbato *et al.*, 2015).

Genomic studies consistently reveal a declining trend in effective population size (Ne) across Indian cattle breeds. Using the BovineHD 777K SNP BeadChip, Strucken *et al*. (2021) reported high historical Ne values (>3000 in most breeds) that have drastically reduced in recent generations, with current estimates as low as 1 in Haryana and 6 in Red Sindhi and moderate values in major breeds such as Gir (197), Sahiwal (68) and Tharparkar (21). Similar reductions were observed in other studies: Chhotaray *et al*. (2021) estimated Ne at 40 in Vrindavani crossbreds using SNP50K data, while Dash *et al*. (2023) reported an Ne of 70 for Sahiwal, Tharparkar and Gir about 13 generations ago. Bhradwaj *et al*. (2024) found an Ne of 52 in Karan Fries and Saravanan *et al*. (2022) documented a decline in Tharparkar cattle from 581 (150 generations ago) to 64 (10 generations ago), depending on MAF thresholds. Consistent with these findings, Rajawat *et al*. (2024) also reported a recent Ne of 74 in Sahiwal cattle using ddRAD sequencing. Collectively, these studies highlight the ongoing reduction in Ne across Indian cattle populations, underscoring concerns about genetic erosion and the need for conservation strategies.

2.4.2.3 Inbreeding Coefficient

The inbreeding coefficient (F) represents the probability that two alleles at a locus are identical by descent (IBD), thereby indicating the degree of genetic relatedness within a population (Wright, 1922). Inbreeding occurs when related individuals mate, leading to increased homozygosity, reduced heterozygosity and negative consequences such as inbreeding depression, reduced fitness and expression of recessive disorders in livestock (Hedrick & Garcia-Dorado, 2016; Leroy, 2014; Doekes *et al*., 2019; Gutierrez-Reinoso *et al*., 2022). Traditionally, F has been estimated from pedigree records, but this approach can be limited by incomplete or inaccurate genealogies. The advent of genomic tools has made it possible to estimate F more accurately, capturing genome-wide variation and overcoming pedigree-based limitations (VanRaden, 2008).

Different genomic estimators of F are based on distinct definitions, such as the correlation between gametes (Wright, 1922), the probability of alleles being IBD (Malecot, 1948), or changes in variability relative to a base population (Villanueva *et al*., 2021). Commonly used

methods include SNP-based measures such as F_{SNP} (Bjelland *et al.*, 2013), F_{HOM} (Wright, 1948), F_{UNI} (Yang *et al.*, 2010) and F_{GRM} (VanRaden, 2008), along with segment-based approaches like runs of homozygosity (F_{ROH}; McQuillan *et al.*, 2008). SNP-based methods analyze loci individually, while ROH-based measures detect extended homozygous tracts, often considered more informative for identifying recent inbreeding (Zhang *et al.*, 2015; Kardos *et al.*, 2015; Dadousis *et al.*, 2022). Despite advances, there is no consensus on the most accurate estimator for capturing genetic variability and inbreeding depression. Yengo *et al.* (2017) suggested weighting rare alleles more heavily in SNP-based approaches, whereas Nietlisbach *et al.* (2019) argued that ROH-based measures provide greater accuracy. Caballero *et al.* (2020) further emphasized that the reliability of F estimates depends strongly on the population under study. Overall, genomic estimators offer greater resolution and reliability than pedigree-based approaches and are therefore essential for maintaining genetic diversity and minimizing inbreeding in livestock populations.

Alemu *et al.* (2021) evaluated different genomic estimators of inbreeding in Dutch Holstein cattle using whole-genome and SNP50K data. They observed that methods such as F_{UNI} and F_{GRM}, which emphasize rare alleles, were highly correlated with homozygosity at rare variants, while estimators like F_{HOM}, F_{ROH} and F_{PED}, which treat all alleles equally, more accurately reflected overall genome-wide homozygosity. The choice of estimator was shown to depend on research goals: rare allele–sensitive measures are useful for conservation and demographic studies, whereas homozygosity-based measures are more suitable for assessing mating patterns and managing genetic diversity in small populations. Among these, F_{ROH} generally outperformed F_{HOM} and F_{PED}, consistent with earlier findings (Keller *et al.*, 2011).

Breed-specific studies further illustrate this variability. In Dutch Holsteins, Alemu *et al.* (2021) reported F_{ROH}, F_{HOM}, F_{GRM} and F_{UNI} values of 0.066, –0.022, –0.022 and –0.020, respectively. Rajawat *et al.* (2024) estimated corresponding values of 0.0249, 0.289, 0.0725 and 0.0403 in Indian Sahiwal cattle using ddRAD data. Saravanan *et al.* (2022) found F_{ROH}, F_{HOM}, F_{GRM} and F_{UNI} values of 0.0589, 0.0215, 0.0532 and 0.0160 in Tharparkar cattle, while Illa *et al.* (2024) reported values of 0.009 (F_{PED}), 0.091 (F_{ROH}), 0.035 (F_{HOM}) and –0.104 (F_{GRM}) for indigenous Sahiwal cattle. These results demonstrate that both the choice of

estimator and the genetic background of the population influence the magnitude and interpretation of genomic inbreeding coefficients.

2.4.2.4 Runs of Homozygosity (ROH)

Runs of Homozygosity (ROH) are continuous homozygous segments of the genome that arise when individuals inherit identical haplotypes from related ancestors. Their length distribution reflects the number of recombination events since a common ancestor and thus provides insights into demographic history and inbreeding. The proportion of the genome covered by ROH, expressed as F_{ROH}, has become a widely used genomic measure of inbreeding in both humans and domestic animals (McQuillan *et al*., 2008; Kardos *et al*., 2016).

ROH detection methods are broadly divided into rule-based approaches, such as those implemented in PLINK (Purcell *et al*., 2007), which rely on user-defined thresholds for consecutive homozygous SNPs and model-based approaches, such as hidden Markov models, which use probabilistic frameworks to infer identity by descent (Bertrand *et al*., 2019; Renaud *et al*., 2019). Accurate detection requires sufficiently dense and well-mapped markers, with whole-genome sequencing (WGS) generally providing the resolution necessary to capture both short and long ROH (Lavanchy & Goudet, 2023). However, detection outcomes are sensitive to SNP density, sequencing depth and parameter choice, which can influence estimates of F_{ROH}.

Beyond measuring inbreeding, ROH are increasingly applied in mapping deleterious recessive alleles and reconstructing demographic events. For instance, longer ROH often indicate recent inbreeding or population bottlenecks, whereas shorter ROH are characteristic of historically larger effective population sizes (Ceballos *et al*., 2018; Thompson, 2013). This makes ROH a powerful tool in conservation genetics, livestock breeding and studies of wild populations where understanding genomic health and demographic history is crucial.

A comparative assessment of runs of homozygosity (ROH) in two groups of Holstein-Friesian cattle-one subjected to over four decades of selection for milk yield and the other maintained as a control without targeted selection-showed broadly similar ROH distributions within the 1.5–5.0 Mb range. Notably, ROH fragments exceeding 10 Mb were observed only

in the selection group, reflecting the impact of intensive breeding practices (Kim *et al*., 2013). In Retinta bulls, contrasting ROH patterns were reported between highly inbred and outbred animals, with the former group displaying a marked increase in both the number and length of ROH, particularly segments larger than 8 Mb (Goszczynski *et al*., 2018). For Gyr dairy cattle in Brazil, analysis of 582 individuals identified an average of 55.12 ± 10.37 ROH tracts with a mean size of 3.17 Mb. Short ROH (1–2 Mb) were predominant, accounting for 60% of all detected segments and approximately 7.01% of the genome (175.28 Mb) was estimated to be autozygous (Peripolli *et al*., 2018). Similarly, Dash (2016) investigated Indian dairy breeds and reported 1,762, 1,138 and 1,528 ROH regions in Gir, Sahiwal and Tharparkar, respectively. Among the studied breeds, Kangayam showed the highest ROH-based inbreeding (F_{ROH} > 1 Mb = 0.113 ± 0.059), whereas Hariana (F_{ROH} > 1 Mb = 0.042 ± 0.031) and Sahiwal (F_{ROH} > 1 Mb = 0.043 ± 0.048) exhibited comparatively lower levels.

2.5 Identification of novel breed specific informative SNPs

Breed identification plays a significant role in various biological applications. It typically involves two key steps: identifying breed-specific SNPs and assigning breeds based on these markers (Zhao *et al*., 2024). Next-generation sequencing (NGS) has revolutionized genome-wide SNP discovery and genotyping, offering advanced, cost-effective methods like Restriction Site-Associated DNA Sequencing (RAD-seq) and its refined variant, double-digest RAD-seq (ddRAD-seq). These methods enable efficient polymorphism genotyping and discovery by targeting specific genomic regions with high precision. ddRAD-seq, developed by Peterson *et al*., 2012, enhances accuracy through a two-enzyme digestion process, surpassing traditional GBS in complex genomes. While whole-genome sequencing (WGS) is comprehensive, its high-cost limits accessibility. In contrast, ddRAD-seq provides a practical alternative, covering over 40% of the genome, making it both efficient and affordable (Shirasawa *et al*., 2013).

Given their ability to target specific genomic regions and efficiently identify polymorphisms, ddRAD-seq and RAD-seq are particularly well-suited for discovering novel, breed-specific SNPs. These methods allows to focus on genetic variations that distinguish breeds, supporting breed identification, population studies and the development of breed-

specific genomic tools essential for livestock improvement (Mehravi *et al.*, 2021). Wang *et al.* (2018) conducted the first study in China to establish a reference set of SNPs which were specific to indigenous cattle in Sichuan. This research utilized the Restriction Site-Associated DNA Sequencing (RAD-seq) method to identify genome-wide SNPs across six Sichuan cattle breeds. Another study developed a cost-effective breed-specific SNP panel for breed identification using SNP data from 178 animals across 10 breeds from SNPs detected via Bovine 50K SNP beadchip using 5 methods - Delta, Pairwise Wright's F_{ST}, informativeness for assignment, frequent item feature selection (FIFS) and minor allele frequency–linkage disequilibrium (MAF-LD), resulting in 591 annotated breed-specific SNPs (Kumar *et al.*, 2021a).

Several studies have demonstrated that only a limited number of carefully selected SNPs are sufficient to discriminate between cattle breeds and verify their genetic origin. For example, Strucken *et al.* (2017) and Wilkinson *et al.* (2011) found that reduced SNP panels can still reliably identify an animal's breed. In India, Kumar *et al.* (2019) proved that 470 SNPs were enough to separate six indigenous and European dairy breeds, while Hulsegge *et al.* (2019) used PCA and random forest to find only 133 SNPs that could clearly distinguish Dutch local breeds. Likewise, Bertolini *et al.* (2018) designed a 96-SNP panel to differentiate cosmopolitan and Italian cattle and Fisher *et al.* (2009) showed that even just 40 SNPs could work for parentage testing when combined with mating and birth records. The ISAG cattle panel also supports this idea-it started with 100 SNPs for parentage checks and later increased to 200 for greater accuracy across different breeds (Morrin and Boscher, 2012). Different methods like delta, Wright's F_{ST}, Weir & Cockerham's F_{ST} and PCA have been used to pick the most informative SNPs. Wilkinson *et al.* (2011) showed that depending on how closely related the breeds are, 60–140 SNPs could already achieve more than 95% success and going beyond 200 SNPs didn't add much value.

2.5.1 Toolbox for Ranking and Evaluation of SNPs (TRES)

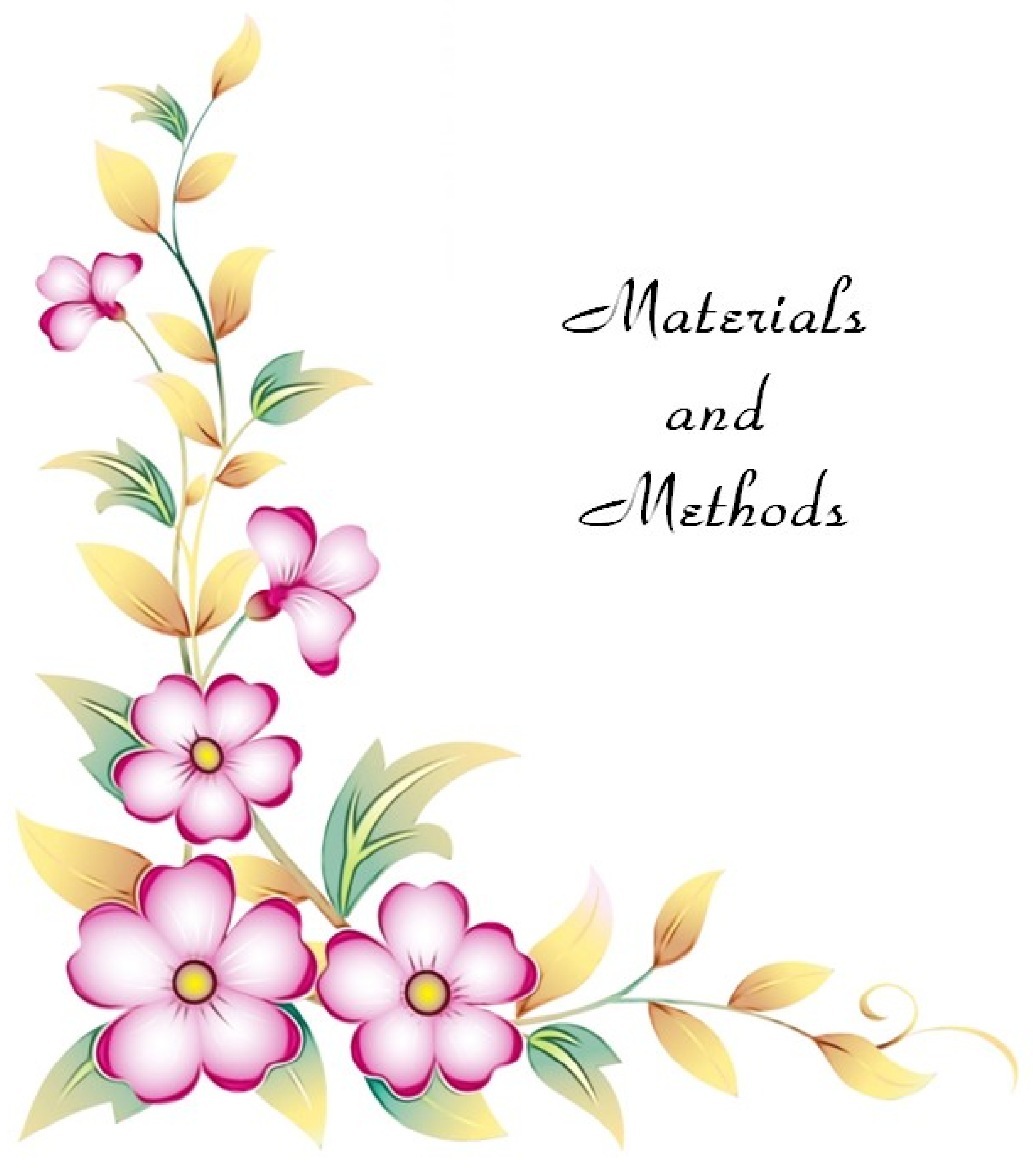
TRES software (Kavakiotis *et al.*, 2015) provides a built-in collection of algorithms to rank SNPs based on their informativeness, specifically designed to calculate scores for

SNP subsets using three established methods: Delta, Pairwise Wright's F_{ST} and Informativeness for Assignment. (1) Delta Method, widely used approach for marker informativeness, originally applied in human population studies. For biallelic markers, Delta is calculated as the difference in allele frequencies between two populations ä = $|P^i_A – P^j_A|$, i.e., difference between frequency of allele A in i^{th} and j^{th} population. In cases involving multiple populations, Delta values are calculated for every pair of populations and their average is used to assign a value to each SNP. (2) Pairwise Wright's F_{ST} method measures genetic differentiation between populations. It is calculated for biallelic markers using the formula $F_{ST} = (H_T - H_S)/H_T$, where H_T represents the total population's expected heterozygosity and H_S is the average heterozygosity within subpopulations. (3) Informativeness for Assignment measure (Rosenberg *et al.*, 2003) calculates SNP scores using a mutual information-based algorithm that incorporates ancestry reports. The informativeness (In) is determined by summing the probabilities of allele occurrences in each population and averaging them across all populations.

2.5.3 Minor Allele Frequency – Linkage Disequilibrium (MAF-LD) method

The MAF-LD method integrates MAF and LD to filter and refine SNPs by focusing on SNPs with intermediate allele frequencies that exhibit variability across breeds, while LD pruning ensures these SNPs are independent and non-redundant, resulting in markers that are both specific and highly informative for a given breed (Kumar *et al.*, 2021a).

✍ ✍ ✍



Materials and Methods

3 MATERIALS AND METHODS

3.1 Animals, Sampling and DNA Extraction

For the present investigation, a total of 48 Tharparkar cattle were randomly selected for sampling from the herd maintained at the Livestock Research Station (LRS), Beechwal, operating under Rajasthan University of Veterinary and Animal Sciences (RAJUVAS), Bikaner, Rajasthan. The LRS herd, established in the year 2000, occupies an area of 233 hectares and was systematically developed by procuring foundation stock from diverse locations within the breed's native tract in western Rajasthan, thereby ensuring a broad genetic base for conservation and improvement. Geographically, the farm is situated at 28.09°N latitude and 73.36°E longitude, with an altitude of 224 m above mean sea level and functions as a key cooperating unit of the Tharparkar cattle breeding project implemented under the state genetic improvement programme (https://rajuvas.org/livestock-research-station/).

From each animal, 5 mL of whole blood was collected aseptically from the jugular vein using sterile needles and transferred into heparin-coated vacutainer tubes to prevent coagulation. All procedures involving animals were carried out in accordance with the ethical standards and guidelines approved by the Institutional Animal Ethics Committee (IAEC). Genomic DNA was subsequently extracted from the blood samples using the QIAamp® DNA Blood Kit (QIAGEN N.V., Hilden, Germany), following the manufacturer's standard protocol to ensure high-quality DNA suitable for downstream applications. The concentration and purity of extracted DNA were quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA), while DNA integrity was cross-verified by electrophoresis on 0.8% agarose gel, ensuring the absence of major degradation or contamination prior to sequencing.

DNA isolation procedure

For each sample, 20 µl of Proteinase K was combined with 200 µl of whole blood in separate microtubes. To this mixture, 200 µl of AL buffer was added, followed by gentle vortexing for about 15 seconds. The tubes were then placed upright in the incubator and incubated at 56 °C for 10 minutes. After incubation, a brief centrifugation was carried out to collect any liquid condensed on the lids.

Next, 200 µl of absolute ethanol was added to each tube, mixed by vortexing for around 15 seconds and centrifuged again. Labeled minispin columns were prepared and the contents of each tube were carefully transferred to the corresponding column. The columns were then centrifuged at 8000 rpm for 1 minute, after which the flow-through was discarded and the DNA-containing columns were placed into fresh collection tubes.

A wash step was performed by adding 500 µl of AW1 buffer to each column, followed by centrifugation at 8000 rpm for 1 minute. This was repeated with 500 µl of AW2 buffer, with centrifugation at 14,000 rpm for 3 minutes. The filtrate was discarded and the rims of the collection tubes were gently dried. To eliminate traces of ethanol and wash buffer, the columns were centrifuged once more at 14,000 rpm for 1 minute.

For elution, the spin columns were transferred to clean 1.5 ml microtubes and 200 µl of AE buffer was added to each column. After a 5-minute incubation at room temperature, DNA was eluted by centrifugation at 8000 rpm for 1 minute.

The extracted DNA was then evaluated for quality and purity using a Nanodrop spectrophotometer, recording absorbance values at 260 nm and 280 nm. Samples with an OD260/280 ratio between 1.7 and 2.0 and a minimum yield of 30/ ng/µL were considered suitable for downstream applications (Table 1). Additionally, integrity of the DNA was confirmed through electrophoresis on a 0.8% agarose gel, followed by visualization using a gel documentation system (Fig. 1).

Table 1: Quality parameters of genomic DNA isolated from 48 samples, showing concentration, purity (A260/280), eluted volume dispatched and total yield

Sl No.	Sample ID	Concentration (ng/µl)	260/280	Eluted Volume Dispatched (µl)	Total Yield (µg)
1	101	35.2	1.78	100	3520
2	102	35.6	1.92	100	3560
3	103	32.4	1.86	100	3240
4	104	44.6	1.7	70	3122
5	105	42.1	1.87	80	3368
6	106	40.6	1.86	75	3045
7	107	34.8	1.7	100	3480
8	108	33.2	1.71	100	3320
9	109	30.9	1.7	100	3090
10	110	33.2	1.84	100	3320
11	111	36.1	1.7	100	3610
12	112	34.1	1.78	100	3410
13	113	39.7	1.83	80	3176
14	114	54.2	1.7	60	3252
15	116	30.5	1.79	100	3050
16	117	55.2	1.77	60	3312
17	118	33.4	1.89	100	3340
18	119	31.6	1.7	100	3160
19	120	33.6	1.71	100	3360
20	121	47.8	1.75	65	3107
21	122	35.2	1.71	100	3520
22	123	32.6	1.7	100	3260
23	124	42.1	1.71	75	3157.5
24	125	34.2	1.81	100	3420
25	126	36.5	1.8	90	3285
26	127	33.4	1.79	100	3340
27	128	32.7	1.95	100	3270
28	129	46.4	1.76	70	3248
29	130	38.2	1.7	85	3247
30	131	34.5	1.71	100	3450
31	132	32.3	1.85	100	3230
32	133	34.2	1.85	100	3420
33	134	55.6	1.75	55	3058
34	135	33.5	1.86	100	3350

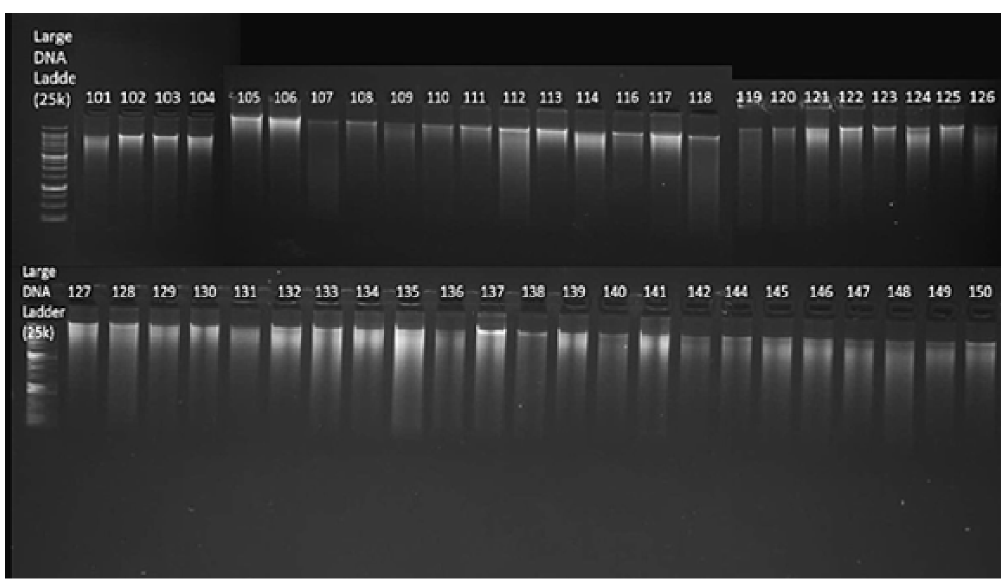


Fig. 1: Gel band image of isolated genomic DNA of 48 Tharparkar cattle

35	136	31.9	1.77	100	3190
36	137	44.4	1.71	70	3108
37	138	32.5	1.7	100	3250
38	139	34.7	1.75	100	3470
39	140	34.5	1.98	100	3450
40	141	45.7	1.73	65	2970.5
41	142	30.9	1.84	50	1545
42	144	31.5	1.75	100	3150
43	145	34.6	1.74	85	2941
44	146	32.1	1.92	100	3210
45	147	34.5	1.72	100	3450
46	148	33.4	1.74	100	3340
47	149	32.3	1.86	100	3230
48	150	34.6	1.87	100	3460

3.2 ddRAD Library Preparation and Sequencing

Genomic DNA samples were outsourced to N2Jenomics Lab Private Limited, New Delhi, for genotyping through the double-digest restriction site-associated DNA (ddRAD) sequencing protocol, a reduced-representation sequencing approach that enables cost-effective SNP discovery and genotyping across large populations (Peterson *et al*., 2012). In this protocol, genomic DNA was digested using two restriction enzymes-MluCI, a frequent cutter generating numerous fragments and SphI, a rare cutter producing larger fragments-to achieve optimal genome coverage and complexity reduction. The digested DNA fragments were then size-selected, ligated with adapters and subjected to library preparation before sequencing on the Illumina HiSeq 2000 platform, which generated high-throughput paired-end reads. For the 48 samples analyzed, the sequencing yielded paired-end FASTQ files with an average read length of 159 base pairs, ensuring sufficient resolution for downstream SNP calling.

The bioinformatic analyses of raw sequencing data were performed in-house to ensure precise data handling and quality control. Analyses were carried out on a Dell Precision 5820 Tower workstation (Dell Technologies, USA) equipped with an Intel® Xeon® W-2265 processor (24 cores, 3.50 GHz), 64 GB RAM and a 64-bit operating system, providing adequate computational power for handling large genomic datasets and executing computationally intensive pipelines.

3.3 Quality Control, Read Processing and Variant Calling

The raw sequencing data file in FASTQ format generated in paired-end format, consisting of forward (R1) and reverse (R2) reads, were first subjected to a comprehensive quality check to assess the overall sequencing performance and data integrity. Quality assessment was carried out using FASTQC v0.12.1 (Andrews *et al.*, 2010), which provided detailed reports on per-base sequence quality, GC content distribution, sequence duplication levels, presence of overrepresented sequences and potential adapter contamination. This preliminary step ensured that downstream analyses were performed only on reads meeting acceptable quality thresholds.

Following quality assessment, reads containing low-quality bases and residual adapter sequences were processed using Trimmomatic v0.39 (Bolger *et al.*, 2014). Trimming was performed to remove bases with low Phred quality scores at the ends of reads and to excise sequencing adapters introduced during library preparation. This step enhanced the accuracy of downstream mapping by reducing errors caused by poor-quality base calls or adapter contamination. The resulting set of high-quality, cleaned reads was retained for alignment.

The filtered paired-end reads were then mapped against the Bos taurus reference genome assembly ARS-UCD 2.0 (Rosen *et al.*, 2020) using the highly efficient alignment software Bowtie2 v2.5.4 (Langmead & Salzberg, 2012). Bowtie2 employs an FM-index based on the Burrows–Wheeler Transform (BWT), which enables rapid and memory-efficient alignment of millions of short reads against a large reference genome. The output of this alignment process was generated in SAM (Sequence Alignment Map) format, a text-based representation of alignments containing essential information such as read mapping coordinates, mapping quality scores and alignment flags.

Since SAM files are large and computationally demanding to process, they were converted into their compressed binary equivalent, BAM files, using the SAMtools v1.21 package (Li *et al.*, 2009). This step not only reduced storage requirements but also improved the efficiency of downstream analyses. The BAM files were subsequently sorted by genomic coordinates to facilitate rapid data retrieval and were indexed to allow efficient access during variant calling and visualization.

Variant discovery was conducted using the bcftools v1.21 package (Li, 2011), which integrates both read-based likelihood calculations and probabilistic variant calling methods. Initially, the command bcftools mpileup was employed to generate genotype likelihoods at each genomic position based on the alignment data. These likelihoods were then processed using bcftools call, which applied Bayesian models to identify variant sites and assign genotypes. This step produced individual Variant Call Format (VCF) files for each of the 48 sequenced samples, containing detailed information on SNPs and other types of genetic variants.

To create a unified dataset suitable for comparative analysis across individuals, the sample-specific VCF files were merged using the bcftools merge function. The resulting consolidated VCF file contained genotype information for all 48 samples in a standardized format. Further, the merged dataset was filtered to retain only SNPs, while excluding insertions, deletions and other complex variants.

Additional quality control filtering steps were applied to ensure high-confidence variant calls. Specifically, only SNPs with a Phred quality score ≥ 30 (corresponding to a base-calling accuracy of 99.9%) and a minimum read depth of 10 were retained. These thresholds minimized the risk of false-positive variant calls due to sequencing errors or insufficient coverage.

3.4 Variant Filtering and Annotation

Following the initial quality control of sequencing data using BCFtools (Danecek *et al.*, 2021), a merged VCF file was generated and carefully filtered to remove non-autosomal contigs, ensuring that only high-confidence autosomal variants were retained for downstream analyses. Each variant was then annotated with chromosome information based on the reference genome assembly available from the National Center for Biotechnology Information (NCBI), providing standardized genomic coordinates. The filtered VCF file was subsequently converted into PLINK-compatible formats (.map and .ped) using PLINK v1.9 (Purcell *et al.*, 2007), which facilitated population genetics and diversity analyses. To ensure robustness, additional quality control filters were implemented: (i) exclusion of SNPs with a minor allele frequency (MAF) < 0.05, reducing the influence of rare alleles with limited statistical power; (ii) removal of SNPs deviating from Hardy–Weinberg equilibrium (HWE) at $p < 0.0001$, which may indicate

genotyping errors, population stratification, or selection; and (iii) elimination of SNPs missing in more than 25% of individuals, minimizing bias introduced by incomplete genotyping. Missing genotypes were then imputed using Beagle v5.4 (Browning *et al.*, 2018), which leverages haplotype information to accurately infer unobserved alleles, thereby producing a more complete and reliable dataset. Finally, the resulting high-confidence, bi-allelic SNP dataset was subjected to structural and functional annotation with SnpEff v5.2e (Cingolani *et al.*, 2012), which classified variants into genomic regions (e.g., exonic, intronic, intergenic) and predicted their potential functional consequences, such as synonymous, missense, or loss-of-function mutations.

3.5 Retrieval of Publicly Available Genomic Data

Genotypic data for other indigenous cattle breeds - Gir, Sahiwal, Red Sindhi and Rathi - were obtained from our laboratory's internal repository. These datasets were incorporated for breed-specific SNP mining and validation analyses.

3.6 Genetic Diversity Analyses

Heterozygosity

Heterozygosity, defined as the presence of different alleles at a genetic locus within an individual, serves as an important indicator of genetic variability within populations (Brito *et al.*, 2017). It provides valuable insights into the level of genetic variation maintained within a population and serves as an indicator of evolutionary potential and adaptive capacity. In this study, both observed heterozygosity (Ho) and expected heterozygosity (He) were estimated to evaluate genetic variability. These parameters were computed using the **-**hardy function implemented in PLINK v1.9 (Purcell *et al.*, 2007), which generates allele frequency-based estimates consistent with Hardy–Weinberg equilibrium assumptions. Observed heterozygosity (Ho) reflects the proportion of individuals in the population that are heterozygous at a given locus, whereas expected heterozygosity (He) represents the probability of heterozygosity predicted under Hardy–Weinberg equilibrium, based on allele frequencies. A comparison of Ho and He values provides a measure of the extent to which the population conforms to equilibrium expectations, thereby offering insights into processes such as inbreeding, selection, or genetic drift.

Average Pairwise Genetic Distance (D)

Average pairwise genetic distance (D) quantifies the genetic divergence among individuals within a population by estimating the proportion of alleles shared between each pair. Higher values of D indicate greater genetic differentiation, which may reflect isolated or divergent populations. The average proportion of shared alleles (D_{ST}) was calculated using the formula:

$$D_{ST} = \frac{IBS2 + (0.5*IBS1)}{N}$$

where IBS1 and IBS2 denote the number of loci sharing one and two alleles identical by state, respectively and N represents the total number of loci analyzed. The average pairwise genetic distance was then derived as 1 - D_{ST} for all pairwise combinations of individuals using PLINK v1.9 (Purcell *et al.*, 2007).

Minor Allele Frequency (MAF)

The minor allele frequency (MAF) refers to the frequency of the less common allele at a genetic locus within a population (Dash *et al.*, 2018). It provides valuable insight into genetic diversity and helps differentiate between common and rare alleles. In this study, the distribution of MAF across all autosomal SNPs was estimated using the –freq function implemented in PLINK v1.9 (Purcell *et al.*, 2007). For downstream analysis, MAF values were categorized into six different classes: 0–0.05, 0.05–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4 and 0.4–0.5. SNPs with a MAF greater than 0.1 were considered polymorphic and included in subsequent diversity analyses, while loci with lower frequencies were regarded as rare variants and excluded to minimize potential biases.

Runs of Homozygosity (ROH)

Runs of homozygosity (ROH), which represent uninterrupted stretches of homozygous genotypes in the genome, were estimated using the detectRUNS package (Biscarini *et al.*, 2019) implemented in R software (R Core Team, 2022). ROH detection was carried out under a sliding window approach of 20 SNPs with a homozygosity threshold of 0.05. To

ensure reliable identification, a minimum of 20 consecutive SNPs was required to define a ROH, while allowing for the presence of one heterozygous and one missing SNP within each window, along with a maximum inter-marker distance of 1 Mb. Stringent quality parameters were also applied, including a minimum ROH length of 1 Mb, a SNP density of at least 1 marker per 50 kb and allowance of no more than one heterozygous and two missing SNPs within an entire ROH (Xu *et al*., 2019). The detected ROHs were subsequently categorized into five length classes: 0–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb and >16 Mb, enabling the assessment of both short and long homozygous tracts, which provide insight into ancient versus recent inbreeding events.

Genomic Inbreeding Coefficient estimation

The inbreeding coefficient (F) measures the degree of genetic relatedness within a population and quantifies the probability that two alleles at a locus are identical by descent (Wright, 1922). Genomic inbreeding was estimated using four complementary approaches: 1. F_{UNI} - based on the correlation between uniting gametes (Yang *et al*., 2011). 2. F_{HOM} - based on the excess of homozygosity in the population (Li & Horvitz, 1953). 3. F_{ROH} - based on the proportion and length of runs of homozygosity (McQuillan *et al*., 2008). 4. F_{GRM} - based on the genomic relationship matrix (Yang *et al*., 2011).

The F_{UNI}, F_{HOM} and F_{GRM} coefficients were calculated using PLINK v1.9 (Purcell *et al*., 2007), while F_{ROH} was estimated using the detectRUNS package (Biscarini *et al*., 2019) in R software (R Core Team, 2022).

Effective Population Size (Ne) estimation

Effective population size (Ne) refers to the number of breeding individuals in an idealized population that would experience the same level of genetic drift and inbreeding as the actual population (Falconer & Mackay, 1996; Wright, 1969). Contemporary Ne was estimated using NeEstimator v2.01 (Do *et al*., 2014) under a random mating model with default parameters and multiple MAF thresholds. Historical trends in Ne were inferred using the SNeP package (Barbato *et al*., 2015), which applies linkage disequilibrium-based methods to estimate Ne across past generations by analyzing correlations in allele frequencies (Sved, 1971; Hill, 1981).

Nucleotide Diversity (π)

Nucleotide diversity (π) is a measure of genetic variation within a population, calculated as the average pairwise nucleotide differences among sampled individuals (Nei & Li, 1979). In this study, nucleotide diversity was estimated using the Pixy software package (Korunes & Samuk, 2021) with 10 kb non-overlapping windows across the genome and the mean π value for the population was subsequently determined.

3.7 Identification of Breed-Specific SNPs Using TRES

3.7.1 Preparation of Reference and Validation Datasets

The ddRAD-based genotypic data were used of five distinct cattle breeds-Tharparkar, Sahiwal, Red Sindhi, Gir and Rathi and was randomly partitioned into two subsets: a reference dataset and a validation dataset. This stratified partitioning was carried out using the Toolbox for Ranking and Evaluation of SNPs (TRES), a dedicated software package designed for SNP selection and performance evaluation (Kavakiotis *et al.*, 2015). To maintain balanced representation and sufficient statistical power, approximately 70% of the individuals from each breed were allocated to the reference dataset, while the remaining 30% were assigned to the validation dataset. The reference dataset served as the primary resource for breed-specific SNP discovery, where multiple complementary approaches were employed, including the TRES algorithm, minor allele frequency–linkage disequilibrium (MAF-LD) pruning (Kumar *et al.*, 2019) and breed allocation analyses. These methodologies ensured a robust and unbiased identification of informative SNPs that could reliably distinguish among the studied breeds. In contrast, the validation dataset functioned as an independent test set to evaluate and confirm the reliability, reproducibility and discriminatory power of the candidate breed-specific SNPs derived from the reference dataset. The numerical distribution of animals within both datasets across the five breeds is summarized in Table 1.

3.7.2 Toolbox for Ranking and Evaluation of SNPs (TRES)

The Toolbox for Ranking and Evaluation of SNPs (TRES) software (Kavakiotis *et al.*, 2015) provides a suite of algorithms for ranking SNPs based on their informativeness. It is specifically designed to score SNP subsets using three established metrics: Delta, pairwise Wright's F_{ST} and Informativeness for Assignment (In).

Delta Method

The Delta (δ) statistic quantifies the difference in allele frequencies between two populations and is calculated as:

$$\delta = |P^i_A - P^j_A|$$

where P^i_A and P^j_A are the frequencies of allele A in the ith and j^{th} populations, respectively. For analyses involving multiple populations, Delta values are computed for every pairwise comparison and the mean value across all comparisons is assigned to each SNP. In this study, pairwise Delta estimates were first computed separately for all five breeds and subsequently averaged to determine the final Delta value for each SNP.

Pairwise Wright's F_{ST} Method

The fixation index (F_{ST}) measures genetic differentiation among populations and is widely applied to biallelic markers. It is defined as:

$$F_{ST} = (H_T - H_S)/H_T$$

where H_T represents the expected heterozygosity of the total population and H_S denotes the average heterozygosity within subpopulations (Weir & Cockerham, 1984). In this study, pairwise F_{ST} values were computed for each SNP by comparing Tharparkar cattle against each of the other four indigenous breeds. The same procedure was repeated for the remaining breeds to assess breed-specific differentiation.

Informativeness for Assignment measure

The Informativeness for Assignment (In) statistic (Rosenberg *et al.*, 2003) estimates the discriminatory power of SNPs based on mutual information principles. It accounts for the probability of allele occurrences across populations and averages these probabilities to determine SNP informativeness. SNPs with higher In values are more informative for distinguishing between breeds and were prioritized in subsequent analyses.

The top-ranked SNPs identified by all three methods were intersected using VENNY 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/) to obtain the common SNPs for each breed.

3.7.3 Minor Allele Frequency - Linkage Disequilibrium (MAF-LD) Method

The MAF-LD method, as described by Kumar *et al*. (2019), leverages SNPs with high minor allele frequencies (MAF > 0.36) to maximize variance and effectively capture breed-specific differences. To further minimize redundancy, LD pruning is applied to remove SNPs in high linkage disequilibrium, thereby retaining only independent and informative markers. In this study, SNP datasets obtained from the pre-selection strategies (Delta, F_{ST} and Informativeness for Assignment) were initially filtered in PLINK v1.9 (Purcell *et al*., 2007) using the -maf (minor allele frequency), -hwe (Hardy-Weinberg equilibrium), -geno (genotyping rate) and -indep-pairwise (LD pruning) functions. This filtering was performed separately for each breed, after which the datasets from all five breeds were merged. A final LD pruning step was then applied to the combined dataset to produce a non-redundant SNP panel containing breed-specific markers across all five breeds (Kumar *et al*., 2019).

3.7.4 STRUCTURE analysis for breed assignment

STRUCTURE v2.3.4 software (Pritchard *et al*., 2000) was employed to estimate the probability of breed assignment for individuals using the final set of breed-specific SNPs. STRUCTURE operates on a Bayesian framework and employs a Markov Chain Monte Carlo (MCMC) algorithm for clustering genotypes into populations based on allele frequencies. The analysis was performed using a burn-in period of 50,000 iterations followed by 50,000 MCMC repetitions under a no-admixture model, as no crossbred individuals were included in the study. To determine the most appropriate number of clusters (K), the analysis was conducted for K values ranging from 3 to 6 and the optimal K was identified using STRUCTURE Harvester following the method described by Evanno *et al*. (2005), based on the lowest cross-validation error.

3.7.5 Annotation of breed-specific SNPs in Tharparkar

Breed-specific SNPs identified in Tharparkar cattle were systematically annotated to uncover the functional significance of the genomic regions harboring these variants and to establish potential associations with economically important traits. Gene annotations corresponding to the Bos taurus ARS-UCD 2.0 genome assembly were obtained through the

Ensembl BioMart tool (https://asia.ensembl.org/biomart/martview/), which served as a reliable resource for retrieving high-quality gene feature data. The genomic positions of Tharparkar-specific SNPs were first converted into BED file format, ensuring compatibility with downstream analyses. These BED files were then intersected with gene coordinate files, also downloaded from BioMart under the same genome build, using the bedtools intersect function (Quinlan & Hall, 2010). This step enabled precise mapping of breed-specific SNPs to nearby or overlapping genes, thereby linking raw variant data to biologically meaningful genomic elements. To further explore the biological consequences of these SNPs, the Variant Effect Predictor (VEP) pipeline was employed (McLaren *et al.*, 2016). VEP classified SNPs into genic categories, such as exonic, intronic and UTR variants, as well as intergenic variants, thereby providing insights into whether a SNP could directly influence coding sequences, gene regulation, or occur in non-coding but potentially regulatory regions. Such categorization is critical for distinguishing between neutral polymorphisms and variants with functional or adaptive relevance. In addition, to connect breed-specific SNPs to known phenotypic outcomes, the annotated genes were cross-referenced against the Cattle QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/BT/search), a curated database that catalogs trait-associated QTLs across bovine populations (Hu *et al.*, 2013). By identifying overlaps between SNP-harboring genes and QTL regions, this integrative strategy allowed the linking of Tharparkar-specific variants to traits such as milk production, fertility, disease resistance and adaptability.

✍✍✍

Results

4 RESULTS

4.1 Dataset Characteristics and SNP Annotation

A total of 119.18 million raw paired-end reads were generated from the 48 ddRAD genotypes, corresponding to approximately 18.95 GB of sequencing data (Table 2). Following the initial quality control, which involved removal of adaptor sequences and low-quality bases, the majority of reads (99.43%) were retained, yielding 119.18 million clean reads (Table 3). However, one sample was excluded as it exhibited a very low genotyping rate. These high-quality reads demonstrated a mapping efficiency of 97.03% to the Bos taurus ARS-UCD 2.0 reference genome. Variant calling from the merged dataset, applying stringent thresholds of Phred quality score > 30 and minimum read depth $\geq$ 10, initially identified 2,264,935 variants. However, when considering the complete variant discovery pipeline prior to progressive filtering, a total of 9,063,145 variants were recovered. Systematic quality control was subsequently performed using bcftools v1.21 (Li, 2011) and PLINK v1.9 (Purcell *et al.*, 2007), applying multiple filtering layers to retain only robust and biologically meaningful variants. Of the recovered variants, 2,264,935 passed the primary thresholds for quality score and depth. Restricting the dataset to SNPs only reduced this number to 1,781,143, while the removal of variants associated with mitochondrial DNA and sex chromosomes eliminated a further 50,148 markers, leaving only autosomal SNPs for analysis. Subsequent filtering to exclude rare alleles with minor allele frequency (MAF $\leq$ 0.05) removed 175,789 SNPs and the exclusion of markers with high levels of missing genotypes (retaining only those with genotyping rate > 90%) further eliminated 1,338,702 SNPs. In addition, SNPs exhibiting significant deviations from Hardy–Weinberg equilibrium ($p < 0.0001$) were discarded (94,950 SNPs removed). After these successive

quality-control steps, a final dataset consisting of 121,554 autosomal biallelic SNPs was obtained (Fig. 2a), with an overall genotyping rate (GR) of 94.316% were retained for further downstream analyses. Chromosome-wise distribution of variants is shown in Fig. 3.

In terms of genome coverage, the total length of the bovine genome was estimated at 2,770,686,120 bp, while the effective genome length after filtering was 2,489,385,779 bp, corresponding to a variant density of one SNP per 20,479 bases. Functional annotation revealed that the majority of variants were located within intronic and intergenic regions, consistent with previous observations in cattle genomes where non-coding regions accumulate higher levels of polymorphism (Fig. 2b).

The transition-to-transversion (Ts/Tv) ratio was estimated at 2.47 , reflecting the predominance of transitions over transversions in cattle genomic variation. Specifically, the dataset comprised 654,741 transitions and 264,136 transversions, with A/T and T/A substitutions occurring at relatively lower frequencies compared to other substitution types (Fig. 2c).

Table 2. Raw sequencing data statistics of paired-end FASTQ files generated from ddRAD sequencing of 48 Tharparkar cattle

Sample	Read Orientation	Bases (bps)	Reads	Q20%	Q30%	Read Length	GC%
101	R1	83,55,52,791	52,55,049	98.15	95.24	159	52.63
101	R2	83,55,52,791	52,55,049	98.24	95.12	159	52.94
102	R1	75,81,14,703	47,68,017	96.03	91.68	159	54.24
102	R2	75,81,14,703	47,68,017	97.42	93.59	159	55.97
103	R1	25,06,24,863	15,76,257	97.16	93.53	159	51.11
103	R2	25,06,24,863	15,76,257	97.57	93.96	159	52.75
104	R1	50,59,19,238	31,81,882	97.92	94.85	159	49.64
104	R2	50,59,19,238	31,81,882	98.06	94.74	159	50.68
105	R1	49,52,33,643	31,14,677	97.94	94.96	159	53.41
105	R2	49,52,33,643	31,14,677	98.11	94.86	159	54.3
106	R1	68,52,66,786	43,09,854	97.96	95	159	52.55
106	R2	68,52,66,786	43,09,854	97.88	94.33	159	53.35
107	R1	50,81,96,118	31,96,202	95.69	90.93	159	58.88

107	R2	50,81,96,118	31,96,202	97.04	93.13	159	51.37
108	R1	55,96,93,674	35,20,086	96.4	92.15	159	54.47
108	R2	55,96,93,674	35,20,086	97.36	93.79	159	56.72
109	R1	62,44,86,015	39,27,585	98.49	95.9	159	51.85
109	R2	62,44,86,015	39,27,585	98.5	95.71	159	51.74
110	R1	39,47,28,948	24,82,572	98.46	95.81	159	51.57
110	R2	39,47,28,948	24,82,572	98.52	95.71	159	51.54
111	R1	1,74,53,81,796	1,09,77,244	98.81	96.58	159	51.09
111	R2	1,74,53,81,796	1,09,77,244	98.68	96.07	159	51.13
112	R1	15,40,31,727	9,68,753	98.59	96.11	159	47.56
112	R2	15,40,31,727	9,68,753	98.55	95.67	159	48.28
113	R1	13,41,38,601	8,43,639	98.55	96.03	159	49.28
113	R2	13,41,38,601	8,43,639	98.45	95.47	159	49.53
114	R1	1,66,92,01,875	1,04,98,125	98.59	96.2	159	49.86
114	R2	1,66,92,01,875	1,04,98,125	98.28	95.04	159	50.1
116	R1	38,37,57,312	24,13,568	98.55	96.05	159	51.03
116	R2	38,37,57,312	24,13,568	98.45	95.62	159	51.03
117	R1	88,96,29,804	55,95,156	98.62	96.2	159	49.96
117	R2	88,96,29,804	55,95,156	98.56	95.77	159	50.22
118	R1	29,81,14,506	18,74,934	98.16	94.81	159	53.9
118	R2	29,81,14,506	18,74,934	96.19	90.45	159	51.3
119	R1	3,04,67,421	1,91,619	98.1	94.6	159	50.53
119	R2	3,04,67,421	1,91,619	97.53	93.01	159	47.49
120	R1	7,25,90,178	4,56,542	98.38	95.13	159	50.55
120	R2	7,25,90,178	4,56,542	97.86	93.62	159	47.57
121	R1	33,24,77,109	20,91,051	98.1	94.6	159	48.47
121	R2	33,24,77,109	20,91,051	96.42	90.79	159	47.89
122	R1	34,44,25,323	21,66,197	98.16	94.83	159	51.26
122	R2	34,44,25,323	21,66,197	96.12	90.31	159	47.85
123	R1	39,48,14,172	24,83,108	98.01	94.49	159	52.4
123	R2	39,48,14,172	24,83,108	95.88	89.92	159	48
124	R1	39,55,34,124	24,87,636	98.11	94.71	159	52.64
124	R2	39,55,34,124	24,87,636	96.38	90.87	159	48.19
125	R1	5,25,60,153	3,30,567	98.26	95.01	159	48.95
125	R2	5,25,60,153	3,30,567	97.73	93.4	159	47.23
126	R1	47,52,02,664	29,88,696	98.59	96.27	159	51.6

126	R2	47,52,02,664	29,88,696	98.45	95.56	159	50.48
127	R1	11,46,85,269	7,21,291	96.14	91.68	159	54.09
127	R2	11,46,85,269	7,21,291	96.36	92.25	159	56.46
128	R1	13,95,47,145	8,77,655	98.61	96.37	159	52.83
128	R2	13,95,47,145	8,77,655	98.21	95.33	159	52.41
129	R1	10,98,68,682	6,90,998	98.53	96.08	159	47.5
129	R2	10,98,68,682	6,90,998	97.83	94.63	159	48.44
130	R1	22,45,30,101	14,12,139	98.65	96.44	159	50.15
130	R2	22,45,30,101	14,12,139	98.23	95.22	159	50.17
131	R1	6,26,92,110	3,94,290	98.09	95.38	159	55.07
131	R2	6,26,92,110	3,94,290	96.97	93.62	159	56.58
132	R1	69,09,78,384	43,45,776	98.3	95.77	159	53.59
132	R2	69,09,78,384	43,45,776	98.08	94.99	159	53.45
133	R1	40,14,24,279	25,24,681	98.44	96.08	159	52.5
133	R2	40,14,24,279	25,24,681	98.17	95.2	159	52.7
134	R1	16,53,88,143	10,40,177	98.73	96.38	159	51.66
134	R2	16,53,88,143	10,40,177	98.63	95.88	159	50.48
135	R1	7,77,59,427	4,89,053	98.52	95.89	159	51.12
135	R2	7,77,59,427	4,89,053	98.48	95.5	159	50.21
136	R1	47,96,90,916	30,16,924	98.14	94.81	159	53.44
136	R2	47,96,90,916	30,16,924	96.43	90.94	159	48.02
137	R1	59,54,01,258	37,44,662	98.44	95.31	159	53.97
137	R2	59,54,01,258	37,44,662	96.87	91.7	159	54.41
138	R1	27,08,74,149	17,03,611	98.01	94.44	159	53.15
138	R2	27,08,74,149	17,03,611	96	90.16	159	48.34
140	R1	61,93,20,264	38,95,096	98.08	94.59	159	53.33
140	R2	61,93,20,264	38,95,096	96.29	90.69	159	48.09
141	R1	38,87,03,484	24,44,676	98.69	96.33	159	50.74
141	R2	38,87,03,484	24,44,676	98.51	95.67	159	50.03
142	R1	10,28,26,572	6,46,708	98.74	96.43	159	49.38
142	R2	10,28,26,572	6,46,708	98.62	95.79	159	49.02
144	R1	33,52,99,836	21,08,804	98.55	95.98	159	51
144	R2	33,52,99,836	21,08,804	98.5	95.6	159	50.44
145	R1	13,85,74,701	8,71,539	98.6	96.05	159	51.06
145	R2	13,85,74,701	8,71,539	98.56	95.68	159	50.53
146	R1	39,98,97,084	25,15,076	98.67	96.28	159	50.74

146	R2	39,98,97,084	25,15,076	98.5	95.64	159	50.48
147	R1	21,67,64,064	13,63,296	98.51	95.97	159	47.55
147	R2	21,67,64,064	13,63,296	98.46	95.43	159	47.88
148	R1	17,57,66,073	11,05,447	98.58	96.05	159	48.81
148	R2	17,57,66,073	11,05,447	98.42	95.29	159	48.72
149	R1	9,08,64,525	5,71,475	98.57	96.11	159	49.45
149	R2	9,08,64,525	5,71,475	98.2	94.81	159	49.3
150	R1	26,71,75,809	16,80,351	98.72	96.33	159	50.31
150	R2	26,71,75,809	16,80,351	98.54	95.67	159	49.9

Table 3. Summary of clean paired-end FASTQ reads of 48 Tharparkar cattle after adapter trimming and quality filtering

Sample	Read Orientation	Bases (bps)	Reads	Q20%	Q30%	Avg. Read Length	GC%
101	R1	64,37,17,030	52,46,070	98.83	96.46	122.705	51.31
101	R2	64,30,74,808	52,46,070	98.93	96.5	122.582	50.78
102	R1	47,27,78,869	47,54,405	96.69	92.35	99.44	50.81
102	R2	46,77,68,661	47,54,405	97.41	92.92	98.386	50.35
103	R1	16,31,25,759	15,69,887	97.95	94.65	103.909	49.89
103	R2	16,16,49,100	15,69,887	97.87	94.02	102.969	49.62
104	R1	37,80,34,412	31,66,857	98.68	96.19	119.372	45.06
104	R2	37,73,23,545	31,66,857	98.91	96.3	119.148	45.29
105	R1	36,97,51,243	31,02,768	98.6	96.09	119.168	48.47
105	R2	36,87,52,838	31,02,768	98.8	96.11	118.846	48.34
106	R1	51,24,69,725	42,91,768	98.68	96.26	119.408	48.67
106	R2	51,15,25,945	42,91,768	98.59	95.6	119.188	48.48
107	R1	27,49,77,189	31,74,079	96.12	90.98	86.632	49.18
107	R2	27,03,38,005	31,74,079	96.9	91.84	85.171	48.9
108	R1	32,98,13,516	34,96,238	97.18	93.03	94.334	47.71
108	R2	32,54,93,796	34,96,238	97.76	93.83	93.098	47.53
109	R1	49,32,62,078	39,22,903	99.05	96.94	125.739	51.78

109	R2	49,32,43,299	39,22,903	99.19	97.15	125.734	51.2
110	R1	31,53,98,849	24,64,648	99.06	96.89	127.969	50.98
110	R2	31,53,75,779	24,64,648	99.18	97.08	127.96	50.3
111	R1	1,39,21,42,846	1,09,66,158	99.27	97.47	126.949	50.9
111	R2	1,39,20,90,996	1,09,66,095	99.38	97.54	126.945	50.48
112	R1	12,29,29,441	9,67,733	99.12	97.14	127.028	46.27
112	R2	12,29,19,620	9,67,733	99.24	97.14	127.018	46.72
113	R1	10,85,28,276	8,42,134	99.01	96.9	128.873	48.52
113	R2	10,85,20,294	8,42,134	99.06	96.75	128.863	48.41
114	R1	1,32,80,93,404	1,04,85,417	99.11	97.19	126.661	49.33
114	R2	1,32,80,22,624	1,04,85,437	98.86	96.25	126.654	49.17
116	R1	30,81,62,456	24,10,062	99.02	96.92	127.865	50.65
116	R2	30,81,40,319	24,10,062	99.09	96.94	127.856	50.23
117	R1	70,72,65,260	55,89,210	99.17	97.26	126.541	49.43
117	R2	70,72,24,338	55,89,210	99.24	97.2	126.534	49.27
118	R1	29,33,13,240	18,57,910	98.31	95	157.873	53.82
118	R2	29,32,85,194	18,57,910	96.57	90.88	157.858	51
119	R1	2,95,63,338	1,91,146	98.29	94.86	154.664	50.4
119	R2	2,95,61,257	1,91,146	97.77	93.34	154.653	47.22
120	R2	7,05,20,794	4,55,906	98.07	93.94	154.683	47.36
121	R1	32,73,11,678	20,72,705	98.26	94.8	157.915	48.36
121	R2	32,72,72,633	20,72,705	96.82	91.23	157.896	47.57
122	R1	33,93,12,136	21,45,448	98.3	95.02	158.154	51.18
122	R2	33,92,78,063	21,45,448	96.54	90.77	158.139	47.55
123	R1	38,94,25,365	24,59,008	98.16	94.67	158.367	52.35
123	R2	38,93,98,356	24,59,008	96.3	90.39	158.356	47.74
124	R1	39,11,68,773	24,70,300	98.25	94.88	158.349	52.58
124	R2	39,11,72,493	24,70,300	96.68	91.21	158.35	47.99
125	R1	5,11,34,690	3,30,176	98.41	95.22	154.871	48.82

125	R2	5,11,34,338	3,30,176	97.92	93.67	154.87	47.04
126	R1	40,60,14,169	29,65,078	99.08	97.12	136.932	50.92
126	R2	40,59,69,122	29,65,078	99.05	96.67	136.917	49.27
127	R1	5,36,58,350	6,84,502	96.48	91.4	78.39	52.58
127	R2	5,28,51,470	6,84,502	95.28	88.57	77.212	49.93
128	R1	11,43,91,101	8,56,509	99.14	97.29	133.555	50.2
128	R2	11,43,10,570	8,56,509	99.04	96.63	133.461	48.88
129	R1	9,36,20,982	6,72,581	99.02	96.9	139.197	45.07
129	R2	9,36,07,750	6,72,581	98.88	96.2	139.177	45.81
130	R1	18,94,02,298	13,88,234	99.11	97.22	136.434	48.19
130	R2	18,93,48,393	13,88,234	98.97	96.44	136.395	47.63
131	R1	4,61,93,138	3,60,172	98.81	96.58	128.253	48.35
131	R2	4,61,45,254	3,60,172	98.45	95.45	128.12	47.37
132	R1	53,53,28,872	42,67,461	98.99	96.95	125.444	49.82
132	R2	53,49,01,733	42,67,461	98.86	96.3	125.344	48.62
133	R1	32,19,71,492	24,67,125	99.05	97.13	130.505	48.61
133	R2	32,17,16,070	24,67,125	98.87	96.32	130.401	47.75
134	R1	14,39,23,291	10,38,875	99.11	97.11	138.538	51.37
134	R2	14,39,16,642	10,38,875	99.03	96.68	138.531	50.04
135	R1	6,64,65,912	4,87,234	99.03	96.86	136.415	50.37
135	R2	6,64,62,501	4,87,234	98.94	96.44	136.408	49.24
136	R1	47,19,43,018	29,97,856	98.3	95.01	157.427	53.37
136	R2	47,19,38,521	29,97,856	96.72	91.28	157.425	47.79
137	R1	58,73,58,762	37,25,710	98.55	95.45	157.65	53.94
137	R2	58,73,02,756	37,25,710	97.11	91.98	157.635	54.29
138	R1	26,55,54,021	16,87,188	98.16	94.63	157.394	53.09
138	R2	26,55,28,712	16,87,188	96.41	90.6	157.379	48.03
140	R1	61,17,69,770	38,72,628	98.22	94.77	157.973	53.28
140	R2	61,17,70,689	38,72,628	96.57	91.01	157.973	47.93
141	R1	33,15,70,315	24,41,505	99.12	97.18	135.806	50.15

141	R2	33,15,52,679	24,41,505	98.97	96.61	135.798	49.32
142	R1	8,92,61,577	6,45,942	99.14	97.22	138.188	48.69
142	R2	8,92,56,737	6,45,942	99.03	96.63	138.181	48.27
144	R1	27,14,03,378	21,04,881	99.05	96.89	128.94	50.38
144	R2	27,13,89,569	21,04,881	99.13	96.9	128.933	49.71
145	R1	11,48,12,858	8,67,985	99.04	96.82	132.275	50.25
145	R2	11,47,98,013	8,67,985	99.12	96.84	132.258	49.52
146	R1	31,33,75,503	25,10,669	99.24	97.37	124.818	49.87
146	R2	31,33,53,265	25,10,669	99.32	97.32	124.809	49.49
147	R1	17,04,02,591	13,61,455	99.16	97.21	125.162	45.71
147	R2	17,03,86,301	13,61,455	99.24	97.09	125.15	46.09
148	R1	14,12,53,779	11,03,481	99.09	97.02	128.007	47.46
148	R2	14,12,34,987	11,03,481	99.07	96.66	127.99	47.34
149	R1	7,34,84,355	5,70,361	99.05	97	128.838	48.23
149	R2	7,34,70,237	5,70,361	98.76	95.93	128.814	47.99
150	R1	22,15,79,914	16,77,306	99.09	97.01	132.105	49.5
150	R2	22,15,50,528	16,77,306	99.09	96.8	132.087	48.97

4.2 Genetic Diversity analysis

Heterozygosity

The mean observed heterozygosity of the population was 0.291 ± 0.078, while the mean expected heterozygosity was 0.347 ± 0.103 (Fig. 4). The observed heterozygosity values were consistently lower than the expected heterozygosity values across the population. The chromosome-wise observed and expected heterozygosity values are summarized in Table 4 and their graphical representation is shown in Fig. 5.

Table 4: Summary of observed and expected heterozygosity values across autosomes

Chromosome No.	Ho (Mean ± S.D.)	He (Mean ± S.D.)
1	0.283 ± 0.078	0.348 ± 0.102
2	0.297 ± 0.076	0.343 ± 0.103
3	0.293 ± 0.075	0.348 ± 0.103
4	0.289 ± 0.077	0.347 ± 0.103
5	0.292 ± 0.081	0.344 ± 0.104
6	0.293 ± 0.076	0.349 ± 0.102
7	0.295 ± 0.083	0.350 ± 0.104
8	0.294 ± 0.078	0.348 ± 0.103
9	0.29 ± 0.077	0.346 ± 0.102
10	0.291 ± 0.078	0.348 ± 0.104
11	0.298 ± 0.075	0.347 ± 0.104
12	0.291 ± 0.075	0.343 ± 0.103
13	0.292 ± 0.079	0.347 ± 0.103
14	0.292 ± 0.077	0.347 ± 0.102
15	0.287 ± 0.085	0.350 ± 0.103
16	0.293 ± 0.081	0.346 ± 0.102
17	0.289 ± 0.073	0.344 ± 0.101
18	0.296 ± 0.084	0.348 ± 0.104
19	0.287 ± 0.076	0.344 ± 0.103
20	0.293 ± 0.074	0.340 ± 0.103
21	0.291 ± 0.082	0.350 ± 0.103
22	0.293 ± 0.076	0.344 ± 0.102
23	0.291 ± 0.082	0.348 ± 0.103
24	0.289 ± 0.078	0.348 ± 0.103
25	0.290 ± 0.073	0.345 ± 0.103
26	0.291 ± 0.075	0.348 ± 0.101
27	0.294 ± 0.089	0.346 ± 0.103
28	0.287 ± 0.074	0.351 ± 0.103
29	0.287 ± 0.079	0.347 ± 0.101
Mean	**0.291 ± 0.078**	**0.347 ± 0.103**

Minor Allele Frequency (MAF)

The highest proportion of SNPs (23.74%) was observed in the 0.10-0.20 MAF range, while the lowest proportion (9.84%) occurred in the 0-0.05 range (Fig. 6). Approximately

Type	Total
SNP	121,554
MNP	0
INS	0
DEL	0
MIXED	0
INV	0
DUP	0
CNV	0
BND	0
INTERVAL	0
Total	121,554

(a)

Type (alphabetical order)	Count	Percent
DOWNSTREAM	28,751	4.944%
EXON	5,115	0.88%
INTERGENIC	66,283	11.399%
INTRON	221,590	38.107%
SPLICE_SITE_ACCEPTOR	10	0.002%
SPLICE_SITE_DONOR	22	0.004%
SPLICE_SITE_REGION	419	0.072%
TRANSCRIPT	225,777	38.827%
UPSTREAM	29,115	5.007%
UTR_3_PRIME	2,792	0.48%
UTR_5_PRIME	1,621	0.279%

(b)

Fig. 2: (a) Classification of variants based on type (b) Single Nucleotide Polymorphism effect distribution

	A	C	G	T
A	0	5,392	16,920	3,837
C	6,754	0	5,850	21,069
G	22,440	5,760	0	6,095
T	4,360	17,248	5,829	0

Fig. 2: (c) Number of transitions and transversions in annotated variants

Type (alphabetical order)	Count	Percent
HIGH	82	0.014%
LOW	2,881	0.495%
MODERATE	789	0.136%
MODIFIER	577,743	99.355%

Fig. 2: (d) Number of effects by impact

Type (alphabetical order)	Count	Percent
MISSENSE	790	36.83%
NONSENSE	49	2.284%
SILENT	1,306	60.886%

Fig. 2: (e) Number of effects by functional class

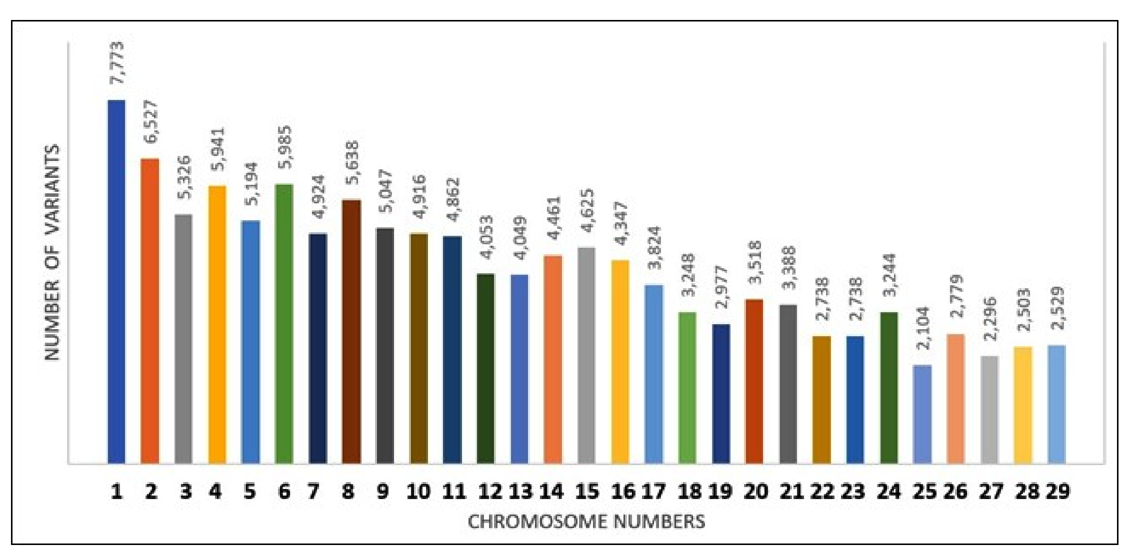


Fig. 3: Diagrammatic representation of Chromosome wise number of variants

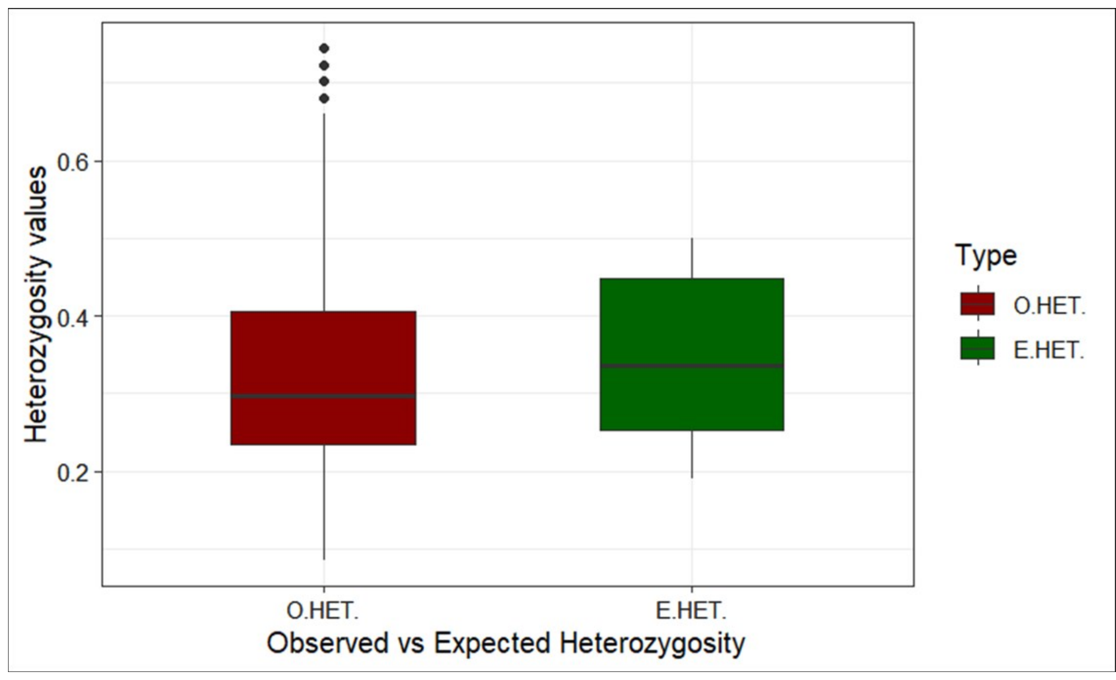


Fig. 4: Boxplot depicting observed (O.HET) and expected (E.HET) heterozygosity values

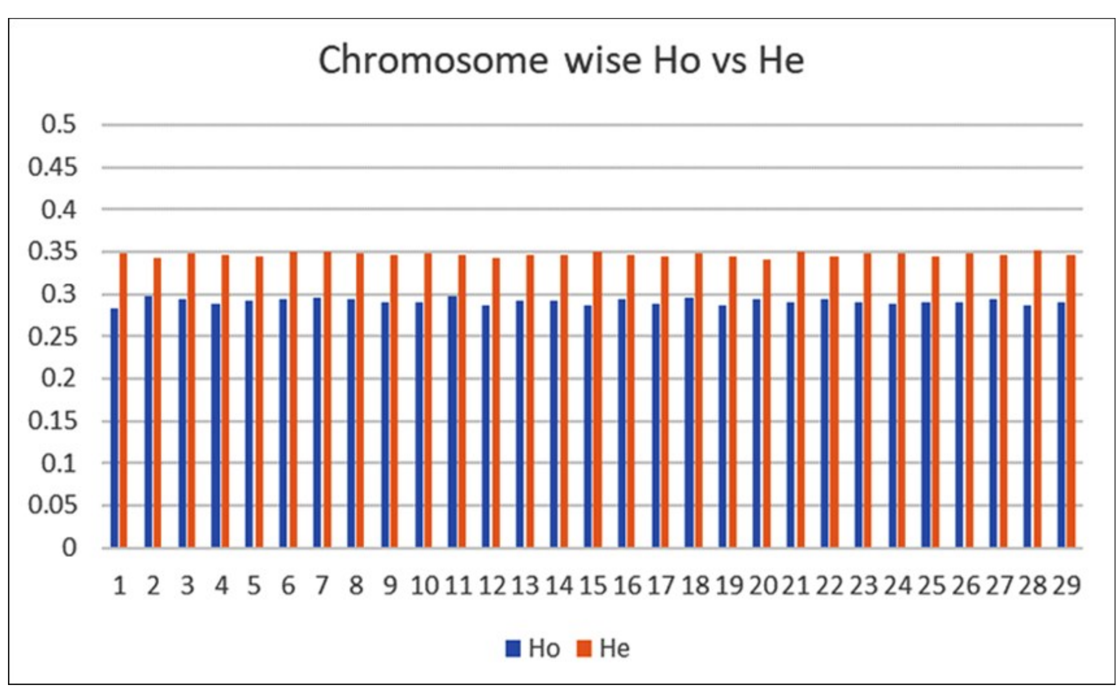


Fig. 5: **Diagrammatic representation of chromosome wise observed (Ho) and expected (He) heterozygosity**

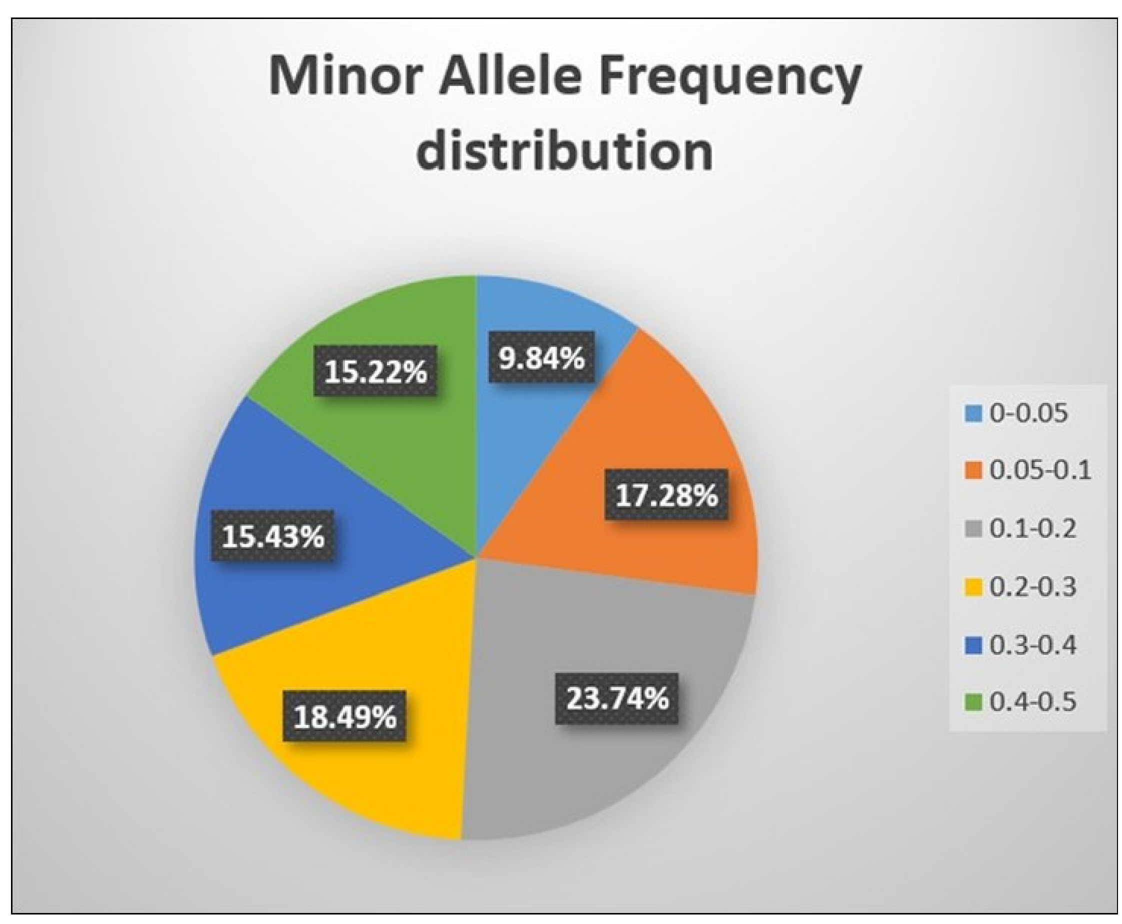


Fig. 6: **Pie chart showing the distribution of Minor Allele Frequency (MAF) across different class intervals**

49.14% of SNPs exhibited a minor allele frequency (MAF) greater than 0.20. The overall MAF across all autosomal chromosomes was 0.216 ± 0.143. Furthermore, the polymorphic marker ratio was 0.728, indicating that about 72.88% of SNP sites were polymorphic (MAF > 0.1).

Runs of Homozygosity (ROH)

A total of 4,373 ROH segments were detected across all autosomal chromosomes of the Tharparkar genome using the sliding window approach. The mean total length of ROHs per individual was 154.66 Mb, accounting for 6.22% of the autosomal genome. The majority of ROHs were classified into the 0-2 Mb length category (81.56%), followed by 2-4 Mb (14.56%), 4-8 Mb (2.88%), 8-16 Mb (0.66%) and only 0.02% falling in the >16 Mb category (Fig. 7a). The F_{ROH} estimates according to various class lengths are shown in Fig. 7b. The genome-wide F_{ROH} (ROH-based inbreeding coefficient) was estimated at 0.1067 ± 0.042, with the highest F_{ROH} value observed on chromosome 13 (Fig. 8).

Genomic Inbreeding Coefficient estimation

The genomic inbreeding coefficient based on F_{HOM} (excess of homozygosity), F_{UNI} (correlation between uniting gametes) and F_{GRM} (genomic relationship matrix) was estimated to be 0.112 $\pm$ 0.074, 0.125 $\pm$ 0.089 and 0.125 $\pm$ 0.082 (Fig. 12).

Effective Population Size (Ne) estimation

Using the random mating model at a critical MAF cutoff of 0.05, the effective population size (Ne) was estimated to be 82.6 with the software NeEstimator v2.01 (Do *et al.*, 2014). Further historical Ne analysis performed with SNeP (Barbato *et al.*, 2015) indicated a continuous decline, dropping from roughly 2,519 individuals about 150 generations in the past to nearly 226 individuals around 13 generations ago (Fig. 13).

Average Pairwise Genetic Distance (D) and Nucleotide Diversity (π)

The mean genetic distance among individuals was determined to be 0.671 ± 0.028, while genomic diversity, assessed using nucleotide diversity (π) with 200 bp sliding windows across the genome, was estimated at 0.344 ± 0.073.

4.3 Mining Breed-Specific SNPs using TRES

4.3.1 Dataset characteristics and preparation of reference and validation datasets

The final Tharparkar dataset comprised 121,554 autosomal biallelic SNPs. The ddRAD genotype data from four additional cattle breeds (Sahiwal, Red Sindhi, Gir and Rathi) were converted into VCF format using the previously described pipeline and aligned to the ARS-UCD 2.0 reference genome. Quality control filters were applied to each breed-wise dataset, retaining SNPs with read depth (RD) > 10, Phred score > 30, minor allele frequency (MAF) > 0.05 and Hardy–Weinberg equilibrium (HWE) p-value > 0.0001. Subsequently, the Tharparkar dataset was merged with these four breed-wise datasets, resulting in a combined dataset of 368 animals with 24,163 common SNPs. To assess population structure, Principal Component Analysis (PCA) was performed on the merged dataset. The first two principal components explained a combined variance of 55.29% (PC1 + PC2) and clearly separated the individuals into five distinct clusters, corresponding to the respective breeds (Fig. 14), thereby confirming the presence of five pure breeds in the dataset. For downstream analysis, the combined dataset was randomly partitioned into two subsets using TRES (Kavakiotis *et al*., 2015): a reference set (70%) consisting of 260 animals and a validation set (30%) consisting of 108 animals. The breed-wise distribution of animals in the reference and validation sets is presented in Table 5.

Table 5: Details of Number of Animals per breed in reference and validation dataset

Breed of Cattle	Number of Animals per breed (Total = 368)	Reference Dataset (Total = 260)	Validation Dataset (Total = 108)
Tharparkar	47	33	14
Gir	48	34	14
Rathi	96	68	28
Red Sindhi	96	68	28
Sahiwal	81	57	24

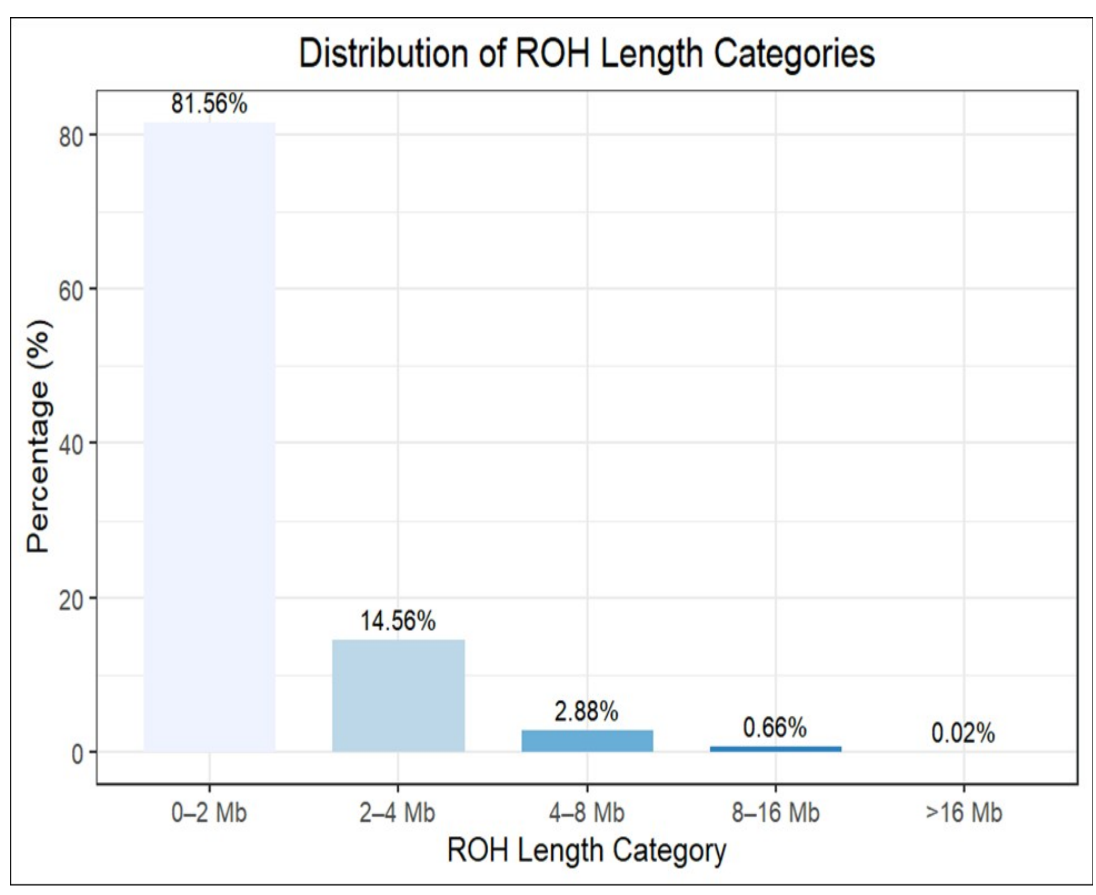


Fig. 7a: Class wise distribution of ROH count

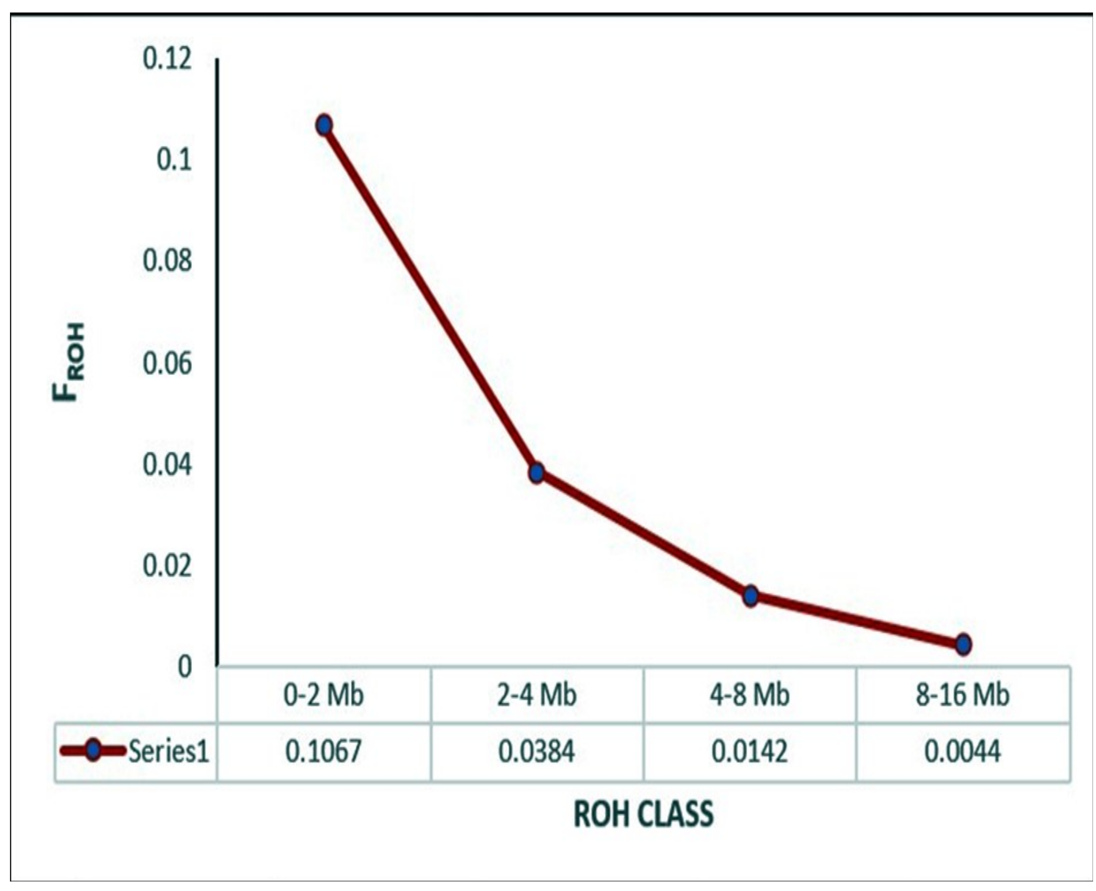
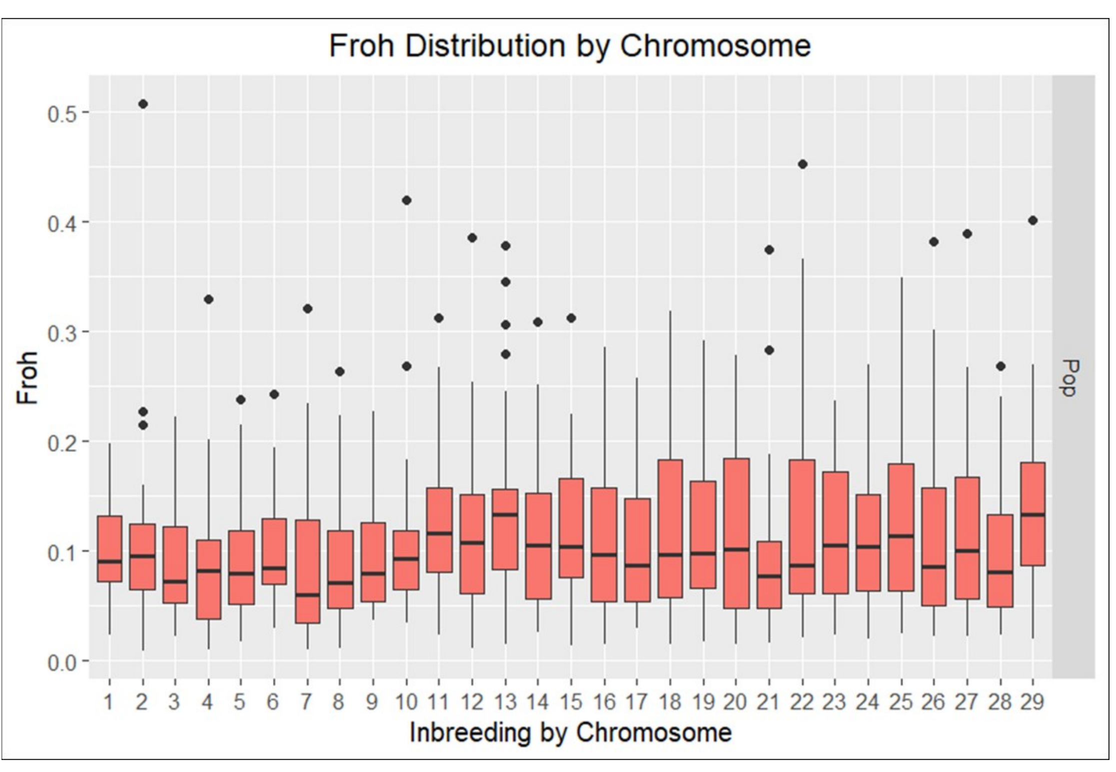


Fig. 7b: F_{ROH} estimates according to various ROH class lengths



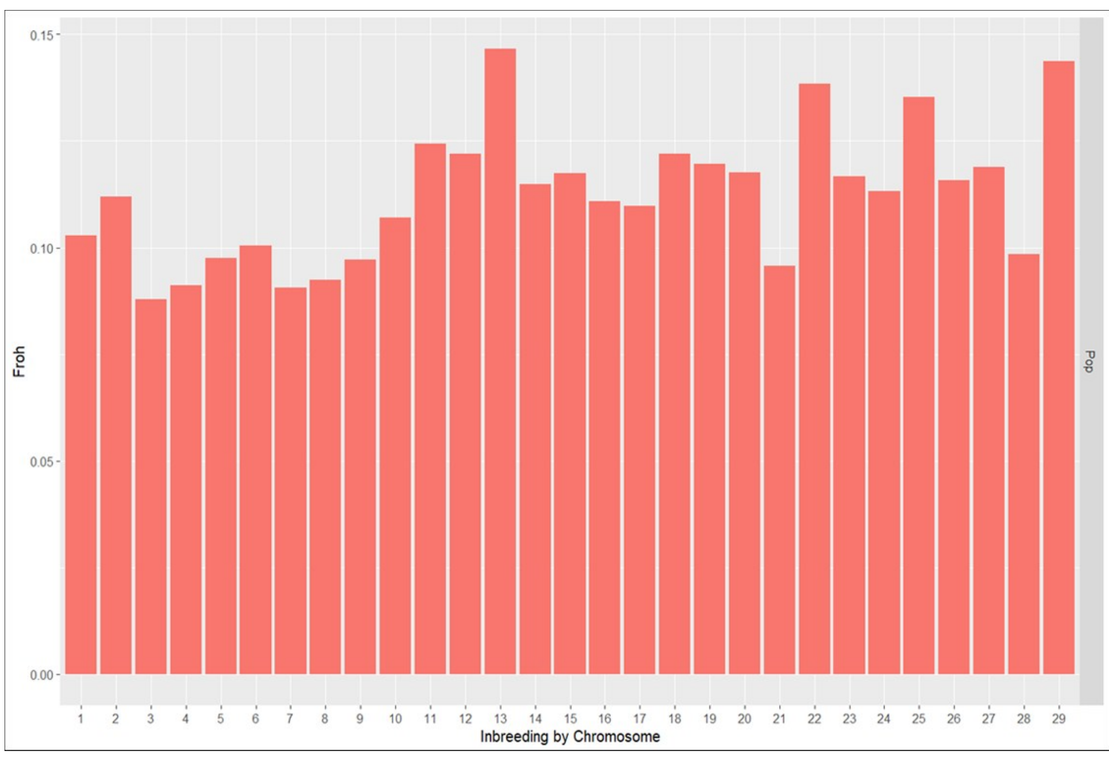
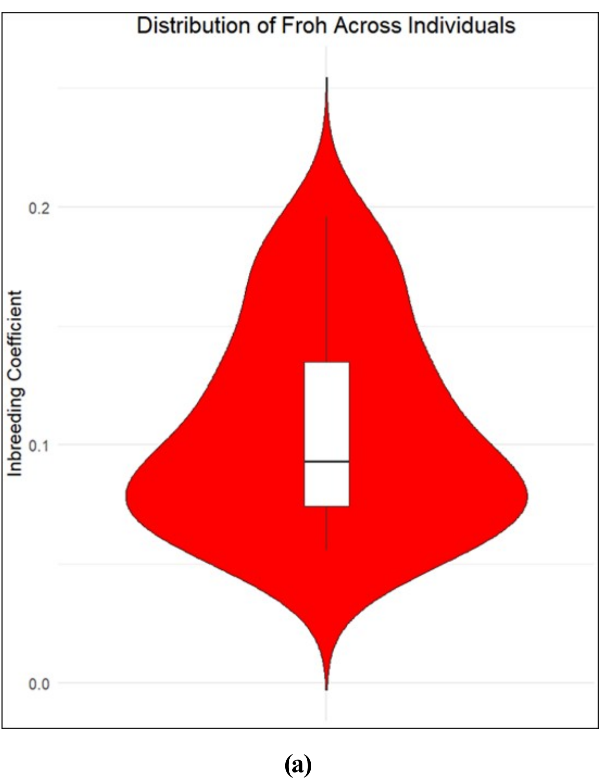
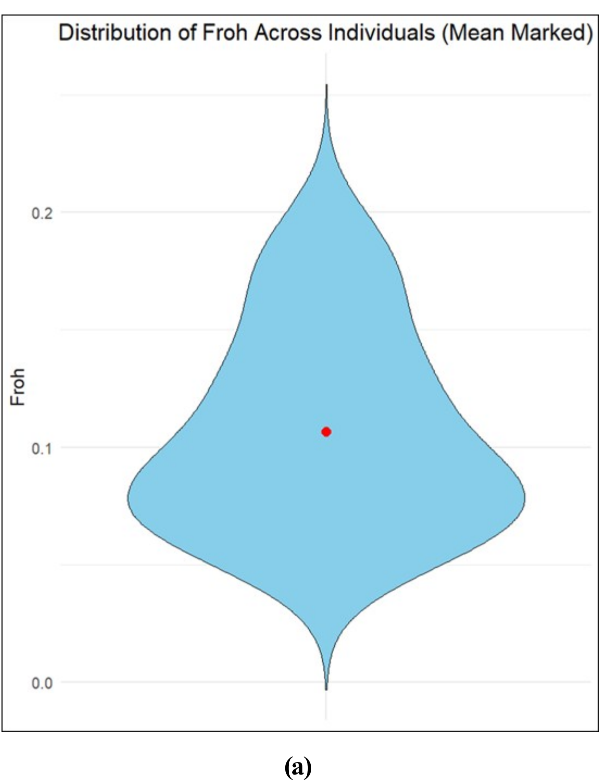


Fig. 8: Chromosome wise F_{ROH} values in Tharparkar cattle



(a)



(a)

Fig. 9: Violin plot depicting genome-wide F_{ROH} distribution. The central red line in (a) signifies the median value of the population's F_{ROH} while the upper and lower boundaries of the central black box denote the upper and lower quartiles of the F_{ROH} distribution within the population. In (b) the central red dot signifies the mean value of the population's $F_{ROH.}$ The width of the violin plot characterizes the probability density distribution of the population's $F_{ROH.}$

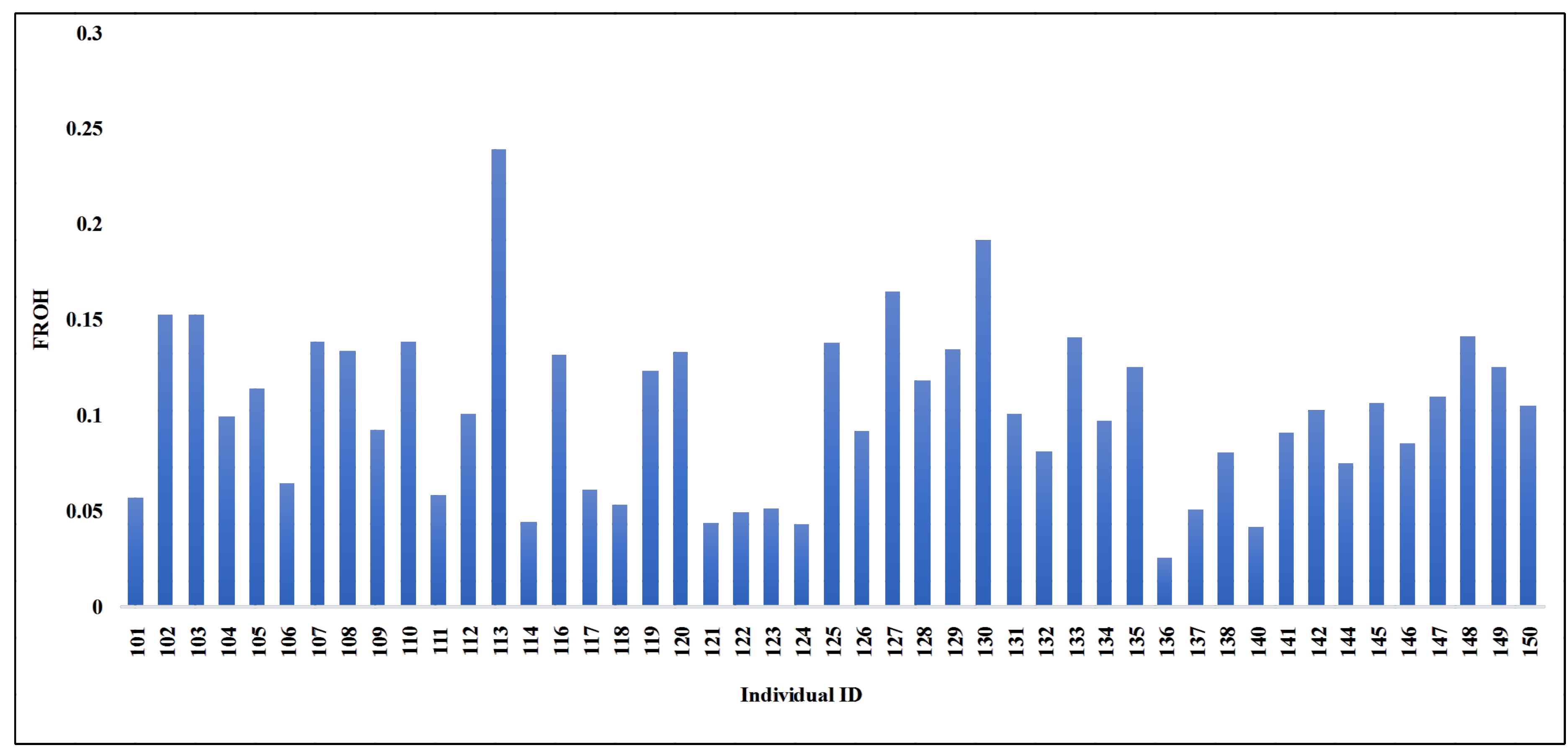


Fig. 10: Bar plot of F_{ROH} for each individual in the population

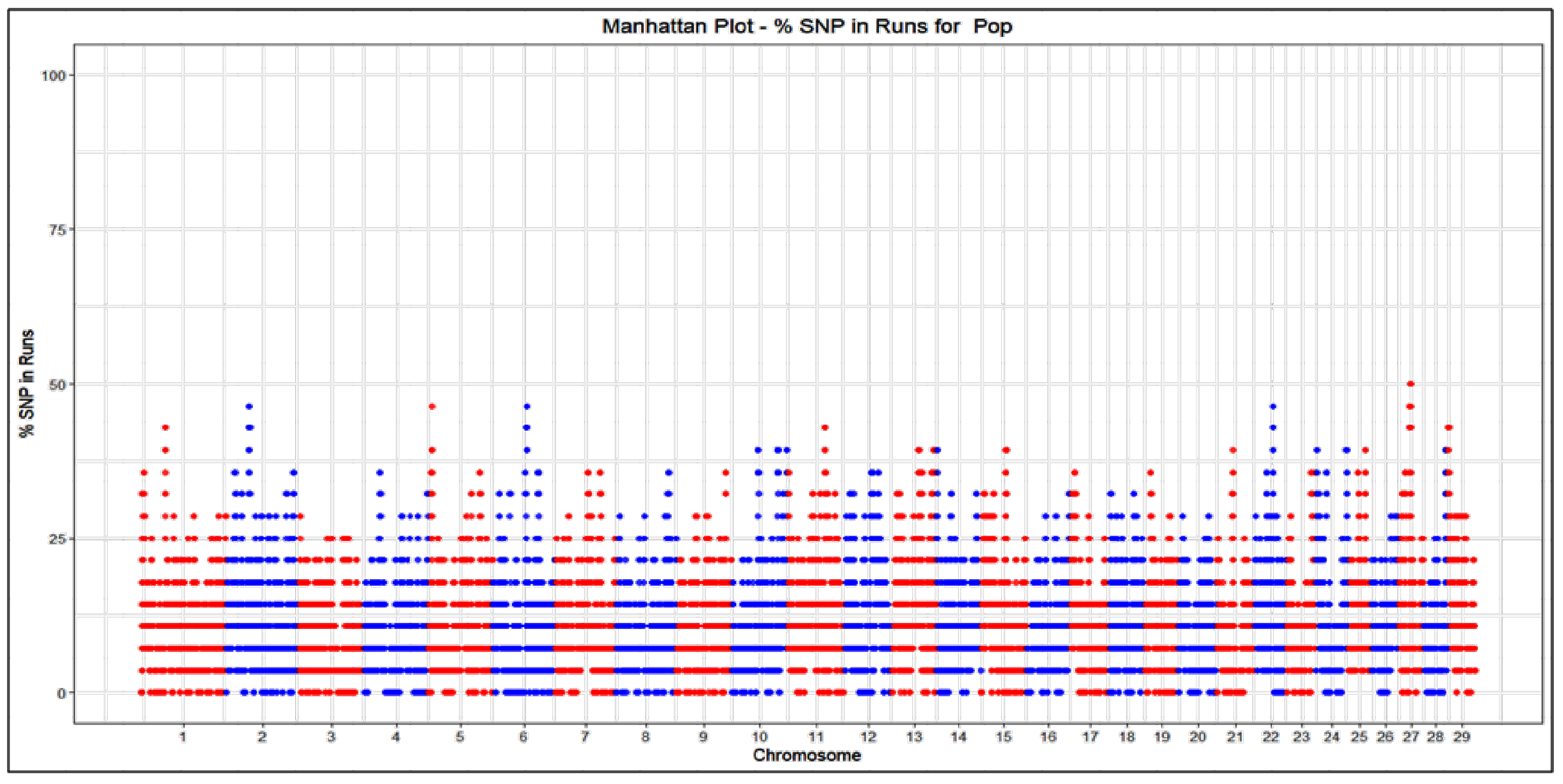


Fig. 11: Manhattan plot showing the percentage of SNPs included in Runs of Homozygosity (ROH) across autosomal chromosomes in Tharparkar cattle. Each dot represents a SNP, with alternating colors (red and blue) distinguishing adjacent chromosomes. The y-axis represents the proportion of SNPs occurring within ROH, while the x-axis corresponds to the chromosome number.

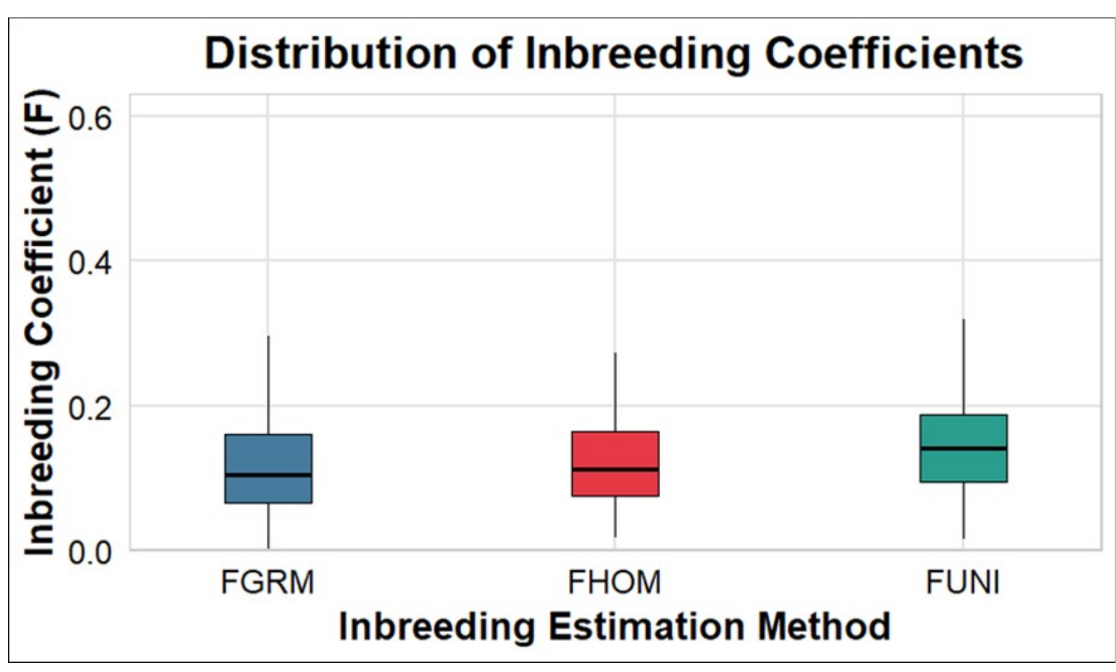


Fig. 12: Boxplot depicting genomic inbreeding coefficients

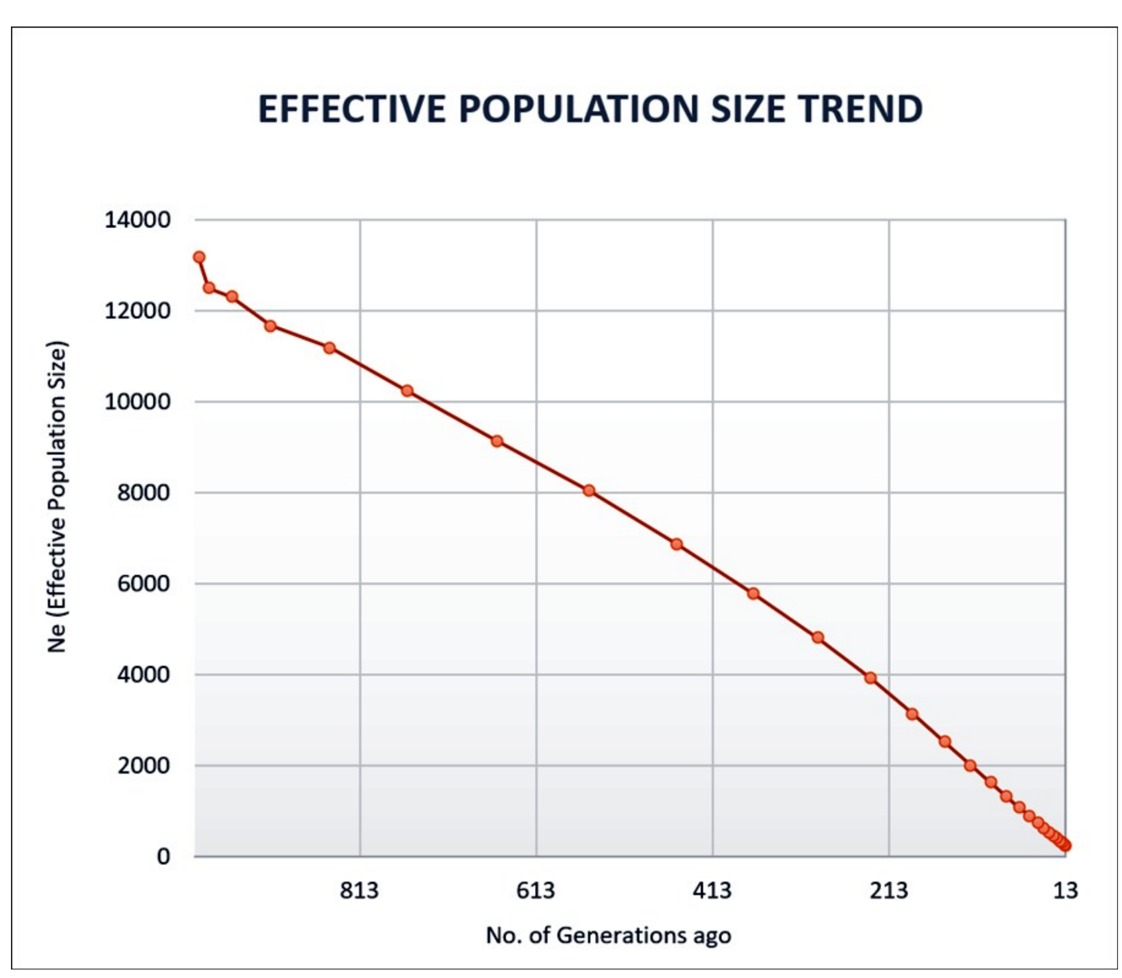


Fig. 13: Declining trend of effective population size (Ne) over the past 150 generations

4.3.2 Breed-specific SNP Mining results

Each breed was compared against all others using TRES (Kavakiotis *et al.*, 2015), which generated a large set of SNPs shared across the three algorithms implemented in the software. To eliminate redundancies, the results were cross-checked using VENNY 2.1.0 and only SNPs consistently identified by all three methods were retained. This filtering step ensured that only high-confidence SNPs were considered. The number of SNPs obtained per breed after this step is presented in Table 6. From these results, the common SNPs across the three algorithms of TRES were extracted for each breed (Fig. 15a, 15b, 15c, 15d, 15e). The curated breed-wise SNP sets were then subjected to additional pruning using the MAF–LD method (Kumar *et al.*, 2019), where SNPs with MAF > 0.36 were retained, followed by LD pruning with $r^2 < 0.01$ to obtain the final breed-specific SNP sets (Table 6). These sets were subsequently merged and another round of MAF–LD filtering was performed to remove redundant or highly correlated SNPs from the combined dataset. This multistage filtering process resulted in a final panel of 309 breed-specific SNPs across 260 individuals representing the five breeds. Of these, 77 SNPs were unique to Tharparkar cattle, while Red Sindhi showed the highest number (84 SNPs) and Sahiwal the lowest (22 SNPs), as summarized in Table 6.

Table 6: Details of SNPs selected through pre-selection methods and retained after the first MAF–LD filtering step, along with the final set of breed-specific SNPs

Breed of Cattle	Combined list of SNPs from 3 preselection methods (Delta, F_{ST}, IFN)	SNPs remaining after 1st MAF-LD step	Breed - specific SNPs
Tharparkar	4118	77	77
Gir	3789	52	30
Rathi	4051	96	96
Red Sindhi	4054	84	84
Sahiwal	4400	29	22

4.3.3 Accuracy of SNP Panel for Breed Assignment

The dataset comprising 309 SNPs from 260 individuals in the reference set was analyzed to assess breed assignment accuracy using STRUCTURE v2.3.4. Population clustering was evaluated for K values ranging from 2 to 6 and the highest ΔK value was observed at K = 5, which corresponded with the plateau of mean log-likelihood values [LnP(K)]. This concordance, determined through STRUCTURE Harvester (Earl and vonHoldt, 2012), confirmed the presence of five genetically distinct clusters, consistent with the five studied breeds. The population structure derived from these clusters is illustrated in Fig. 16a and 16b. Genetic differentiation among the clusters was further evaluated using F_{ST} values, which ranged from 0.8864 to 0.995, signifying substantial divergence. Specifically, Tharparkar cattle showed an F_{ST} of 0.9001 when compared with the pooled populations of the other breeds, highlighting its strong genetic distinctiveness (Fig. 18). From the analysis, 77 informative SNPs specific to Tharparkar were identified, distributed across all autosomes (chromosomes 1–29). The number of SNPs per chromosome ranged from one to four, with chromosome 2 harboring the maximum (four SNPs). In contrast, chromosomes 28 and 29 carried only one SNP each, while the remaining chromosomes contained two to three SNPs.

4.3.4 Validation of Breed-Specific SNPs

Individual allocation was further validated using an independent dataset of 108 animals in STRUCTURE v2.3.4. The 309 SNP panel showed strong power to separate breeds, with all animals correctly assigned to their respective clusters. The mean F_{ST} values across clusters ranged from 0.8155 to 0.9961, showing clear genetic differences among breeds. Tharparkar cattle had an F_{ST} value of 0.8269, confirming their distinctiveness from other populations (Fig. 19). All fourteen Tharparkar animals in the validation set were correctly assigned to their cluster, demonstrating the reliability of the identified breed-specific SNPs. These results highlight the usefulness of the SNP panel for accurate breed identification and its potential role in conservation and genomic-based breeding programs. The assignment accuracy for all breeds is shown in Fig. 17a and 17b, providing a clear view of the clustering efficiency and precision of individual allocation.

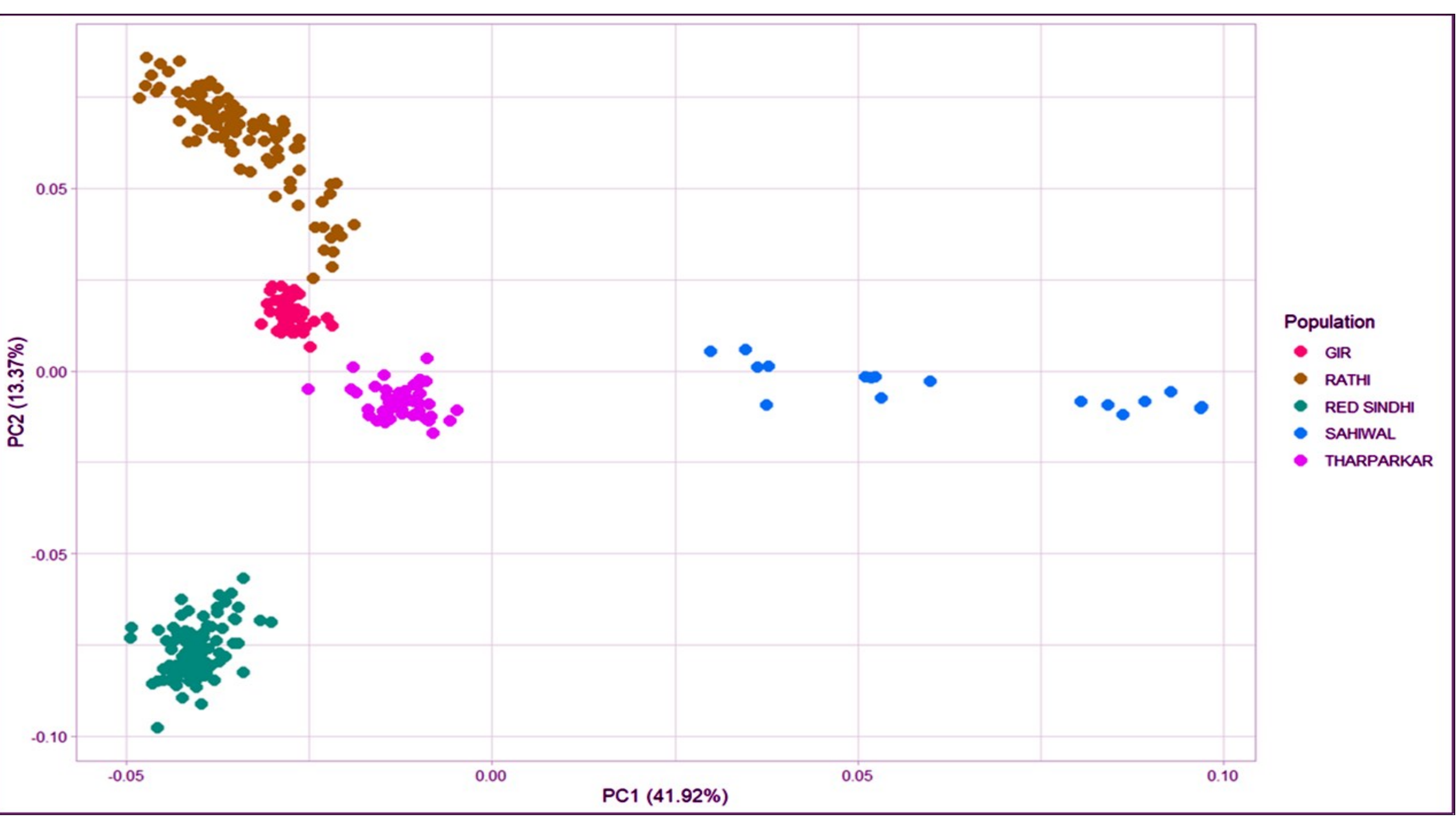


Fig. 14: Principal Component Analysis (PCA) plot showing the genetic variation among individuals, confirming the presence of five distinct breed clusters

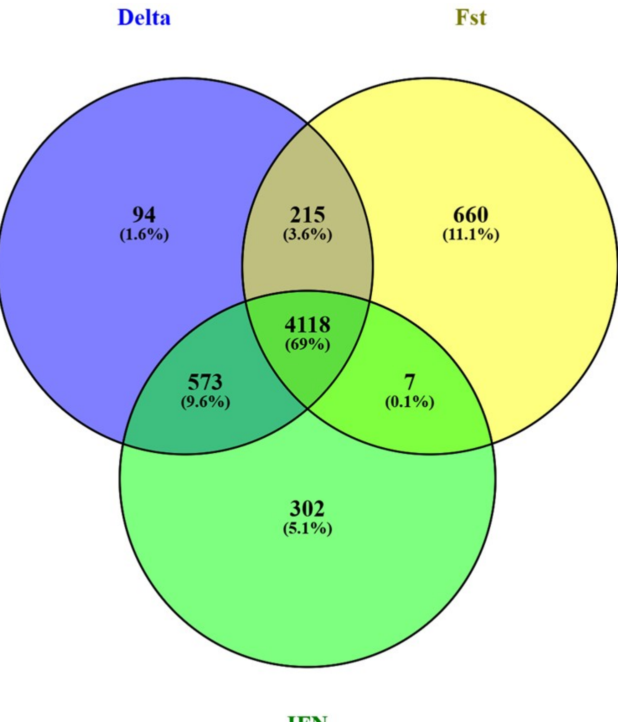


Fig. 15a: SNPs selected by TRES for distinguishing Tharparkar breed from other breeds using three pre-selection methods - Delta, pairwise Wright's FST, and Informativeness for Assignment - with common SNPs identified by intersecting the results across methods through VENNY 2.1.0

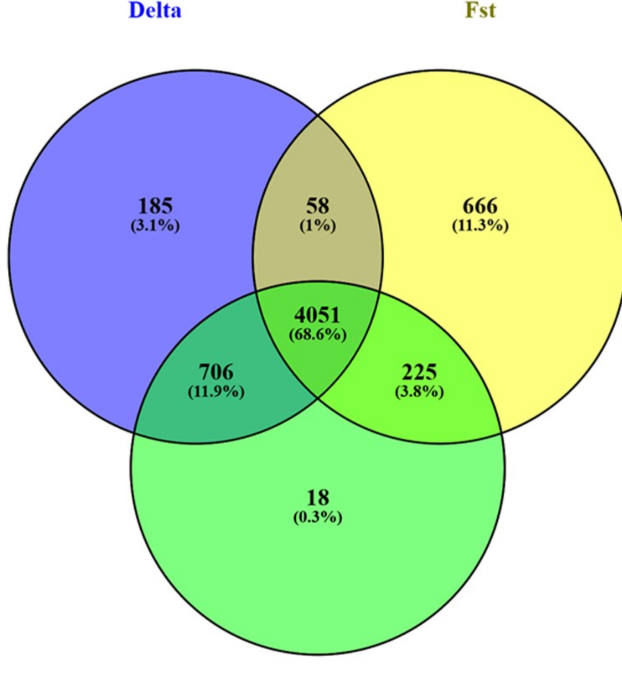


Fig. 15b: Rathi breed versus other breeds

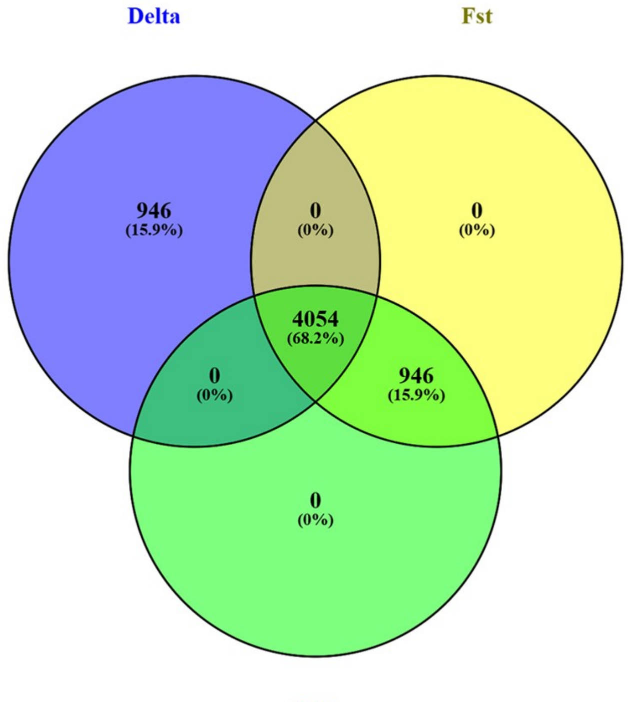


Fig. 15c: Red Sindhi breed versus other breeds

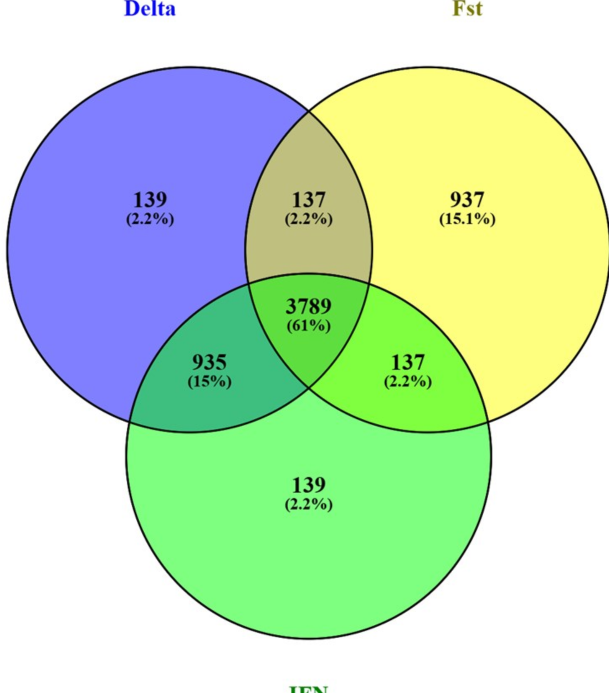


Fig. 15d: Gir breed versus other breeds

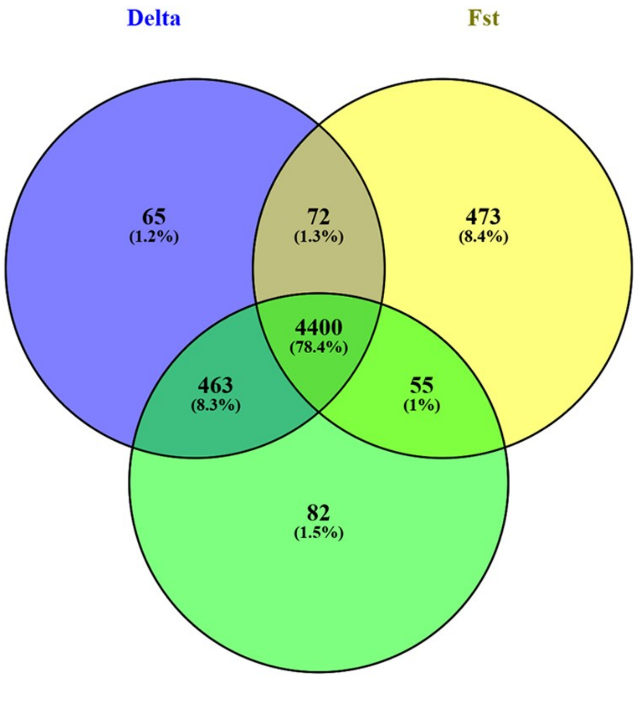


Fig. 15e: Sahiwal breed versus other breeds

4.3.5 Annotation of Tharparkar Specific SNPs

Variant Effect Predictor (VEP) annotation of the 77 breed-specific SNPs identified in Tharparkar cattle revealed that 30 loci (39%) represented novel polymorphisms, while 47 loci corresponded to previously reported variants. Collectively, these SNPs overlapped with 47 genes, 143 transcripts and one regulatory feature, indicating widespread genomic distribution. Functional consequence analysis indicated a predominance of non-coding variants, with intronic variants constituting 59% of the total, followed by intergenic (14%), upstream (8%), downstream (7%) and non-coding transcript variants (6%). Coding region alterations comprised a minor proportion, where synonymous and stop-gained variants were equally represented (50% each) (Fig. 20). The impact classification suggested that most of the breed-specific polymorphisms were of modifier impact, implying a potential influence on gene regulation or transcript processing rather than direct protein-altering effects.

Further annotation using Ensembl BioMart and BEDtools identified 35 protein-coding genes associated with these SNPs (listed in Table 7). In these some key genes like ADAMTS17 and PKHD1 were also reported by Kumar *et al*. (2019) in Tharparkar cattle and NRXN1 was reported by Li *et al*. (2014). Comparative analysis with the Cattle QTLdb revealed overlap of these SNPs and associated key genes like PPM1H, CHST11, NRXN1, ATP6V1E2, SUSD4, ADAMTS17, PKHD1, LAMA3 and SORC1 with QTL regions primarily associated with milk production traits, including milk fat yield, milk yield, milk protein yield, milk stearic acid content, milk C18 index, milk C14 index and milk linolenic acid content. Additional overlaps were observed for reproductive traits (inseminations per conception, first service conception, age at puberty) and growth-related traits (average daily gain, strength, carcass weight) (Table 8). Gene Ontology (GO) enrichment analysis performed using DAVID Bioinformatics Resources (https://davidbioinformatics.nih.gov/) highlighted multiple biological pathways associated with these genes, suggesting their potential roles in key metabolic and physiological processes relevant to Tharparkar cattle performance and adaptation.

Table 7. Genes overlapping with genomic regions harboring Tharparkar-specific SNPs, with chromosome position, SNP ID, gene symbol and gene name

Chromosome	SNP ID	GENE	GENE name	Reference
chr2	2:26444958	UBR3	ubiquitin protein ligase E3 component n-recognin 3	
chr2	2:15179664	ITGA4	Integrin subunit alpha 4	
chr2	2:119073416	ARMC9	armadillo repeat containing 9	
chr3	3:23091510	SEC22B	SEC22 homolog B, vesicle trafficking protein	
chr3	3:29934414	MAGI3	membrane associated guanylate kinase, WW and PDZ domain containing 3	
chr4	4:8420324	CDK14	cyclin dependent kinase 14	
chr5	5:50753881	PPM1H	protein phosphatase, Mg2+/Mn2+ dependent 1H	
chr5	5:68178886	CHST11	carbohydrate sulfotransferase 11	
chr6	6:10748116	NDST4	N-deacetylase and N-sulfotransferase 4	
chr10	10:73134372	SLC38A6	Solute carrier family 38 member 6	
chr10	10:89944298	NRXN1	Neurexin 1	Li *et al.*, 2014
chr11	11:28976459	ATP6V1E2	ATPase H+ transporting V1 subunit E2	
chr12	12:13559774	ENOX1	Ecto-NOX disulfide-thiol exchanger 1	
chr13	13:61913394	NOL4L	Nucleolar protein 4 like	
chr14	14:55087857	NUDCD1	NudC domain containing 1	
chr14	14:55087857	ENY2	ENY2 transcription and export complex 2 subunit	
chr15	15:79655039	OR8J3	Olfactory receptor family 8 subfamily J member 3	
chr16	16:949441	ADORA1	Adenosine A1 receptor	
chr16	16:26685253	SUSD4	Sushi domain containing 4	
chr17	17:51322010	NCOR2	Nuclear receptor corepressor 2	
chr19	19:61442032	ABCA6	ATP binding cassette subfamily A member 6	
chr20	20:37129408	CPLANE1	Ciliosis and planar polarity effector complex subunit 1	
chr21	21:6618185	ADAMTS17	ADAM metallopeptidase with thrombospondin	Kumar *et al.*, 2021

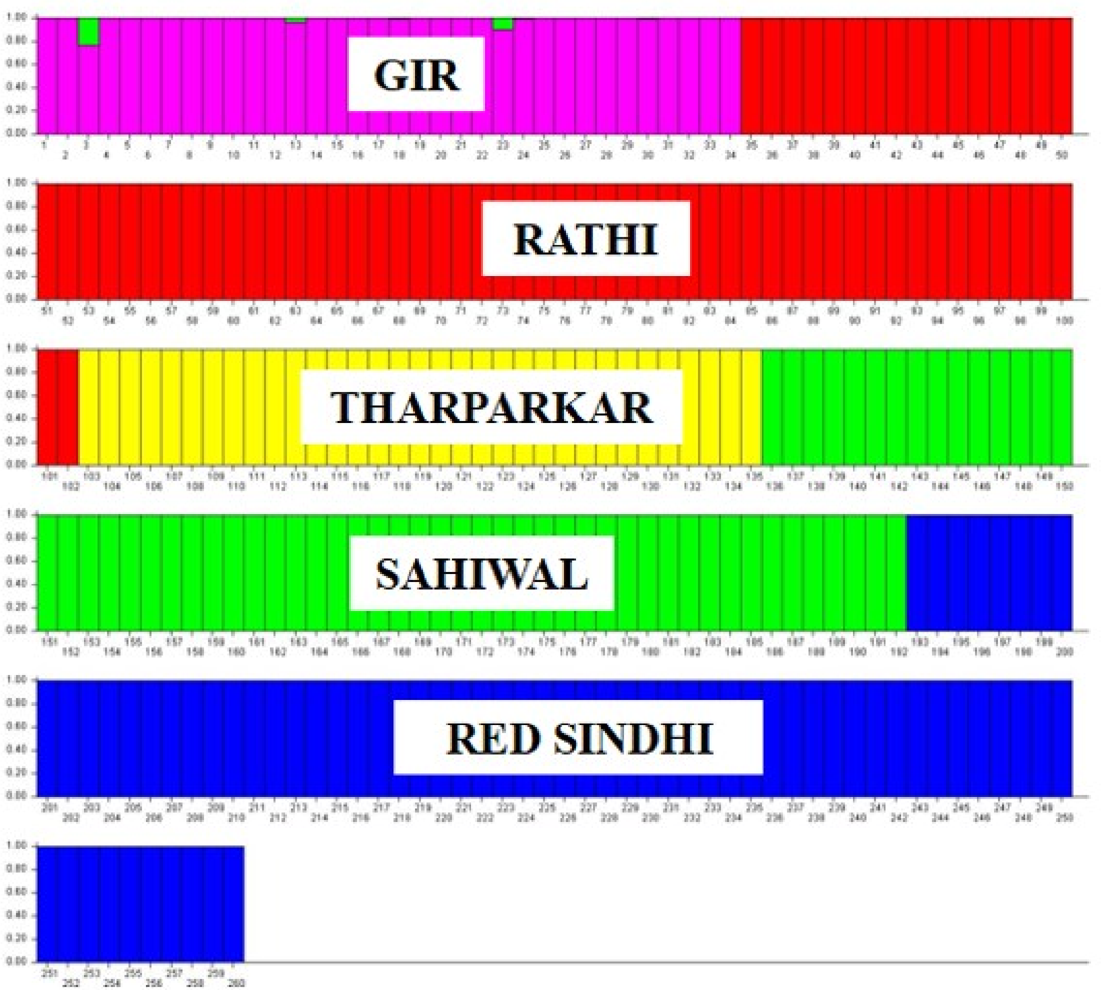


Fig. 16a: **Open bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Reference dataset**

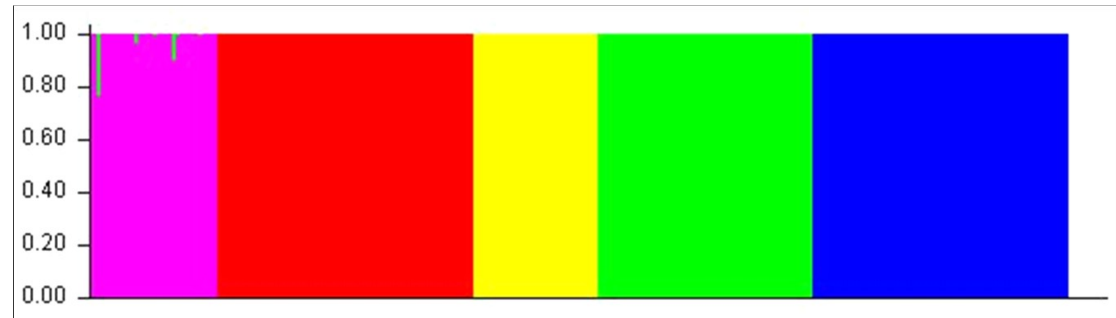


Fig. 16b: **Close bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Reference dataset**

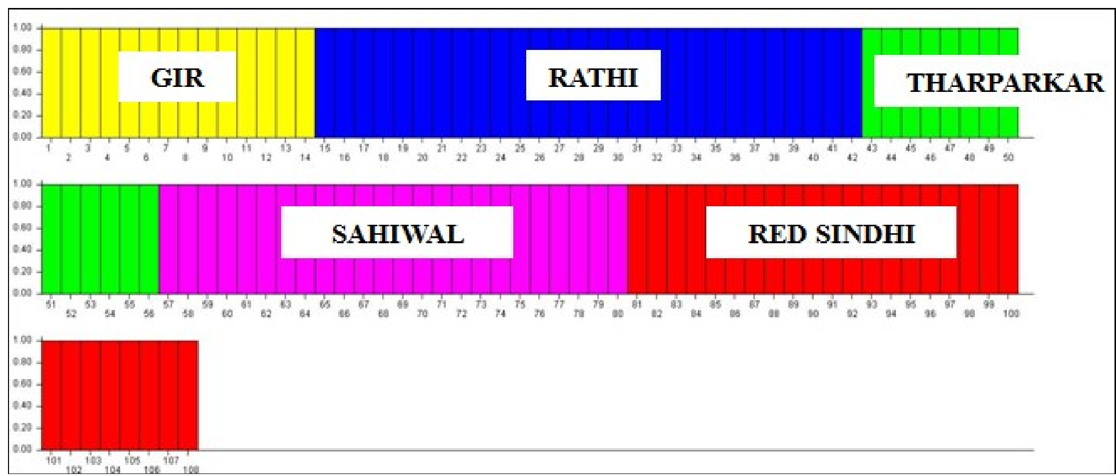


Fig. 17a: Open bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Validation dataset

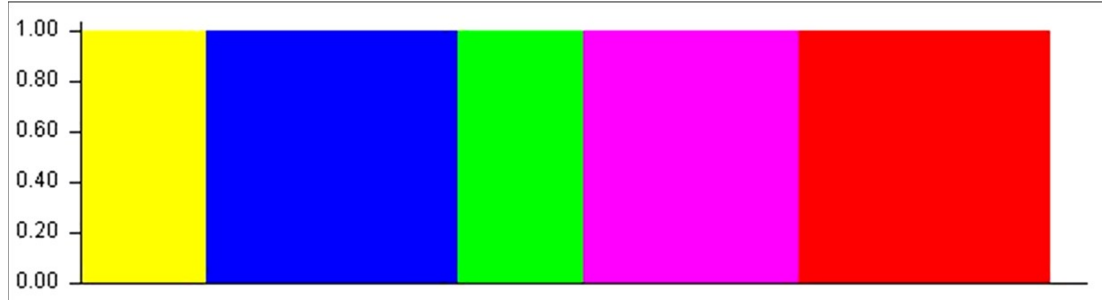


Fig. 17b: Close bar plot showing clustering of Tharparkar cattle along with four other breed populations through STRUCTURE v2.3.4 using 309 markers in the Validation dataset

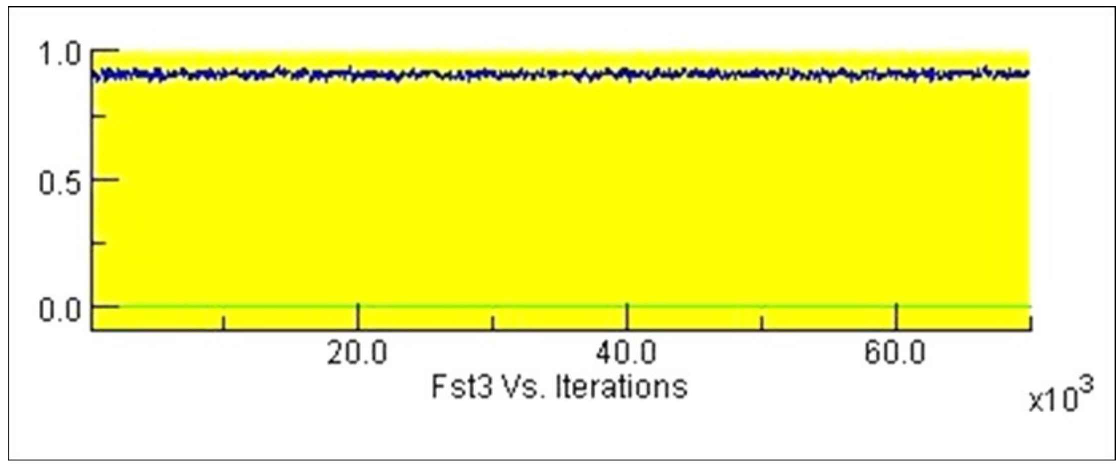


Fig. 18: **F_{ST} plot generated using STRUCTURE v2.3.4, illustrating the differentiation of Tharparkar breed from the other breeds in reference dataset**

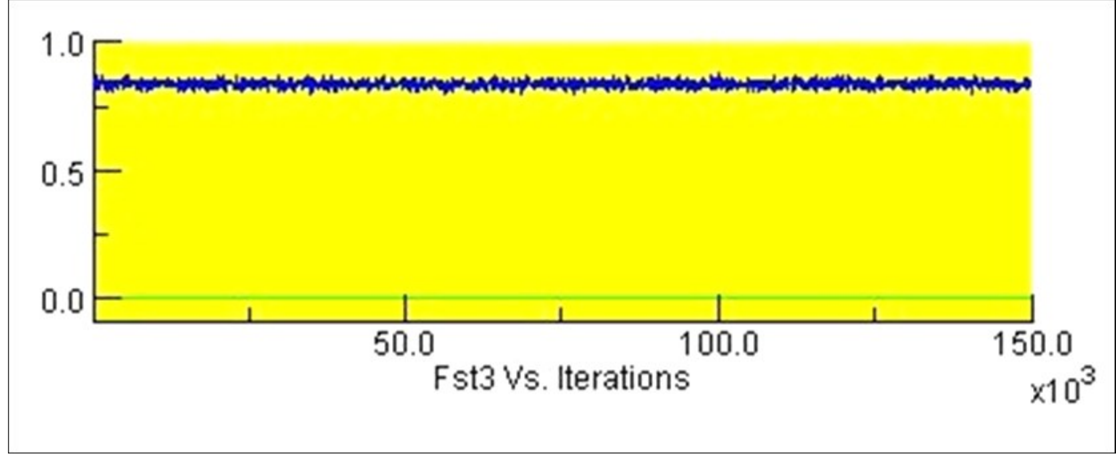
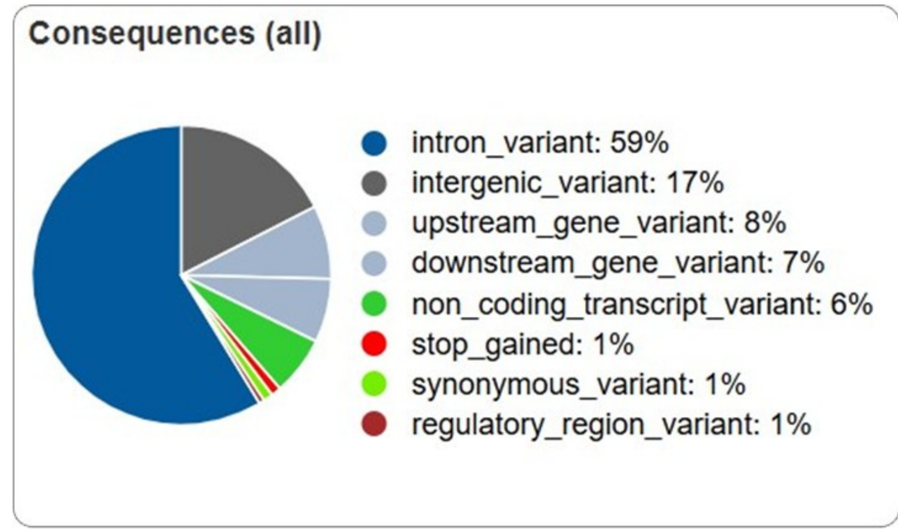


Fig. 19: **F_{ST} plot generated using STRUCTURE v2.3.4, illustrating the differentiation of Tharparkar breed from the other breeds in validation dataset**

Category	Count
Variants processed	77
Variants filtered out	0
Novel / existing variants	30 (39.0) / 47 (61.0)
Overlapped genes	47
Overlapped transcripts	143
Overlapped regulatory features	1



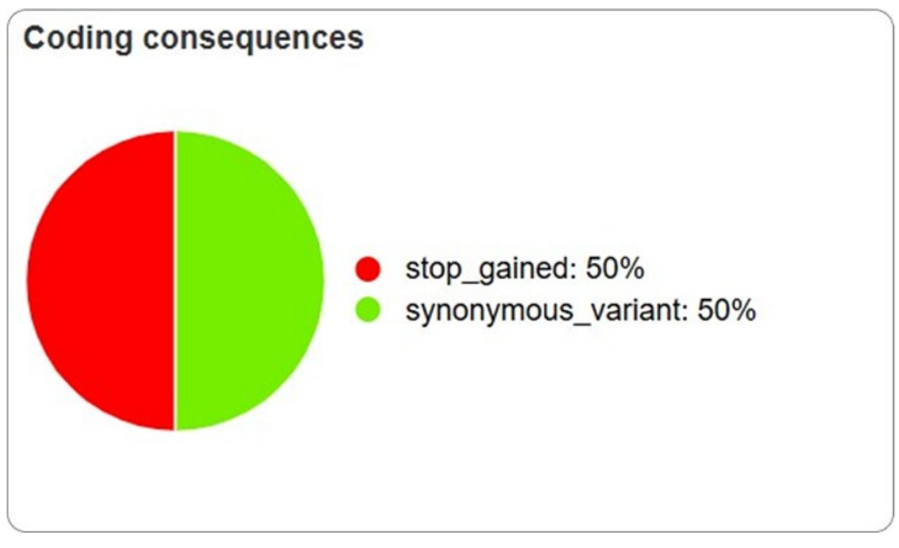


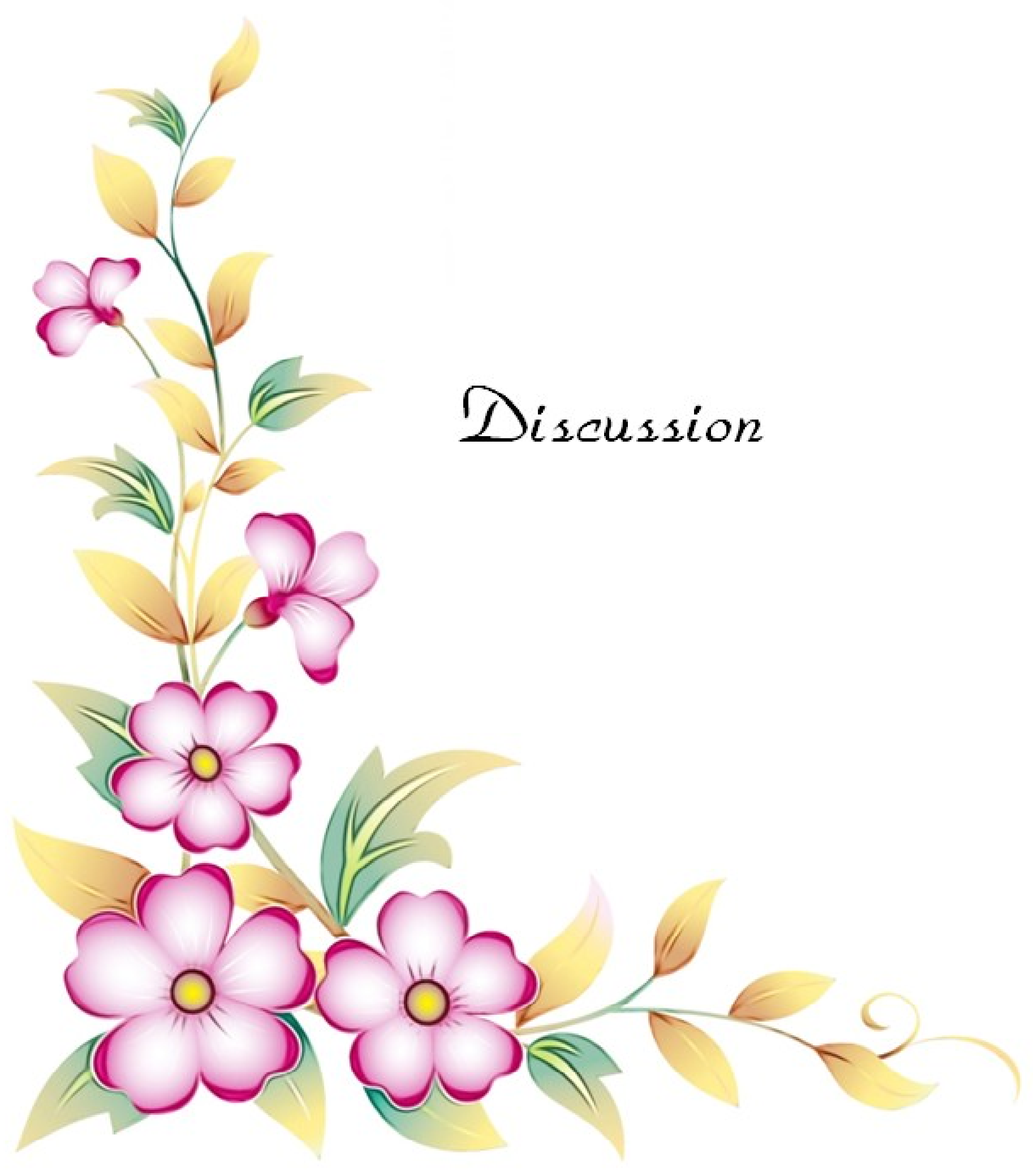
Fig. 20: Summary of functional annotation of 77 breed-specific SNPs in Tharparkar cattle obtained using the Variant Effect Predictor (VEP)

			type 1 motif 17	
chr21	21:19156282	NTRK3	Neurotrophic receptor tyrosine kinase 3	
chr21	21:57267050	RIN3	Ras and Rab interactor 3	
chr22	22:39410266	PTPRG	Protein tyrosine phosphatase receptor type G	
chr23	23:24355061	PKHD1	PKHD1 ciliary IPT domain containing fibrocystin/polyductin	Kumar *et al.*, 2021
chr24	24:33016163	ANKRD29	Ankyrin repeat domain 29	
chr24	24:33016163	LAMA3	Laminin subunit alpha 3	
chr24	24:33016163	ANKRD29	Ankyrin repeat domain 29	
chr25	25:3295294	ADCY9	Adenylate cyclase 9	
chr25	25:41224481	MAD1L1	Mitotic arrest deficient 1 like 1	
chr26	26:28093314	SORCS1	Sortilin related VPS10 domain containing receptor 1	
chr27	27:28434059	NRG1	Neuregulin 1	
chr27	27:38851572	PSD3	Pleckstrin and Sec7 domain containing 3	

Table 8. Summary of QTLs overlapping with breed-specific Tharparkar SNPs and their associated genes

Trait	Chromosome	QTLID	Reference	Gene	
Milk Fat yield	5	QTL #175872	Jiang *et al.*, 2019	PPM1H	
Milk Protein yield	5	QTL #176243	Jiang *et al.*, 2019	PPM1H	
Milk Yield	5	QTL #175964	Jiang *et al.*, 2019	PPM1H	
Inseminations per conception	5	QTL #212557	Galliou *et al.*, 2020	PPM1H	
Milk stearic acid content	5	QTL #32588	Li *et al.*, 2014	CHST11	
Milk C18 index	11	QTL #32360	Li *et al.*, 2014	NRXN1	
Average daily gain	11	QTL #153993	Abo-ismail *et al.*, 2018	ATP6V1E2	
Strength	16	QTL #48260	Cole *et al.*, 2011	SUSD4	6 traits associated and 7 QTL associations

Milk Protein percentage	21	QTL #50977	Cole *et al*., 2011	ADAMTS1	21 traits associated and 33 QTL associations
Age at Puberty	21	QTL #172255	Melo *et al*., 2019	ADAMTS2	
Insemination per conception, Milk fat yield, First service conception	23	QTL #212939, QTL #223344, QTL #212358	Galliou *et al*., 2020, Kim *et al*., 2021	PKHD1	9 traits associated and 11 QTL associations
Inseminations per conception	24	QTL #212955	Galliou *et al*., 2020	LAMA3	
Milk yield	24	QTL #218683	da Cruz *et al*., 2021	LAMA4	
Carcass weight, Milk C14 index, Milk Linoleic acid content	26	QTL #151379, QTL #33452, QTL #34740	Edea *et al*., 2017, Buitenhuis *et al*., 2014	SORCS1	



Discussion

5 DISCUSSION

The primary objective of this study was to provide a comprehensive genome-wide assessment of genetic diversity and develop a breed-specific single nucleotide polymorphism (SNP) panel for Tharparkar, an important indigenous cattle breed of India. This was achieved using double-digest Restriction Site-Associated DNA sequencing (ddRAD-seq) data from 48 Tharparkar cattle. Native to the arid regions of western Rajasthan and the Rann of Kutch in Gujarat, Tharparkar cattle are valued as a dual-purpose breed, contributing both to milk production and draught power. In light of concerns over declining population size and unregulated crossbreeding, this study aimed to quantify genetic variation, assess inbreeding and effective population size and identify breed-specific markers to aid in conservation and genetic improvement efforts.

5.1 Genome-wide diversity analysis

Heterozygosity serves as an important indicator of genetic variability within populations and provides insights into population structure and evolutionary history. In this study, the mean observed heterozygosity and mean expected heterozygosity were estimated at 0.291 ± 0.078 and 0.347 ± 0.103, respectively. These values are broadly comparable to the estimates reported by Saravanan *et al*. (2022) for Tharparkar cattle using the Bovine SNP50 array. However, the lower observed heterozygosity compared to the expected value in this study suggests a heterozygote deficit, possibly due to factors such as assortative mating, population bottlenecks, or recent inbreeding within the sampled farm population. A similar heterozygote deficit was also observed in the Kankrej breed using the ddRAD approach (Venkatesan *et al*., 2025), attributed to small effective population size and unplanned breeding practices.

The mean minor allele frequency (MAF) in this study was 0.249 ± 0.117, aligning with reported MAF estimates (0.21-0.25) in other Indian breeds such as Sahiwal and Nellore (Rajawat *et al*., 2024; Espigolan *et al*., 2013). The polymorphic marker ratio was 0.728, indicating that 72.88% of SNP loci were polymorphic (MAF > 0.1). The presence of adequate polymorphism is crucial for implementing genomic selection, improving breeding programs and conducting reliable genome-wide diversity assessments.

Inbreeding occurs when genetically related individuals mate, leading to an increase in homozygosity (pairing of identical alleles at a locus). This process can substantially reduce genetic diversity and may result in the expression of deleterious recessive traits, adversely affecting health, reproductive efficiency and production traits. Traditionally, inbreeding coefficients were estimated from pedigree records using methods such as path coefficients or probabilities of identity by descent (Wright, 1922). However, these approaches were limited by the accuracy and completeness of pedigree information.

The development of genome-wide tools such as SNP arrays, RAD-seq and whole-genome sequencing has enabled more precise estimation of inbreeding coefficients using genomic data, minimizing errors caused by incomplete or inaccurate pedigree records (Kardos *et al*., 2015). In this study, genomic inbreeding coefficients were estimated using four complementary metrics: F_{HOM} (excess homozygosity), F_{UNI} (correlation between uniting gametes), F_{GRM} (genomic relationship matrix) and F_{ROH} (runs of homozygosity). The estimated values were 0.112 ± 0.074, 0.125 ± 0.089, 0.125 ± 0.082 and 0.1067 ± 0.042, respectively, showing a high degree of correlation among methods. It is noteworthy that, unlike F_{ROH}, the other estimates are influenced by allele frequencies in the base population (Zhang *et al*., 2015). The average genomic inbreeding coefficient obtained in this study is comparable to the estimates reported by Pryce *et al*. (2014) in Holstein Friesian (F_{GRM} = 0.134) and Jersey populations (F_{GRM} = 0.144). However, it is higher than the estimates reported by Masharing *et al*. (2023) using ddRAD data on six indigenous cattle breeds, including Tharparkar (F_{IS} ranging from "0.253 to 0.0513) and by Saravanan *et al*. (2020) using the Bovine SNP50 array on 24 Tharparkar cattle (ranging from 0.0160 to 0.0589).

Runs of Homozygosity (ROH) are continuous stretches of homozygous segments within diploid genomes that arise when individuals inherit chromosomal regions identical by descent (IBD) from a shared ancestor (Broman and Weber, 1999; Gibson *et al*., 2006). These regions capture both recent and ancient inbreeding events and their number and length serve as indicators of population history, autozygosity and demographic processes such as bottlenecks or selection pressures (McQuillan *et al*., 2008; Curik *et al*., 2014). Consequently, ROH have emerged as a powerful genomic tool for estimating inbreeding levels and monitoring genetic diversity in livestock populations.

In the present study, ROH were employed to estimate genomic inbreeding coefficients (F_{ROH}) and to examine the distribution of autozygosity across the Tharparkar cattle genome. A predominance of short homozygous segments within the 0-2 Mb category (81.56%) was observed, suggesting the presence of ancestral or distant inbreeding. This pattern is consistent with findings reported by Saravanan *et al*. (2020) and was further validated here using approximately twice the sample size employed in their study. The F_{ROH} estimate in our analysis was 0.1067 ± 0.042, which is comparable to results from Dixit *et al*. (2020) in indigenous Bos indicus cattle (Kangayam; F_{ROH} = 0.113 ± 0.059) but higher than that reported for Tharparkar cattle (F_{ROH} = 0.054 ± 0.048; n = 17) in the same study. Similarly, Mulim *et al*. (2025) reported a comparable value (F_{ROH} = 0.13) in Brazilian Angus cattle. However, our F_{ROH} estimate exceeded that reported by Rajawat *et al*. (2024) in Sahiwal cattle (0.0249 ± 0.013) and Saravanan *et al*. (2020) using the BovineSNP50K data (0.0589).

Effective population size (Ne) refers to the number of breeding individuals in an idealized population that would exhibit the same level of genetic drift and inbreeding as observed in the actual population (Falconer and Mackay, 1996; Wright, 1969). It plays a crucial role in understanding the dynamics of genetic diversity loss, the fixation of deleterious alleles and the risk of inbreeding depression (Gasca-Pineda *et al*., 2013). In livestock, Ne can be estimated through demographic data, pedigree records, or molecular markers. Among these, Linkage Disequilibrium (LD)-based approaches are commonly employed, particularly when pedigree information is lacking. These methods rely on SNP genotype data to estimate Ne by assessing correlations between alleles at different loci (Sved, 1971; Hill, 1981; Barbato *et al*., 2015).

According to FAO guidelines, populations with an Ne below 50 are considered at risk due to insufficient fitness and increased inbreeding. For long-term adaptability and sustainable breeding programs, an Ne of at least 100 is recommended (Franklin and Frankham, 1998). In the present study, the estimated Ne showed a notable decline-from approximately 2,519 individuals around 150 generations ago to 226 individuals about 13 generations ago and further down to 82.6 in the current generation. These findings align with the results reported by Strucken *et al*. (2021), who observed a similar downward trend in Ne across 13 Indian cattle breeds. Specifically, in the Tharparkar breed (n = 15), Ne estimates declined from 82 in generation 5 to 21 in generation. They also noted that these values might be underestimated due to the limited sample size.

The average pairwise genetic distance (D) provides an estimate of genetic differentiation among individuals within a population, with higher values reflecting greater genetic variability. In the present study, the mean D value was 0.671 ± 0.028, which is higher than the values reported in earlier studies (Freitas *et al*., 2021; Makina *et al*., 2014) but slightly lower, though comparable, to that reported by Rajawat *et al*. (2024) in Sahiwal cattle (0.7642 ± 0.227).

Nucleotide diversity (π), defined as the average number of nucleotide differences per site between any two randomly chosen DNA sequences from a population (Nei and Tajima, 1981), also serves as an indicator of genetic variation. In this study, π was estimated at 0.344 ± 0.073, which is lower than the value reported by Masharing *et al*. (2023) in four Tharparkar cattle (0.458), but comparable to Rajawat *et al*. (2024) in Sahiwal cattle (0.327). However, this estimate was higher than that observed in Chinese Lianshan cattle (0.227) using RADseq data (Mao *et al*., 2021).

5.2 Identification of Breed-specific SNPs

The development of breed-specific SNP panels has been widely explored across livestock species as a strategy to improve genetic characterization, parentage verification and breed assignment accuracy. Several studies have proposed different statistical and computational approaches for selecting informative SNPs from high-density datasets. For instance, Tortereau *et al*. (2017) demonstrated that stringent linkage disequilibrium (LD)-based pruning of pairwise

SNPs using correlation thresholds could effectively reduce redundancy while retaining discriminatory power. Similarly, Ding *et al*. (2011) employed multiple metrics-including Fisher Information Content (FIC), Shannon Information Content (SIC), F-statistics and Absolute Allele Frequency Differences (Δ)-to identify 100-200 highly informative markers capable of discriminating populations. Later, Kavakiotis *et al*. (2015, 2017) developed the Toolbox for Ranking and Evaluation of SNPs (TRES) and the Frequent Item Feature Selection (FIFS) algorithm, which were successfully applied to large-scale genomic datasets such as pigs (446 individuals, 14 sub-populations, 59,436 SNPs), further proving the utility of systematic marker selection.

In addition, STRUCTURE-based evaluations from previous works consistently emphasized the role of F_{ST} as a crucial parameter in assessing breed assignment accuracy. Negrini *et al*. (2009) and Ramos *et al*. (2011) reported that subsets of highly informative markers, when carefully filtered, achieved precise clustering of individuals into their respective populations. Comparative analyses have further shown that while methods such as delta and pairwise Wright's F_{ST} may not always yield consistent clustering patterns, Informativeness for Assignment and FIFS approaches provide more robust results (Kavakiotis *et al*., 2017). Importantly, the Minor Allele Frequency-Linkage Disequilibrium (MAF-LD) method has been highlighted as particularly effective for selecting discriminatory SNPs by balancing allele frequency variation with LD constraints (Kumar *et al*., 2019). Such evidence underscores the growing consensus that breed-specific SNP panels can be reliably constructed using rigorous filtering pipelines.

Building on these earlier studies, the objective of this study was to identify breed-specific SNPs in Tharparkar cattle using the ddRAD-seq approach, with the goal of developing a cost-effective set of informative markers capable of differentiating the breed. Such breed-specific panels serve as reliable DNA-based tools for accurate parentage verification, helping to correct pedigree errors caused by missing or inaccurate records. By incorporating markers tailored to the breed, these panels improve pedigree accuracy, strengthen selection decisions and support effective breeding programs (Strucken *et al*., 2014; Fontanesi *et al*., 2010). The ddRAD-seq method provides a cost-efficient and high-resolution alternative to conventional

genotyping, enabling precise SNP discovery in targeted genomic regions (Peterson *et al*., 2012). Unlike SNP arrays-which are often affected by ascertainment bias and typically designed using taurine cattle SNPs (DeDonato *et al*., 2013) ddRAD-seq can detect novel and rare variants, while allowing genotyping of a larger number of animals at a lower cost. This broader coverage of the reference population helps capture greater within-breed variability, enhancing the robustness of downstream analyses.

Several approaches have been used to identify markers capable of differentiating breeds, with Principal Component Analysis (PCA) being one of the most common. PCA effectively reduces high-dimensional SNP datasets to manageable sizes while retaining sufficient analytical resolution (Bertolini *et al*., 2015; Andrews *et al*., 2018). Kumar *et al*. (2019) applied five different methods-Delta, Pairwise Wright's F_{ST}, Informativeness for Assignment, Frequent Item Feature Selection (FIFS) and Minor Allele Frequency-Linkage Disequilibrium (MAF-LD) to select breed-discriminatory SNPs and reduce marker numbers for developing breed-specific panels.

A SNP is considered breed-specific if it is polymorphic within the target breed but has one allele fixed in all other breeds. The MAF-LD method (Kumar *et al*., 2019) has proven effective for reducing SNP numbers, as higher MAF values are often associated with breed differentiation under balancing selection, which maintains different alleles in different populations to enhance genetic divergence. Incorporating LD analysis further helps to avoid selecting SNPs in high linkage disequilibrium, improving panel efficiency. However, relying solely on LD-based filtering risks excluding highly informative markers.

In the present study, for the first time, the ddRAD-seq method was used to mine SNPs in Tharparkar cattle and attempt the development of a breed-specific panel. Pre-selection strategies were employed to minimize redundancy within breeds and reduce SNP density by nearly one-third while preserving discriminatory capacity. Using this approach, 309 SNPs were identified across 260 individuals, demonstrating strong performance on a validation set of 108 animals, with mean F_{ST} values across clusters ranging from 0.8155 to 0.9961. The Tharparkar breed showed an F_{ST} of 0.8269 and 77 breed-specific SNPs were successfully identified for this breed. Variant Effect Predictor (VEP) annotation revealed that 39% of these

SNPs were novel, distributed across 47 genes, 143 transcripts and one regulatory feature. Most variants were intronic (59%), followed by other non-coding classes, with a small proportion in coding regions. Several SNPs overlapped with QTLs for milk production, reproduction and growth traits, underlining their potential utility in breeding and selection strategies.

Taken together, the integration of ddRAD-seq-based SNP discovery with rigorous filtering approaches has proven successful in constructing a cost-effective and highly informative SNP panel for Tharparkar cattle. The results corroborate earlier studies on informative marker selection while extending the methodology to an indigenous zebu breed. The breed-specific SNPs identified here hold promise for practical applications in genetic resource management, conservation of indigenous cattle and implementation of precision breeding strategies.

✍✍✍

Summary and Conclusions

6 SUMMARY AND CONCLUSIONS

Amid growing concerns of an impending mass extinction, conservation genetics has become an essential approach for maintaining genetic diversity and protecting species against environmental challenges. Cattle, which are vital for human sustenance, are increasingly threatened by intensive breeding practices that prioritize productivity over genetic variation. Despite existing conservation efforts, many livestock breeds, including cattle, are experiencing significant population declines.

Recognizing the critical role of genetic diversity in breed preservation and improvement, this study aimed to comprehensively assess the genetic landscape of Tharparkar cattle—a breed widely valued for its adaptability to arid conditions and economic importance in dairy farming. Using the ddRAD sequencing technique, we sought to evaluate genetic diversity parameters and identify breed-specific SNPs that can guide informed breeding strategies and promote sustainable livestock management, thereby ensuring the long-term resilience of this breed.

For this study, 48 Tharparkar cattle samples were analyzed through ddRAD sequencing. Several intra-population genetic diversity parameters were calculated, including observed and expected heterozygosity (Ho and He), polymorphic SNP proportion (PSNP), average pairwise genetic distance (D), nucleotide diversity (π), Runs of Homozygosity (ROH), genomic inbreeding coefficients (F_{ROH}, F_{GRM}, F_{UNI}, F_{HOM}) and minor allele frequency (MAF).

Raw paired-end reads (R1 and R2) were generated for all 48 individuals and processed through FastQC v0.12.1 for quality assessment, followed by trimming for base quality and

adapter removal. High-quality reads were aligned to the Bos taurus reference genome (ARS-UCD 2.0), producing SAM files, which were subsequently converted to BAM and VCF formats for downstream analyses. Only SNPs meeting stringent criteria (Phred score $\geq$ 30, Read Depth $\geq$ 10, missingness $\leq$ 10% and MAF $\geq$ 0.05) were retained using PLINK v1.9, resulting in 121,554 high-quality SNPs. Structural and functional annotation using SnpEff revealed 581,495 effects across the genome, with the majority of variants located in intronic and intergenic regions. The transition/transversion (Ts/Tv) ratio was 2.47, showing a higher proportion of transitions.

Within the population, observed and expected heterozygosity values were 0.291 $\pm$ 0.078 and 0.347 $\pm$ 0.103, respectively, indicating slightly lower observed heterozygosity compared to expectations. The mean MAF across autosomes was 0.216 $\pm$ 0.143, with most SNPs falling in the 0.1-0.2 range (23.74%), while the lowest proportion (9.84%) was observed in the 0-0.05 range. The polymorphic marker ratio was 0.728, demonstrating that about 73% of SNP sites were polymorphic (MAF > 0.1). The average genetic distance (D) between individuals was 0.671 $\pm$ 0.028, reflecting substantial heterogeneity. Nucleotide diversity (π) was estimated at 0.344 $\pm$ 0.073.

Effective population size (Ne), estimated using the NeEstimator v2.01 software under a random mating model, was 82.6 at a critical MAF threshold of 0.05. Historical trend analysis using the SNeP package revealed a sharp decline in Ne—from approximately 2519 nearly 150 generations ago to just 226 around 13 generations ago.

A total of 4373 ROHs were identified across autosomes using a sliding window approach, spanning 154.66 Mb (6.22% of total autosomal length). Most ROHs were short (0-2 Mb, 81.56%), followed by 2-4 Mb (14.56%), 4-8 Mb (3.53%) and 8-16 Mb (0.1%). The genome-wide inbreeding coefficient (F_{ROH}) was 0.1067 $\pm$ 0.042, while other estimates (F_{HOM}, F_{UNI} and F_{GRM}) were 0.112 $\pm$ 0.074, 0.125 $\pm$ 0.089 and 0.125 $\pm$ 0.082, respectively, suggesting low to moderate levels of inbreeding.

In recent years, SNPs have increasingly replaced microsatellites in livestock genetics for applications such as parentage testing, breed composition analysis, individuality assessment

and genomic selection. However, cost-effective SNP resources tailored for cattle breed identification remain scarce. Developing such tools could significantly enhance breeding programs and conservation strategies.

To address this, the analysis was expanded by incorporating ddRAD genotypic data of 368 animals from five cattle breeds—Gir (48), Sahiwal (81), Red Sindhi (96), Tharparkar (47) and Rathi (96)—sourced from our in-house repository. The dataset was divided into a reference set (260 individuals) and a validation set (108 individuals). The TRES tool was applied for SNP pre-selection using multiple methods including Delta, pairwise Wright's F_{ST}, Informativeness-for-assignment and MAF-LD.

The final SNP panel, consisting of 309 SNPs across 260 individuals, was evaluated using STRUCTURE software which achieved the highest clustering accuracy. Breed-specific markers successfully allocated individuals to their respective breeds in both reference and validation datasets. From this panel, 77 SNPs specific to Tharparkar cattle were annotated in detail. These markers overlapped with 47 genes in the Tharparkar reference genome.

Comparative analysis with the Cattle QTLdb revealed associations between these SNPs and key genes such as PPM1H, CHST11, NRXN1, ATP6V1E2, SUSD4, ADAMTS17, PKHD1, LAMA3 and SORC1, which are linked to traits like milk yield, milk fat and protein content, milk fatty acid composition (C18, C14 and linolenic acid indices), reproductive efficiency (age at puberty, insemination success) and growth performance (carcass weight, average daily gain and strength).

Overall, this study highlights the rich genetic diversity and moderate inbreeding levels of Tharparkar cattle, while identifying informative SNPs with potential applications in breed identification, conservation and genomic selection. These findings provide a strong foundation for implementing cost-effective and sustainable breeding strategies to secure the genetic future of this valuable indigenous breed.

CONCLUSION

- Inbreeding coefficients (0.1067 to 0.125) indicate low to moderate inbreeding, highlighting the need for conservation measures and structured breeding plans.

- ROH analysis showed that 81.56% were short (0-2 Mb), suggesting predominantly distant inbreeding and historical relationships among sampled individuals.
- The average inbreeding coefficient (F_{ROH}) was 0.1067 ± 0.042, while other estimates were F_{HOM} = 0.112 ± 0.074, F_{UNI} = 0.125 ± 0.089 and F_{GRM} = 0.125 ± 0.082, confirming low to moderate levels of inbreeding.
- Effective population size (Ne) has declined from 2519 to 226, with a recent estimate of 82.3, likely due to intense selection, genetic drift and selective breeding.
- The declining Ne reflects domestication's shift from larger native populations to smaller breed-specific groups, underscoring the need to preserve sufficient Ne to mitigate negative effects of inbreeding.
- A panel of 309 SNPs from five indigenous cattle breeds was developed using TRES and MAF-LD pre-selection methods and these SNPs were breed-discriminatory when tested on both reference and validation datasets using STRUCTURE software.
- For Tharparkar cattle, 77 breed-specific SNPs were identified that clearly differentiated it from other breeds in both reference and validation datasets.

✍✍✍

Mini Abstract

7 MINI ABSTRACT

Tharparkar, an indigenous milch breed of India, is recognized for its superior milk production and remarkable adaptability to arid and desert climates, playing a crucial role in supporting the livelihoods of rural communities. This study aimed to estimate the genetic diversity using double digest restriction site-associated DNA (ddRAD) sequencing and to identify breed-specific single nucleotide polymorphisms (SNPs) for developing an ultra-low-density SNP panel tailored to Tharparkar cattle. A total of 48 animals from the Livestock Research Station (LRS), Beechwal, RAJUVAS, Bikaner, Rajasthan, were genotyped, resulting in 121,554 high-quality biallelic SNPs after stringent filtering. The population showed an average minor allele frequency of 0.216 ± 0.143, with expected and observed heterozygosity values of 0.347 ± 0.103 and 0.291 ± 0.078, respectively. The proportion of polymorphic SNPs was 0.728. Nucleotide diversity was estimated to be 0.344 ± 0.073 and the average pairwise genetic distance (D) was 0.671 ± 0.028. Genomic inbreeding coefficients estimated using different methods (FHOM, FUNI, FGRM and FROH) ranged from 0.106 to 0.125, indicating a low to moderate level of inbreeding. The predominance of short runs of homozygosity (0–2 Mb) suggests limited recent inbreeding, with inbreeding primarily attributable to ancient shared ancestry, likely caused by genetic drift or founder effects. The effective population size in recent generations was estimated at 82.6. The application of a cost-effective tool for breed identification would enable better implementation of breeding methodologies and livestock conservation strategies. To identify breed-specific informative SNPs, pre-selection statistics and the TRES (Toolbox for Ranking and Evaluation of SNPs) tool were applied. TRES uses Delta, pairwise Wright's Fst and the informativeness-for-assignment algorithm, followed by the application of the MAF-LD method across a dataset of 260 individuals from five different Indian cattle breeds. This approach led to the identification of a total of 309 SNPs across 260 individuals in the reference dataset, including 77 unique SNPs specific to Tharparkar. STRUCTURE analysis confirmed that these informative SNPs could accurately cluster individuals into their respective breeds in both reference and validation datasets, thereby effectively distinguishing Tharparkar from other breeds. These findings provide valuable insights into the genetic diversity of Tharparkar cattle, demonstrate the utility of ddRAD sequencing for identifying breed-specific SNPs and underscore the need for targeted conservation strategies to ensure the breed's long-term sustainability.

लघु सारांश

8 लघु सारांश

थारपारकर भारत की एक देशी दुग्धार्जक नस्ल है, जो अधिक दूध उत्पादन और शुष्क व रेगिस्तानी जलवायु में ढलने की क्षमता के लिए जानी जाती है। यह नस्ल ग्रामीण परिवारों की आजीविका में अहम योगदान देती है। इस अध्ययन का उद्देश्य डबल डाइजेस्ट रेस्ट्रिक्शन साइट–असोसिएटेड डीएनए (ddRAD) सीक्वेंसिंग तकनीक का उपयोग करके थारपारकर नस्ल में आनुवंशिक विविधता का अनुमान लगाना और इस नस्ल के लिए कम घनत्व SNP पैनल तैयार करना था। इसके लिए नस्ल विशिष्ट SNPs की पहचान की गई। बीकानेर, राजस्थान के लाइवस्टॉक रिसर्च स्टेशन से 48 पशुओं का चयन किया गया और उनका जीनोटाइप किया गया। जांच के बाद 1,21,554 उच्च गुणवत्ता वाले SNP's प्राप्त हुए। औसत माइनर एलील फ्रक्वेंसी (MAF) 0.216±0.143 पाई गई। अपेक्षित हेटेरोजायगोसिटी 0.347±0.103 और प्रेक्षित हेटेरोज़ायगोसिटी 0.291±0.078 रही। पॉलिमॉर्फिक SNPs का अनुपात 0.728 था। न्यूक्लियोटाइड विविधता 0.344±0.073 और औसत आनुवंशिक दूरी 0.671±0.028 दर्ज की गई। इनब्रिडिंग गुणांक अलग–अलग तरीकों से 0.106 से 0.125 के बीच रहा, जो कम से मध्यम स्तर का इनब्रिडिंग दर्शाता है। छोटे रन ऑफ होमोसायगोसिटी (0-2 Mb) अधिक मात्रा में पाए गए, जिससे पता चला कि हाल के समय में इनब्रिडिंग सीमित रही। अधिकांश इनब्रिडिंग पुरानी साझा दशानुगति से जुड़ी थी, जो संभवतः जेनेटिक ड्रिफ्ट या फाउंडर इमेक्ट्स के कारण हुई। हाल की पीढ़ियों में प्रभावी जनसंख्या का आकार (Ne) 82.6 पाया गया। नस्ल की पहचान के लिए सस्ता और प्रभावी तरीका प्रजनन और संरक्षण योजनाओं को बेहतर बना सकता है। नस्ल विशिष्ट SNPs की पहचान करने के लिए प्री–सेलेक्शन सांख्यिकी और TRES टूल का प्रयोग किया गया। इसमें Delta, Wright's Fst और Informativeness for assignment जैसे तरीके शामिल थे। इसके बाद MAF-LD विधि को पांच भारतीय नस्लों के 260 पशुओं के डेटा पर लागू किया गया। इस प्रक्रिया से कुल 309 SNPs मिले, जिनमें से 77 सिर्फ थारपारकर में पाए गए। STRUCTURE विश्लेषण ने दिखाया कि ये SNPs अलग–अलग नस्लों के पशुओं को सही ढंग से पहचान सकते हैं और थारपारकर को अन्य नस्लों से अलग कर सकते हैं। इस अध्ययन से थारपारकर नस्ल की आनुवंशिक विविधता के बारे में महत्वपूर्ण जानकारी मिली। यह भी साबित हुआ कि ddRAD सीक्वेंसिंग नस्ल–विशिष्ट SNPs की पहचान में उपयोगी है। परिणाम बताते हैं कि नस्ल की स्थिरता और संरक्षण के लिए उचित रणनीतियाँ बनाना जरूरी है।

References

9 REFERENCES

Abo-Ismail, M.K., Lansink, N., Akanno, E., Karisa, B.K., Crowley, J.J., Moore, S.S., Bork, E., Stothard, P., Basarab, J.A. and Plastow, G.S., 2018. Development and validation of a small SNP panel for feed efficiency in beef cattle. Journal of animal science, **96**(2): 375-397.

Ahmad, S.F., Panigrahi, M. and Bhushan, B., 2018. Breed clustering analysis of six Indian cattle breeds using genome-wide SNP data. WORLDWIDE RESEARCH INITIATIVES FOR AGRICULTURE, SCIENCE & TECHNOLOGY.

Ajmone Marsan, P., Garcia, J. F., & Lenstra, J. A. (2010). On the origin of cattle: how aurochs became cattle and colonized the world. Evolutionary Anthropology: Issues, News and Reviews, **19**(4): 148-157.

Alemu, S.W., Kadri, N.K., Harland, C., Faux, P., Charlier, C., Caballero, A. and Druet, T., 2021. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. Heredity, **126**(3): 410-423.

Alhussien, M., Manjari, P., Mohammed, S., Sheikh, A.A., Reddi, S., Dixit, S. and Dang, A.K., 2016. Incidence of mastitis and activity of milk neutrophils in Tharparkar cows reared under semi-arid conditions. Tropical Animal Health and Production, **48**: 1291-1295.

Amos, W. and Harwood, J., 1998. Factors affecting levels of genetic diversity in natural populations. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, **353**(1366): 177-186.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. and Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, **17**(2): 81-92.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C. and Wingett, S., 2022. FastQC: a quality control tool for high throughput sequence data [Online]. 2010 [online] http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Baker, C.A. and Manwell, C. 1980. Chemical classification of cattle. 1.Breed groups. Animal Blood Groups Biochem. Genet. **11**(2): 127-150.

Barbato, M., Orozco-terWengel, P., Tapio, M. and 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.

Bejarano, D., Martínez, R., Manrique, C., Parra, L.M., Rocha, J.F., Gómez, Y., Abuabara, Y. and Gallego, J., 2018. Linkage disequilibrium levels and allele frequency distribution in Blanco Orejinegro and Romosinuano Creole cattle using medium density SNP chip data. Genetics and Molecular Biology, **41**: 426-433.

Berry, D.P., 2018. Symposium review: Breeding a better cow—Will she be adaptable?. Journal of Dairy Science, **101**(4): 3665-3685.

Bertolini, F., Galimberti, G., Calo, D.G., Schiavo, G., Matassino, D. and Fontanesi, L., 2015. Combined use of principal component analysis and random forests identify population informative single nucleotide polymorphisms: application in cattle breeds. Journal of Animal Breeding and Genetics, **132**(5): 346-356.

Bertolini, F., Galimberti, G., Schiavo, G., Mastrangelo, S., Di Gerlando, R., Strillacci, M.G., Bagnato, A., Portolano, B. and Fontanesi, L., 2018. Preselection statistics and Random Forest classification identify population informative single nucleotide polymorphisms in cosmopolitan and autochthonous cattle breeds. Animal, **12**(1): 12-19.

Bhanuprakash, V., Chhotaray, S., Pruthviraj, D.R., Rawat, C., Karthikeyan, A. and Panigrahi, M., 2018. Copy number variation in livestock: a mini review. Veterinary World, **11**(4): p.535.

Bharati, J., Dangi, S.S., Chouhan, V.S., Mishra, S.R., Bharti, M.K., Verma, V., Shankar, O., Yadav, V.P., Das, K., Paul, A. and Bag, S., 2017. Expression dynamics of HSP70 during chronic heat stress in Tharparkar cattle. International journal of biometeorology, **61**: 1017-1027.

Bhardwaj, S., Togla, O., Mumtaz, S., Yadav, N., Tiwari, J., Muansangi, L., Illa, S.K., Wani, Y.M., Mukherjee, S. and Mukherjee, A., 2024. Comparative assessment of the

effective population size and linkage disequilibrium of Karan Fries cattle revealed viable population dynamics. Animal Bioscience, **37**(5): 795.

Bhat, S., Kumar, P., Kashyap, N., Deshmukh, B., Dige, M.S., Bhushan, B., Chauhan, A., Kumar, A. and Singh, G., 2016. Effect of heat shock protein 70 polymorphism on thermotolerance in Tharparkar cattle. Veterinary world, **9**(2): p. 113.

Biscarini, F., Cozzi, P., Gaspa, G. and Marras, G., 2019. Detect runs of homozygosity and runs of heterozygosity in diploid genomes.

Bjelland, D.W., Weigel, K.A., Vukasinovic, N. and Nkrumah, J.D., 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. Journal of dairy science, **96**(7): 4697-4706.

Bohmanova, J., Sargolzaei, M. and Schenkel, F.S., 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC genomics, **11**: 1-11.

Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, **30**(15): 2114-2120.

Bollongino, R., Burger, J., Powell, A., Mashkour, M., Vigne, J.D. and Thomas, M.G., 2012. Modern taurine cattle descended from small number of near-eastern founders. Molecular biology and evolution, **29**(9): 2101-2104.

Brito, L.F., Kijas, J.W., Ventura, R.V., Sargolzaei, M., Porto-Neto, L.R., Cánovas, A., Feng, Z., Jafarikia, M. and Schenkel, F.S., 2017. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC genomics, **18**: 1-20.

Broman, K.W. and Weber, J.L., 1999. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. The American Journal of Human Genetics, **65**(6): 1493-1500.

Browning, B.L., Zhou, Y. and Browning, S.R., 2018. A one-penny imputed genome from next-generation reference panels. The American journal of human genetics, **103**(3): 338-348.

Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics, **4** (11): 900-910.

Buitenhuis, B., Janss, L.L., Poulsen, N.A., Larsen, L.B., Larsen, M.K. and Sørensen, P., 2014. Genome-wide association and biological pathway analysis for milk-fat

composition in Danish Holstein and Danish Jersey cattle. BMC genomics, **15**(1): 1112.

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M. and Wilson, J.F., 2018. Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics, **19**(4): 220-234.

Chagunda, M.G., Mujibi, F.D., Dusingizimana, T., Kamana, O., Cheruiyot, E. and Mwai, O.A., 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. Frontiers in genetics, **9**: p.438.

Chhotaray, S., Panigrahi, M., Pal, D., Ahmad, S.F., Bhanuprakash, V., Kumar, H., Parida, S., Bhushan, B., Gaur, G.K., Mishra, B.P. and Singh, R.K., 2021. Genome-wide estimation of inbreeding coefficient, effective population size and haplotype blocks in Vrindavani crossbred cattle strain of India. Biological Rhythm Research, **52**(5): 666-679.

Choudhary, G., Pannu, U., Gahlot, G.C., Nehara, M. and Poonia, N.K., 2018. Estimation of Genetic, Phenotypic and Environmental Trends in Various Production and Reproduction Traits of Tharparkar Cattle at Organized Farm at Bikaner. Journal of Animal Research, **8**(6): 1027-1034.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. fly, **6**(2): 80-92.

Cole, J.B., Wiggans, G.R., Ma, L., Sonstegard, T.S., Lawlor Jr, T.J., Crooker, B.A., Van Tassell, C.P., Yang, J., Wang, S., Matukumalli, L.K. and Da, Y., 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. BMC genomics, **12**(1): 408.

Curik, I., Ferenèakoviæ, M. and Sölkner, J., 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livestock science, **166**: 26-34.

Da Cruz, A.S., Silva, D.C., Minasi, L.B., de Farias Teixeira, L.K., Rodrigues, F.M., da Silva, C.C., do Carmo, A.S., da Silva, M.V.G.B., Utsunomiya, Y.T., Garcia, J.F. and da Cruz, A.D., 2021. Single-Nucleotide Polymorphism variations associated with

specific genes putatively identified enhanced genetic predisposition for 305-day milk yield in the Girolando crossbreed. Frontiers in Genetics, **11**: 573344.

Dadousis, C., Ablondi, M., Cipolat-Gotet, C., Van Kaam, J.T., Marusi, M., Cassandro, M., Sabbioni, A. and Summer, A., 2022. Genomic inbreeding coefficients using imputed genotypes: Assessing different estimators in Holstein-Friesian dairy cows. Journal of dairy science, **105**(7): 5926-5945.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. and Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience, **10**(2): 008.

Dash, S. 2016. Genome wide SNP mining and diversity analysis of indigenous dairy cattle breeds. ICAR-NDRI PhD thesis.

Dash, S., Singh, A., Bhatia, A.K., Jayakumar, S., Sharma, A., Singh, S., Ganguly, I. and Dixit, S.P., 2018. Evaluation of bovine high-density SNP genotyping array in indigenous dairy cattle breeds. Animal biotechnology, **29**(2): 129-135.

Dash, S.R., Sahoo, H.K., Padhy, P.K. and Pattanaik, B.R., 2022. Characteristics of Motu cattle, a draft cattle breed of Odisha, India. Journal of Livestock Science, (**13**).

Davey, J.W., Cezard, T., Fuentes Utrilla, P., Eland, C., Gharbi, K. and Blaxter, M.L., 2013. Special features of RAD Sequencing data: implications for genotyping. Molecular ecology, **22**(11): 3151-3164.

De Donato, M., Peters, S.O., Mitchell, S.E., Hussain, T. and Imumorin, I.G., 2013. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PloS one, **8**(5): p.e62137.

Dikmen, S.E.R.D.A.L., Alava, E., Pontes, E., Fear, J.M., Dikmen, B.Y., Olson, T.A. and Hansen, P.J., 2008. Differences in thermoregulatory ability between slick-haired and wild-type lactating Holstein cows in response to acute heat stress. Journal of dairy science, **91**(9): 3395-3402.

Ding, L., Wiener, H., Abebe, T., Altaye, M., Go, R.C., Kercsmar, C., Grabowski, G., Martin, L.J., Khurana Hershey, G.K., Chakorborty, R. and Baye, T.M., 2011. Comparison of measures of marker informativeness for ancestry and admixture mapping. BMC genomics, **12**(1): 622.

Dixit, S.P., Singh, S., Ganguly, I., Bhatia, A.K., Sharma, A., Kumar, N.A., Dang, A.K. and Jayakumar, S., 2020. Genome-wide runs of homozygosity revealed selection signatures in Bos indicus. Frontiers in Genetics, **11:** 92.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J. and Ovenden, J.R., 2014. NeEstimator v2: re implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular ecology resources, **14**(1): 209-214.

Doekes, H.P., Veerkamp, R.F., Bijma, P., de Jong, G., Hiemstra, S.J. and Windig, J.J., 2019. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. Genetics Selection Evolution, **51**(1): 54.

Earl, D.A. and VonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources, **4**(2): 359-361.

Edea, Z., Jeoung, Y.H., Shin, S.S., Ku, J., Seo, S., Kim, I.H., Kim, S.W. and Kim, K.S., 2017. Genome-wide association study of carcass weight in commercial Hanwoo cattle. Asian-Australasian journal of animal sciences, **31**(3): 327.

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. and Schenkel, F.S., 2013. Study of whole genome linkage disequilibrium in Nellore cattle. BMC genomics, **14**(1): 305.

Evanno, G., Regnaut, S. and Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology, **14**(8): 2611-2620.

Falconer, D.S., 1996. Introduction to quantitative genetics. Pearson Education India.

Fisher, P.J., Malthus, B., Walker, M.C., Corbett, G. and Spelman, R.J., 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. Journal of dairy science, **92**(1): 369-374.

Flury, C., Tapio, M., Sonstegard, T., Drögemüller, C., Leeb, T., Simianer, H., Hanotte, O. and Rieder, S., 2010. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. Journal of Animal Breeding and Genetics, **127**(5): 339-347.

Fontanesi, L., Scotti, E. and Russo, V., 2010. Analysis of SNPs in the KIT gene of cattle with different coat colour patterns and perspectives to use these markers for breed traceability and authentication of beef and dairy products. Italian Journal of Animal Science, **9**(2): 42.

Franklin, I.R. and Frankham, R., 1998, February. How large must populations be to retain evolutionary potential?. In Animal Conservation Forum (Vol. 1, No. 1, pp. 69-70). Cambridge University Press.

Freitas, P.H., Wang, Y., Yan, P., Oliveira, H.R., Schenkel, F.S., Zhang, Y., Xu, Q. and Brito, L.F., 2021. Genetic diversity and signatures of selection for thermal stress in cattle and other two Bos species adapted to divergent climatic conditions. Frontiers in genetics, **12:** 604823.

Galliou, J.M., Kiser, J.N., Oliver, K.F., Seabury, C.M., Moraes, J.G., Burns, G.W., Spencer, T.E., Dalton, J. and Neibergs, H.L., 2020. Identification of loci and pathways associated with heifer conception rate in US Holsteins. Genes, **11**(7): 767.

Gasca-Pineda, J., Cassaigne, I., Alonso, R.A. and Eguiarte, L.E., 2013. Effective population size, genetic variation and their relevance for conservation: the bighorn sheep in Tiburon Island and comparisons with managed artiodactyls. PLoS One, **8**(10): p.e78120.

Gebrehiwot, N.Z., Strucken, E.M., Marshall, K., Aliloo, H. and Gibson, J.P., 2021. SNP panels for the estimation of dairy breed proportion and parentage assignment in African crossbred dairy cattle. Genetics Selection Evolution, **53**(1): 21.

Ghildiyal, K., Nayak, S.S., Rajawat, D., Sharma, A., Chhotaray, S., Bhushan, B., Dutt, T. and Panigrahi, M., 2023. Genomic insights into the conservation of wild and domestic animal diversity: a review. Gene, **886**: p.147719.

Ghoreishifar, S.M., Moradi-Shahrbabak, H., Parna, N., Davoudi, P. and Khansefid, M., 2019. Linkage disequilibrium and within-breed genetic diversity in Iranian Zandi sheep. Archives Animal Breeding, **62**(1): 143-151.

Gibson, J., Morton, N.E. and Collins, A., 2006. Extended tracts of homozygosity in outbred human populations. Human molecular genetics, **15**(5): 789-795.

Goszczynski, D., Molina, A., Terán, E., Morales-Durand, H., Ross, P., Cheng, H., Giovambattista, G. and Demyda-Peyrás, S., 2018. Runs of homozygosity in a selected cattle population with extremely inbred bulls: Descriptive and functional analyses revealed highly variable patterns. PLoS One, **13**(7): 0200069.

Götherström, A. anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D. G., & Ellegren, H. (2005). Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. Proceedings of the Royal Society of London B: Biological Sciences, **272**(1579): 2345-2351.

Gregory, T.R., 2009. Understanding natural selection: essential concepts and common misconceptions. Evolution: Education and outreach, **2**: 156-175.

Groeneveld , K.R., Adams, J.R., Cassirer, E.F., Plowright, R.K., Gardner, C., Dwire, M., Hohenlohe, P.A. and Waits, L.P., 2018. A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RAD seq data. Molecular ecology resources, **18**(6): 1263-1281.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E.J.A.G. and Weigend, S., 2010. Genetic diversity in farm animals-a review. Animal genetics, **41**: 6-31.

Gutiérrez-Reinoso, M.A., Aponte, P.M. and García-Herreros, M., 2022. A review of inbreeding depression in dairy cattle: current status, emerging control strategies, and future prospects. Journal of Dairy Research, **89**(1): 3-12.

Halvoník, A., Moravèíková, N., Chalupková, M. and Kasarda, R., 2024. Commonly used genomic estimators of individual inbreeding in livestock. Czech Journal of Animal Science, **69**(7).

Hedrick, P.W. and Garcia-Dorado, A., 2016. Understanding inbreeding depression, purging and genetic rescue. Trends in ecology & evolution, **31**(12): 940-952.

Hill, W.G., 1981. Estimation of effective population size from data on linked genes. Advances in Applied Probability, **13**(1): 4-4.

https://bioinfogp.cnb.csic.es/tools/venny/

https://dahd.gov.in/document/annual_report

https://dahd.gov.in/sites/default/files/2024-11/BAHS-2024

https://nbagr.icar.gov.in/en/registered-cattle/

https://rajuvas.org/livestock-research-station/

https://www.ebi.ac.uk/ena/browser/home (European Nucleotide Archive)

https://www.indiabudget.gov.in/economicsurvey/

https://www.ncbi.nlm.nih.gov/sra (NCBI Sequence Read Archive)

Hu, Z.L., Park, C.A., Wu, X.L. and Reecy, J.M., 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic acids research, **41**(D1): D871-D879.

Hulsegge, B., Calus, M.P.L., Windig, J.J., Hoving-Bolink, A.H., Maurice-van Eijndhoven, M.H.T. and Hiemstra, S.J., 2013. Selection of SNP from 50K and 777K arrays to predict breed of origin in cattle. Journal of Animal Science, **91**(11): 5128-5134.

Hulsegge, I., Schoon, M., Windig, J., Neuteboom, M., Hiemstra, S.J. and Schurink, A., 2019. Development of a genetic tool for determining breed purity of cattle. Livestock Science, **223**: 60-67.

Jiang, J., Ma, L., Prakapenka, D., VanRaden, P.M., Cole, J.B. and Da, Y., 2019. A large-scale genome-wide association study in US Holstein cattle. Frontiers in genetics, **10**: 412.

Jingar, S.C., Mehla, R.K. and Singh, M., 2014. Climatic effects on occurrence of clinical mastitis in different breeds of cows and buffaloes. Archivos de zootecnia, **63**(243): 473-482.

Kardos, M., Luikart, G. and Allendorf, F.W., 2015. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity, **115**(1): 63-72.

Kashyap, N., Kumar, P., Deshmukh, B., Dige, M.S., Sarkar, M., Kumar, A., Chauhan, A. and Singh, G., 2014. Influence of ambient temperature and humidity on ATP1A1 gene expression in Tharparkar and Vrindavani cattle. Indian Journal of Animal Research, **48**(6): 541-544.

Kavakiotis, I., Samaras, P., Triantafyllidis, A. and Vlahavas, I., 2017. FIFS: A data mining method for informative marker selection in high dimensional population genomic data. Computers in biology and medicine, **90**: 146-154.

Kavakiotis, I., Triantafyllidis, A., Ntelidou, D., Alexandri, P., Megens, H.J., Crooijmans, R.P., Groenen, M.A., Tsoumakas, G. and Vlahavas, I., 2015. TRES: identification of discriminatory and informative SNPs from population genomic data. Journal of Heredity, **106**(5): 672-676.

Keller, M.C., Visscher, P.M. and Goddard, M.E., 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics, **189**(1): 237-249.

Kim, E.S., Cole, J.B., Huson, H., Wiggans, G.R., Van Tassell, C.P., Crooker, B.A., Liu, G., Da, Y. and Sonstegard, T.S., 2013. Effect of artificial selection on runs of homozygosity in US Holstein cattle. PloS one, **8**(11): 80813.

Kim, S., Lim, B., Cho, J., Lee, S., Dang, C.G., Jeon, J.H., Kim, J.M. and Lee, J., 2021. Genome-wide identification of candidate genes for milk production traits in Korean Holstein cattle. Animals, **11**(5): 1392.

Koopaee, H.K. and Koshkoiyeh, A.E., 2014. SNPs Genotyping technologies and their applications in farm animals breeding programs. Brazilian Archives of Biology and Technology, **57**: 87-95.

Korunes, K.L. and Samuk, K., 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Molecular ecology resources, **21**(4): 1359-1368.

Kristensen, T.N., Hoffmann, A.A., Pertoldi, C. and Stronen, A.V., 2015. What can livestock breeders learn from conservation genetics and vice versa?. Frontiers in genetics, **6**: p.38.

Kumar, A., Ashraf, S., Goud, T.S., Grewal, A., Singh, S.V., Yadav, B.R. and Upadhyay, R.C., 2015. Expression profiling of major heat shock protein genes during different seasons in cattle (Bos indicus) and buffalo (Bubalus bubalis) under tropical climatic condition. Journal of Thermal Biology, **51**: 55-64.

Kumar, H., Panigrahi, M., Chhotaray, S., Pal, D., Bhanuprakash, V., Saravanan, K.A., Shandilya, R., Parida, S. and Bhushan, B., 2019. Identification of breed-specific SNP panel in nine different cattle genomes. Biomed Res, **30**(1): 78-81.

Kumar, H., Panigrahi, M., Chhotaray, S., Parida, S., Chauhan, A., Bhushan, B., Gaur, G.K., Mishra, B.P. and Singh, R.K., 2021a. Comparative analysis of five different methods to design a breed-specific SNP panel for cattle. Animal biotechnology, **32**(1): 130-136.

Kumar, H., Panigrahi, M., Saravanan, K.A., Parida, S., Bhushan, B., Gaur, G.K., Dutt, T., Mishra, B.P. and Singh, R.K., 2021b. SNPs with intermediate minor allele frequencies facilitate accurate breed assignment of Indian Tharparkar cattle. Gene, **777**: p.145473.

Kumar, H., Panigrahi, M., Saravanan, K.A., Rajawat, D., Parida, S., Bhushan, B., Gaur, G.K., Dutt, T., Mishra, B.P. and Singh, R.K., 2023. Genome-wide detection of copy number variations in Tharparkar cattle. Animal biotechnology, **34**(2): 448-455.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods, **9**(4): 357-359.

Leroy, G., 2014. Inbreeding depression in livestock species: review and meta analysis. Animal genetics, **45**(5): 618-628.

Li, C., Sun, D., Zhang, S., Wang, S., Wu, X., Zhang, Q., Liu, L., Li, Y. and Qiao, L., 2014. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. PloS one, **9**(5): 96186.

Li, C.C. and Horvitz, D.G., 1953. Some methods of estimating the inbreeding coefficient. American journal of human genetics, **5**(2): 107.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, **27**(21): 2987-2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. bioinformatics, **25**(16): 2078-2079.

Linck, E. and Battey, C.J., 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Molecular Ecology Resources, **19**(3): 639-647.

Liu, Y., Li, D., Li, H., Zhou, X. and Wang, G., 2011. A novel SNP of the ATP1A1 gene is associated with heat tolerance traits in dairy cows. Molecular biology reports, **38**: 83-88.

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., & Cunningham, P. (1994). Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences U S A, **91**: 2757-2761.

Mackay, I. and Powell, W., 2007. Methods for linkage disequilibrium mapping in crops. Trends in plant science, **12**(2): 57-63.

Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E., MacNeil, M.D. and Maiwashe, A., 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Frontiers in genetics, **5**: 333.

Malécot, G., 1948. The mathematics of heredity.

Mao, C., Zhang, T., Ren, A., Jia, X., Lai, S.J. and Chen, S.Y., 2021. Genome-wide SNP discovery and genetic diversity evaluation of Liangshan cattle in China. Animal Biotechnology, **32**(6): 671-675.

Marth, G.T., Czabarka, E., Murvai, J. and Sherry, S.T., 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics, **166**(1): 351-372.

Masharing, N., Sodhi, M., Chanda, D., Singh, I., Vivek, P., Tiwari, M., Kumari, P. and Mukesh, M., 2023. ddRAD sequencing based genotyping of six indigenous dairy cattle breeds of India to infer existing genetic diversity and population structure. Scientific Reports, **13**(1): 9379.

Mason, I.L., 1996. A world dictionary of livestock breeds, types and varieties.

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. and Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS one, **4**(4): p.e5350.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F., 2016. The ensembl variant effect predictor. Genome biology, **17**(1): 122.

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A. and MacLeod, A.K., 2008. Runs of homozygosity in European populations. The American Journal of Human Genetics, **83**(3): 359-372.

Mehravi, S., Ranjbar, G.A., Mirzaghaderi, G., Severn-Ellis, A.A., Scheben, A., Edwards, D. and Batley, J., 2021. De novo SNP discovery and genotyping of Iranian Pimpinella species using ddRAD sequencing. Agronomy, **11**(7): p.1342.

Melo, T.P., Fortes, M.R., Fernandes Junior, G.A., Albuquerque, L.G. and Carvalheiro, R., 2019. RAPID COMMUNICATION: Multi-breed validation study unraveled genomic regions associated with puberty traits segregating across tropically adapted breeds. Journal of Animal Science, **97**(7): 3027-3033.

Min, L., Zhao, S., Tian, H., Zhou, X., Zhang, Y., Li, S., Yang, H., Zheng, N. and Wang, J., 2017. Metabolic responses and "omics" technologies for elucidating the effects of heat stress in dairy cows. International journal of biometeorology, **61**: 1149-1158.

Mishra, D.C., Sikka, P., Yadav, S., Bhati, J., Paul, S.S., Jerome, A., Singh, I., Nath, A., Budhlakoti, N., Rao, A.R. and Rai, A., 2020. Identification and characterization of trait-specific SNPs using ddRAD sequencing in water buffalo. Genomics, **112**(5): 3571-3578.

Morrin, R. and Boscher, M., 2012, July. Cattle molecular markers and parentage testing workshop. In ISAG Conference 1-7

Mulim, H.A., Campos, G.S., Cardoso, F.F. and Rojas de Oliveira, H., 2025. Exploring inbreeding depression in Brazilian Angus cattle population using pedigree and genomic data. Frontiers in Genetics, **16**: 1613820.

Negrini, R., Nicoloso, L., Crepaldi, P., Milanesi, E., Colli, L., Chegdani, F., Pariset, L., Dunner, S., Levéziel, H., Williams, J.L. and Ajmone Marsan, P., 2009. Assessing SNP markers for assigning individuals to cattle populations. Animal genetics, **40**(1): 18-26.

Nei, M. and Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences, **76**(10): 5269-5273.

Nei, M. and Tajima, F., 1981. Genetic drift and estimation of effective population size. Genetics, **98**(3): 625-640.

Nietlisbach, P., Muff, S., Reid, J.M., Whitlock, M.C. and Keller, L.F., 2019. Nonequivalent lethal equivalents: Models and inbreeding metrics for unbiased estimation of inbreeding load. Evolutionary applications, **12**(2): 266-279.

Notter, D.R., 1999. The importance of genetic diversity in livestock populations of the future. Journal of animal science, **77**(1): 61-69.

Panigrahi, M., Kumar, H., Saravanan, K.A., Rajawat, D., Nayak, S.S., Ghildiyal, K., Kaisa, K., Parida, S., Bhushan, B. and Dutt, T., 2023. Trajectory of livestock genomics in South Asia: a comprehensive review. Gene, **843**: p.146808.

Peripolli, E., Stafuzza, N.B., Munari, D.P., Lima, A.L.F., Irgang, R., Machado, M.A., Panetto, J.C.D.C., Ventura, R.V., Baldi, F. and da Silva, M.V.G.B., 2018. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. BMC genomics, **19**(1): 34.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS one, **7**(5): 37135.

Pierce, M.D., Dzama, K. and Muchadeyi, F.C., 2018. Genetic diversity of seven cattle breeds inferred using copy number variations. Frontiers in genetics, **9**: p.163.

Porto-Neto, L.R., Kijas, J.W. and Reverter, A., 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genetics Selection Evolution, **46**:1-5.

Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics, **155**(2): 945-959.

Pryce, J.E., Haile-Mariam, M., Goddard, M.E. and Hayes, B.J., 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genetics Selection Evolution, **46**(1): 71.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, **81**(3): 559-575.

Putra, W.P.B., Hartati, H., Aditama, R., Handiwirawan, E., Margawati, E.T. and Elieser, S., 2024. Runs of homozygosity analysis and genomic inbreeding estimation in Sumba Ongole cattle (Bos indicus) using a BovineSNP50K BeadChip. Veterinary World, **17**(8): 1914.

Quinlan, A.R. and Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, **26**(6): 841-842.

Rajar, A.W., Herani, G.M. and Dhakan, A.A., 2007. Demographic, social and economic changes in Tharparkar (1988-2006)-An Analysis.

Rajawat, D., Ghildiyal, K., Nayak, S.S., Sharma, A., Parida, S., Kumar, S., Ghosh, A.K., Singh, U., Sivalingam, J., Bhushan, B. and Dutt, T., 2024. Genome-wide mining of diversity and evolutionary signatures revealed selective hotspots in Indian Sahiwal cattle. Gene, **901**: p.148178.

Ramos, A.M., Megens, H.J., Crooijmans, R.P.M.A., Schook, L.B. and Groenen, M.A.M., 2011. Identification of high utility SNPs for population assignment and traceability purposes in the pig using high throughput sequencing. Animal genetics, **42**(6): 613-620.

Rosen, B.D., Bickhart, D.M., Schnabel, R.D., Koren, S., Elsik, C.G., Tseng, E., Rowan, T.N., Low, W.Y., Zimin, A., Couldrey, C. and Hall, R., 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience, **9**(3): 021.

Rosenberg, N. A., Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of genetic markers for inference of ancestry. The American Journal of Human Genetics, **73**(6): 1402-1422.

Saravanan, K.A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T. and Mishra, B.P., 2020a. Selection signatures in livestock genome: A review of concepts, approaches and applications. Livestock Science, **241:** p.104257.

Saravanan, K.A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G.K., Kumar, P., Dutt, T., Mishra, B.P. and Singh, R.K., 2022. Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. Animal biotechnology, **33**(2): 297-311.

Shafer, A.B. and Kardos, M., 2025. Runs of homozygosity and inferences in wild populations. Molecular Ecology, **34(**3): 17641.

Sharma, R., Ahlawat, S., Sharma, H., Bhagat, R. L., Singh, P. K., & Tantia, M. S. (2020). Identification of a new potential native Indian cattle breed by population differentiation based on microsatellite markers. Molecular Biology Reports, **47** (8): 6429-6434.

Sharma, R., Kishore, A., Mukesh, M., Ahlawat, S., Maitra, A., Pandey, A. K., & Tantia, M. S. (2015). Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers. BMC Genetics, **16**(1): 73.

Sheriff, O. and Alemayehu, K., 2018. Genetic diversity studies using microsatellite markers and their contribution in supporting sustainable sheep breeding programs: A review. Cogent Food & Agriculture, **4**(1): p.1459062.

Shirasawa, K. and Hirakawa, H., 2013. DNA marker applications to molecular genetics and genomics in tomato. Breeding science, **63**(1): 21-30.

Slatkin, M., 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics, **9**(6): 477-485.

Srivastava, A.K., Patel, J.B., Ankuya, K.J., Chauhan, H.D., Pawar, M.M. and Gupta, J.P., 2019. Conservation of indigenous cattle breeds. Journal of Animal Research, **9**(1): 1-12.

Strucken, E.M., Al-Mamun, H.A., Esquivelzeta-Rabell, C., Gondro, C., Mwai, O.A. and Gibson, J.P., 2017. Genetic tests for estimating dairy breed proportion and parentage assignment in East African crossbred cattle. Genetics Selection Evolution, **49**(1): 67.

Strucken, E.M., Gebrehiwot, N.Z., Swaminathan, M., Joshi, S., Al Kalaldeh, M. and Gibson, J.P., 2021. Genetic diversity and effective population sizes of thirteen Indian cattle breeds. Genetics Selection Evolution, **53**(1): 47.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto Neto, L.R., Lee, S.H. and Gondro, C., 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, **45**(4): 572-575.

Sved, J.A., 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theoretical population biology, **2**(2): 125-141.

Tesfa, D., Feyissa, T. and Assefa, K., 2023. Genetic diversity and population structure of selected tef core germplasm lines based on microsatellite markers. Molecular Biology Reports, **50**(10): 8603-8613.

Tortereau, F., Moreno, C.R., Tosser-Klopp, G., Servin, B. and Raoul, J., 2017. Development of a SNP panel dedicated to parentage assignment in French sheep populations. BMC genetics, **18**(1): 50.

Van Tassell, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C. and Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature methods, **5**(3): 247-252.

VanRaden, P.M., Tooker, M.E. and Gengler, N., 2008. Effects of allele frequency estimation on genomic predictions and inbreeding coefficients. Journal of Dairy Science, **91**.

Venkatesan, R.T., Rani, A., Umesh, S., Sushil, K., Kumar, D.A., Sowmya, P., Kesavan, M., Singh, R.B., Panchasara, H.H., Kumar, S.A. and Chhaya, R., 2025. Genome-wide scan for SNPs and selective sweeps reveals candidate genes and QTLs for milk production and reproduction traits in Indian Kankrej cattle. 3 Biotech, **15**(4): 1-16.

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics selection evolution, **34**(3): 275-305.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E., Toro, M.A. and Pong-Wong, R., 2021. The value of genomic relationship matrices to estimate levels of inbreeding. Genetics Selection Evolution, **53**(1): 42.

Visser, C., Lashmar, S.F., Van Marle-Köster, E., Poli, M.A. and Allain, D., 2016. Genetic diversity and population structure in South African, French and Argentinian Angora goats from genome-wide SNP data. PloS one, **11**(5): p.e0154353.

Wang, W., Gan, J., Fang, D., Tang, H., Wang, H., Yi, J. and Fu, M., 2018. Genome-wide SNP discovery and evaluation of genetic diversity among six Chinese indigenous cattle breeds in Sichuan. PLoS One, **13**(8): p.e0201534.

Wang, Z., Wang, G., Huang, J., Li, Q., Wang, C. and Zhong, J., 2011. Novel SNPs in the ATP1B2 gene and their associations with milk yield, milk composition and heat-resistance traits in Chinese Holstein cows. Molecular biology reports, **38**: 1749-1755.

Waples, R.K., Larson, W.A. and Waples, R.S., 2016. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity, **117**(4): 233-240.

Weir, B.S. and Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. evolution, 1358-1370.

Wilkinson, S., Wiener, P., Archibald, A.L., Law, A., Schnabel, R.D., McKay, S.D., Taylor, J.F. and Ogden, R., 2011. Evaluation of approaches for identifying population informative markers from high density SNP chips. BMC genetics, **12**(1): 45.

Willi, Y. and Hoffmann, A.A., 2009. Demographic factors and genetic variation influence population persistence under environmental change. Journal of evolutionary biology, **22**(1): 124-133.

Worku, A. and Tadesse, Y., 2017. A review on population structure, genetic diversity analysis, genetic distance between population and genetic singularity in livestock. Advances in Life Science and Technology, **54**: 1-6.

Wright, S., 1922. Coefficients of inbreeding and relationship. The American Naturalist, **56**(645): 330-338.

Wright, S., 1969. Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies.

Xu, L., Zhao, G., Yang, L., Zhu, B., Chen, Y., Zhang, L., Gao, X., Gao, H., Liu, G.E. and Li, J., 2019. Genomic patterns of homozygosity in Chinese local cattle. Scientific reports, **9**(1): 16977.

Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics, **88**(1): 76-82.

Yang, W., Kang, X., Yang, Q., Lin, Y. and Fang, M., 2013. Review on the development of genotyping methods for assessing farm animal diversity. Journal of animal science and biotechnology, **4**: 1-6.

Yengo, L., Zhu, Z., Wray, N.R., Weir, B.S., Yang, J., Robinson, M.R. and Visscher, P.M., 2017. Detection and quantification of inbreeding depression for complex traits from SNP data. Proceedings of the National Academy of Sciences, **114**(32): 8602-8607.

Yousefi, S., Abbassi-Daloii, T., Kraaijenbrink, T., Vermaat, M., Mei, H., van ‘t Hof, P., van Iterson, M., Zhernakova, D.V., Claringbould, A., Franke, L. and ‘t Hart, L.M., 2018. A SNP panel for identification of DNA and RNA specimens. BMC genomics, **19**: 1-12.

Zeuner, F. E. (1963). A History of Domesticated Animals. London, UK: Hutchinson.

Zhang, Q., Calus, M. P., Guldbrandtsen, B., Lund, M. S., & Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genetics, **16** (1): 1-11.

Zhao, C.H., Wang, D., Yang, C., Chen, Y., Teng, J., Zhang, X.Y., Cao, Z., Wei, X.M., Ning, C., Yang, Q.E. and Lv, W.F., 2024. Population structure and breed identification of Chinese indigenous sheep breeds using whole genome SNPs and InDels. Genetics Selection Evolution, **56**(1): p.60.

✍ ✍ ✍

VITAE

Name	:	Dr. Anal Bose
Father's name	:	Mr. Pradip Kumar Bose
Mother's name	:	Mrs. Soma Bose
Address	:	Flat no. 48, Parvati Condominium Phase-3, Parsudih, Jamshedpur, East-Singhbhum, Jharkhand - 831002
Mobile number	:	+91-8210874046
Email	:	boseanal77@gmail.com
Date of birth	:	18-11-1998
Citizenship	:	Indian
Gender	:	Male

Academic qualification:

Degree	Board/university	Year of passing	OGPA
B.V.Sc. & A.H.	Birsa Agricultural University, Ranchi, Jharkhand	2022	8.413
MVSc	Deemed University, ICAR-IVRI, Bareilly, Uttar Pradesh	2025	8.596

Awards and Recognition

- University Gold Medal for securing highest OGPA in BVSc & AH batch 2017 from BAU.
- Received Subject Gold Medals in Veterinary Physiology, Animal Genetics and Breeding, Veterinary Gynaecology and Obstreitics, Veterinary Medicine and in overall Clinical Subjects in BVSc & AH programme.
- 2nd Position in All India National Quiz Competition by VIRBAC Animal Health Private ltd.
- 1st prize in National Level Essay Competition 2020 by AGRIVISION BHARAT.
- Conferred with the certificate of "Outstanding contribution to Veterinary Profession" by Indian Veterinary Association.
- 3rd position in All India Science Exhibition in the year 2013 issued by Atomic Energy Education Society for "Innovative Car Parking Solution".

This document was created with the Win2PDF “print to PDF” printer available at
http://www.win2pdf.com

This version of Win2PDF 10 is for evaluation and non-commercial use only.

This page will not be added after purchasing Win2PDF.

http://www.win2pdf.com/purchase/