EXPLORING EVIDENCE OF UNIQUE GENOMIC FOOTPRINTS IN THARPARKAR CATTLE BREED

Thesis

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

Dr. Divya Rajawat
Roll No. M-6038

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Veterinary Science
(Animal Genetics and Breeding)

2021

Dedicated To...

My Beloved Parents
and
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.
Scientist (Sr. Scale)

Dated: 19/7/2021

Certificate

This is to be certified that the research work embodied in this thesis entitled **"Exploring evidence of unique genomic footprints in Tharparkar cattle breed"** *submitted by* **Dr. Divya Rajawat, Roll No. M-6038,** *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 herself under my supervision and guidance.*

It is further certified that **Dr. Divya Rajawat, Roll No. M-6038,** *has worked for more than 21 months in the Institute and has put in more than 150 days 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.*

Manjit Panigrahi
(MANJIT PANIGRAHI)
Chairman
Advisory Committee

Certificate

We the undersigned members of Advisory Committee of **Dr. Divya Rajawat, Roll No. M-6038** *a candidate for the degree of* **Master of Veterinary Science** *with the major discipline* **Animal Genetics and Breeding**, *agree that the thesis entitled* **"Exploring evidence of unique genomic footprints in Tharparkar cattle breed"** *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. A. K. Ghosh
External Examiner

Date : 27.08.2021

Manjit Panigrahi
(Manjit Panigrahi)
Chairman
Advisory Committee

Date : 19/7/2021

MEMBERS OF STUDENT'S ADVISORY COMMITTEE

Dr. Bharat Bhushan, Principal Scientist & Head
Division of Animal Genetics, ICAR-IVRI, Izatnagar

Dr. A.K. Sharma, Principal Scientist
Division of Animal Genetics, ICAR-IVRI, Izatnagar

Dr. S.K. Bhanja, Principal Scientist
Division of Avian Genetics and Breeding, ICAR-CARI, Izatnagar

ACKNOWLEDGEMENTS

ऊँ भूर्भुवः स्व तत्सवितुर्वरेण्यं भर्गो देवस्यधीमहि धियो यो नः प्रचोदयात।।

First and foremost, I would like to thank God. You have given me the power to believe in myself and pursue my dream. I could have never have done this without the faith I have in you.

I express my deep sense of thankfulness to my honourable Chairman of Advisory Committee **Dr. Manjit Panigrahi**, Scientist, AG Division, ICAR-IVRI, Izatnagar, for allowing me the space and freedom I needed to work and for continued support and guidance from day one. His friendly supervision, analytical philosophy and expert advice have been invaluable throughout all stages of the work.

I would like to thank my parents **Mr. Arvind Singh** and **Mrs. Gayatri Devi,** My Grandmother **Mrs. Ramkali,** my uncle **Mr. R. B. Singh, Mr. Jitendra Singh, Mr. Devendra Singh**, my sisters **Mrs. Mandvi Bhadouria, Rohini Singh, Pragya Rajawat, Simran Singh,** My brother in Law **Mr. Ashvnee K. Bhadouria, Mr. Vivek Jadon**my brothers **Pushpendra Singh, Yatendra Singh, Pankaj Singh, Ankit Singh, Anshul Singh**for their encouragement, love and support throughout my life. I would like to thank my fiancé **Dr. Brijesh Singh** for always making everything easier for me and invaluable support throughout all stages of the work.

I am highly indebted to my advisory committee member, **Dr. Bharat Bhushan** (Principal Scientist and Head, Division of Animal Genetics). I extend my esteemed and profound sense of gratitude for his constructive counselling, erudite suggestions and efforts for the execution of this study.

I would also wish to express my gratitude to all the other members of my Advisory Committee, **Dr. A.K. Sharma**, Principal Scientist, Division of Animal Genetics, ICAR-IVRI and **Dr. S.K. Bhanja**, Principal Scientist, ICAR-CARI. I wholeheartedly thank, **Dr. T. Dutt**, Director, ICAR-IVRI, for their kind support and providing the adequate resources and laboratory facilities because of which I could accomplish my research in time.

I would like to convey my sincere gratitude to the faculty of Animal Genetics Division, ICAR-IVRI, **Dr. G.K. Gaur**, Principal Scientist, **Dr. A.K. Sharma**, Principal Scientist, **Dr. Sanjeev Kumar**, Principal Scientist, **Dr. A.K. Pandey**, Sr. Scientist, **Dr. Ran Vir Singh**, Principle Scientist, **Dr. Subodh Kumar**, Principal Scientist, **Dr. Amit Kumar**, Sr. Scientist, **Dr. Anuj Chauhan**, Scientist, **Dr. Sheikh Firdous Ahmad** for their cooperation, insightful comments and help during the course of work.

Words are less to express the help, guidance, cooperation and care of TECHNICAL and NONTECHNICAL STAFF OF THE Division of Animal Genetics, ICAR-IVRI. I must express my special gratitude to **Mr. Rajiv Kumar**, Technical staff, AG Division, ICAR-IVRI, for his immense help, support and parental care throughout the work.

I thank my fellow lab mates **Dr. Harshit Kumar, Dr. Kaiho Kaisa, Dr. Saravanan K.A** for their valuable inputs in conducting this research work. I would like to thank my seniors in AG division, for their help and support.

It is my pleasure to thank my colleagues (**Drs**) **Anuradha Panwar, Munish Gangwar, Chhaya Rani, Sakshi Vaishnav, Amit Verma, Pranisha Chhetri** for their kind support throughout my research work.

I shall always remember the deep concern, selfless help directly or indirectly by my senior **Dr. Shivam Bharadwaj,** my friends (**Drs**). **Shivani Khare, Megha Gaur** and **Ruchi Bansal,** during my research work.

I acknowledge the financial support provided to me by Indian Council of Agricultural Research, New Delhi in the form of **ICAR-JRF** fellowship during my Master's degree.

Divya

Date: 19|07|2021

(Divya Rajawat)

Place: *ICAR-IVRI, Izatnagar*

ABBREVIATIONS

%	Percentage
AI	Artificial Insemination
A_{492}	Absorbance at 492 nm
BTA	Bos Taurus
Chr	Chromosome
CLR	Composite Likelihood Ratio
DAHDF	Department of Animal Husbandry, Dairying and Fisheries
DNA	Deoxyribose Nucleic Acid
ET	Embryo Transfer
F	Inbreeding Coefficient
FAO	Food and Agriculture Organization
Fig.	Figure
FLK	Lewontin and Krakauer statistics
GDV	Gross Domestic Production
GO	Gene Oncology
GWAS	Genome-Wide association study
H_p	Pooled Heterozygosity
HSP	Heat Shock Protein
HWE	Hardy Weinberg equilibrium
ICAR	Indian Council of Agriculture Research
iHS	Integrated haplotype Score
Kb	Kilo base
LD	Linkage Disequilibrium
LRH	Long Range Haplotype
MAF	Minor Allele frequency
Mb	Mega base
NBAGR	National Bureau of Animal Genetics Resources
NCBI	National Centre for Biotechnology Information
N_e	Effective Population Size
NGS	Next Generation Sequencing
NPBBDD	National Programme on Bovine Breeding & Dairy Development
QC	Quality Control
QTL	Quantitative Trait Loci
RAPD	Random Amplification of Polymorphic DNA

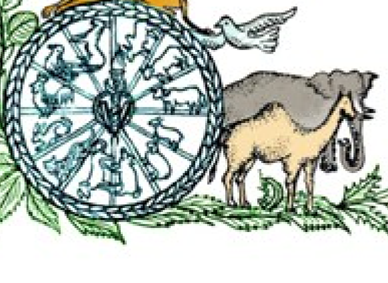
rEHH	Relative Extended haplotype homozygosity
RFLP	Restriction Fragment Length Polymorphism
ROH	Runs of Homozygosity
SFS	Site Frequency Spectrum
SNP	Single Nucleotide Polymorphism
SweeD	Sweep Detector
TRES	Toolbot for Ranking and Evaluation of SNPs
XP-EHH	Cross Population extended haplotype homozygosity
µl	Microlitre

LIST OF TABLES

LIST OF FIGURES

CONTENTS

Introduction



1 INTRODUCTION

Animal domestication was a significant phase in the demographic and cultural growth of mankind. The key evolutionary forces and genetic drift have produced an immense genetic diversity of populations over the successive growth, development, and expansion of livestock (Diamond, 2002). The extent of the biological distinction between or within animals is shown by diversity. Based on genetic variation in the species, animal adaptation is either natural or manmade. One of the key purposes of theoretical and applied evolutionary genetics is the recognition of regions that have been selected or under selection. These studies can also give knowledge to the biological evolution involved in genome shaping, morphological, and physiological gene/genomic regions.

The growth and intensified emphasis on more productive breeding programs have driven genetic progress in a variety of breeds in recent years. Furthermore, advancements in transportation, feed technology, and communication strategy have resulted in stable and strictly regulated production environments. Different evolution forces like selection, migration, mutation and random drift have guided the extended genetic variations in the population (Lenstra, 2010). Natural selection creates significant biodiversity and opted individuals often show more receptiveness in their natural environment while artificial selection creates variations between species such as better livestock with desired expressiveness.

India’s livestock industry accounts for nearly 28.4% of the agriculture and 4.9% of the Allied Industry’s Gross Value Added. While the share of the agriculture sector and related industries to the country’s GDP decreased from 23% in 1999-2000 to 17.4% in 2017-18.

However, the share of the livestock contribution in the country stayed unchanged at around 4% over the same duration. India is still a developing country and bovines are the most prevalent animals among all livestock species (DAHDF, 2019). Our country is the place to a diverse range of livestock, and the share of indicine animals is about 79% (Bollongino *et al*., 2012). According to the latest Census (NBAGR, 2020), there is a total of 192.49 million cows, which comprise almost 35.94% of the whole domestic animal population in India. Modern breeds were originated and adapted from Wild (Bos primigenius) in 8000-9000 BC. The bulk of the Indian cattle breeds produce less milk, but they have more resilience to the hot and harsh climatic conditions, consume less feed intake, and also have high adaptability to diseases. In India, there are 50 cattle breeds documented by the National Bureau of Animal Genetics Resources (NBAGR, 2020).

Among the cattle breeds Tharparkar is one of the most important dual-purpose breeds. Tharparkar cattle breed is also called Kutchhi, Thari and grey Sindhi and has distinct characteristics and is a dual-purpose breed. Tharparkar has good adaptability to hot climatic conditions and consists of disease resilience properties (Herani, 2008). This breed is mainly bred because of its ability to milking. With an overall population of around 732443 Tharparkar shares 0.44% of the whole population of cattle in India. Regarding concerns about its purity, only 197295 (26%) pure and 535135 (74%) improved with exotic breeds. It is the premier milk-producing cattle in the western desert region of India and, where temperature varies from below 0 to 50 °C, is well suited to the hostile climate. Tharparkar has drawn the interest of breeders worldwide because of its unique characteristics and has been utilized extensively in breeding programs (Choudhary, 2018). It has proven to be a very successful milk producer with its crosses with different exotic milch breeds. In Indian cattle breeds, owing to unmethodical reproduction and foreign germplasm debuting, the population has decreased largely due to accelerated genetic loss and dissipation. Different factors such as non-selective crossbreeding, new agricultural methods, focused planting, atmospheric variability, and industrialization are worsening the situation (Shrivastava *et al*., 2019). Thus, there is an urgent need to upgrade and preserve indigenous cattle resources.

With the emergence of next-generation sequencing, SNP genotyping, sophisticated statistical tools, and different computational software in a vast range of species with insufficient time and energy, the accuracy of animal genetic variation and selection signature research have been revolutionized. The subject of attention has currently been transferred to marker-assisted selection (MAS) and DNA marker technologies. This area primarily focuses on the locus of genes and variations that have been discovered to be related exclusively to beneficial traits. Owing to its accessibility, accuracy, and cost-effectiveness in animals, Single nucleotide genotyping (SNP) genotyping is an important desirable approach to evaluate genetic diversity (Hamed *et al.*, 2014). With the breakthrough discovery of SNP estimation, the expense of genotyping was minimized and the ability was provided to examine genetic variation and recognize genomic regions. As single nucleotide polymorphism markers are hereditary secure, copious in amount, and suitable for next-generation sequencing, SNP genotyping is the most frequent approach for exploring genetic diversity.

In 2008, 50K SNP chip was designed and after that, these SNP chips were designed for other animals (Suekawa et al., 2010). Among all the livestock species, the goat was the last species where in 2012 the SNP array was produced (Klopp *et al.*, 2014). This Chip is a multi-sampled and broadly used SNP chip, consist of 53,714 uniformly distributed and exceedingly revealing SNPs (Ding *et al.*, 2009). This BeadChip was designed by Illumina, and it is authenticated in 19 cattle breeds embracing three Indian cattle breeds. This bead chip is developed in collaboration with United State Department of Agriculture (USDA), Alberta University, and Missouri University. In the present scenario, these chips are utilized for various applications such as recognition of quantitative traits loci (QTLs) (Fels *et al.*, 2014), genome-wide studies (MacEachern *et al.*, 2013), assessment of individual genetic merits (Arad *et al.*, 2014), and comparative genetic studies (Lee *et al.*, 2013; Lamichhaney *et al.*, 2019).

A number of studies have been performed in our laboratory employing Bovine SNP50k chip viz. analysis of levels of breed admixture in Vriandavani cattle and population structure in Frieswal and Vrindavani cattle (Ahmad *et al.*, 2018; Ahmad *et al.*, 2019). With the aid of the Bovine SNP50 array, haplotype blocks, genome-wide inbreeding coefficient, and ROH have been analysed (Chhotaray *et al.*, 2019). The next study was accomplished utilizing a 50K chip to establish a Breed-specific SNP panel (Kumar *et al.*, 2019a; Kumar *et al.*, 2019b). This Bovine SNP50K chip data is being used to analyse the genetic diversity parameters, linkage disequilibrium, and structure of haplotype block in Tharparkar cattle (Saravanan *et al.*, 2020). Research team of our lab has also published a brief review on selection signature in cattle genome (Saravanan *et al.*, 2020) and genomic analysis of selective sweep in adaptive and production traits of different Sheep (Saravanan *et al.,* 2020) cattle breeds (Saravanan *et al.*, 2021). Recently, Kumar *et al* Identified 63 unique Tharparkar-specific SNPs near intermediate gene frequencies by combining FIFS, F_{ST}, Delta (Kumar *et al.,* 2021). Study on genome-wide copy number variations was also performed in Tharparkar cattle using 50K SNP chip (Kumar *et al.,* 2021).

Natural selection, Domestication, and artificial selection in animals resulted in the development of breeds with separate phenotypic attributes, including morphological, physiological, and adaptability characteristics (Hagger *et al.*, 2005), and directed to increase genetic variability across breeds in India (Engelsma, 2012). The selection process thrust some influence on certain genomic areas that set distinct breed attributes, this remarkable genetics footprints persistent in an individual's genome, which is under selection is described as "Selection Signature".

In any particular individual, these genomic footprints, which are under selection (may be natural or artificial) have been identified based on SNP chips and these particular genetic characteristics might be beneficial to the assessment of unique breed characters (Kreitman, 2000). The numerous statistical methods have been evolved for the analysis of these footprints utilizing SNP data in livestock. These tools are based on single population (within-population) analysis and between-population analyses. Single population statistics include the following main techniques based on site frequency statistics, linkage disequilibrium, and reduced local variability (Qanbari and Simianer, 2014).

The study of selection signature is a substitute for the GWAS study for recognition of genomic areas and spots of genetic determinants related to particular phenotypes (Vittie *et al.*, 2013). The spots might have genes that are responsible for certain economic important traits and show selection in the influence of different environmental factors.

In recent times, genomic applications in animal breeds have become more famous. Yet there is no sufficient knowledge about signatures of selections which is responsible for different characteristics in different breeds (Oleksy *et al.*, 2010).

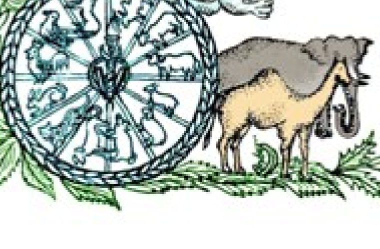
In a tropical country like India, summer temperatures increase by around 40-48 °C, which is certainly beyond the animal comfort zone and induces discomfort. Stress reflects the body's response to stimuli that disrupt homeostasis, often with adverse effects that are expressed in the inability to attain genetic capability for production characteristics (Harris and Meyer, 2006; Kashyap *et al.*, 2014). During their distinct emergence from exotic cattle, indigenous cattle have gained genes of thermotolerance at cellular and physiological levels (Hansen, 2004). The study of selection signature may help to gain knowledge about genes which is responsible to counteract the adverse effect of high temperature. These studies are also very helpful in the understanding gene responsible for disease resistance and also the process of adaptive evolution, breed development, population structure, etc. (Sandip *et al.*, 2016). With the above facts keeping in mind the present study will be undertaken with the following objective:

OBJECTIVE

To identify and annotate the unique genomic footprints in Tharparkar cattle breed using advanced statistical approaches.

* * *

Review of Literature



2

REVIEW OF LITERATURE

Among all the livestock species cattle are one of the utmost important, diverse, and valuable domestic ungulates. Because of the combined influence of natural and artificial selection, the genomes of domesticated cattle breeds retain the history of evolution.

2.1. Domestication and selective breeding of cattle

Since domestication, being a cultural and economic legacy, cattle have been dispersed to the whole world, with their owners adding a great deal to human society (Bollongino *et al.* 2012). Since being an economic and atheistic heritage, cattle have spread to every part of the world. Because of all these causes, cattle have been manifested to new habitats and foods, disease and parasites, and numerous climatic conditions. The cattle breeds of these days originated into subspecies of Bos taurus (exotic) and Bos indicus (indigenous) from wild aurochs (Bos primigenius) in 8000-9000 BC (Ajmone *et al.*, 2010; Wright, 2015). However, interbreeding between them has led to several other groups. Domestication resulted in a variety of cattle breeds and the creation of distinct phenotypes along with evolutionary and demographic factors, such as selection, mutation and migration, random genetic drift, non-random breeding. There are currently a huge number of cattle breeds in the world (>1200) with major within-population and inter-population differences in their morphology (coat color, body confirmation) and adaptation characteristics (heat tolerance, disease resistance) (The Bovine HapMap Consortium, 2009; Groeneved *et al.*, 2010, Decker *et al.*, 2014).

Along with domestication, natural and artificial selection endow a large effect on the distinction of the predominant cattle breeds today. In animal breeding, the evolution of breeds through human intervention is selection. According to Darwin, we can classify artificial selection into unconscious and methodical selection. The unconscious selection was common in past and used to occur without any thought of changing the breed, as time proceeds humans start to select and breed animals methodologically. Rather than most other groups of domestic animals, humans have changed the genotypes of cattle more for their advantage (Albertí *et al.*, 2008, Höglund *et al.*, 2015). Along with the selection, crossbreeding, molecular breeding, and reproductive technologies (AI and ET) take part in the development of breeds with the required specialty (Baruselli *et al.*, 2006)

2.2. Difference between indicine and exotic cattle

The hump is a unique feature of the Indian cattle and makes it easy for them to bear loads without effort, whereas, in exotic cattle, there is an absence of hump. The Native Cattle have a greater tolerance to heat due to their substantial Dewlaps below the jaw, which their European cattle lack. The Indicine cattle are accommodated in tropical environments, while the exotic groups are adapted to temperate climates (Muralidhar *et al.*, 2004). Increased immunity of indigenous cattle to pathogens, ensures that they stay healthy and do not fall sick periodically, saves their owners from needless expenses, and reports 80 to 90% survival. Their increased globulin and lower levels of creatinine enable them to remain healthy naturally (Malik, 2018). With an inadequate level of globulin and an increased level of creatinine, exotic cattle are often prone to diseases, reducing their survival rate in rural areas. Native cows stay healthy in extreme ambient environments, with no detrimental effects on their production, and the fascinating fact is that they can conceive even in adverse weather conditions. The A2 milk rendered by the humped Indian cow, is the ideal food of nature, with a multitude of useful nutrients including high-quality carbohydrates, proteins, and micronutrients that help enhance immune resistance and Indian breeds have a gene with unique genetic variation (bovine HSP70 gene family) that makes them competent to the tropical climate (Sodhi *et al.,* 2013). Exotic cattle, on either hand, cannot tolerate humid and tropical conditions, resulting in decreased productivity, increased mortality, and decreased reproductive performance in exotic and crossbred cattle. Because of rigorous and methodological selection, exotic cattle breeds are very good in milk production compared to native cattle breeds.

Bos indicus (Zebu), also called indicine or humped cattle was originated in India and then spread to Africa and Southern Asia. It can be identified by a prominent hump on the shoulders, drooping ears, and pendulous dewlap (Gregory *et al.*, 1985). Indicine cattle are also imported to other countries mainly in Africa and Brazil. In Africa, Zebu cattle were interbred and by genome analysis of African cows, it has been founded that their genome comprises a high concentration of Zebu cattle genome. In the initial stage of the 20th century, Indian cattle which can withstand high temperature were imported into Brazil. Brazil imported Ongole, Gir, and Kankrej (Pereira *et al.*, 2006). According to the present scenario, Brazil is the solitary largest exporter of the zebu cattle breed with improved germplasm in the World and comprises 80% of the total indicine cattle population. Due to the peculiar characteristics, breeders made different crosses namely Australian Friesian Sahiwal (Sahiwal× Holstein Friesian), Jamaica Hope (Jersey and Holstein Friesian×Sahiwal), Australian milking Zebu (Sahiwal×Jersey), etc.

2.3. Cattle population in India

India is the having highest livestock population in the World, accounting for nearly 535.78 million. The livestock sector in India contributes 4.9% of GDP and 28.4% of overall GDP to agriculture. India, one of the world's 12 mega-biodiversity nations, has a huge variety of the genetic resources for cattle. It constitutes approximately 16% of the overall global population of cattle. In 2019, the 20th livestock census, the population of cattle in India is 192.49 million, which is 16% of the total World's cattle population. Along with a huge number of non-descript cattle population, currently, there are 50 registered indigenous cattle breeds (NBAGR, 2020), which can be categorized into three categories based on utility, namely milch breeds, draught breeds, and dual-purpose breeds. Milch breeds are mainly Sahiwal, Gir, and Red Sindh. Deoni, Krishna Valley, Mewati, Hariana, Ongole, Rathi, Kankrej, and Gaolao are dual purpose breeds. However, Nagori, Nimari, Khillari, Kenkatha, Malvi are used for draught purpose.

2.4. Crossbreeding in India and depletion of indigenous germplasm

The National Dairy Development Board (NDDB) was established in the year 1965, which leads to the initiation of operation flood in the country. Crossbreeding in India was started in 1865 but it had incorporated in the policies after 1961. Since 1965, the crossbreeding of the indigenous breed with imported exotic breeds has been accomplished extensively to improve milk production in the country (Venkatasubraman *et al.,* 2010), however such programs lead to much controversy. Although milk production increased tremendously due to crossbreeding but at the same time it also leads to dilution of pure indigenous germplasm. Compare to the 19th census, the total exotic or crossbred cattle population has increased by around 27% in the current census.

Some other factors are also responsible for the exhaustion of indigenous resources like modern farming, specialized farming, reduced herd size etc. If the same pattern persists, we will lose some important indigenous cattle breeds in the next recent years (Felius *et al.,* 2015). According to Food and Agriculture Organization (FAO) (2015), the report of animal genetic resources indicated that around 54% of the total bovine breed population comes under “unknown status”, 12.3% comes at a risk and 13% of cattle breeds were extinct and only 20.2% of breeds population is safe or in the zone of ‘not at risk, that is a disturbing scenario for the World (FAO, 2015). In India, three breeds (Kumauni, Punganur, and Kasargod) are registered as endangered, two breeds (Tarai and Krishna Valley) are reported as vulnerable and Vechur is considered critical. Many of the local breeds like Tharparkar, Mewati, Red Sindhi, Hallikar, Hariana, etc. showed a decreasing trend and need to be conserved for their improvement and reliable use (Sujatha and Ramabrahman, 2016). The protection and adequate documentation of these substantial animal genetic resources is a demanding challenge that needs to be taken up on a high priority. These protection programs need to be focused on appreciable results, both in the long and short terms. These can be dependent on national survey initiatives, followed by different plans and legislations. The statistically validated ranges should be restricted to exotic inheritance (i.e., 52-62.5% for semi-intensive and 62-75% for intensive production systems). Overall, the biodiversity of livestock in India is enormous and we must make effective use of this gift (Ahmad *et al.,* 2016).

2.5. Tharparkar cattle

Tharparkar is a domestic cattle breed that is utilized as dual-purpose cattle and recognized for both its milk quality and draught ability. It is also named White Sindhi, Kutchi, and Thari. Tharparkar is considered as one of the most important cattle breeds of the hot desert region of India covering Rann of Kutch in Gujrat and Western Rajasthan but its origination is believed to be in Tharparkar district in Sindh of Pakistan. They are resistant to various diseases and well suited to their home tract's harsh climate, unfavourable environment, and feed deprivation.

2.5.1. Characteristics of Tharparkar cattle

Tharparkar is a medium-sized cattle breed with a convex forehead, long tapering face, medium-sized and lyre-shaped horns that curve outward, large and pendulous ears, and whitish-grey. They have a prominent hump, strong legs, medium dewlap, strong udder, and black twitch of the tail. The average body weight of bull and cow ranges between 400 – 450 kg and 300 – 350 kg respectively (Reddy *et al.,* 2018). Amid all the indigenous breeds, Tharparkar has a distinct reputation among the Indian breeds of cattle and is well known for its resistance to disease and heat tolerance. Tharparkar is bred mainly because of its ability for milking. With an overall population of around 7,32,473, Tharparkar shares 0.44% of the total bovine population in the country. At the time of endemic scarcity of fodder, these cattle are excellent foragers and can feed bushy vegetation. They are effective milk producer in arid conditions. The bull is also perfect for draught purposes. The Tharparkar cattle are highly suited to adverse climatic conditions with extreme temperatures ranging from below zero to 50°C. Tharparkar appears to consist of comparatively more resistant to mastitis than other breeds. The occurrence of mastitis seemed to be low in Tharparkar (12.5%) than Gir (26.99%). The typical indicator of mastitis (somatic cell count) in Tharparkar cattle was quite low under filed conditions (Choudhary *et al.,* 2018).

2.5.2. Heat regulation effect in Tharparkar

In a tropical climate like India, the temperature in the summer goes up to 40-48 °C, which undoubtedly will be out of the comfort zone of livestock and causing issues. Stress reflects the response of the body to triggers that disrupt homeostasis, frequently with adverse effects that are responsible for failure to achieve the genetic potential for different economic traits like growth (Tao *et al*., 2012), production, and reproduction (Min *et al*., 2016). When cellular function experiences environmental and physiological stress, it affects various original functions of cells, such as replication, transcription, translation, and other transport functions. However, some proteins ideally express themselves in such restricting conditions as the slick hair gene, ATP1B2 (Wang *et al.*, 2011), and heat shock protein (Ellis, 1987). These are a group of proteins generated by the cellular response to reactions to adverse environments. They were identified concerning heat shock, but are now understood to be often expressed during other pressures, including cold exposure, Ultraviolet, and during tissue repair or tissue remodeling (Maleki *et al.*, 2014). These proteins are highly conserved and can be described namely, Hsp90, Hsp70, and Hsp60 refer molecular weight families of 90, 70, and 60 kilo Daltons, respectively. Numerous HSPs are used as intracellular chaperones for several other proteins (Wu, 1995). They (mainly HSP70) have an important role in interactivity between two proteins such as dimensional configuration, helping to develop proper protein conformation (shape) and avoid unnecessary protein aggregation. By helping to stable unfolded proteins, HSPs are helping to transport proteins across cell membranes (Walter and Buchner, 2002).

The ATP1A1 gene encodes the á1 chain of Na+/K+ - ATPase which carries the catalyzing unit of this enzyme (Zicha *et al.*, 2001). This enzyme is responsible for the regulation and maintaining the balance of sodium (Na+) and potassium (K+) ions in the internal homeostatic environment. In Tharparkar cattle, heat shock proteins were expressed during adverse conditions which helps them to adjust to the unfavorable environment (Das *et al.*, 2015) and the biphasic pattern of these proteins provides a protective source to Tharparkar during long term heating effects.

2.5.3. Conservation and Genetic Improvement

The bulk of the genetic capital of cattle is currently held *in-situ* by pastoralists and farmers as an integral part of the method of agricultural production. The conservation efforts in India began with the creation of NBAGR in 1984 under the Indian Council of Agricultural Research (ICAR). NBAGR is located in Karnal (Haryana) (Srivastava *et al.,* 2019) and under NBAGR, a Network project is executing, which is aimed to characterize various important indigenous breeds/germplasm. Some *in-situ* programs have been developed by ICAR-NBAGR and also provided some technical sources and inducements to the farmers in home niche of that particular breed (Gandhi, 2016). These models were started mainly to conserve Tharparkar and Krishna Valley. The Central Herd Registration program was started in 1963, with the inclusion of certain cattle breeds such as Ongole, Kankrej, Gir, Haryana cattle breeds (Kumar, 2016). Certificate and awards/bonuses for the encouragement of the survival of indigenous breeds and the development of high-quality cows are given for the owners of registered animals in the scheme (Kachhawaha *et al.,* 2015). The Government of India (GOI) has initiated different programs and schemes to build-up and strengthen the dairy sector. One program National Programme for Bovine Breeding (NPBB) is responsible for improvement and conservation of indicine cattle population included 33 cattle breeds and is an integral part of National Programme for Bovine Breeding and Dairy Development (NPBBDD). Another technical program National Gokul Mission is executing under NPBBDD is dedicated to the improvement of three breeds namely, Tharparkar, Gir, and Sahiwal, and sanctioned in the 12th Five-Year Plan with a budget of 500 Crores. Central Frozen Semen Development and Training Institute, Government Livestock Farms, Central Cattle Breeding Farms are effectively doing the preservation and maintenance of indigenous livestock. Numerous state governing agencies are also sustaining farms for the sustainable development of Tharparkar. For successful conservation, strategic alignment between the various organizations and the execution of the "National Partners Consortium" with a comprehensive approach with effective communication is necessary.

2.6. Genetic markers

It is a sequence of DNA that has situated on a particular position on a chromosome with a known location. Genetic markers can enable to interconnect of the inherited disease to the responsible gene. These can be in two forms; first is biochemical markers and the second is molecular markers (Schwartz and Cavalli, 2017). The biochemical markers detect variations at the stage of transcription and molecular markers detect the variation at DNA level. Genetic markers are utilized to monitor the inheritance of neighbouring genes that have not been identified but can be easily accessed. There are various type of genetic markers such as restriction fragment length polymorphism (RFLP), simple sequence length polymorphism (SSLP), random amplification of polymorphic DNA (RAPD), variable number of tandem repeat (VNTR), Amplified fragments length polymorphism (AFLP), etc. The genetic marker itself may be part of a gene without playing any role (Masseurs *et al.,* 2012; Hueng *et al.,* 2014). Different bioinformatics tools, software technologies, and next-generation sequencing have allowed widespread use of single nucleotide polymorphism markers to analyse genetic diversity (Sun *et al.,* 2014).

2.7. SNP polymorphism and genotyping

Polymorphism is a common term that means ‘many shapes’. Single nucleotide polymorphism (also called “snips”) is the most usual type of genetic variation. It constitutes the nucleotide difference. Single nucleotide polymorphism is described as a locus of the particular allele that varies at a single nucleotide, with the incorporation of frequency of rarer allele at least more than 1% in a random population. SNPs are found mostly in noncoding regions and occur with a very high frequency (1 in 1000 bases) (Phillips *et al.,* 2007b; Bovine HapMap Consortium, 2009). This ample amount of polymorphisms and facilities for easy measurement make it most efficient to act as a predominant marker type to detect genetic variations (Butler, 2012; Nisha *et al.,* 2015). Many bioinformatics tools and high throughput sequencing techniques have been implemented on the identification and development of SNPs like PCR-based sequencing and random shotgun sequencing (Mitchelson, 2005). However, a specific portion of these SNPs has been transformed as genetic markers for evolutionary biology, Linkage Disequilibrium (LD) analysis (Khatkar *et al.,* 2008), and genome-wide interaction findings. A pioneer class of genetic markers has been more desirable in genome-related studies these days. These polymorphisms, however, have been evolved as genetic markers for the study of positive and negative selection sweep (Fariello *et al.,* 2014). Due to rapidly evolving high-throughput SNP genotyping facilities, there has been a drastic reduction in genotyping expenses. SNP markers are very efficient in genomic selection and QTL analysis (Beuzen *et al.,* 2000, Dikmen *et al.,* 2013).

There are mainly two methods by which we can perform SNPs genotyping namely, the traditional method and modern high throughput methods (Wiggans *et al.,* 2009). Tradition methods were mainly gel-based methods such as restricted fragments length polymorphism (RFLP), amplification of refractory mutation system (ARMS), etc., (Zhao *et al.,* 2019). Modern high throughput methods include allele-specific hybridization, single Base Primer Extension, oligonucleotide ligations, DNA arrays, etc., (Eldridge *et al.,* 2002). One of the preferred methods for investigating genetic variation is SNP genotyping, as SNPs are genetically stable, widespread in the genome, capable of sequencing, and cost-effective analysis. In 2008, Genome-wide SNP chip data for bovine (cattle) was developed (Tietjen *et al.,* 2005, Rasmussen *et al.,* 2015). The Bovine SNP50 Bead Chip is a multi-sampled and most widely used SNP array, consist of 53,714 uniformly distributed and extremely informative SNPs. In collaboration with USDA-ARS, the University of Missouri, and the University of Alberta, this chip is utilized for various genomic analysis such as genome-wide studies (Qiao *et al.,* 2017), assessment of individual genetic merits (Arad *et al.,* 2014), and comparative genetic studies ((Sermyagin *et al.,* 2018). In the recent past, study-related to identifying selective sweep regions were also performed with the utilization of SNP chips (Li *et al.,* 2015; Iso-Touru *et al.,* 2016). Currently, eleven commercial SNP chip arrays evolved for cattle and three major companies Illumina, Neogen, and Affymetrix are involved with its development. The Bovine SNP50 Bead Chip v3 is the latest and the most extensive genotyping array (Rohilla *et al.,* 2020). It consists of the following properties like fabulous call rates and more 99% accuracy; uniformly covering of evenly distributed polymorphic SNPs; ligation-free protocol and simple PCR based workflow (Hayes *et al.,* 2009).

2.8. Selection in livestock

Cattle are amongst the most popular and abundant domestic mammals. The domesticated breed's genome harbours the heritage of domestication and breed development due to the combined influence of natural and artificial selective factors. Natural selection, Domestication, and artificial selection in animals resulted in the development of breeds with separate phenotypic attributes, including morphological, physiological, and adaptability characteristics (Hagger *et al.,* 2005) and directed to increased genetic diversity across the breeds in India. The first step of selection is domestication. Domestication is the process of progressive change in the behaviour of an animal so that they can be fitted according to the need of people. The result of domestication was the tameness in animals which means to develop the quality in an animal being welcoming towards the human presence (Albert *et al.,* 2009; Wright, 2015). Natural selection is a very complex process and depends on many factors such as differences in the degree of fertility of parents and viability of offspring, differences in the mortality in the population, and differences in the sexual activity. Natural selection and artificial selection up to an optimum level led to specialized cattle breeds. There may be different nature of natural selection, namely, positive selection, negative selection, and balancing selection. In positive selection, the farthest phenotype is preferred causing the frequency of allele to convey in the direction of that particular phenotype and eventually become fit that's why it is also known as directional selection (Biswas, 2006; Vittie *et al.,* 2013; Gouveia *et al.,* 2014). Negative selection is the process that causes the selective discarding of the lethal alleles from the whole population. Balancing selection is the process in which two alleles are regulated at the heterozygous frequency in the population. The most common cause of balancing selection is heterozygote advantage (Ajmone Marsan *et al.,* 2010; Mirkena *et al.,* 2010; Wright, 2015).

2.9. Genomic footprints

Before going to a description of selection, there is a necessity to understand selective sweep **(Fig. 1)**. When a derived allele or variant with a fitness advantage proceed in a population under positive selection, this causes other linked and nearby alleles to be carried forward along with the selected variant. Genetic hitchhiking, also known as the genetic draft, occurs when an allele changes its frequency not because it is under natural selection, but because it is nearer (linked) to another gene that undergoes selective sweeping and is on the same DNA chain (Rege, 1999). A selective sweep in a particular genomic spot causes a reduction in genetic variation, skewness in the allele frequency spectrum, the elevation of linkage disequilibrium, and finally elevation of inter-species divergence and reduction in the extent of inter-population differentiation (Mirkina *et al.*, 2010; Okeyo *et al.*, 2015). The selective sweep may be of two types, hard sweep, and soft sweep. A hard sweep is a phenomenon where a specific haplotype nurtured a specifically beneficial allele increases in frequency, while, in a soft sweep, several haplotypes carrying favorable mutations will spike in frequency. Thus, selective sweeps reflect a wide and non-homogenous range of genomic signatures. Through dispersive forces like random drift, variants that do not cause a fitness gain to the organism carrying them may increase in occurrence in the population. In both cases, natural selection

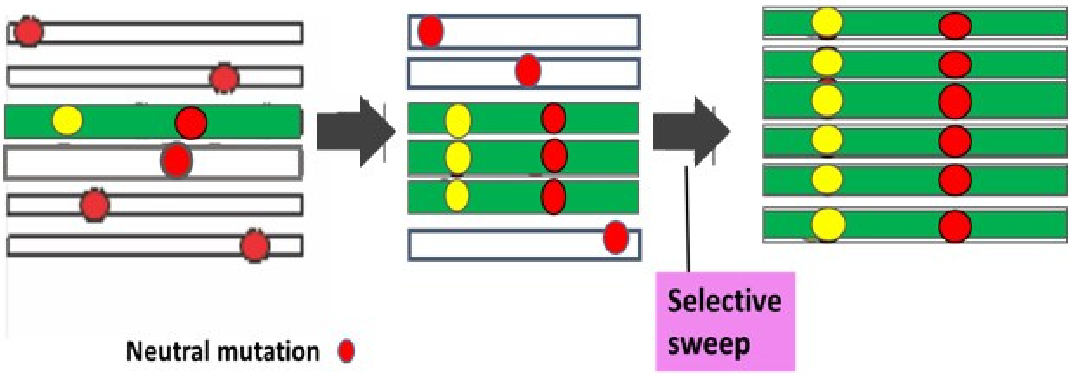


Fig. 1: Selective sweep: (A) region of six chromosomes with red alleles (neutrally segregating), in the green chromosome one yellow allele (adaptive); (B) at the time of the selective sweep, showing the decline in the differentiation of the region around the favourable allele; and (C) the result after the complete sweep with the fixation of beneficial alleles in the region.

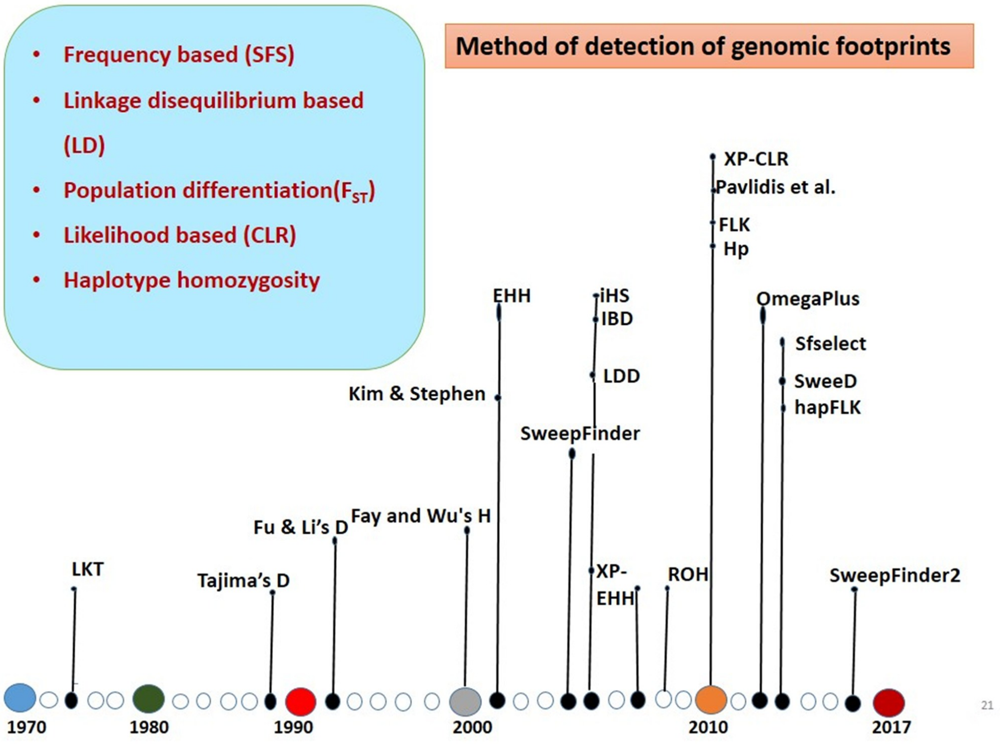


Fig. 2: Different statistical methods of detection of Genomic Footprints.

and artificial selection, the variant frequency provides fitness advantage for carrier or affecting a trait of choice by breeder together with the hitchhiking neutral variants will increase in population (Utsunomiya *et al.*, 2015).

In any particular individual, these footprints, which are under selection (may be natural or artificial) have been identified on the account of SNP chips and these particular genetic characteristics that might be beneficial to the assessment of unique breed characters (Kreitman, 2000). To analyse these footprints, several statistical tools have been developed and used by different breeders. These methods are widely categorized into the following classes, inter-population statistics and intra-population statistics (Gouveia *et al.*, 2014). Intra-population statistics include three methods according to site or allele frequency statistics, linkage disequilibrium, and reduction in local variability (Qanwari and Simianer, 2014).

2.10. Methods of detection of these genomic footprints

Statistical tools based on a single population are further classified into Tajima's D, Fay and Wu's H, and CLR. Linkage disequilibrium includes rEHH, iHS, XP-EHH., and reduced local variability compile runs of homozygosity (ROH) and pooled heterozygosity (Hp). Inter-population statistics include FLK, hapFLK, and F_{ST} **(Fig. 2)**.

2.10.1. Statistics based on the single population

2.10.1.1. Site frequency spectrum

In evolutionary genetics, it is the scattering of the allele in a given group of loci / SNPs. It is a compilation of different tests. Site frequency statistics further classified into Tajima's D statistics, Fay and Wu's H Statistics, and CLR.

2.10.1.1.1. Tajima's D statistics

It is a standard tool for detecting selection signature, Tajima's D considers the difference in the mean pair-wise difference and the total number of different sites of segregation in nucleotide polymorphism data (Carlson *et al.,* 2005).

Mean pairwise difference or pi is given by:

$$\hat{\theta}_{\mathrm{T}} = \frac{\sum_{i<j} dij}{n(n-1)/2}$$

Where n represents the total sequences that need to be compared, and dij is the number of difference between the two sequences.

Watterson's estimator is given by:

$$\hat{\theta}_{\mathrm{W}} = \frac{S}{\sum_{i=1}^{n-1} 1/i}$$

Where S is the total number of sites that are segregating in the sample.

Tajima's D:

$$\mathrm{D} = \frac{\hat{\theta}\mathrm{T} - \hat{\theta}\mathrm{W}}{\sqrt{\hat{V}}(\hat{\theta}\mathrm{T} - \hat{\theta}\mathrm{W})}$$

It identifies regions of selective sweep which cause the fixation of infrequent alleles in the population (Tajima, 1989). The negative value shows the fixation of rare alleles which means a smaller value (i.e., more negative) of Tajima's D suggests an abundance of rare allele and an indication of positive selection (Korneliussen *et al.*, 2013). Positive D value represents the balancing selection which shows an abundance of intermediate allele frequencies. The zero value of D is the indication of neutral variation.

2.10.1.1.2. Fu and Li's D* test

This test uses information from intraspecific data and is derived from SFS analysis (site frequency spectrum). Under the neutral model, estimates the number of polymorphic sites and the mean number of spectral nucleotides are correlated. Tajima's D test is not based on coalescent but Fu and Li's test is supposed based on coalescent. In the presence of purifying or negative selection, the excess of mutations may be present in the superficial (external) branches (new mutations) because harmful alleles are present in low frequencies. On the contrary, if balancing selection is operating some alleles might be remote and there might be a low occurrence in mutational events in the external branches.

$$\mathrm{D}* = \frac{\left(\frac{n}{n-1}\right)\eta - an}{\sqrt{uD} * \eta + vD * \eta 2}$$

Where, n is the number of samples, η is the aggregated number of mutations and η_s is the number of single mutations.

2.10.1.1.3. Fay and Wu's H statistic

This method is developed by Justin Fay and Chung Wu in the year 2000. The objective of the test is to differentiate between the genes that are evolving neutrally and others that are generating under positive selection (Fay and Wu, 2000). It is utilized to distinguish neutrally evolving sequences from the sequences which evolve non-randomly. In this test, it is assumed that the ancestor allele is known and used for mainly medium to high non-ancestor allele frequency (Hedrick, 2005). This test emphasizes the identification of recent positive selection. A notable small (negative) value of H statistics designates a high-frequency excess derived SNPs (Sterken, 2009). This test is commonly used to classify sequences that have experienced selective sweeps in their evolutionary history. It is utilized to distinguish neutrally evolving sequences from the sequences which evolve non-randomly. In this test, we assume that the ancestor allele is known and used for mainly medium to high non-ancestor allele frequency (Hedrick, 2005). This test emphasizes the identification of recent positive selection. A notable small (negative) value of H statistics designates a high-frequency excess Derived SNPs (Sterken, 2009). This test is commonly used to classify sequences that have experienced selective sweeps in their evolutionary history.

2.10.1.1.4. Composite Likelihood Ratio (CLR) test

The Composite Likelihood Ratio CLR statistics is a method of detection of the likelihood for the selective sweep that describes the differentiation of allele frequency of multiple-locus between two populations. This method is operated to analyze the multi-loci allele frequency spectrum skewness and includes recombination rate to evaluate selection from different population-based events (Kim and Nielsen, 2004). It is the more advanced method than the above two and provides observed evidence of the ongoing or "incomplete" selective sweep. It is firstly estimated in Drosophila melanogaster at the Sod locus. This method is highly sensitive to find out positive selective sweep regions in a population (Nielsen *et al.*, 2005). This modern method is derived by Kim and Stephan in 2002. He estimates CLR by divided the maximum composite likelihood in neutrality (without selection sweep) and maximum likelihood under selective sweep (Eriksson *et al.*, 2008). This method exploits the complete advantage of the demographic pattern of variation of SFS (Bersagliri *et al.*, 2004). This method is used to detect ongoing selective sweep on the basis of the probability of joint sampling for allele frequency as the strength of recombination rate and selection. The CLR method showed consistent structures of the haplotype expected under incomplete selective sweeps. Along with the evaluation of frequency spectrum skewness in all the multiple loci, this method also analyses information on the recombination rate to detect selective events (Ma *et al.*, 2015).

These SFS methods are not suited for spotting genome-wide SNPs and it is the main drawback of these methods. So the next step towards selection signature was linkage disequilibrium-based methods.

2.10.1.2. Linkage disequilibrium-based methods

Linkage disequilibrium (LD) is the association between adjacent variants in such a way that the polymorphic alleles (observed in a few numbers of alleles) are correlated more frequently than if they were not associated within a group (Robinson, 1998). An interaction between genetic differentiation and a phenotype reveals that either the genetic differences at that locus significantly influence the desired phenotype or the locus with the specific mutation is in LD (Mueller, 2004). The linkage disequilibrium targeted methods utilize high-frequency haplotypes along with homozygous spots (Ramakrishnan, 2013). There are various computational approaches using linkage disequilibrium, which are discussed below:

2.10.1.2.1. Long-range haplotype (LRH)

This method is used to find out the high-frequency alleles with long-range linkage disequilibrium, which infer that haplotype quickly got up to high frequency before the disruption of association with nearby markers due to recombination (Sabeti *et al.*, 2002).

2.10.1.2.2. Extended haplotype homozygosity (EHH)

It is the expectation that a chromosomal pair bears a homozygous core haplotype. This method is based on LD to identify genomic footprints in a population. A positive selection signature based on EHH is described as a region characterized by a long and strong range of linkage disequilibrium and an allele within an exceptionally high-frequency haplotype. Due to heterogeneous recombination frequencies along the genome, EHH is expected to produce a significant number of false findings. The problem of rigorous inference, e.g. the need to differentiate between real and spurious signs, is an additional downside dispersed by all signature processes, including EHH (Kemper *et al.*, 2014). These statistics constructed to use imputed genotypes to detect presumed ongoing and novel positive selective regions in the genome of a particular individual. Sabeti and his team in the year 2002 introduced the relatively extended haplotype homozygosity (rEHH) approach to partly account for these shortcomings, applying an observational approach to determine the importance of cues. Compared with the EHH value of other haplotypes at the identical loci of the core haplotype, the rEHH of a core haplotype (i.e. short area in high LD around the genome) is used as a monitor for local variance in recombination frequency. After detecting the core of haplotypes, the age of every single core is evaluated by the linkage disequilibrium as per the distance (Wang *et al.*, 2014). In particular, it is appropriate for SNP data analysis since it is less prone than other approaches of ascertainment bias (Neilson *et al.*, 2007).

2.10.1.2.3. Integrated haplotype score (iHS)

To resolve the effect of heterogeneous recombination rates all over the genome, iHS method has been introduced, which is an advancement of the EHH approach and is focused on the comparison of EHH between derived and remote alleles in the selected population. This method attains its threshold potential when a chosen allele sets apart at mid frequencies in a particular population. iHS is a method to find out the unusual haplotype around a particular SNP comparing to the whole genome. It is used to find out a standard measure of decay in EHH from the developed allele compared to the ancestral allele at SNP. Therefore, this method needs all the information about the ancestor and derived allele for each SNP (Voigt *et al.*, 2006).

Unstandardized iHS was estimated as follows:

$$\text{Unstandardized iHS} = \ln\left(\frac{iHHA}{iHHD}\right)$$

Standardized iHS (mean =0, variance =1), was estimated as:

$$\text{iHS} = \frac{ln\left(\frac{iHHA}{iHHD}\right) - Ep\left[ln\left(\frac{iHHA}{iHHD}\right)\right]}{SDp\left[ln\left(\frac{iHHA}{iHHD}\right)\right]}$$

Where iHH_A is the combined EHH values for the inherited ancestral alleles and iHHD is the beneficial selected alleles, respectively. This method required imputation and haplotype phasing and detect selection signature at an intermediate frequency (Maiorano *et al.,* 2018, Chen *et al.,* 2018, Zhhu *et al.,* 2015).

2.10.1.3. Methods on the basis of diminished local variation

This method is used to evaluate the composition of genomic regions which has reduced variation. Reduced local variability statistics consist of different methods, namely ROH (runs of homozygosity) and Hp (pooled heterozygosity) (McCQuillan *et al.,* 2008).

2.10.1.3.1. Homozygosity runs (ROH)

ROHs are contiguous regions of homologous genotypic form occurring in a diploid organism. This homozygosity occurs due to parents pass similar haplotypes to the next generation. This method is broadly utilized to estimate the inbreeding level of the genome in animals. High selective pressure genomic regions exhibit characteristic runs of homozygosity that reflect low genetic diversity (Fortan *et al.*, 2018). ROH identifications that contribute to output reduction may offer useful insight into understanding the genetic framework of complex characteristics (Guoyao Zhao *et al.*, 2020). ROHs can be grouped into two categories, long runs, and short runs. In general, the presence of the longer segments shows, that there is recent inbreeding (Kirain *et al.*, 2010). The short ROH dispersion may guide us to find out more remote affinity and less inbreeding (McQuillan *et al.*, 2008).

2.10.1.3.2. Pooled heterozygosity (HP)

This method is used to estimate heterozygosity by the use of allele count. This method calculates the divergence of expected heterozygosity decline from the genomic average heterozygosity (Rubin *et al.* 2010; Elferink *et al.*, 2012).

Here we have completed the methods of intra-population selection signatures. The next part is the selection signature statistics among the various populations.

2.10.2. Statistics based on more than population (inter-population)

Recognition of the relatively recent genomic footprints in livestock can provide evidence on genomic regions that are influenced by natural selection. These statistics correlate SNP genomic information among more than two populations to the identification of hotspots under the selective process. The approaches primarily depend on the extent of divergence among the populations due to fixed allele frequencies. We can classify these statistics into two different classes; differentiation in single sites and differentiation in haplotypes. The first one is further divided into FST and FLK (Wright, 1949; Bonnhomme *et al.*, 2010) and the second method is further classified into XP-EHH and hapFLK ((Sabeti *et al.*, 2007; Fariello *et al.*, 2013).

2.10.2.1. Single-site differentiation

2.10.2.1.1. Pairwise F_{ST}

F_{ST} is another way to analyse the regions of selective sweep and based on the population divergence because of allele frequencies at specified locus between populations. F_{ST} is frequently analysed from polymorphic data i.e. SNP and microsatellite (Berli, 1998). It is developed by Wright, it is among the most frequently used statistics in evolutionary genetics (Holsinger *et al.*, 2009; Laercio *et al.*, 2013). It is measured on the basis of the frequency variations of an allele. It can be calculated as follows:

$$F_{ST} = \frac{\sigma S^2}{\sigma T^2}$$

Where σ_T^2 and σ_S^2 is the variance in the total population and various subpopulation respectively.

The value of F_{ST} ranges from 0 to 1. 0 means no differentiation and 1 means fixation between populations. A higher F_{ST} value detects positive selection and a low value of F_{ST} indicates negative selection.

2.10.2.1.2. FLK

FLK test is estimated based on the single-site population differentiation and reports for the heterozygosity of the population. It is an extended form of the Lewontin and Krakauer (LK) test and accounts for the hierarchical setup between the populations. FLK test reckons global F_{ST} for each SNP. Before accounting for the analysis for each SNP, this statistic utilized the phylogenetic analysis of the kinship matrix (F). Under a neutral evolution, the theory of the FLK test is to correlate, the distributions of variations between allele frequencies in many populations to their expectations. The neutral population (null hypothesis) presumes a tree-like structure with a length of branch corresponding to the quantity of genetic drift in every population. This tree called Neighbour-joining (NJ) tree, is analyzed from Reynold's distance matrix, utilizing the Neighbour-joining algorithm. The measured differentiation in the allele frequency is over-burdened in the population if the estimated population is small or a population with more remote ancestors. This method is outstanding when operating on alike or closely related populations. F matrix is estimated from the measurement of expected genetic drift and expected covariance.

Even under normality, the distribution of the chi-square test is imprecise, under normality, the composite analysis of allele frequencies is completely specified by initial frequency p0 and then by the F matrix.Let, $\hat{p}_0$ matrix from the SNP data, as follows: Assuming normal condition, the p0 is estimated from the F matrix and SNP data:

$$\hat{p}_0 = \frac{1'n\ F - 1p}{1'n\ F{-}1\ 1n}$$

Where, for SNP, p is the frequency of allele in n number of populations Then, T_{FLK} is estimated as:

$$T_{FLK}\ (\hat{p}_0) = (p - \hat{p}_0\ 1n)'\ V(p)^{-1}\ (p - \hat{p}_0\ 1n)$$

Where, $(p)^{-1}$ is the co-variance of p vector and calculated as:

$$(p)^{-1} = F\hat{p}_0\ (1-\hat{p}_0)$$

This test is strong measurement for analysis of numerous SNPs among diverse populations.

2.10.2.2. Extended Haplotype Homozygosity Cross Population Statistic (XP-EHH)

This method is a test of cross-population for directional (positive selection) which means that it identifies SNPs that are undergoing in selective process in a particular population and unselected for another population. The Extended Haplotype Homozygosity Cross Population Statistic (XP-EHH) measures the sum of diverse haplotype homozygosity at every locus relative to a reference population found at each locus. XP-EHH is a method of haplotype differentiation. To estimate XP-EHH between populations, first, for each population, iHH values are estimated individually by the incorporation of the EHH of the whole population. This statistic follows the standard normal distribution (Sabeti *et al.*, 2007). It was very difficult to detect selection signals in the population so this statistic was developed. A comparison between two populations may permit the detection of feeble signals.

If we consider population A and B, SD represent standard deviation and E represent mean then, XP-EHH is estimated as:

$$XPEHH = \frac{ln\left(\frac{iHHA}{iHHB}\right) - E\left[ln\left(\frac{iHHA}{iHHB}\right)\right]}{SD\left[ln\left(\frac{iHHA}{iHHB}\right)\right]}$$

This statistic is the best for the detection of selection signature which is close to fixation. Although (XP-EHH) is operated with the help of EHH and it is estimated between the subpopulations. XP-EHH is estimated as the ratio between these subpopulation-specific intrinsic.

2.10.2.3. hapFLK

As we all know that FLK integrates the hierarchical population characteristics, but the test is generalized to take into consideration the haplotype structure. On the contrary, hapFLK reflects the haplotype frequency variations between species. hapFLK is a powerful statistic regarding migration and bottlenecks ((Bonhomme *et al.*, 2010). It also upgrades concerning many other statistics in many situations. hapFLK combines the utilization of haplotypes details and hierarchical composition of populations, that's why this method is more accurate than other methods.

To estimate hapFLK, there is a necessity to incorporate haplotype information in the FLK framework, and based on posterior probability we need to transform SNP into multiallelic genotypes. Kinship matrix (F) is used in the hapFLK method and this method is applied to assess unphased SNP data. Between populations, Reynold's genetic distances are used to analyze the Hierarchical population structure (Manunza *et al.*, 2016). For each SNP, between populations, the pairwise Reynold's distances are estimated and converted into a kinship matrix. hapFLK utilizes increment in computation cost as compare to F_{ST} and FLK because it needs to estimate the LD model on the data. Studies related to genomic footprints/selection signatures in various livestock species are incorporated in **Table 1.**

2.11. Significance of these footprints

The study of genomic footprints and analysing the affected genetic variants are of vital importance in the area of population genetics and genomics. The prime aim of identifying the selection signature in livestock is to grasp the idea of the evolutionary genetics shaping the genome structurally and functionally (Akay *et al.,* 2002). It allows specific key adaptive events that have induced the tremendous phenotypic diversity observed today among different cattle breeds. Furthermore, it helps to understand and verify causal mutations in domains previously reported by different genomic related studies (Aquadro *et al.,* 2001; Zhou *et al.,* 2016). Besides, it enables us to comprehend the biological activities of selection-affected genes that respond to variations in adaptation and production characteristics and help to plan breeding programs to further develop these attributes and to plan strategies for the survival and utilization of endangered breeds (Qanberi *et al.,* 2011). The analysis of the selection signature goes further and explores the true explanation for that diversification. These studies are a sustainable alternative to genome-wide association studies (Maria *et al.,* 2015). Recent advancements in genome-wide scanning technologies and computing resources have expanded the number of resources available to scientists attempting to nucleotide variations in animal genomes. These modern methods of selection signature have improved the identification and analysis of selected regions, and especially for newly mixed breeds, it remains a difficult task to distinguish between actual selection signals or those that seemingly arise via drift (Talanti *et al.,* 2017). The analysis of selection signature is a relevant subject due to its ability to detect genes and beneficial

Table 1. Studies related to genomic footprints/selection signatures in various livestock species

Study	Author	Software	Methods
The origin of selection signatures on bovine chromosome 6	Haye's *et al.*, 2008	• PLINK • Beagle • VCFtools • rehh	his Tajima's D
Selection Signatures in Worldwide Sheep Populations	Fariello *et al.*, 2014	• PLINK • Fastphase • Admixture • hapFLK	FLK hapFLK
Analysis of genomic signatures in broiler chicken	Weixuan Fu *et al.*, 2016	• PLINK • Beagle • rehh • XP-CLR 1.0	ZHp XP-EHH CLR
Analysis of haplotype based methods in Spanish ovine breeds	Manunza *et al.,* 2016	• PLINK • ADMIXTURE • hapFLK	hapFLK FLK
Evidences related to genomic footprints in cattle breed for different traits	Taye *et al.*, 2017	• PLINK • Beagle • VCFtools • rehh • SweepFinder2	Tajima's D, CLR, XPEHH
History of selective sweep in domestic sheep	Rochus *et al.*, 2018	• PLINK • Fastphase • hapFLK	FLK hapFLK
Post-domestic evolution of goat genome.	Bertolini *et al.*, 2018	• SHAPEIT2 • rehh v2.0 R package • PLINK • Beagle	ROH, F_{ST} XP-EHH
Analysis of selection signature for meat traits in Qinchuan cattle breed	Chugang *et al.,* 2018	• PLINK • SweeD • SAMtool • DAVID • VCFtools	Tajima's D F_{ST}
Analysis of Selection signatures in Admixed Dairy Cattle breed of Tanzania	Cheruiyot *et al.,* 2018	• PLINK • ADMIXTURE • fastPHASE • GENESIS package • rehh R package	iHS XP-EHH

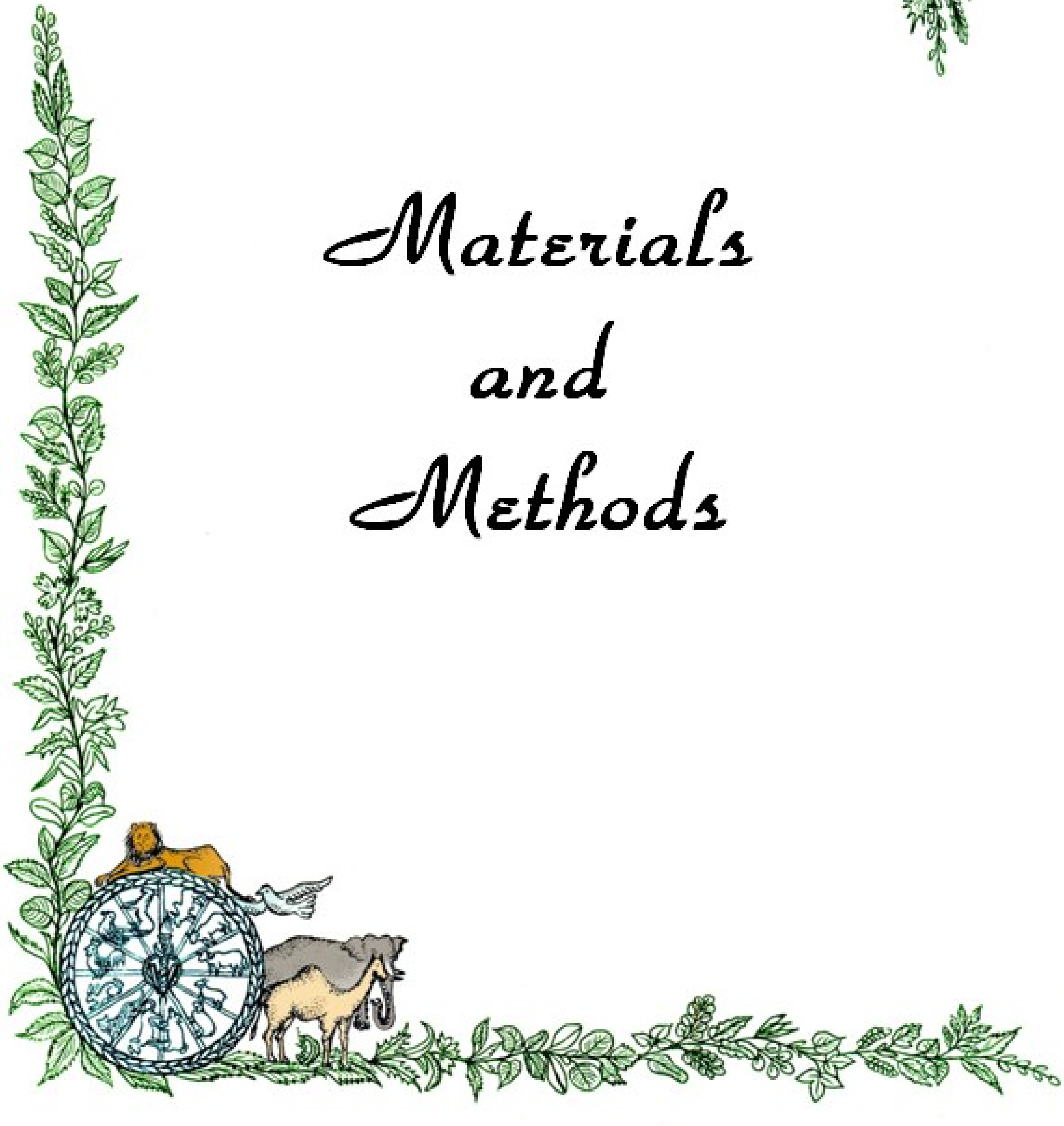
Table 1. Contd...

Study	Author	Software	Methods
Tracing evolutionary study in modern sport horse by analysis of genomic footprints	Nolte *et al.*, 2019	• Beagle 4.0 • PLINK • rehh	ROH XP-EHH iHS
Analysis of selection signatures and breeding history in Braunvieh cattle	Bhati *et al.*, 2020	• PLINK • VCFtools • BCFtools • Eagle2.0 • R package • SweepFinder2	CLR iHS
Scanning of whole genome and genomic selection signature in Yanbian Cattle	Shen *et al.*, 2020	• PLINK • Beagle • VCFtools • rehh	Fst CLR θΠ

mutations induced with phenotypic characteristics without any phenotypic data. In the analysis of genetic divergences between and within populations, the advancement of genomic strategies has significantly contributed. At the extent of the entire genome, high-resolution studies may detect selection signatures that clarify phenotypic differences among species, and thus theoretically specific genetic diversity underlying traits of economic interest.

* * *

Materials and Methods



3 MATERIALS AND METHODS

3.1. Selection of Animals and Data collection

A total of 72 Tharparkar cattle were chosen from the Cattle and Buffalo Farm, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly (UP), India. A total of 10 ml blood sample was collected from jugular vein of each animal. The farm is located at 28^oN and 79^oE with an altitude of 172 meters above sea level. The collection of blood samples was accomplished randomly to acquire a precise idea about the population without any biases.

3.2. Isolation of genomic DNA from blood sample

The genomic DNA was isolated from the fresh whole blood using a kit of DNA isolation (QIAamp^R kit, QIAGEN, Germany) strictly following the manufacturer's guidelines. The following steps were used for the extraction of genomic DNA.

I. A microcentrifuge tube was taken and 20µl QIAGEN protease was put at its bottom.

II. With QIAGEN protease, 200 µl of blood was added to the microcentrifuge tube.

III. From the previous step, 200 µl of buffer AL was incorporated and the mixture was mixed by putting it into the process of pulse-vortexing for 15 sec to minimizing the breakage chances of DNA strands.

IV. The mixture was then incubated at 56^0C for 10 min. The mixture was then further processed to avoid any sticking drops to the lid.

V. After incubation, 200 µl of ethanol was added to the incubated sample and careful mixing was performed again by pulse-vortexing for 15 sec. Again, the mixture was then further processed to avoid any sticking drops to the lid.

VI. The mixture was then applied to the QIAamp Spin Column (available in the kit). Precaution has to be taken to avoid wetting of the rim. The mixture was centrifuged at 8000 rpm for 1 min. The QIAamp Spin Column was then kept in a new 2 ml tube and the previous tube with filtrate was discarded.

VII. The QIAamp Spin Column was opened and 500 µl of AW1 Buffer was added again without wetting the rim and centrifuge at 8000 rpm for 1 min. Again, the QIAamp Spin Column was kept into a new 2ml collection tube.

VIII. The QIAamp Spin Column was opened and 500 µl of AW2 Buffer was added again without wetting the rim and centrifuge at the highest speed 14000 rpm for 3 min. This process was repeated to avoid any chance of AW2 buffer carries over.

IX. The QIAamp Spin Column was kept in a new 1.5 ml micro-centrifuge tube after the discard of filtrate. 125 µl of Buffer AE was then added after the opening of the QIAamp Spin Column. The elute was then incubated at 25^0C (room temperature) for 5 min and further centrifuged at 8000 rpm for 1 min.

X. The elute-containing tube was carefully labelled and paraffin was applied to the cap.

XI. The extracted sample was stored at 4^0C for 24 h, DNA quality was checked and then stored at -80^0C until outsourced.

3.3. Checking of Quality and genotyping of DNA

The concentration and quality of DNA were checked by Spectrophotometer and gel electrophoresis to ensure that the quality of DNA is up to the mark for the aim of genotyping. The DNA samples were refined using 0.8% agarose gel with the process of gel electrophoresis. The gel was visualized under the gel documentation system. After quality checking, the samples were genotyped with Illumine Bovine SNP50k BeadChip version 3 (Gautier *et al.,* 2011). The Bovine SNP50K Bead Chip is a multi-sampled and most widely used SNP array, consist of 53,714 uniformly distributed and extremely informative SNPs with a median spacing of 50.6 kb (Rohilla *et al.,* 2020).

3.4. Preparation of Data-set

We accessed through Dryad Data Repository, WIDDE online web portal (Sempéré *et al.,* 2015), and Krashi Data Inventory Repository (Pawar *et al.,* 2015) and extracted a total of 212 animals' dataset from different bovine breeds, namely Ayrshire (N=18), Brown Swiss (N=24), Gir (N=24), Guernsey (N=21), Hariana (N=10), Holstein Friesian Friesian (N=30), Jersey (N=28), Kankrej (N=10), Ongole (N=20), Red Sindhi (N=10), Sahiwal (N=17). Out of all these breeds Gir, Hariana, Kankrej, Ongole, Red Sindhi, and Sahiwal are indigenous and Ayrshire, Brown Swiss, Guernsey, Holstein Friesian Friesian (HF), and Jersey are exotic. We have merged this extracted dataset with generated Tharparkar genotypic data and obtained a final dataset of 284 animals.

3.5. Quality control Analysis

Only autosomal SNPs have been considered so that there will be no chance of gender biases and we have also excluded SNPs related to mitochondria. The parameters of SNP filtering were incorporated Call rates (>90%), Hardy-Weinberg Equilibrium (HWE) (p<0.0001), polymorphic SNPs, minor allele frequency (MAF) (>5%), coverage of genotypes of markers and individuals, etc. SNP filtering and quality control were performed using PLINKv1.90 to make sure high-quality genotypic data (Purcell *et al.,* 2007). We have avoided the overabundance of SNPs due to linkage Disequilibrium due to widely spaced markers in SNP50K chip array. We have also considered SNPs of known coordinated according to UMD3.1 bovine genome assembly. After quality control, the remaining SNPs were used for further analysis.

3.6. Analysis of Genomic Footprints

Evidence of genomic footprints was investigated with the utilization of multiple statistical tools. In our study, we implemented analysis using different methods including Tajima's D, Fu and Li's D* test, Composite Likelihood Ratio test (CLR), Runs of Homozygosity (ROH), integrated heplotype score (iHS), FLK, hapFLK, and pairwise F_{ST}.

3.6.1. Tajima's D

Tajima's D is a standard tool for detecting selection signature, Tajima's D considers the difference between the mean pair-wise difference ($\widehat{\theta}_{T)}$ and the number of diverse values of segregation/Watterson's estimator ($\widehat{\theta}_{w}$) in nucleotide polymorphism data (Tajima, 1989). Due to the effect of selective sweep, there is an alteration in the value of $\widehat{\theta}_{T}$ but the value of Watterson's estimator ($\widehat{\theta}_{w}$) remains unchanged (Hartl and Clark, 2010). Tajima's D is estimated as;

$$D = \frac{\widehat{\theta}\mathrm{T} - \widehat{\theta}\mathrm{w}}{\sqrt{\widehat{V}}(\widehat{\theta}\mathrm{T} - \widehat{\theta}\mathrm{w})}$$

It identifies regions of the selective sweep due to which the rare novel alleles become fixed in the population (Tajima, 1989). The negative value shows the fixation of rare alleles which means a smaller value (i.e., more negative) of Tajima's D suggests an abundance of rare alleles and an indication of positive selection (Korneliussen *et al.*, 2013). Positive D value represents the balancing selection which shows an abundance of intermediate allele frequencies. The zero D value is the indication of neutral variation.

To estimate the Tajima's D values, we used the software VCFtools with a window size of 500kb (Denecek *et al.,* 2011). We choose the bottom 1% of empirical Tajima's D values as candidate genomic regions for the selective sweep.

3.6.2. Fu and Li's D* test

Fu and Li's test is directly influenced by the coalescent but Tajima's D test is not much affected on coalescent. This test uses information from intraspecific data and is based on the site frequency spectrum. Under the neutral model, the analysis of the mean number of nucleotide differences and the number of segregating/polymorphic sites are interrelated (Fu and Li, 1993). We estimated Fu and Li's D* test using DnaSP Software which was implemented with the sliding window approach of 500 kb window Size and 25 bps step size. Similar to the above method, the bottom 1% was fixed as a cut-off limit, and the selected region was considered a region of interest.

3.6.3. Composite likelihood Ratio (CLR) Test

This method is estimated by divided the maximum likelihood under neutrality (without selection sweep) and maximum likelihood under selective sweep (Eriksson *et al.,* 2008). This method exploits the complete advantage of the demographic pattern of variation of the site frequency spectrum (Bersagliri *et al.,* 2004). We estimated the CLR test by using the information of allele frequencies to scan the regions of the complete sweep. We calculated CLR using the SweeD with a grid size of 40kb. The above 1% of the likelihood values with a cutoff limit ofr 500 bp upstream and downstream were selected as candidate regions and overlapped genes were considered as candidate genes (Chen *et al.,* 2018).

3.6.4. Runs of homozygosity (ROH)

ROHs can be grouped into two categories, long runs, and short runs. In general, the presence of the longer segments shows that there is recent inbreeding. The short ROH dispersion may guide us to find out more remote affinity and less inbreeding (McQuillan *et al.,* 2008). With the help of the detectRUNS R package, ROHs were estimated in a sliding -window based approach. The following threshold parameters were fixed for ROH analysis: (i) (max-oppWindoow-1) maximum number of heterozygous SNPs; (ii) (minSNP-20) minimum number of SNPs; (iii) (maxMissWindow-1) number of missing called per window. The threshold level of ROH estimation was imposed at 0.6 and passed SNPs were selected as candidates for selective sweep (Almeida *et al.,* 2019).

3.6.5. Integrated haplotype score (iHS)

This method is used to find out the unusual haplotype around a particular SNP comparing to the whole genotype. It is a standard measure of decay in EHH from the derived allele compared to the inherited allele at SNP. So, this method needs all the information about the ancestor and derived allele for each SNP. This test was implemented to scanning of ongoing or recent selective sweep regions within a particular part of the population. To impute missing genotype and haplotype phasing, we have utilized Beagle Software with a recent version with its default settings (Browing *et al.,* 2018). iHS values were calculated using Selscan Software (Szpiech *et al.,* 2014). We have also utilized the Norm function to normalize the selected SNP sites and the top 1% normalized values were chosen as candidate regions for the selective sweep (Saravanan *et al.,* 2021).

3.6.6. Fixation index (pairwise F_{ST})

F_{ST} is another way to analyse the genomic footprints and elicit from the population divergence because of allele frequencies at specified locus between the populations. It is used to analyse divergent selection between different exotic cattle breed and Tharparkar was used as a control group for exotic breeds. The values were calculated utilizing the TRES software (Kavakiotis *et al.,* 2015). Low F_{ST} values recommend less genetic differentiation while high values suggest strong genetic differentiation. Negative values were converted to zero as there is no interpretation for negative F_{ST} values. After performing the analysis on each exotic breed, we have imposed a threshold level of 0.1% and these top regions were selected as candidates for selective sweep (Zhao *et al.,* 2015).

3.6.7. FLK and hapFLK (Haplotype Differentiation)

FLK estimation is derived for the single-site population differentiation and it is an extended form of the Lewontin and Krakauer (1973) test. It accounts for the pedigreed set-up between the populations and heterogeneity in effective population size due to genetic drift which is not facilitated by F_{ST} (Reynolds *et al.,* 1983). This method was applied to unphased genotype data by analysing the neighbour-joining population tree and kinship matrix based on the genetic distances (Reynolds') among the breeds. FLK integrates the hierarchical population characteristics, but this test accounts only for a single SNP. To take into consideration the haplotype structure, another method hapFLK was performed which reflects the haplotype frequency variations between species. hapFLK combines the utilization of haplotypes details and the hierarchical composition of populations.

For each SNP, between populations, the pairwise Reynold's distances are estimated (including an outgroup species *Bos grunniens*) and converted into a kinship matrix with the help of R script assembled with hapFLK package (https://forge-dga.jouy.inra.fr/projects/hapflk). We then used the hapFLK package to perform analysis of FLK (Bonhomme *et al.,* 2010) and hapFLK (Fariello *et al.,* 2013) using the genotypic data and kinship matrix. For hapFLK analysis, we have additionally added an assumption of haplotype cluster in the FASTPHASE model (k=15). With the help of generated FLK and hapFLK values, corresponding *p*-values and chi-square density were estimated and the regions with $p<0.005$ and $p<0.001$ were selected as candidate regions for FLK and hepFLK, respectively (Rochus *et al.*, 2018).

3.7. Composite analysis of selective sweep

Composite analysis of selective sweep has immerged as a recent concept to avoid predefined errors and enhance the detection power of candidate selective sweeps (Ma *et al.*, 2015). We amalgamated all the eight methods (Tajima's D, Fu & Li's D*, CLR, ROH, iHS, $F_{ST,}$ FLK, and hapFLK) in DCMS Framework and later estimated the significant correlation between all the methods (Boitard *et al.*, 2016). We amalgamated all the eight methods in DCMS Framework by combining their *p* values and later estimated the significant correlation between all the methods. To estimate the statistics, we used the R MINOTAUR package. *stat to pvalue* function of R MINOTAUR was used to convert statistics values to *p* values and then, covariance analysis was done between all the statistics. Finally, the DCMS statistic was calculated for each method using the MASS package (rlm function).

3.8. Annotation of relevant genomic regions

After the detection of significant candidate regions, we performed gene annotation by using the NCBI database (https://www.ncbi.nlm.nih.gov) utilizing information on gene locations of UMD3.1. DAVIDv6.7 (Database for Annotation, Visualization, and Integrated Discovery) gene ontology PANTHER was used to further analysis of the biological functional activities and selected genes pathway. Important Gene Oncology terms disperse information into the functional characteristics of annotated genes. We have also explored the quantitative trait loci (QTL) database (https://www.animalgenome.org/QTLdb) to find out any overlapping of selected regions with previously published QTL in cattle breeds.

3.9. Different statistical and Bioinformatics software

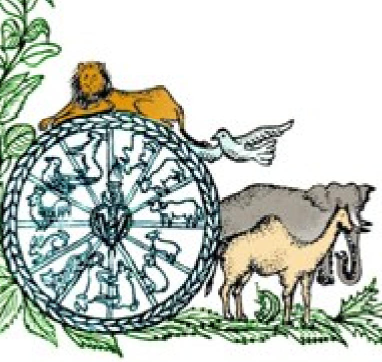
For analysis of such diverse and complementary approaches in huge genotypic data, several bioinformatics tools and statistical software were aimed to use. Some of them is described in **Table 2** along with their features and function.

* * *

Table 2. Various bioinformatics and statistical software

Aim	Software to be used	Specification
➢ **SNP filtration** ➢ **Data management** ➢ **Distance matrix** ➢ **Input file conversion**	**PLINK** Package: PLINKv1.9	It is a free software, developed to analyze basic range to large-scale analysis. It's a command-line interface tool.
➢ **Getting rehh input file format**	**PLINK 2.00** Developer: Christopher Chang	Command line interface tool to get .hap and .map files
➢ **Getting VCF files** ➢ **Calculating Tajima'sD, nucleotide diversity** ➢ **Calculating F_{ST}**	**VCFtools** Developer: Adam Auton	Consists of two parts, a Perl module and a binary executable.It enables the conversion of common PLINK output files into VCF files.
➢ **Imputation and phasing**	**BEAGLE5.1** Developer: Brian Browning	Beagle is a software package utilized for phasing genotypes and for imputing ungenotyped markers
➢ **Getting Fu and Li test**	**DnaSPv6.12.03** Universitat de Barcelona	DnaSP can also carry out several tests of neutrality: Hudson, Kreitman and Aguadé (1987), Tajima (1989), McDonald and Kreitman (1991), Fu and Li (1993), and Fu (1997) tests.
➢ **Imputation and phasing**	**FastPHASE** Developer: P. Scheet	Utilized to infer missing genotypes and haplotype phasing in large scale population data.
➢ **CLR estimation**	**SweeD** Developer: Pavlos Pavlidis	SweeD analyze a composite likelihood ratio test (CLR) of Site Frequency Spectrum (SFS) patterns of single-nucleotide polymorphisms (SNPs)
➢ **F_{ST}**	**TRES (Toolbox for Ranking And Evaluation of SNPs)**	Between population differentiation statistics and ranking
➢ **FLK/hapFLK Analysis**	**hapflk 1.4** Developer: Fariello	hapflk is a software for the analysis of hapFLK and FLK tests for the detection of selection signatures.

Results



4

RESULTS

4.1. Description of Dataset and markers used in the study

With the help of Illumina BovineSNP50 Bead Chip, genotyping of 53212 SNPs was performed of 72 Tharparkar animals and the genotyping rate was 98.169%. We have performed quality control and excluded 997 SNPs due to missing genotype, 759 due to unknown position, 14389 SNPs due to minor allele frequency, 39 variants due to failure of Hardy-Weinberg test, and 3902 SNPs were associated with mtDNA and sex chromosome. In the final data set, a total of 33126 SNPs were present with a genotyping rate of 99.19%.

The summary statistics of genomic Data-set and SNP numbers were incorporated in **Table 3** that was used in footprints analysis. For within breed statistical analysis, we used a

Table 3. Description of Data-set and markers used for different cattle breeds used for genomic footprint analysis

Sr. No.	Name	Abbreviation	Size	Origin	Markers before QC	Markers after QC
1	Ayrshire	**AYR**	18	Exotic	52497	34555
2	Brown Swiss	**BSW**	24	Exotic	51998	35275
3	Gir	**GIR**	24	Indigenous	51998	21780
4	Guernsey	**GNS**	21	Exotic	51998	35413
5	Hariana	**HAR**	10	Indigenous	52497	15606
6	Holstein Friesian	**HOL**	30	Exotic	51998	39488
7	Jersey	**JER**	28	Exotic	51998	34884
8	Kankrej	**KAN**	10	Indigenous	52497	21282
9	Ongole	**ONG**	20	Indigenous	52497	27474
10	Red Sindhi	**RSI**	10	Indigenous	52497	23174
11	Sahiwal	**SAH**	17	Indigenous	52497	20574
12	Tharparkar	**THAR**	72	Indigenous	53212	33126

separate dataset of each breed and for between population statistical analyses, we have merged the dataset of all breeds and found a total of 284 individuals, 12 breeds, and 47,515 markers. For FLK and hapFLK analysis, we additionally used an out-group population of *Bos grunniens* (Yak) and obtained 296 individuals and 23887 common markers.

4.2. Footprints detected by Tajima's D

Using software VCFtools with a constant window size of 500kb (Denecek *et al.,* 2011), We selected a total of 135 candidate regions in all the individuals of 12 different breeds by choosing the bottom 1% of empirical Tajima's D values. These regions were distributed across 29 chromosomes **(Table 4).** The mean Tajima's D across the selected loci was -1.72140. We identified a total of 568 candidate genes overlapped with these 135 regions under positive selective sweep in different breeds. In breed-wise analysis total of 24, 27, 64, 7, 24, 23, 29, 24, 66, 110 and 170 genes were found in Ayrshire, Brown Swiss, Gir, Guernsey, Hariana, Jersey, Kankrej, Ongole, Red Sindhi, Sahiwal, and Tharparkar breeds, respectively. Further, these genes were analysed to find out the information if any of these genes were related to QTL in different cattle breeds. We found few important genes related to important traits viz. for somatic cell score *(NCOA7)* in Ayrshire, paratuberculosis susceptibility (*CLEC7A*), immune response *(KLRF2, KLR9A)* in Brown Swiss, milk fat percentage (*HERC5*), and fat yield (*HERC6*) in Gir, heat tolerance in Tharparkar (*HSP90AB1*), immune response and intramuscular fat (*IGFBP6*) in Tharparkar **(Table 4).** Fig 3 shows the Manhattan plot of the distribution of Tajima's D across all breeds.

4.3. Footprints detected by Fu and Li's D* test

We have applied Fu and Li's D* test by utilizing a window-based tool DnaSP6 (www.ub.edu/dnasp) with window-size 500 kb and step-size 25 bps to find out the sweep regions in the different breed of cattle. As a result, we found a total of 109 candidate regions and total of 413 candidate genes in different cattle breeds. Fig. 4 shows the distribution of Fu and Li's D* test across all breeds. After analysis, we performed functional annotation and extracted out numerous functionally important genes were identified such as *A2ML1, SL24A2, ADARB2, MOCS1* in Ayrshire, *THSD7B* in Brown Swiss, *LRRC8D* in Jersey, *PTPRR* in Hariana, *CDS2* in Red Sindhi were associated with production trait. Like-wise *SLC1A3* in

Table 4. Genomic footprints and candidate genes detected using Tajima's D method in Tharparkar cattle

CHR	Position (Mb)	SNPs	Tajima's D	Genes
1	82.5-83	7	-1.8353	*VPS8,C1H3orf70,TRANK-UUU*
	115-115.5	5	-1.64781	*RAB2B, LOC100295503*
2	25-25.5	5	-1.69779	*TLK1,METTL8,OCAF17*
	47.5-48	5	-1.64134	*EPC2*
	130-130.5	7	-1.56926	*LUZP1,HTR1D,KDM1A,ZNF436,TCEA3*
3	39.5-40	8	-1.74032	*AMY2B*
	94-94.5	8	-1.81437	*ZYG11B,LOC520J18,ZCCH11,GPX7*
4	57.5-58	5	-1.58457	*AKKIB1,TMBIMIB,PEX1,FM13B,RBM48, GATA1*
4	11-11.5	8	-1.84089	***CALCR**, TFPI2, GNGT1, GNG11*
5	27-27.5	8	-1.69439	*E1F4B,SPRYD3,EIF4B,KRTT9,KPT77,KRT73, KRT78,SOAT2,KT8,**IGFBP6***
6	22-22.5	6	-1.88178	*NCLN,SIPR4, GNA15,TRNAM-CAU,GNA11, AES,TLE2, TLE6,ZNF77,THOP1,SGTA*
7	45.5-46	8	-1.86458	*TCF3,PCSK4,REEP6,ADAMTSL5,PLK5, MEX3D,M BD3,UQCR11,ATP8B3,ADAT3*
8	53.5-54	6	-1.68805	*LOC101907250,**GNA14,**VPS13A,**GNAQ***
	61-61.5	9	-1.84848	*MELK,RNF38,PAX5,TRNAC-GCA*
	96-96.5	10	-1.55574	*OR13C3,LOC100140461**,ABCA1**,NIPSNA3A*
12	27.5-28	6	-1.78345	*STARD13*
	47.5-48	11	-1.80851	*DIS3, MZT1,PIBF,BORA*
13	46-46.5	8	-1.71361	***ADARB2**, LOC109569109*
	47-47.5	5	-1.63538	*TAF10,ILK,RRP8,**DNHD1**,TIMM10B,ARFIP2, **TRIM3**,HPX,**A PBB1,**MR2316,PRKCD3P,CNGA, FAM160A2*
16	62.5-63	5	-1.61079	*TORIAIP2,TORIA1P1,CEP350,**LHX4**, ACBD6, QSOX1*
17	68-68.5	10	-1.52662	*MYO18B,TRNAS-GGA,SEZ6L,ASPHD2,HPS4, SRRD, TFIP11,TPST2*
17	70.5-71	5	-1.55681	*ZNRF3,KREMEN1,EMD1,RHBDD3,EWSR1, GAS2L1,RASL10A,AP1B1,NEFH, THOC5, NPSNAP1*
18	11.5-12	8	-1.75871	*KIAA0182,GINS2,C18H160orf74,EMC8,IRF8, COX4I1,MIR2326,MIR2325A*
19	44.5-45	7	-1.88606	*LSM12,G6PC3,HDAC5,ASB16,TMUB2, ATXN7L3, UBTF,SLC4A1,R UNDC3A,SLC25A39, GRN,FAM171A2,ITGA28,GPATCH8*

Table 4. Contd...

CHR	Position (Mb)	SNPs	Tajima's D	Genes
19	45-45.5	9	-1.9535	*C19H17orf105, MPP2, DUSP3, SOST, PPY, NAGS, LSM12, HDACS, ASB16, TUMB2, C19H170orf52*
22	50.5-51	7	-1.81824	*MAPKAPK3, HEMK1, CACNA2D, CISH, C22H3orf18*
22	53-53.5	6	-1.58268	*LOC109575891, MAP4, SMARCC1, ELP6, CSPG5*
23	17.5-18	**7**	-1.53429	*TCTE1,CAPN11,SLC29A1,**HSP90AB1**, SLC35B2, TMEM151B,NFKBIE*

Ayrshire, *PTPRS, IGFBP5, SEC14L1* in Brown Swiss, *RPE65, PARN* in Holstein Friesian Friesian, *MYO10* in Ongole, *KIA1217, CYP2C87* in Red Sindhi, *MTMR7* in Tharparkar was found to be associated with reproductive traits. We also found few genes related to body confirmation such as *PMS2* gene in Holstein Friesian, *MGST1* in Jersey, *MTNR1A* in Gir for rump angle and chest depth (Cochran *et al.*, 2013). A detailed description of an overview of footprints on the genome across different breeds is provided in the table **(Table 5).**

Table 5. Genomic footprints and candidate genes detected using Fu & Li's D method in Tharparkar cattle

CHR	Position (Mb)	Fu & Li's D	Candidate genes
2	13.3-133.5	-1.2406	*VBXN10,PLA2G2C,PLA2G2F,PLA2G2D4, PLA2G2D1, PLA2G5,PLA2GA,OTUD3,PNF186*
5	3-3.5	-0.89554	-
8	65-65.5	-1.91312	*STX17,LOC539818,NR4A3*
12	27-27.5	-1.72592	*RFC3*
16	39-39.5	-1.96577	*LOC1019001943,PRRX1,MROH9*
18	16.5-17	-2.066437	*ABCC12,ABCC11,LONP2,SIAH1*
23	7-7.5	-0.928754	***TAP1**,TAP2,PSMB8,PSMB9,BRD2,RXRB, HSD17B8*
27	19-19.5	-1.69876	***MTMR7**,VPS37A,CNOT7,ZDHHC2,MICU3, FGF20*

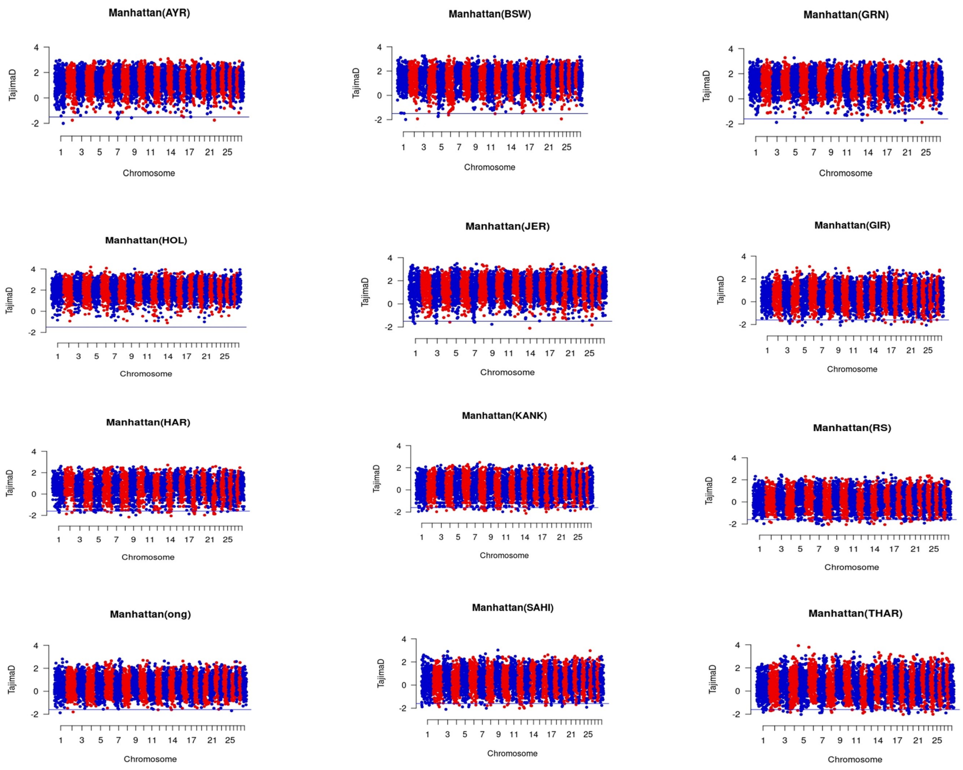


Fig. 3: Genome-wide distribution of Tajima's D values for taurine (5) and indicine (7) cattle breeds. The horizontal line shows the cut-off (Tajima's D) value to call SNP outliers.

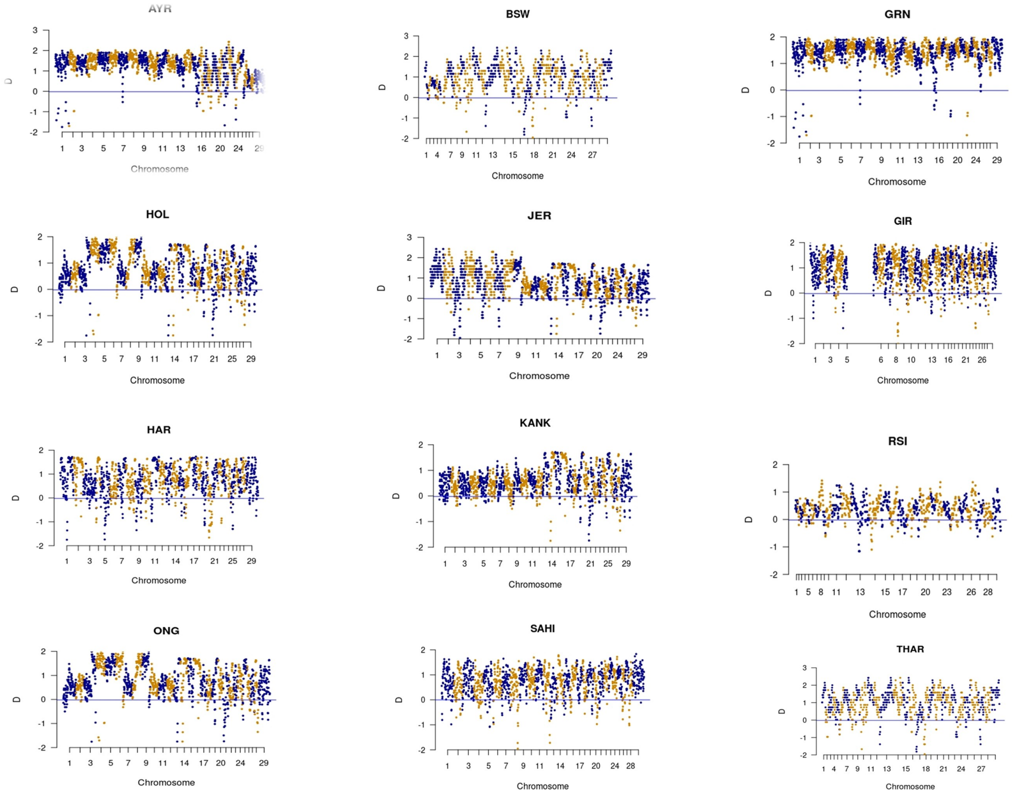


Fig. 4: Genome-wide distribution of Fu & Li's D* values for taurine (5) and indicine (7) cattle breeds. The horizontal line shows the cut-off (Fu & Li's D*) value to call SNP outliers.

4.4. Footprints detected by CLR

We have executed a parallel and optimized tool SweeD (**S**weep **D**etector) with a grid size of 40 kb to detection of likelihood regions in the different cattle breeds (Pavlidis *et al.*, 2017). After the estimation of CLR, we have annotated candidate regions with a distance cut-off limited to be 500 bp upstream and downstream. We have obtained 470 candidate regions in 12 cattle breeds and genes found within the interval of the candidate regions were considered as candidate genes. Total 237 candidate genes were annotated in all the breeds and 31 genes were found common to two or more breeds. Fig. 5 shows the Manhattan plot of the distribution of CLR across all breeds. However, 13, 12, 22, 21, 19, 19, 9, 21, 13, 12, 22 and 23 genes were found unique in Ayrshire, Brown Swiss, Gir, Guernsey, Hariana, Holstein Friesian Friesian, Jersey, Kankrej, Ongole, Red Sindhi, Sahiwal, and Tharparkar respectively. Further, in QTL analysis using tool cattleQTL database v46, we identified genes related to important phenotypes, for example, *CACHD1* gene for calving ease and *PPP1R14C* for body conformation in

Table 6. Genomic footprints and candidate genes detected using CLR method in Tharparkar cattle

CHR	Position (Mb)	CLR values	Candidate genes
2	87889336	8.443875	*LOC109568856*
3	86724288	6.481862	***CACHD1***
5	26018248	9.560096	*UBE2N*
	69220144	6.607312	*NR1H4*
6	17149704	9.126207	*MCUB*
8	458919	8.97096	*PALLD*
	89172488	8.48973	***PHF2***
9	90521600	9.44881	*PPP1R14C*
11	76624216	8.938512	*ADCY3, CENPO*
14	48966900	7.885748	*CSMD3*
15	55178040	7.392399	*UVRAG*
16	40719312	9.88586	*CTNNBIP1*
17	10809466	9.823114	*NR3C2*
19	50369556	7.240045	*SMURF2*
24	40369048	8.027663	*EPB41L3*
25	30637782	6.971668	*CALN1*
28	23117922	7.656583	*CTNNA3*
29	29700248	8.214736	*VSIG2*

Tharparkar, *EEFSEC* gene for muscle taurine content in Kankrej, *CNTN5* gene for milk fat and protein quality and KLHL20 gene for somatic cell score in Gir, *PHF2* gene for milk fat percentage in Tharparkar and Red Sindhi, *GRIP1* for marbling score in Sahiwal and Red Sindhi, *KLHL3* gene for oleic acid content in Hariana and Ongole, *TTC7A* gene for phosphatidylinositol metabolic process, *CADPS* gene for conception rate and *PHF2* gene for fat percentage in Gir and Sahiwal, *MYH10* gene for body weight in Ayrshire and Holstein Friesian, *SMG6* gene for shear force in Brown Swiss, *RHAG* gene for productive life length in Guernsey, and *HSPA12A* gene for milk protein yield in Guernsey and Brown Swiss and also related to heat tolerance **(Table 6).**

4.5. Footprints detected by ROH

With the help of the detectRUNS R package, utilizing the sliding window approach, a total of 28326 ROH were detected in 284 animals of 12 breeds of cattle **(Table 7).** Total 23827 ROH were shorter than 5 Mb, 2724 ROH were found in the range of 5-10 Mb, 1186 ROH were found in the range of 10-20 Mb, 457 ROH were found in 20-40 Mb and 132 ROH were longer than 40 Mb. In the shorter-range category (<5 Mb), the maximum number of ROH was found in Tharparkar. However, in a longer range (>40 Mb) maximum ROHs were identified in Jersey. Some breeds like Kankrej and Ongole did not have any ROH in the category of more than 40 Mb. No indigenous cattle breed was identified with more than 13 ROH in the category of larger than 20 Mb which means there is no recent inbreeding in these breeds. Afterward, we have selected genomic regions that resulted in ROH in more than 80% of individuals, and 34 regions were extracted in all 12 breeds. Total 267 genes were found overlapped in selected regions in various breeds and numerous functionally important genes were identified such as *ERC2* (Ayrshire), *EFCAB* (Brown Swiss), *FST* and *KLHL20* (Jersey), *RGS9* (Kankrej), *QKI* (Sahiwal) for reproduction, *CACNA2D3* (Ayrshire), *MRC2, ANKRD17* and *RASSF6* (Brown Swiss), *PTRH1* and (Holstein Friesian Friesian), *NDUSF4, PELO, ITGA1, SLC27A7* (Jersey), *ZBTB7* (Ongole), *FAM134B* and *TMEM36B* (Red Sindhi) for production, *METTL2A* and *TLK* (Ayrshire), *RAB2A* (Holstein Friesian Friesian), *SORCS1* (Kankrej) for carcass trait, *MYO10* (Red Sindhi) for Bovine tuberculosis susceptibility, *HSP90AB1* (Red Sindhi) for heat tolerance, and *CXCL1* (Brown Swiss), *ARID5A* (Ongole), *LOC100295221* (Sahiwal), *TBK1* (Kankrej) for innate immune response. Fig. 6 shows the genome wide distribution of ROH across all breeds.

Table 7. Number of Runs of Homozygosity (Mb) per class in different cattle breeds

BREEDS	0-5	5-10	10-20	20-40	>40	TOTAL
AYR	1411	186	55	22	7	1681
BSW	1509	286	165	56	9	2025
GIR	1838	234	70	18	8	2168
GNS	2028	300	130	43	9	2510
HAR	142	123	107	73	21	466
HOL	2483	247	145	53	12	2940
JER	2778	488	226	64	17	3573
KAN	742	80	24	4	-	850
RSI	1064	76	32	18	13	1203
ONG	3995	218	17	1	-	4231
SAH	1078	128	62	30	6	1304
THAR	4759	358	153	75	3	5375
TOTAL	23827	2724	1186	457	132	28326

4.6. Footprints detected by iHS

This method exhibits maximal power when there is segregation of selected alleles at intermediate frequencies (Quanbari *et al.,* 2014). With a constant win size of 500kb and top 1% of normalized iHS values, we have obtained a total of 63 putative regions under selective sweep in seven indigenous breeds of cattle. Total 194 genes were found overlapping to 63 selective regions in different breeds. However, in breed-wise analysis, 16,11,38,46,11,39,33 genes were present in GIR, HAR, KANK, ONG, RS, SAH, and THAR respectively. Various functionally important genes were identified, for example, *ARL6* for heat tolerance, *NLRC4* and *BIRC6* for immune response, *EPHA6* for quality of milk, and *LRRIQ3* for body weight in Tharparkar, *DIS3L* for milk production in Red Sindhi, *DNAJB4* for heat tolerance in Ongole, *SEC14L1* for conception rate in Gir, *APBB1* for milk yield in Sahiwal, *FAM49B* for immune response in Hariana **(Table 8).**

Table 8. Genomic footprints and candidate genes detected using iHS method in Tharparkar cattle

CHR	Position (Mb)	Peak SNP	iHS	Candidate genes
1	41.5-42	rs43227622	4.66596	*CRYBG3, EPHA6, ARL6, MINA*
	75.5-76	rs110789800	4.94164	*FGF12*
2	12.5-13	rs43097333	2.70147	*LOC101907972, LOC787311*
3	70.5-71	rs42417908	4.40797	*TINNI3K, FPGT,* ***LRRIQ3***
7	29-29.5	rs43710068	2.73363	*LOC104969134*
	77-77.5	rs41613282	3.35925	*CCNG1,NUDCD2,HMMR,MAT2B*
11	14.5-15	rs110050664	3.0863	*MEMO1, DPY30,* ***SPAST****, SLC30A6, NLRC4, YIPF4, BIRC6*
15	45-45.5	rs110470148	3.35121	*RIC3, TUB, EIF3F, NLRP10, OR10A6, OR5P3, OVCH2*
20	9.5-10	rs41933725	3.07067	*CARTPT, MCCC2, BDP1*
23	18.5-19	rs110204420	3.52594	-

4.7. Footprints detected by F_{ST}

Genomic footprints were scanned for each exotic breed (AYR, BS, GNS, HOL, JER) using pairwise Wright's F_{ST} statistic by choosing Tharparkar as the control population. By using TRES software, Animals were ranked according to their F_{ST} values and a total of 58 candidate regions were identified in 5 exotic breeds by choosing the top 0.1% of empirical F_{ST} values. These regions were considered as signals of positive selection. We found several loci with moderate to low genetic differentiation in different breeds and when we performed gene annotation, a distance cut-off of 500 bp upstream and downstream were considered as candidate regions. We have obtained a total of 52 overlapping genes within the interval associated with the regions under positive selection. Fig. 7 shows the Manhattan plot of the distribution of F_{ST} statistics across all exotic breeds. Finally, we performed functional analysis of all the candidate genes and obtained functionally important genes such as in Ayrshire cattle breed (AYR vs THAR) *TNNI3K* gene for regulation of heart rate, *KIAA1468* for abomasal displacement, *SP110* gene for paratuberculosis, in Brown Swiss (BS vs THAR) *TLR4* gene for calving interval, *PTRH1* gene fat percentage, in Guernsey (GS vs THAR) *ZNF740* gene for milk protein and fat percentage, Holstein Friesian (HOL vs THAR) *CNNM2* gene for milk myristoleic acid content, *HNF1B* gene for feed intake, *C10H15orf41* gene for feet morphology, for Jersey (JER vs THAR) *HNF1B* gene for feed intake, *PCCB* gene for calving rate. Various non-coding genes were also present in different breeds under selective regions **(Table 9).**

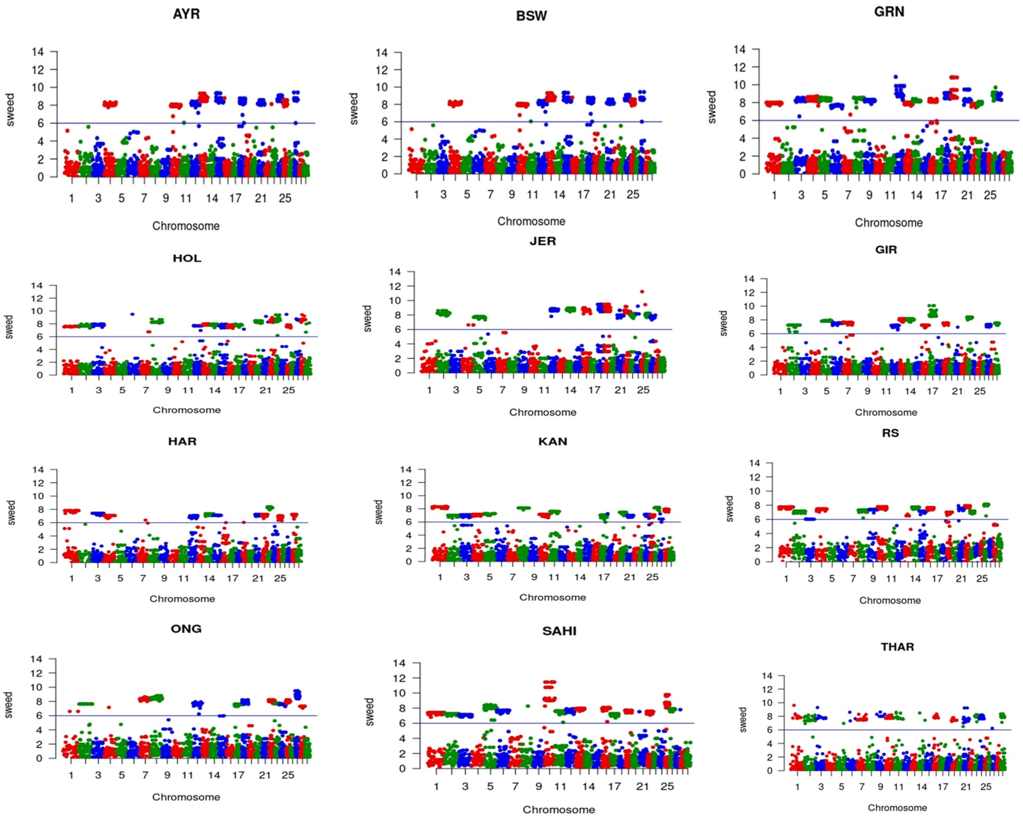


Fig. 5: Genome-wide distribution of CLR values for taurine (5) and indicine (7) cattle breeds. The horizontal line shows the cut-off (CLR) value to call SNP outliers.

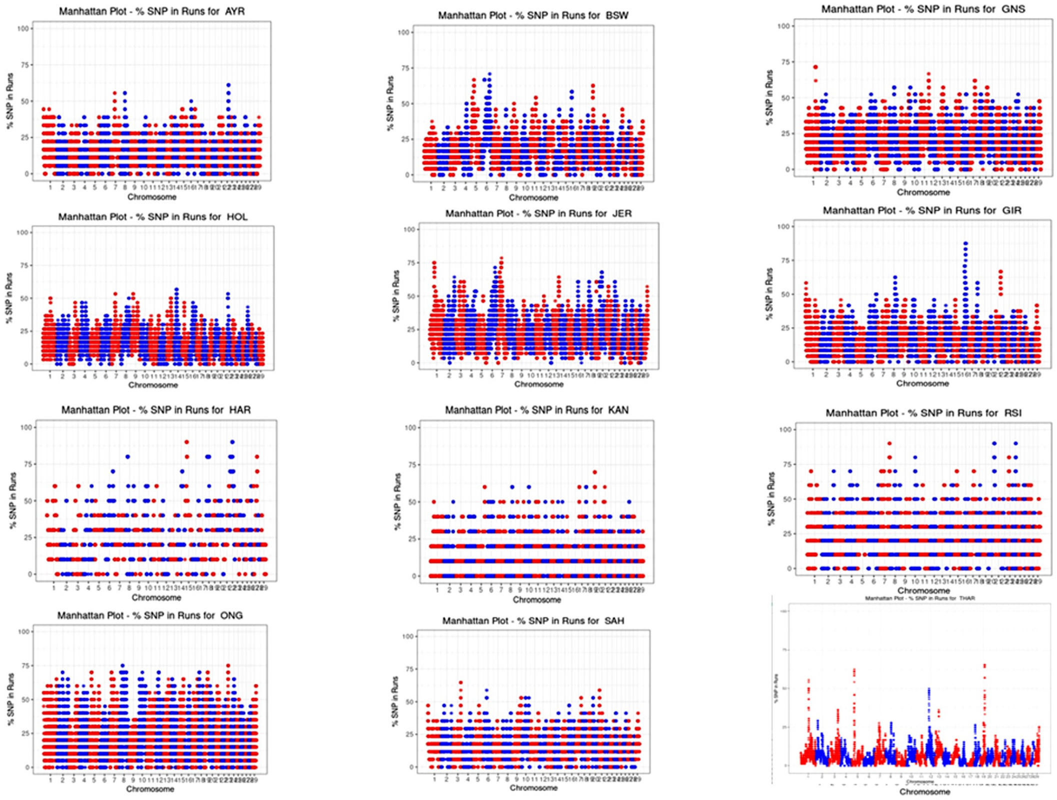


Fig. 6: Genome-wide distribution of ROH values for taurine (5) and indicine (7) cattle breeds.

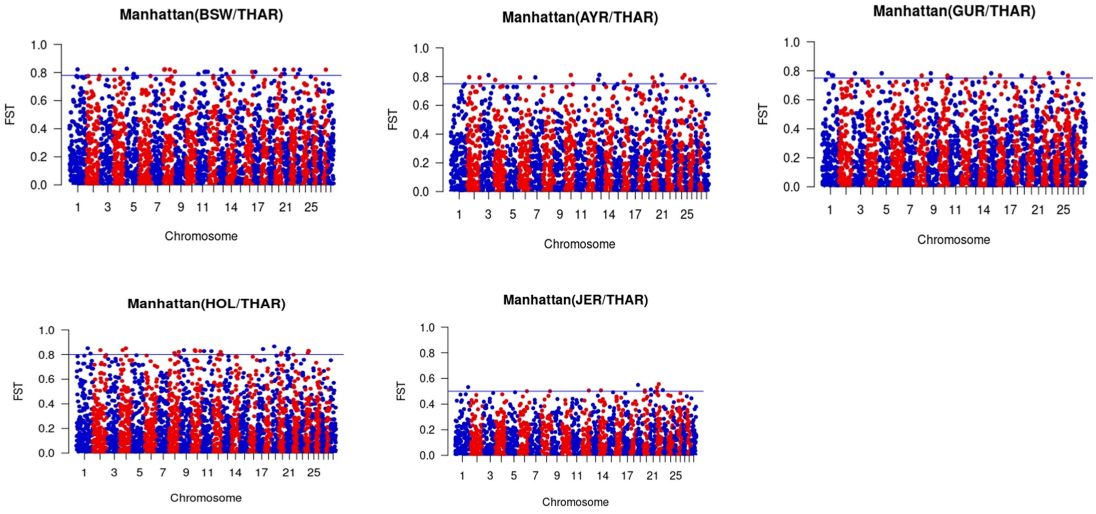


Fig. 7: Genome-wide distribution of FST values for taurine (5) breeds with Tharparkar as control. The horizontal line shows the cut-off (FST) value to call SNP outliers.

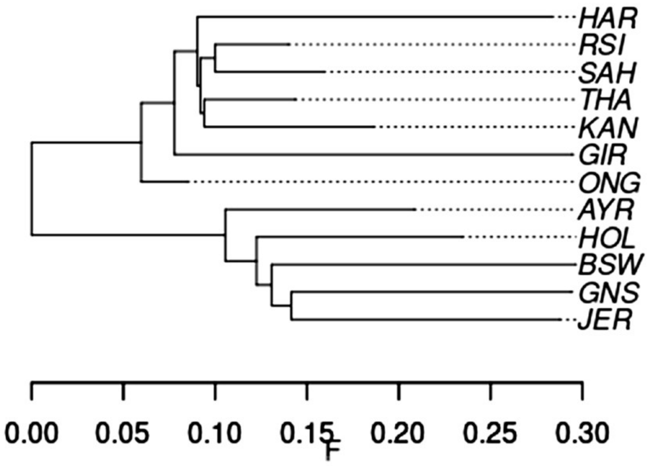


Fig. 8: Illustration of the population tree among all the breeds.

Table 9. Genomic footprints and candidate genes detected using F_{ST} method in Exotic cattle breed by using Tharparkar cattle as control population

BREED	CHR	F_{ST}	SNP	BP POS	GENES
THAR/AYR	13	0.81179	rs109122489	28149879	*MCM10*
	21	0.8105	rs111031346	22991710	*SLC28A1*
	3	0.8105	rs43312703	70496274	*TNNI3K*
	24	0.8105	rs109777881	61109709	***KIAA1468***
	10	0.8105	rs42300167	53716124	*ZNF280D*
	7	0.79437	rs109555725	40290621	*GRK6*
	2	0.79323	rs109287167	119104115	***SP110***
	20	0.79323	rs41930689	6638385	*FAM169A*
	24	0.79323	rs41584305	42698685	*MCTP2,PIEZO2*
	26	0.77787	rs29022332	11201198	*SLC16A12*
	4	0.77689	rs109762864	13061952	*DYNC1I1*
	28	0.76387	rs109467698	29084057	OIT3
	16	0.7614	rs29024834	13944653	*LOC107133223*
THAR/BS	5	0.82721	rs43424389	6656617	-
	20	0.82229	rs109699650	4597488	*ERGIC1*
	22	0.82026	rs109972213	31052013	*LOC104976682*
	28	0.82026	rs42142909	24065549	*CTNNA3*
	8	0.80704	rs41661224	108976660	***TLR4***
	11	0.8062	rs109904482	78645742	*PUM2*
	11	0.80535	rs110440617	98407974	*CFAP157,**PTRH1***
	14	0.80448	rs41632146	63827753	*LOC107133128*
	17	0.80448	rs41635656	17969614	*MAML3*
THAR/GRN	1	0.78452	rs41575366	57141557	*PHLDB2*
	5	0.78312	rs110457668	26986116	***CSAD**,ZNF740*
	9	0.78312	rs41614223	27356886	*NKAIN2*
	10	0.76716	rs41589008	81549606	*SLC39A9*
	26	0.76716	rs41589654	23876476	***CNNM2***
	14	0.75195	rs42378135	50815043	*TRPS1*
	11	0.75195	rs42687393	9790265	*HK2,LOC112448750*
THAR/HOL	19	0.86633	rs110009994	23477516	***HNF1B***
	21	0.85116	rs109276791	28771478	*FAM189A1*
	17	0.84512	rs41585612	55519655	*MLXIP*
	10	0.83648	rs43620139	32518725	***C10H15orf41***
	10	0.82914	rs41594273	84977901	*RBM25*
	24	0.82914	rs43364879	34380600	*LOC101904579,* *LOC104970025*

Table 9. Contd...

BREED	CHR	F_{ST}	SNP	BP POS	GENES
	10	0.82835	rs110240563	38097257	*STARD9*
	11	0.82835	rs41620554	88596997	*LOC101905845*
	12	0.82239	rs41667815	70163625	*LOC515333*
	8	0.82227	rs42884618	96878869	*SLC44A1*
	24	0.81345	rs111010562	28824671	*LOC112444220*
THAR/JER	22	0.55601	rs109622687	23329129	*IL5RA*
	19	0.54964	rs43709673	14305612	***HNF1B***
	1	0.53229	rs109791845	134124559	***PCCB***
	22	0.53023	rs41996490	5051642	-
	21	0.51424	rs110152531	12018140	*LOC101902131*
	23	0.50892	rs109673019	3468292	*DST*
	12	0.50609	rs41612922	67073994	*GPC5*
	22	0.50215	rs41617909	3141689	-
	5	0.49205	rs41591894	83547814	***ITPR2***
	3	0.48604	rs110338384	98552988	*LOC107132336*

4.8. Footprints detected by FLK

We identified selective sweep using FLK that aims at candidate regions of magnificent genetic divergence between populations based on the single SNP. For this purpose, first, we have estimated the Reynolds' genetic distances among the breeds by using hapFLK package (Bonhomme *et al.,* 2010), Reynolds' genetic distances between different breeds were used to estimate neighbour joining population tree among all the breed (Fig. 8 illustrate the population tree among all the breeds). Finally, we performed FLK analysis on all the breeds and identified a total of 326 significant SNPs (Fig. 9 illustrates the histogram of FLK output among all the breeds and reliability of empirical distribution of FLK). We then used Ensembl Variant Effect Predictor or (VEP) and found 156 overlapped genes under selective sweep (https://grch37.ensembl.org/info/docs/tools/vep). Further, in QTL analysis using tool cattleQTL database v46, functional genes were found which were correlated to various phenotypes such as *TIAM1* gene for first service conception, *KCNMB2* gene for service rate, *EMCN* gene for marbling score, *PK1* gene for somatic cell count, *GALNTL6* gene for disease susceptibility, *CA10* gene for carcass weight, *IPO11* gene for residual feed intake, *EIF5* gene for stature, *MAP4* and *C4A* gene for 305-day milk yield, *RBFOX1* for sperm count, *NTM* gene for confirmation score, etc. Fig. 10 and Fig. 11 shows the distribution of FLK values across all breeds.

4.9. Footprints detected by hapFLK

hapFLK method is based on the differentiation between populations based on haplotype frequency between different breeds. This method combines the utilization of haplotype details and the hierarchical composition of population. We have estimated Reynold Distance between all the breeds (outgroup Species OYK) and Kinship matrix with the help of the R script attached with the python hapflk package. The Neighbour-joining population distance algorithm was used to plot NJ population tree using Reynold's distance. The significant genomic regions under hapFLK were analysed and we found a total of 67 SNPs on the 3rd, 7th, 15th, 17th, and 25th chromosomes under selective sweep. However, when we performed VEP analysis, we found 28 candidate genes, Further, in QTL analysis using tool cattle QTL database (QTLdb) v46, functional genes were found which were correlated to various phenotypes such as *ARRB1* gene for Bovine respiratory disease susceptibility and Milk yield, *DCDC1* gene for milk fat yield, *ADAMTS17* gene for age at puberty, daughter pregnancy rate, and Stature, *SCARB1* gene for milk beta-carotene content, milk fat percentage, and milk protein percentage, and *OTOA* gene for rump confirmation. A detailed description of an overview of footprints on the genome across different breeds is provided in **Table 10**. Fig 12 shows Manhattan plot of distribution of hapFLK values across all breeds.

4.10. Composite analysis of selective sweep

Under the composite analysis of selective sweep, a correlation matrix was calculated for each statistical tool and included in **(Table 11).** In analysis, we found that the correlation coefficient between intra-population statistical tools was slightly higher than the correlation coefficient between intra and inter-population statistical tools **(Table 11).** We extracted out various candidate genes in different cattle breeds, which were common in more than two methods. After finding the values of the combining statistic, the cut-off limit was set to above 0.05 and selected genes were assumed to be candidate genes. After functional analysis, we found that genes were associated with few important phenotypes. A list of candidate genes revealed by the combining approach is incorporated in **Table 12**.

* * *

Table 10. Genomic footprints and candidate genes detected using hapFLK method among different breeds

CHR	POS	SNP	hapFLK value	p value	GENES
7	27497043	rs42228636	185.4993843	0.042474786	*CTXN3,* *LOC107132618*
7	27891275	rs41612553	186.1115138	0.039761779	*MEGF10*
7	27973088	rs42232105	186.3469984	0.038757844	*MEGF10*
7	28182762	rs109544319	187.1151957	0.035630831	*C7H5orf63*
7	28657444	rs41658865	191.8845038	0.020641949	*GRAMD2B*
7	30638795	rs110103530	193.7660987	0.016460171	*LOC112447389*
7	31729491	rs109176483	190.8705015	0.023260571	*CEP120*
7	31946863	rs41605088	189.9851073	0.025779125	*PRDM6*
7	33017653	rs110778387	188.8161582	0.029463915	*LOC781423*
7	33937983	rs42587778	185.2454925	0.04364493	*LOC112447394*
15	55254553	rs41632784	183.7962809	0.050853781	***ARRB1***
15	56135629	rs42581757	185.8337873	0.040973976	*UVRAG*
15	56246427	rs41663593	185.8045197	0.041103518	*GVQW3*
15	56716615	rs41628280	188.085797	0.031989375	*EMSY*
15	57179894	rs41770726	188.6135954	0.03014654	*ACER3*
15	58050245	rs41573721	188.9383648	0.029058555	*MUC15*
15	61958209	rs42331838	193.1661742	0.017704005	*LOC112441584*
15	62672763	rs41628788	190.0233671	0.025665592	***DCDC1***
15	63033190	rs42899521	188.3559184	0.031034509	*IMMP1L*
17	48276271	rs109696031	184.2233547	0.048633147	*TMEM132D*
17	49263153	rs109115143	186.4481787	0.038333125	*GLT1D1*
17	52217521	rs110832725	186.9498838	0.036284981	*ADAMTS17*
17	52577471	rs109320410	186.8939316	0.036508689	*TMEM132B*
17	53067783	rs110947124	186.4064263	0.038507906	*BRI3BP*
17	53230205	rs109716262	186.1020207	0.039802708	***SCARB1***
25	19414368	rs109492393	184.8972689	0.045293776	*LOC524391*
25	19995956	rs109589165	184.8472589	0.045534808	*VWA3A*
25	20398199	rs110730355	184.0110868	0.049726556	***OTOA***

Table 11. Values of correlation coefficient of eight method of Analysis of genomic footprints

	Tajima's D	Fu & Li's D	CLR	iHS	ROH	F_{ST}	FLK	hapFLK
Tajima's D		0.12	0.11	0.03	0.09	0.08	0.16	0.14
Fu & Li's D	0.12		0.16	0.18	0.07	0.05	0.56	0.06
CLR	0.11	0.16		0.01	0.13	0.04	0.02	0.02
iHS	0.03	0.18	0.01		0.17	0.63	0.12	0.15
ROH	0.09	0.07	0.13	0.17		0.09	0.67	0.87
F_{ST}	0.08	0.05	0.04	0.63	0.09		0.75	0.97
FLK	0.16	0.56	0.02	0.12	0.67	0.75		0.57
hapFLK	0.14	0.06	0.02	0.15	0.87	0.97	0.57	

Table 12. A partial list of candidate genes revealed by composite analyses

Gene	Chr	Position (BP)	QTLs	Reference
ADARB2	13	46.2-46.5 Mb	Milk protein percentage	Jiang *et al.*, 2019
PARN	25	13.5-13.7 Mb	First service conception	Galliou *et al.*, 2020
ZBTB20	1	59.7-60 Mb	Temperament,Bovine leukemia virus susceptibility	Dos Santos *et al.*, 2017 Brym *et al.*, 2016
KIAA1217	13	25.4-25.7 Mb	Somatic cell score	Marete *et al.*, 2018
FAM134B	20	56.4-56.7 Mb	Still birth	Cole *et al.*, 2011
WDR70	20	36.9-37.2 Mb	Milk protein percentage	Ning *et al.*, 2018
CA8	14	26-26.3 Mb	Insulin like growth factor 1 level Milk protein yield	Fortes *et al.*, 2012Marques *et al.*, 2011
TRHR	14	55.3-55.5 Mb	Milk protein percentage	Jiang *et al.*, 2019
PTRH1	11	98.4 -100 Mb	Milk fat percentage	Fang *et al.*, 2014
CACNA2D3	22	-	Daughter pregnancy rate	Brym *et al.*, 2016

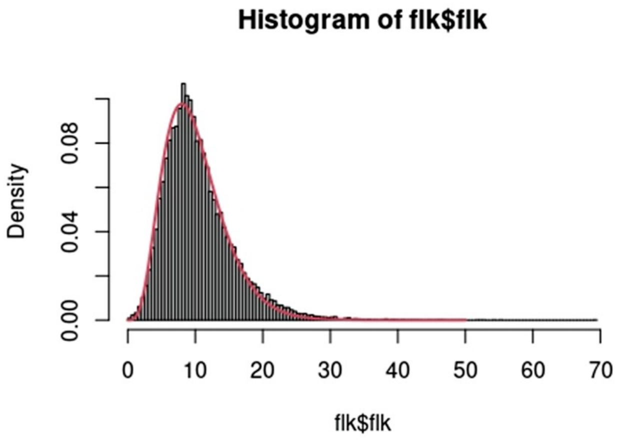


Fig. 9: Illustration of the histogram of FLK output among all the breeds and reliability of empirical distribution of FLK

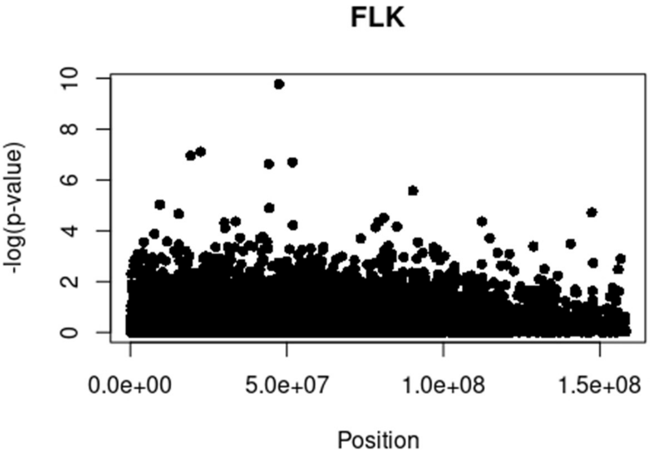


Fig. 10: The distribution of FLK values across all breeds

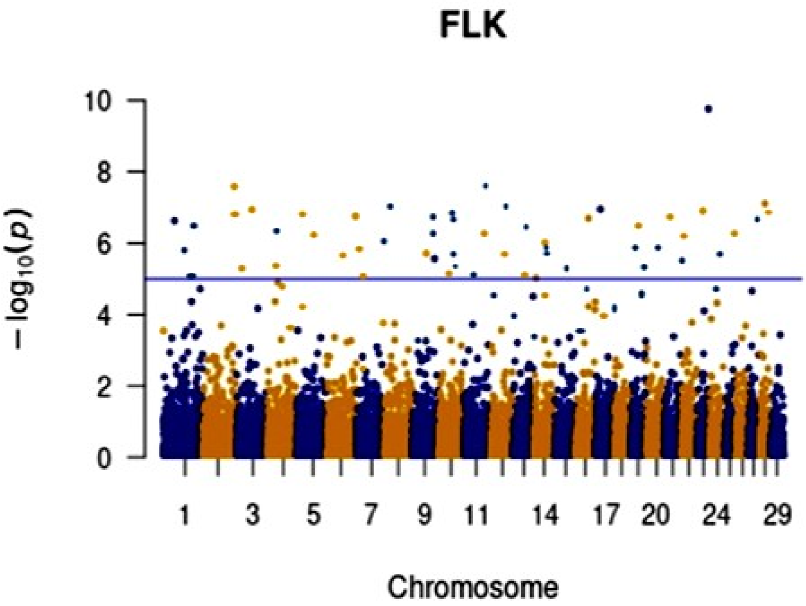


Fig. 11: The Manhattan plot of the distribution of FLK across all breeds.

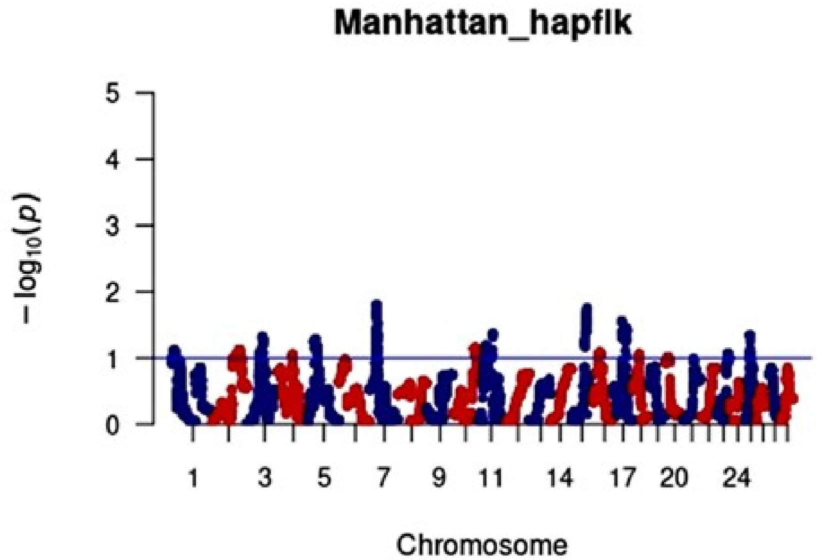
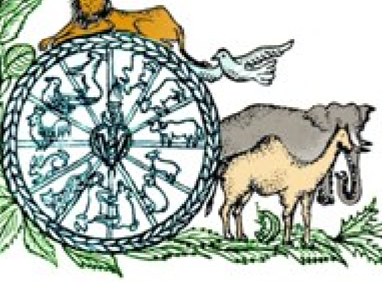


Fig. 12: The Manhattan plot of the distribution of hapFLK across all breeds.

Discussion



5

DISCUSSION

Undoubtedly, the genome of domestic cattle breeds has significantly changed over the generations and subsequently, different breeds have been formed through Domestication and selection (Lofttus *et al.,* 1994). Scanning of genomic footprints due to various evolutionary processes has been a research subject in current years influenced by the wish to recognize the molecular mechanism associated with different adaptive processes and the detection of genomic regions. Identification of these footprints facilitates illustration of the interesting candidate gene identification between and within the population (Akey and Biswas, 2006). There are various statistical tools those have been used to scanning of specific genomic pattern in the livestock genome. However, there is scanty literature available in Zebu cattle related to the analysis of selection signature (Andersson and Georges, 2004). Several studies have been performed and various genes were detected related to reproduction, immune response, coat coloration, tick resistance, immune response, and energy metabolism. In Indigenous cattle, the detection of immune related genes may be relative to selective sweep that has been applying by the pathogen in the region, selective sweep with thermoregulation and reproduction is an adaptation to perform in stressful conditions.

In this study, our objective was to perform an analysis of sweep regions utilizing various statistical tools in 12 breeds of cattle using Bovine SNP 50K Bead Chip. Tajima's D test was used to found out sequences that show divergence from the neutral model and based on the difference between nucleotide differences and site frequency spectrum (SFS) (Achaz, 2009). The likelihood Ratio (CLR) test is also based on SFS but it is not influenced much by different demographical events unlike Tajima's D and thus, shows more accuracy than other SFS-based methods. CLR test can be effective right after when selective alleles have gone to fixation (Stella *et al.,* 2010). Integrated haplotype score is based on the identification of high LD regions and consists of maximal power when there are intermediate frequencies of segregating selective alleles (ongoing Selection) (Makina *et al.,* 2015). F_{ST}, FLK, and hapFLK are used to analyse selection signatures in multiple populations. However, FLK and hapFLK account for the hierarchical pedigreed structure and heterogeneity of different populations. A composite method of analysis was also performed using DCMS approach to find out the strong candidate genes (Lotterhos *et al.*, 2017). Using FLK and hapFLK statistics, we identified various genome regions under positive selection and suggested genes associated with Thyroxin level, fertility and reproduction traits, conformation traits, efficiency of food conversion, and production traits as potentially under selection. Combined analysis of all the methods may provide refined information about selective sweep regions in different breeds. Further, in composite analysis, we extracted out various candidate genes in different cattle breeds which were common in more than two methods. Out of those genes, 10 genes were associated to previously cited QTL genes (*ADARB2, PARN, ZBTB20, KIAA1217, FAM134B, WDR70, CA8, TRHR, PTRH1,* and *CACNA2D3*).

In our study, we summarized identified functional genes overlapped by different traits such as production traits, reproduction traits, thermo-tolerance traits, health traits, carcass traits, growth and exterior traits, we have also compared our results with previous findings of various studies to confirm the reliability of our work.

5.1. Production traits

The genetic make-up, favourable environmental, and fair management practices are the factors on which milk production is majorly dependent. It is critical to strike a balance between selection and economics for production e.g. milk output and composition and functional characteristics. Because it is possible to determine an animal's potential even before the trait is expressed phenotypically techniques used in molecular genetics in conjunction with conventional animal breeding techniques could be used to optimize animal breeding programs.

In our study, we identified various candidate genes associated with production trait using different approaches, for example, *ERC2, NCOA7, MYH10* in Ayrshire, *MRC2,*

ANKRD17, RASSF6, PLXNC1, MOGAT, CLEC2B in Brown Swiss, *PTRH1, RHAG, CSAD, CNNM2* in Guernsey, *ADARB2* in Holstein Friesian (also in Tharparkar) (Fang *et al.,* 2014), *NDUFS4, ITGA1, PELO, SLC27A6* in Jersey. Some of these genes like *ADARB2, ITGA1, KLHL20* have been previously validated in U.S. Holstein Friesian cattle in GWAS studies (Jiang *et al.,* 2019). *ADARB2* gene may play a significant role in BMR, fat decomposition and, insulin hormone regulation and thus, helps in Energy balance (Mei *et al.,* 2017). MRC2 (detected by ROH approach in Brown Swiss) gene may be associated with the udder development and dairy form trait (Cole *et al.,* 2011). RHAG (Chr23: 22-22.5 Mb) gene and CNNM2 (Chr 26: 23.5-24Mb) were associated with ion homeostasis (GO:0006873, GO:0010960), and CSAD was associated with taurine amino acid biosynthesis (GO: 0042412).

We have also performed an analysis in indicine cattle related to production traits and found several important genes such as *HERC5, BTRC, SLC25A48* in Gir, *FAM134B, LY6D* in Red Sindhi, *ZBTB7A, MYO5A, SMCO3, GUCY2C, PLBD1* in Ongole, *CACHD1, EPHA6, TRIM3* in Tharparkar, *ABCA1, ST8SIA1, CDKAL, PTKP2* in Kankrej, *DCHS1, MROH1, ARHGAP39, PP1R116A, TMUM2, CYRH1*(Do Den *et al.,* 2018), *KHDRBS3* (Ning *et al.,* 2018) in Sahiwal, *TRHR* in Hariana, KIAI217 in Sahiwal (Gcebe *et al.,* 2018), *ZBTB7A and MYO5A* genes were associated with 305 days milk production and *SMCO3, GUCY2C, PLBD1* genes were associated with milk fat percentage (Marete *et al.,* 2018; Raschia *et al.,* 2018), *LY6D* (Lymphocyte antigen 6 family member D) gene in Red Sindhi associated with milk protein and fat percentage and previously validated in U. S. Holstein Friesian (Bruitenhuis *et al.,* 2014), *CACHD1* genes found in Tharparkar by CLR approach was also related to stillbirth and marbling score (Ryu *et al.,* 2016).

5.2. Reproduction trait

Reproductive efficiency is a complicated phenomenon affected by genetic and non-genetic variables, with non-genetic influences including environment, diet, and management level. The reproductive efficiency of animals varies not only across species and breeds, but even within the same breed. Even the finest food and care will not be able to push an inferior animal's performance beyond its genetic limit. Improving the genetic qualities of cattle populations is critical at all management levels. A good breeding program is an essential component of the overall animal production system. It is vitally necessary to increase the animal population's productive capability and physical attractiveness.

In livestock, the reproduction trait is one of the most important economic traits. In our study, numerous candidate genes were harboured associated with several reproduction traits such as *SLC25A48* in Ayrshire, EFCAB in Brown Swiss (Galliou *et al.,* 2020), *KLHL20* in Holstein Friesian, *FST, IRPR2* in Jersey, *IQGAP1, QKI, CADPS, IRF2, APBB1, ZBTB20* in Sahiwal, *GNAQ, DNHD1, APBB1* in Tharparkar, *PPA1* and *MYO5A* in Ongole, *KCKN10, TRAPPC9, ABCA1, ZBTB40* in Kankrej, *IRF2, SLC25A48, TMEM181, UTRN, SEC14L1* in Gir. *KLHL20* (involve in somatic cell score*)* gene was detected in more than two methods and found in Holstein, Gir, and Kankrej. *KCKN10* and *TRAPPC9* in Kankrej were found to be associated with interval to the first estrus after calving and daughter pregnancy rate respectively (Cole *et al.,* 2011; Melo *et al.,* 2019). *ZBTB40* was found in Kankrej and associated with sperm motility (Hering *et al.,* 2014). *GMLC14* in Red Sindhi was associated with the calving ease and conception rate and *ABCA1* detected in Kankrej (Chr: 91.2-92 MB) by CLR approach was associated with interleukin beta secretion and cholesterol metabolism (GO: 0008203). *GNAQ* was responsible for the GnRH signalling pathway and Estrogen signalling (Wu *et al.,* 2014). Other than reproduction traits, *TRAPPC9* gene was associated with 13 traits and 180 QTLs, such as stature (Cole *et al*., 2011), Feet and leg conformation, net merit, etc. and associated with several other QTLs like Bone quality and udder texture besides somatic cell count (Kolbehdari *et al.,* 2008). *SEC14L1* in Gir and *MYO5A* in Ongole was associated with conception rate and validated by gene enrichment analysis (Cochran *et al.,* 2013). Few other genes like *HOXC12* and *HOXC13* are indirectly involved in reproduction functions and were formally reported to be under selection in African cattle (Makina *et al.,* 2015).

5.3. Thermo-tolerance trait

Heat stress caused by prolonged periods of high ambient temperature reduces animal production resulting in significant economic losses under the current climate change scenario this disastrous situation for cattle productivity is only becoming worse Strategies aimed at breeding animals with improved thermo tolerance and climatic resilience are being actively pursued these days in order to reduce the stress due to heat. The chaperons are a protein family that is well known for its possible function in thermo tolerance and is frequently used as a cellular thermometer Hsp70 expression is regulated by a variety of variables including cyclic AMPs and others.

The thermotolerance power of an animal is associated with various groups of proteins such as heat shock proteins, heat shock transcription factors, slick hair genes, and *AMPK* (AMP-activation protein kinase). Heat shock proteins (HSPs) proteins enhance the thermotolerant capacity of an animal by the activation of Chaperones and have been a subject of research for several years in Tropical cattle (Belhadj Slimen *et al.*, 2016). In our study, we found few genes related to thermotolerance such as *HSP90AB1* (Heat Shock Protein 90 Alpha Family Class B Member 1) gene in Tharparkar in Red Sindhi (Chr23: 17.5-18 Mb) was associated with binding of heat shock proteins, *HSPA12A* (Heat Shock Protein Family A (Hsp70) Member 12A) in Ongole (Chr26:37130024-37131024) was associated with heat stress mechanism (Dangi *et al.* 2016). *TLR4* was found by the F_{ST} (AYR/ THAR) was previously validated by Bharti *et al.* for reducing the heat stress in 2017 (Sejian *et al.,* 2018; Eslamizad *et al.,* 2020). *ITPR2* gene was also detected by F_{ST} approach (JER/THAR) controls a basic cellular process in sweat glands, and promotes heat loss mechanism (Taye *et al.,* 2018). Thus, by the analysis of genomics of heat tolerance mechanism in indigenous cattle, we can efficiently utilize the magnificent genetic resources of the indigenous germplasm.

5.4. Traits related to health and immune response

Cattle show a wide range of adaptability to harsh climate and adverse conditions especially tropical cattle. Tropical cattle breeds are more resistant to different diseases and adverse climatic conditions but the different diseased conditions may be responsible for mortality and morbidity and thus reduce the productive and reproductive capacity of animals. In our study, we found few candidate genes related to health and immune response, such as *SP110* in Ayrshire, *CXCL2, CLXCL3, CXCL5, CXCL8, GRO1, KLRF2, KLR9A* gene in Brown

Swiss, *NFE2L2* in Jersey, *ARID5A, VAPB, IRF8, MYOM1* gene in Ongole, *NLRC4, MELK, IRF8,* IRF8 in Tharparkar, *TBK1, SETBP1* in Kankrej, *RAG2* in Hariana. *MELK* (Chr18:11.5-12 Mb) and *IRF8* (Chr18:11.5-12 Mb) in Tharparkar (Chr8:61-61.5 Mb) is associated with for intrinsic apoptotic signalling pathway and innate immune response (Sichien *et al.,* 2016). In Ongole *MYOM1* gene for innate immunity (Zimin *et al.,* 2009). *TSNARE1* gene was found in Red Sindhi was associated with Body temperature (Lou *et al.,* 2021). *CSPG5* (chondroitin sulphate proteoglycan 5) in Tharparkar (Chr23:53-53.5 Mb) is associated with positive regulation of defence response to virus by the host (GO: 0002230) and *NLRC4* gene is also related to immune response mechanism (Zimin *et al.,* 2009). *SOCS5, TT7AA* genes were detected in Sahiwal and associated with adaptive immune response and regulation of JAK-STAT pathway (GO: 0007259), *LOC100295221, LOC16868* in Sahiwal associated with innate immune response (GO: 000227), *TBK1* in Kankrej for activation of immune response (GO: 0002218). We found one gene *CACNA2D1* Gene in Ayrshire on BTA4 which was associated with various traits like carcass weight, milk yield, fat percentage, and also associated with clinical mastitis in Sahiwal and Karan Fries (Kumar *et al.,* 2015).

5.5 Carcass trait

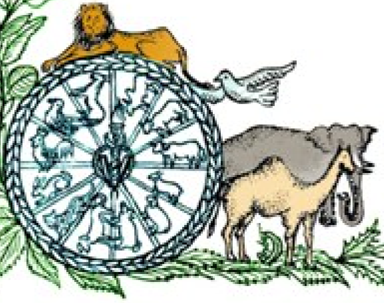
We also found out few genes related to carcass trait and meat quality such as *CACNA2D1* in Ayrshire, *SMG6* in Brown Swiss, *VNN1, HSPA12A* in Guernsey and Brown Swiss, *RAB2A* in Holstein Friesian and Ongole, *ITGA2B* in Jersey, *SORCS1* in Kankrej, *GRIP1, SNX9* in Gir and Sahiwal, *RUNX1T1* in Sahiwal, *CARTPT, IGFBP6, TMUB2* in Tharparkar, *SLC8A1* in Red Sindhi, *CACNA2D1* in Ayrshire (Calcium voltage-gated channel auxiliary subunit alpha-2 / delta) is associated with carcass weight and Fat thickness at the 12th rib (Hou *et al.,* 2010), *METTL2A* and *TLK* gene in Brown Swiss which has been previously reported for juiciness and Myristoleic Acid in Angus and Japanese Black cattle respectively (Sasego *et al.,* 2016; Dos Santos *et al.,* 2017). *EEFSEC* in Kankrej may responsible for muscle taurine content (Sasago *et al.,* 2018) and is also associated with temperament (Chen *et al.,* 2021). *TMEM36B (ANO1)* gene found in Red Sindhi was previously validated as a potential regulator of beef tenderness and calcium chloride infusion in meat (Koohmaraie, 1994).

5.6. Growth and Exterior trait

Growth is directly related to reproduction and production rate in livestock and body weight is directly proportional to the age at puberty in cattle. So, we have emphasized candidate genes related to important growth traits also like *MYH10* in Ayrshire, *MFSD1* in Brown Swiss (Santana *et al.,* 2015), and *CA8* gene in Holstein Friesian. These genes have been previously validated in Guernsey and Brahmin cattle respectability. (Marques *et al.,* 2011; Fortes *et al.,* 2012). Other genes *CLK4, PROP1* in Jersey, *PPROM16, PPP1R14C, LRRIQ3, GNA14, CALCR* in Tharparkar (Wang *et al.,* 2010; Cole *et al.,* 2011). *CALCR* gene in Tharparkar (CHR4: 2-2.5 Mb) was associated with angularity and body condition score (Berkowicz *et al.,* 2011). *PPP1R14C* gene is also associated with teat placement and responsible for udder morphology (Cole *et al.,* 2011). *CA8 and CHD7* gene was detected in Ongole and associated with insulin-like growth factor 1 level and milk protein percentage. These genes have been validated in Brahmin and Holstein Friesian (Marques *et al.,* 2011; Fortes *et al.,* 2012).

* * *

Summary
and
Conclusions



6 SUMMARY AND CONCLUSIONS

Cattle genomic architecture influence by natural and artificial selection for a particular trait that's why various breeds of cattle have been developed with a mosaic of favourable productive, reproductive, and conformational characters. We utilized various exotic and indigenous cattle breeds, which are used usually for milch or dual purpose. Scanning of the genomic footprints affected due to selective sweep would provide valuable insight into the history of selective events for important traits in a particular environment.

For optimum exploitation of this information, we used eight different statistical approaches to find out and verify more genomic footprints for different cattle breeds. We extracted a total of 212 animals' dataset from different bovine breeds, namely Ayrshire (N=18), Brown Swiss (N=24), Gir (N=24), Guernsey (N=21), Hariana (N=10), Holstein Friesian (N=30), Jersey (N=28), Kankrej (N=10), Ongole (N=20), Red Sindhi (N=10), Sahiwal (N=17). We have merged this extracted dataset with generated Tharparkar genotypic data and obtained a final dataset 0f 284 animals. First, we implemented eight different statistical methods separately (Tajima's D, Composite Likelihood Ratio test (CLR), Runs of Homozygosity (ROH), intgrated haplotype score (iHS), FLK, hapFLK, and F_{ST}) and then, finally combined all the methods. Tajima's D estimation was performed in all 12 breeds using the tool VCFtools. The mean Tajima's D across the selected loci was -1.72140 in different breeds and 135 non-overlapping 500kb windows with the bottom 1% of empirical Tajima's D values were selected candidate regions for selection. Composite likelihood ratio (CLR) was performed using SweeD software with grid size 40 and total obtained 470 candidate regions in 12 cattle breeds and genes found within the interval (500bp) spanning the candidate regions were considered as candidate genes. CLR values reflect the probability of likelihood in site frequencies. Fu and Li's D* test focus on coalescence and is estimated using DnaSP. The mean Fu and Li's D* across the selected loci was -1.354277 in different breeds show the positive sweep in different breeds. We have found a total of 109 candidate regions and 413 overlapped genes. iHS (integrated haplotype score) method is based on haplotype information of genome and estimated using selscan software we have found total 63 candidate regions after performing of normalization of iHS values and 194 genes were overlapping to these regions in different indigenous breeds. Runs of homozygosity (ROH) method was performed using detectRUNS software of R package and a total of 28326 ROHs were found in different categories. Afterward, we have selected genomic regions that resulted in ROH in more than 80% of individuals, and 34 regions were extracted in all 12 breeds. Total 267 genes were found overlapped in selected regions in 12 breeds. Inter-population statistics include F_{ST}, FLK, and hapFLK. F_{ST} method is based on population differentiation without hierarchical information and we have estimated F_{ST} in five exotic breeds by fixing Tharparkar as a controlled breed using TRES software. A total of 52 candidate regions were identified in 5 exotic breeds by choosing the top 0.1% of empirical F_{ST} values. Total 58 genes were found to be overlapped within these regions. FLK is based on the LK test and embraces both population differentiation and hierarchical information and is taken into account for each SNP. FLK method was analysed using the python package hapflk. Different scripts were available with the package to find out the population tree and kinship matrix. For analysis, all the 112 breeds were merged and Yak was used as an out-group population. A total of 326 SNPs were found to spanning in candidate regions and 156 overlapped genes under selective sweep. hapFLK method is based on the differentiation between populations based on haplotype frequency between different breeds. This method combines the utilization of haplotype details and the hierarchical composition of the population. We found a total of 67 SNPs on the 3rd, 7th, 15th, 17th, and 25th chromosomes under selective sweep Obtained a total of 28 overlapping genes within the interval associated with the regions under selective sweep. Finally, we performed composite analysis by combining all the methods in 12 breeds by using DCMS function under the composite analysis of selective sweep, a correlation matrix was calculated for each statistical tool. A total of 10 genes were extracted out using combine analysis which may be an indication of strong selective sweep regions in different breeds. The correlation coefficient between intra-population statistical tools was slightly higher than the correlation coefficient between intra and inter-population statistical tools. Functional analysis of selected genes was performed and we identified functional genes overlapped by different traits such as Production traits, Reproduction traits, Thermo-tolerance traits, health traits, Carcass traits, Growth and Exterior traits, we have also compared our results with previous findings of various studies to confirm the reliability of our work.

Conclusion

The following points have been extracted out from the study.

- The mean Tajima's D and mean Fu and Li's D* across the selected loci was -1.354277 and -1.72140 in different breeds indicate the moderate positive selective sweep regions in different cattle breeds.
- The smaller number of ROHs in indigenous cattle may be due to less sample size as compare to Exotic cattle. More number of short ROHs in Tharparkar show a more historical relationship.
- The Composite Likelihood Ratio (CLR) values reflect the most accurate form of site frequency spectrum-based selective sweep without affecting by recombination and demographic variations.
- We found a total of 135, 109, 470, 63, 34 regions using Tajima's D, Fu & Li's D*, CLR, iHS, ROH, and 52, 326, 67 regions using F_{ST}, FLK, hapFLK respectively under selective sweep.
- We found a total of 518, 413, 237, 194, 267 overlapping genes using Tajima's D, Fu & Li's D*, CLR, iHS, ROH, and 58, 156, 28 overlapping genes using F_{ST}, FLK, hapFLK respectively in the annotation.
- Combined analysis of all the methods may provide refined information about selective sweep regions in different breeds.

- Some of the extracted genes from statistical tools and different breeds were overlapped with previously cited QTL regions that may be the source of further evidence for identified genes. However, Non-overlapped genes might be a source of further genomic investigation for QTL analysis.
- Other Evolutionary forces such as random genetic drift, recombination could be the cause of selective sweep affecting the genetic structure of different breeds included in this study. Therefore, other validation phenomena are also required before using these results for different applications of selection and breeding programs. However, these results might be taken as generating a hypothesis but not testing.

✍✍✍

Mini Abstract

7

MINI ABSTRACT

In this study, the genome of different cattle breeds is explored in order to decipher genomic regions affected due to selective events for Productive traits, Reproductive traits, Thermo-tolerant traits, health traits, Quality traits, etc. To find out the genomic modification due to selective sweep, we have applied eight different statistical tools (Tajima's D, Fu & Li's D*, CLR, ROH, iHS, FST, FLK, hapFLK) on 72 Tharparkar animals including six other indigenous (Gir, Hariana, Kankrej, Ongole, Red Sindhi, and Sahiwal) and five exotic cattle (Ayrshire, Brown Swiss, Guernsey, HF, and Jersey) and also performed composite analysis by comparing their covariance matrix. We found a total of 135, 109, 470, 63, 34 regions using Tajima's D, Fu & Li's D*, CLR, iHS, ROH, and 52, 326, 67 regions using F_{ST}, FLK, hapFLK respectively under selective sweep and annotation of these region disclose a total of 518, 413, 237, 194, 267 overlapping genes using Tajima's D, Fu & Li's D*, CLR, iHS, ROH, and 58, 156, 28 overlapping genes using F_{ST}, FLK, hapFLK. The mean Tajima's D and mean Fu and Li's D* across the selected loci was -1.354277 and -1.72140 indicate the moderate positive selective sweep regions in different cattle breeds. From this Study, several candidate genes were extracted that are related to milk traits (*ADARB, WDR70, CA8*), reproductive traits (*PARN, FAM134B2, ZBTB20*), health traits (*SP110, CXCL2, CLXCL3, CXCL5*), immunity traits (*IRF8, MYOM1*). The outcomes discussed here provide a basis for detecting selective sweeps that underlie genetic variation of traits that have functional importance for multiple cattle breeds in different subcontinents. However, further studies are required to rectify the findings using large size genomic data and further phenotypic information.

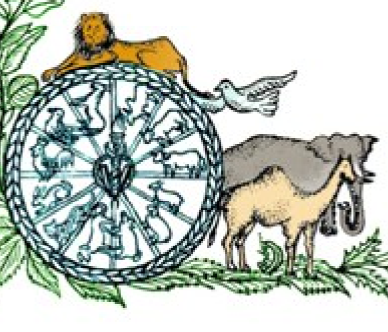
लघु सारांश

8

लघु सारांश

इस अध्ययन में उत्पादक लक्षणों, प्रजनन लक्षण, धर्मो–सहिष्णु लक्षण, स्वास्थ्य लक्षण, गुणवत्ता लक्षण आदि के लिए चयनात्मक घटनाओं के कारण प्रभावित जीनोमिक क्षेत्रों को समझने के लिए विभिन्न पशु नस्लों के जीनोम का पता लगाया जाता है। चयनात्मक स्वीप के कारण जीनोमिक संशोधन का पता लगाने के लिए, हमने आठ विभिन्न सांख्यिकीय उपकरण (Tajima's D, Fu and Li's D*, CLR, ROH, iHS, F_{ST}, FLK, hapFLK) पर छः अन्य स्वेदशी (गिर, हरियाना, कांकेरेज, ओनगोल, रेड सिंधी और साहीवाल) और पांच विदेशी मवेशी (आयरशायर, ब्राउन स्विस, गर्नसे, एचएफ और जर्सी) सहित 72 थारपारकर जानवरों पर और उनके कोवर्न्स संभोग की तुलना करके समग्र विश्लेषण भी किया। हमें Tajima's D, Fu and Li's D*, CLR, iHS, ROH का उपयोग करके कुल 135, 109, 470, 63, 34 क्षेत्र मिले और FST, FLK, hapFLK का उपयोग करके क्रमशः 52, 326, 67 क्षेत्र मिले जो इन क्षेत्रों के चयनात्मक स्वीप और एनोटेशन के तहत कुल 518, 413, 237, 194, 267 ओवरलैपिंग जीन का खुलासा करते हैं। चयनित लोकी में mean Tajima's D and mean Fu and Li's D* -1.354277 और -1.72140 विभिन्न पशु नस्लों में मध्यम सकारात्मक चयनात्मक स्वीप क्षेत्रों का संकेत देता है। इस अध्ययन से, कई उम्मीदवार जीन निकालेगए जो दूध के लक्षण (*ADARB, WDR70, CA8*), प्रजनन लक्षण (*PARN, FAM134B2, ZBTB20*), स्वास्थ्य लक्षण (*SP110, CXCL2, CLXCL3, CXCL5*), प्रतिरक्षा लक्षण (*IRFS, MYOM1*) से संबंधित हैं। यहां चर्चा किए गए परिणाम चुनिंदा स्वीप्स का पता लगाने का आधार प्रदान करते हैं जो विभिन्न उपमहाद्वीपों में कई पशु नस्लों के लिए कार्यात्मक महत्व वाले लक्षणों की आनुवंशिक भिन्नता को रेखांकित करते हैं। हालांकि, बड़े आकार के जीनोमिक डेटा और आगे फेनोटाइपिक जानकारी का उपयोग करके निष्कर्षों को सुधारने के लिए आगे के अध्ययनों की आवश्यकता है।

References



9

REFERENCES

Achaz, G. 2009. Frequency Spectrum Neutrality Tests: One for All and All for One. Genetics, **183(1)**: 249–258.

Ahmad, S. F., Panigrahi, M., Chhotaray, S., Pal, D., Parida, S., Bhushan, B. and Singh, R. K. 2019. Revelation of genomic breed composition in a crossbred cattle of India with the help of Bovine50K BeadChip. Genomics **112(2)**:1531–1535.

Ahmad, S.F., Panigrahi, M., Chhotaray, S., Pal, D., Chauhan, A., Sonwane, A., Parida, S., Bhushan, B., Gaur, G.K., Mishra, B.P. and Singh, R.K. 2018. Population structure and admixture analysis in Frieswal crossbred cattle of India – a pilot study. Animal Biotechnology. **4(3)**: 1532-1547.

Ajmone-Marsan, P., Garcia, J. F. and 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.

Alberti, M., Correa, C., Marzluff, J. M., Hendry, A. P., Palkovacs, E. P., Gotanda, K. M., Hunt, V. M., Apgar, T. M. and Zhou, Y. 2017. Global urban signatures of phenotypic change in animal and plant populations. Proceedings of the National Academy of Sciences. **114(34)**: 8951–8956.

Almeida, O. A. C., Moreira, G. C. M., Rezende, F. M., Boschiero, C., de Oliveira Peixoto, J., Ibelli, A. M. G., Ledur, M. C., de Novais, F. J. and Coutinho, L. L. 2019. Identification of selection signatures involved in performance traits in a paternal broiler line. BMC Genomics. 20(1).

Andersson, L. and Georges, M. 2004. Domestic-animal genomics: deciphering the genetics of complex traits. Nature Reviews Genetics, **5(3)**: 202–212.

Aquadro, C.F., DuMont, V.B. and Reed, F.A. 2001. Genome-wide variation in the human and fruitfly: a comparison. Current opinion in genetics and development, **11(6)**: pp.627-634.

Arad, M., Monserrat, L., Haron-Khun, S., Seidman, J.G., Seidman, C.E., Arbustini, E., Glikson, M. and Freimark, D. 2014. Merits and pitfalls of genetic testing in a hypertrophic cardiomyopathy clinic. The Israel Medical Association journal: IMAJ, **16(11)**: pp.707-713.

Aravamudan, M., Rajasekharan, A. and Ramakrishnan, K.G. 2013. Methods and systems for selecting and presenting content based on a comparison of preference signatures from multiple users. Veveo Inc.

Bahbahani H., Cliffor H., Wragg, D., Mbole-Kariuki M.N., Van Tassell C., Sonstegard T., Woolhouse M. and Hanotte O. 2015. Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. Scientific reports, **5**: p.11729.

Balakrishna, D., Srinivasan Babu, K., Venkatesh Bhat, B., Vinodh, R., Sreedhar, M., Shyam Prasad, G., Pawar, D. B., Shekharappa, P.P. and Ambilwade, J.V. 2015. Improved shoot fly resistant sources by gamma irradiation induced mutations in sorghum (Sorghum bicolor (L.) Moench).

Baruselli, P. S., de Sá Filho, M. F., Martins, C. M., Nasser, L. F., Nogueira, M. F. G., Barros, C. M. and Bó, G. A. 2006. Superovulation and embryo transfer in Bos indicus cattle. Theriogenology, **65(1)**: 77–88.

Belhadj Slimen, I., Najar, T., Ghram, A. and Abdrrabba, M. 2015. Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review. Journal of Animal Physiology and Animal Nutrition, **100(3)**: 401–412.

Berkowicz, E. W., Magee, D. A., Sikora, K. M., Berry, D. P., Howard, D. J., Mullen, M. P., Evans, R. D., Spillane, C. and MacHugh, D. E. 2010. Single nucleotide polymorphisms at the imprinted bovine insulin-like growth factor 2 (IGF2) locus are associated with dairy performance in Irish Holstein-Friesian cattle. Journal of Dairy Research, **78(1)**: 1–8.

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E. and Hirschhorn, J. N. 2004. Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene. The American Journal of Human Genetics. **74(6)**: 1111–1120.

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)**: 113-117.

Biswas, S. and Akey, J. M. 2006. Genomic insights into positive selection. Trends in Genetics. **22(8)**: 437–446.

Boitard, S., Boussaha, M., Capitan, A., Rocha, D. and Servin, B. 2016. Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. Genetics. **203(1)**: 433–450.

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. Genome Biol. Evol. **29(9)**: 2101–2104.

Bollongino, R., Elsner, J., Vigne, J. and Burger, J. 2008. Y SNPs do not indicate hybridisation between European aurochs and domestic cattle. PLoS ONE 3, e3418.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S. and SanCristobal, M. 2010. Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics, **186(1)**: pp.241-262.

Bovine HapMap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 324(5926):528-532. Veterinary Journal, **160(1)**: pp.42-52.

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.

Browning, S. R. and Browning, B. L. 2010. High-Resolution Detection of Identity by Descent in Unrelated Individuals. The American Journal of Human Genetics. **86(4)**: 526–539.

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.

Carlson, S.M. 2005. Developmentally sensitive measures of executive function in preschool children. Developmental neuropsychology, **28(2)**: pp.595-616.

Chen, M., Wang, J., Wang, Y., Wu, Y., Fu, J. and Liu, J. 2018. Genome-wide detection of selection signatures in Chinese indigenous Laiwu pigs revealed candidate genes regulating fat deposition in muscle. BMC Genetics. 19(1).

Cheruiyot, E.K., Bett, R.C., Amimo, J.O., Zhang, Y., Mrode, R. and Mujibi, F.D. 2018. Signatures of selection in admixed dairy cattle in Tanzania. Frontiers in genetics, **9**: p.607.

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. 2019. Genome-wide estimation of inbreeding coefficient, effective population size and haplotype blocks in Vrindavani crossbred cattle strain of India. Biological Rhythm Research, **1(4)**: 0929-1016.

Choudhary, S., Yamini, K., Barman, D., Prasad, C.K., Panchbhai, G. and Kumar, A. 2018. Tharparkar: The pride of desert. Journal of Entomology Zoology Study **6(2)**:1915-1919.

Cochran, S. D., Cole, J. B., Null, D. J. and Hansen, P. J. 2013. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genetics. **14(1)**: 49.

Cochran, S. D., Cole, J. B., Null, D. J. and Hansen, P. J. 2013. Single Nucleotide Polymorphisms in Candidate Genes Associated with Fertilizing Ability of Sperm and Subsequent Embryonic Development in Cattle1. Biology of Reproduction. 89(3).

Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Jr, 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 U.S. Holstein cows. BMC Genomics. 12(1).

DAHDF. 2019. Annual report of department of animal husbandry, dairying and fisheries, ministry of agriculture, Government of India. Retrieved from http://dadf.gov.in/sites/default/filess/Annual%20Report.pdf on 30.08.2019.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G. and Durbin, R. 2011. The variant call format and VCFtools. Bioinformatics. **27(15)**: 2156–2158.

Dangi, S. S., Gupta, M., Dangi, S. K., Chouhan, V. S., Maurya, V. P., Kumar, P., Singh, G., and Sarkar, M. 2014. Expression of HSPs: an adaptive mechanism during long-term heat stress in goats (Capra hircus). International Journal of Biometeorology, **59(8)**: 1095–1106.

Decker, J. E., Taylor, J. F., Kantanen, J., Millbrooke, A., Schnabel, R. D., Alexander, L. J. and MacNeil, M. D. 2016. Origins of cattle on Chirikof Island, Alaska, elucidated from genome-wide SNP genotypes. Heredity. **116(6)**: 502–505.

Decker, J.E., Taylor, J.F., Kantanen, J., Millbrooke, A., Schnabel, R.D., Alexander, L.J. and MacNeil, M.D. 2016. Origins of cattle on Chirikof Island, Alaska, elucidated from genome-wide SNP genotypes. Heredity. **116(6)**: 502.

Diamond, J. 2002. Evolution, consequences and future of plant andanimal domestication. Nature. **418**: 700– 707.

Do, D. N., Schenkel, F., Miglior, F., Zhao, X. and Ibeagha Awemu, E. M. 2020. Targeted genotyping to identify potential functional variants associated with cholesterol content in bovine milk. Animal Genetics. **51(2)**: 200–209.

dos Santos, F. C., Peixoto, M. G. C. D., Fonseca, P. A. de S., Pires, M. de F. Á., Ventura, R. V., Rosse, I. da Cruz., Bruneli, F. A. T., Machado, M. A. and Carvalho, M. R. S. 2017. Identification of Candidate Genes for Reactivity in Guzerat (Bos indicus) Cattle: A Genome-Wide Association Study. PLOS ONE, **12(1)**: e0169163.

Driscoll, E. E., Hoffman, J. I., Green, L. E., Medley, G. F. and Amos, W. 2011. A Preliminary Study of Genetic Factors That Influence Susceptibility to Bovine Tuberculosis in the British Cattle Herd. PLoS ONE, **6(4)**: e18806.

Edea, Z., Dessie, T., Dadi, H., Do, K. T. and Kim, K. S. 2017. Genetic diversity and population structure of Ethiopian sheep populations revealed by high-density SNP markers. Frontiers in Genetics. **8**: 218.

Eriksson, A., Fernstrom, P., Mehlig, B. and Sagitov, S. 2008. An Accurate Model for Genetic Hitchhiking. Genetics, **178(1)**: 439–451.

Eslamizad, M., Albrecht, D. and Kuhla, B. 2020. The effect of chronic, mild heat stress on metabolic changes of nutrition and adaptations in rumen papillae of lactating dairy cows. Journal of Dairy Science, **103(9)**: 8601–8614.

Fang, X., Du, Y., Zhang, C., Shi, X., Chen, D., Sun, J., Jin, Q., Lan, X. and Chen, H. 2011. Polymorphism in a Microsatellite of the Acrp30 Gene and Its Association with Growth Traits in Goats. Biochemical Genetics. **49(7–8)**: 533–539.

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M. and Servin, B. 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics. **193(3)**: 929-941.

Fariello, M.-I., Servin, B., Tosser-Klopp, G., Rupp, R., Moreno, C., Cristobal, M. S. and Boitard, S. 2014. Selection Signatures in Worldwide Sheep Populations. PLoS ONE, **9(8)**: e103813.

Fay, J.C. and Wu, C.I. 2000. Hitchhiking under positive Darwinian selection. Genetics, **155(3)**: pp.1405-1413.

Fels, L. and Distl, O. 2014. Identification and Validation of Quantitative Trait Loci (QTL) for Canine Hip Dysplasia (CHD) in German Shepherd Dogs. PLoS ONE, **9(5)**: e96618.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I. and Sölkner, J. 2013. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. Journal of Animal Breeding and Genetics. **130(4)**: 286-293.

Fernández, M. E., Goszczynski, D. E., Lirón, J. P., Villegas-Castagnasso, E. E., Carino, M. H., Ripoli, M. V. and Giovambattista, G. 2013. Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Genetics and Molecular Biology. **36(2)**: 185-191.

Flink, L.G., Allen, R., Barnett, R., Malmström, H., Peters, J., Eriksson, J., Andersson, L., Dobney, K. and Larson, G. 2014. Establishing the validity of domestication genes using DNA from ancient chickens. Proceedings of the National Academy of Sciences, **111(17)**: pp.6184-6189.

Flori, L., Fritz, S., Jaffrezic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.L. and Gautier, M. 2009. The genome response to artiicial selection: a case study in dairy cattle. PLoS One. 4.

Fortes, M. R. S., Reverter, A., Hawken, R. J., Bolormaa, S. and Lehnert, S. A. 2012. Candidate Genes Associated with Testicular Development, Sperm Quality, and Hormone Levels of Inhibin, Luteinizing Hormone, and Insulin-Like Growth Factor 1 in Brahman Bulls1. Biology of Reproduction, 87(3).

Fortes, M. R. S., Reverter, A., Kelly, M., McCulloch, R. and Lehnert, S. A. 2013. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. Andrology. **1(4)**: 644–650.

Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F. S. and Sargolzaei, M. 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC genomics. **19(1)**: 98.

Fu, Y. X. and Li, W. H. 1993. Statistical tests of neutrality of mutations. Genetics, **133(3)**: 693–709.

Galliou, J. M., Kiser, J. N., Oliver, K. F., Seabury, C. M., Moraes, J. G. N., Burns, G. W., Spencer, T. E., Dalton, J. and Neibergs, H. L. 2020. Identification of Loci and Pathways Associated with Heifer Conception Rate in U.S. 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)**: e78120.

Gautier, M., Laloë, D. and Moazami-Goudarzi, K. 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PloS one. **5(9):** 13038.

Getachew, T., Gizaw, S., Wurzinger, M., Haile, A., Rischkowsky, B., Okeyo, A.M., Sölkner, J. and Mészáros, G. 2015. Survival analysis of genetic and non-genetic factors influencing ewe longevity and lamb survival of Ethiopian sheep breeds. Livestock Science, **176**: pp.22-32.

Gholami, M., Reimer, C., Erbe, M., Preisinger, R., Weigend, A., Weigend, S. and Simianer, H. 2015. Genome scan for selection in structured layer chicken populations exploiting linkage disequilibrium information. PloS one. **10(7)**: e0130497.

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.

Gouveia, V.V., Milfont, T.L. and Guerra, V.M. 2014. The functional theory of human values: From intentional overlook to first acknowledgement—A reply to Schwartz. Personality and Individual Differences, **68**: pp.250-253.

Gowane, G. R., Kumar, A. and Nimbkar, C. 2019. Challenges and opportunities to livestock breeding programmes in India. Journal of Animal Breeding and Genetics. **136(5)**: 329-338.

Gregory, T. R. 2009. Artificial Selection and Domestication: Modern Lessons from Darwin's Enduring Analogy. Evolution Education Outreach. **2(1)**: 5–27.

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

Gurgul, A., Jasielczuk, I., Ropka-Molik, K., Semik-Gurgul, E., Pawlina-Tyszko, K., Szmatoła, T. and Krupiński, J. 2018. A genome-wide detection of selection signatures in conserved and commercial pig breeds maintained in Poland. BMC genetics. **19(1)**: 95.

Hagen, D. E., Unni, D. R., Tayal, A., Burns, G. W. and Elsik, C. G. 2018. Bovine Genome Database: Tools for Mining the Bos taurus Genome. In: Kollmar M. (eds) Eukaryotic Genomic Databases. Methods in Molecular Biology, vol 1757. Human Press, New York, NY. pp. 211–249.

Hagger, C. 2005. Estimates of genetic diversity in the brown cattle population of Switzerland obtained from pedigree information. Journal of Animal Breeding and Genetics. **122(6)**: pp.405-413.

Haneda, M., Utsunomiya, K., Koya, D., Babazono, T., Moriya, T., Makino, H., Kimura, K., Suzuki, Y., Wada, T., Ogawa, S. and Inaba, M. 2015. A new classification of diabetic nephropathy 2014: a report from joint committee on diabetic nephropathy. Journal of diabetes investigation. **6(2)**: pp.242-246.

Hansen, P. J. 2004. Physiological and cellular adaptations of zebu cattle to thermal stress. Animal Reproduction Science. **82–83**: 349–360.

Harris, E.E. and Meyer, D. 2006. The molecular signature of selection underlying human adaptations. American Journal of Physical Anthropology. **131(S43)**: 89-130.

Hayes, B.J., Chamberlain, A.J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., Sethuraman, L. and Goddard, M.E. 2009. A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. Animal genetics. **40(2)**: 176-184.

Hayes, B.J., Lien, S., Nilsen, H., Olsen, H.G., Berg, P., Mac, Eachern, S., Potter, S. and Meuwissen, T.E. 2008 The origin of selection signatures on bovine chromosome 6. Animal Genetics **39:** 105-111.

Hedrick, P.W. 2005. A standardized genetic differentiation measure. Evolution. **59(8)**: pp.1633-1638.

Hering, D. M., Olenski, K. and Kaminski, S. 2014. Genome-wide association study for poor sperm motility in Holstein-Friesian bulls. Animal Reproduction Science, **146(3–4)**: 89–97.

Höglund, M., Sandin, F. and Simonsson, B. 2015. Epidemiology of chronic myeloid leukaemia: an update. Annals of Hematology. **94(S2)**: 241–247.

Hou, G.-Y., Yuan, Z.-R., Gao, X., Li, J.-Y., Gao, H.-J., Chen, J.-B. and Xu, S.-Z. 2010. Genetic Polymorphisms of the CACNA2D1 Gene and Their Association with Carcass and Meat Quality Traits in Cattle. Biochemical Genetics, **48(9–10)**: 751–759.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A. M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R. F., Magwire, M. M., Blankenburg, K., Carbone, M. A., Chang, K., Ellis, L. L., Fernandez, S., Han, Y., Highnam, G., Hjelmen, C. E. and Mackay, T. F. C. 2014. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Research, **24(7)**: 1193–1208.

Hudson, R. R., Slatkin, M. and Maddison, W. P. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics. **132(2)**: 583-589.

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N. and Vellend, M. 2008. Ecological consequences of genetic diversity. Ecology Letters. **11(6)**: 609–623.

Igoshin, A. V., Yurchenko, A. A., Belonogova, N. M., Petrovsky, D. V., Aitnazarov, R. B., Soloshenko, V. A. and Larkin, D. M. 2019. Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. BMC Genetics. **20(S1)**: 26.

Iso-Touru, T., Tapi, M., Vilkki, J., Kiseleva, T., Ammosov Ivanova, Z., Popov, R., Ozerov, M. and Kantanen, J. 2016. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. Animal Genetics. **47(6)**: 647–657.

Jiang, J., Ma, L., Prakapenka, D., VanRaden, P. M., Cole, J. B. and Da, Y. 2019. A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. Frontiers in Genetics. 10.

Kashyap, N., Kumar, P., Deshmukh, B., Dige, M. S., Sarkar, M., Kumar, 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.

Kavakiotis, I., Triantafyllidis, A., Ntelidou, D., Alexandri, P., Megens, H.-J., Crooijmans, R. P. M. A., Groenen, M. A. M., Tsoumakas, G. and Vlahavas, I. 2015. TRES: Identification of Discriminatory and Informative SNPs from Population Genomic Data: Figure 1. Journal of Heredity, **106(5)**: 672–676.

Kemper, K.E., Saxton, S.J., Bolormaa, S., Hayes, B.J. and Goddard, M.E. 2014. Selection for complex traits leaves little or no classic signatures of selection. BMC genomics. **15(1)**: p.246.

Khatkar, M.S., Collins, A. and Cavanagh, J.A. 2006b. A first generation metric linkage disequilibrium map of bovine chromosome 6. Genetics. **174**: 79– 85.

Khatkar, M.S., Thompson, P.C., Tammen, I., Cavanagh, J.A., Nicholas, F.W. and Raadsma, H.W. 2006a. Linkage disequilibrium on chromosome 6 in Australian Holstein-Friesan cattle. Genetics Selection Evolution. **38**: 463–77.

Kim, Y., Nielsen, R. 2004. Linkage disequilibrium as a signature of selective sweeps. Genetics. **167(3)**: pp.1513-1524.

Kim, Y. and Stephan, W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics. **160(2)**: pp.765-777.

Kingan, S. B., Garrigan, D. and Hartl, D. L. 2010. Recurrent Selection on the winters sex-ratio Genes in Drosophila simulans. Genetics. **184(1)**: 253–265.

Kolbehdari, D., Wang, Z., Grant, J. R., Murdoch, B., Prasad, A., Xiu, Z., Marques, E., Stothard, P. and Moore, S. S. 2009. A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. Journal of Animal Breeding and Genetics. **126(3)**: 216–227.

Koohmaraie, M. 1994. Muscle proteinases and meat aging. Meat science. **36(1-2)**: 93-104.

Korneliussen, T. S., Moltke, I., Albrechtsen, A. and Nielsen, R. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics, 14(1).

Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. Annual review of genomics and human genetics, **1(1)**: pp.539-559.

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

Kumar, H., Panigrahi, M., Chhotaray, S., Parida, S., Chauhan, A., Bhushan, B. and Singh, R. K. 2019(b). Comparative analysis of five different methods to design a breed-specific SNP panel for cattle. Animal Biotechnology. 1–7.

Kumar, H., Panigrahi, M., Saravanan, K. A., Parida, S., Bhushan, B., Gaur, G. K., Dutt, T., Mishra, B. P. and Singh, R. K., 2021. SNPs with intermediate minor allele frequencies facilitate accurate breed assignment of Indian Tharparkar cattle. Gene, 777, 145473. https://doi.org/10.1016/j.gene.2021.145473

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., 2021. Genome-wide detection of copy number variations in Tharparkar cattle. Animal Biotechnology, 1–8. https://doi.org/10.1080/10495398.2021.1942027

Kumar, R., Gupta, I. D., Verma, A., Verma, N., Magotra, A. and Vineeth, M. R. 2015. Molecular characterization and polymorphism detection in HSPB6 gene in Sahiwal cattle. Indian Journal of Animal Research, OF.

Lamichhaney, S., Card, D. C., Grayson, P., Tonini, J. F. R., Bravo, G. A., Näpflin, K., Termignoni-Garcia, F., Torres, C., Burbrink, F., Clarke, J. A., Sackton, T. B. and Edwards, S. V. 2019. Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. Philosophical Transactions of the Royal Society B: Biological Sciences. **374(1777)**: 20180248.

Le Bihan-Duval, E., Hennequet-Antier, C., Berri, C., Beauclercq, S.A., Bourin, M.C., Boulay, M., Demeure, O. and Boitard, S. 2018. Identification of genomic regions and candidate genes for chicken meat ultimate pH by combined detection of selection signatures and QTL. BMC Genomics, **19(1)**: p.294.

Le Rouzic, A., Hansen, T.F., Gosden, T.P. and Svensson, E.I. 2015. Evolutionary time-series analysis reveals the signature of frequency-dependent selection on a female mating polymorphism. The American Naturalist. **185(6)**: pp.E182-E196.

Li, Y. and Kim, J.J. 2015. Effective population size and signatures of selection using bovine 50K SNP chips in Korean native cattle (Hanwoo). Evolutionary Bioinformatics, **11**: pp.EBO-S24359.

Lindsay, B. G. 1988. Composite likelihood methods. Contemp. Math. **80(1)**: 221-239.

Llopis-Albert, C. and Capilla, J.E. 2009. Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data: 3. Application to the Macrodispersion Experiment (MADE-2) site, on Columbus Air Force Base in Mississippi (USA). Journal of hydrology. **371(1-4)**: pp.75-84.

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q. and Simianer, H. 2015. Properties of different selection signature statistics and a new strategy for combining them. Heredity. **115(5)**: 426–436.

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q. and Simianer, H. 2015. Properties of different selection signature statistics and a new strategy for combining them. Heredity. **115(5)**: pp.426-436.

Maiorano A.M., Lourenco D.L., Tsuruta S., Ospina A.M.T., Stafuzza N.B., Masuda Y., Filho A.E.V., Cyrillo J.N.D.S.G, Curi R.A. and Silva J.A.I.D.V. 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLoS One. **13(8)**: p.e0200694.

Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., Taylor, J. F., Makgahlela, M. L. and Maiwashe, A. 2015. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genetics Selection Evolution, 47(1).

Marete, A. G., Guldbrandtsen, B., Lund, M. S., Fritz, S., Sahana, G. and Boichard, D. 2018. A Meta-Analysis Including Pre-selected Sequence Variants Associated With Seven Traits in Three French Dairy Cattle Populations. Frontiers in Genetics, 9.

Marete, A., Sahana, G., Fritz, S., Lefebvre, R., Barbat, A., Lund, M. S., Guldbrandtsen, B., and Boichard, D. 2018. Genome-wide association study for milking speed in French Holstein cows. Journal of Dairy Science. **101(7)**: 6205–6219.

Marques, E., Grant, J. R., Wang, Z., Kolbehdari, D., Stothard, P., Plastow, G. and Moore, S. S. 2011. Identification of candidate markers on bovine chromosome 14 (BTA14) under milk production trait quantitative trait loci in Holstein. Journal of Animal Breeding and Genetics, **128(4)**: 305–313.

Mastrangelo, S., Tolone, M., Sardina, M.T., Sottile, G., Sutera, A.M., Di Gerlando, R. and Portolano, B. 2017. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. Genetics Selection Evolution, **49(1)**: pp.1-10.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P. and Van Tassell, C. P. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS one. **4(4)**: e5350.

Mc Vittie, A., Hussain, S. 2013. TEEB valuation database manual.

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., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F. and Wilson, J. F. 2008. Runs of Homozygosity in European Populations. The American Journal of Human Genetics, **83(3)**: 359–372.

McRae, K.M., McEwan, J.C., Dodds, K.G. and Gemmell, N.J. 2014. Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. BMC genomics. **15(1)**: p.637.

Metta, M., Kanginakudru, S., Gudiseva, N. and Nagaraju, J. (2004). BMC Genetics. **5(1)**: 16.

Mirkena T., Duguma G., Haile A., Tibbo M., Okeyo A.M., Wurzinger M. and Sölkner J. 2010. Genetics of adaptation in domestic farm animals: A review. Livestock Science. **132(1-3)**: pp.1-12.

Mueller, J.C. 2004. Linkage disequilibrium for different scales and applications. Briefings in bioinformatics. **5(4)**: pp.355-364.

Nielsen, R. 2005. Molecular signatures of natural selection. Annual Review. Genetics. **39**: pp.197-218.

Ning, C., Kang, H., Zhou, L., Wang, D., Wang, H., Wang, A., Zhang, J. S. and Liu, J. 2017. Performance Gains in Genome-Wide Association Studies for Longitudinal Traits via Modeling Time-varied effects. Scientific reports, **7(1)**: 590.

Odani, M., Narita, A., Watanabe, T., Yokouchi, K., Sugimoto, Y., Fujita, T., Oguni, T., Matsumoto, M. and Sasaki, Y. 2006. Genome wide linkage disequilibrium in two Japanese beef cattle breeds. Animal Genetics. **37**: 139– 44.

Oleksyk, T. K., Smith, M. W. and O'Brien, S. J. 2010. Genome-wide scans for footprints of natural selection. Philosophical Transactions of the Royal Society B: Biological Sciences. **365(1537)**: 185–205.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G. and Simianer, H. 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics. **12(1)**: 318.

Qanbari, S. and Simianer, H. 2014. Mapping signatures of positive selection in the genome of livestock. Livestock Science. **166**: 133–143.

Qiao X., Su R., Wang Y., Wang R., Yang T., Li X., Chen W., He S., Jiang Y., Xu Q. and Wan W. 2017. Genome-wide target enrichment-aided chip design: a 66 K SNP chip for cashmere goat. Scientific reports. **7(1)**: pp.1-13.

Raschia, M. A., Nani, J. P., Maizon, D. O., Beribe, M. J., Amadio, A. F. and Poli, M. A. 2018. Single nucleotide polymorphisms in candidate genes associated with milk yield in Argentinean Holstein and Holstein x Jersey cows. Journal of Animal Science and Technology, 60(1).

Ray, P., Torck A., Quigley L., Wangzhou A., Neiman M., Rao C., Lam T., Kim J.Y., Kim T.H., Zhang M.Q. and Dussor, G. 2018. Comparative transcriptome profiling of the human and mouse dorsal root ganglia: an RNA-seq-based resource for pain and sensory neuroscience research. Pain, **159(7)**: p.1325.

Rege, J.E.O. and Tawah, C.L. 1999. The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains. Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animals. **26**: pp.1-25.

Reynolds, B.S. and Weir, C.C. 1983. Cockerham. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics. **105**: 767-779

Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G. and Servin, B. 2018. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. BMC Genomics, 19(1).

Rutten, C. J., Velthuis, A. G. J., Steeneveld, W. and Hogeveen, H. 2013. Invited review: Sensors to support health management on dairy farms. Journal of Dairy Science. **96(4)**: 1928–1952.

Ryu, D., Mouchiroud, L., Andreux, P. A., Katsyuba, E., Moullan, N., Nicolet-dit-Félix, A. A., Williams, E. G., Jha, P., Lo Sasso, G., Huzard, D., Aebischer, P., Sandi, C., Rinsch, C. and Auwerx, J. 2016. Urolithin A induces mitophagy and prolongs lifespan in C. elegans and increases muscle function in rodents. Nature Medicine, **22(8)**: 879–888.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. and Ackerman, H.C. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature, **419(6909)**: pp.832-837.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R. and Schaffner, S.F. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature. **449(7164)**: pp.913-918.

Santana, M. H. A., Ventura, R. V., Utsunomiya, Y. T., Neves, H. H. R., Alexandre, P. A., Oliveira Junior, G. A., Gomes, R. C., Bonin, M. N., Coutinho, L. L., Garcia, J. F., Silva, S. L., Fukumasu, H., Leme, P. R. and Ferraz, J. B. S. 2015. A genomewide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nellore cattle. Journal of Animal Breeding and Genetics. **132(6)**: 420–427.

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

Saravanan, K. A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G. K., Dutt, T., Mishra, B. P. and Singh, R. K. 2021. Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. Genomics. **113(3)**: 955–963.

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. 2020. Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. Animal Biotechnology. 1–15.

Sasago, N., Abe, T., Sakuma, H., Kojima, T. and Uemoto, Y. 2016. Genome-wide association study for carcass traits, fatty acid composition, chemical composition, sugar, and the effects of related candidate genes in Japanese Black cattle. Animal Science Journal. **88(1)**: 33–44.

Scheet, P. and Stephens, M. 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. The American Journal of Human Genetics. **78(4)**: 629–644.

Schwartz, Y. B. and Cavalli, G. 2017. Three-Dimensional Genome Organization and Function in Drosophila. Genetics. **205(1)**: 5–24.

Sejian, V., Bhatta, R., Gaughan, J. B., Dunshea, F. R. and Lacetera, N. 2018. Review: Adaptation of animals to heat stress. Animal, **12**: s431–s444.

Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M. and Flori, L. 2015. WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. BMC Genomics, 16(1).

Sichien, D., Scott, C. L., Martens, L., Vanderkerken, M., Van Gassen, S., Plantinga, M., Joeris, T., De Prijck, S., Vanhoutte, L., Vanheerswynghels, M., Van Isterdael, G., Toussaint, W., Madeira, F. B., Vergote, K., Agace, W. W., Clausen, B. E., Hammad, H., Dalod, M., Saeys, Y. and Guilliams, M. 2016. IRF8 Transcription Factor Controls Survival and Function of Terminally Differentiated Conventional and Plasmacytoid Dendritic Cells, Respectively. Immunity, **45(3)**: 626–640.

Simianer, H. and Qanbari, S. 2010. A two-step method for detecting selection signatures using genetic markers. Genetics Research. **92(2)**:141–155.

Simoni Gouveia, J.J., Paiva, S.R., McManus, C.M., Caetano, A.R., Kijas, J.W., Facó, O., Azevedo, H.C., de Araujo, A.M., de Souza, C.J.H., Yamagishi, M.E.B. and Carneiro P.L.S. 2017. Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. Livestock Science, **197**: pp.36-45.

Singh, B. D. and Singh, A. K. 2015. Marker-Assisted Plant Breeding: Principles and Practices. Springer India.

Skotte L., Korneliussen T.S. and Albrechtsen A., 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics. **195(3)**: pp.693-702.

Sodhi, M., Mukesh, M., Kishore, A., Mishra, B. P., Kataria, R. S. and Joshi, B. K. 2013. Novel polymorphisms in UTR and coding region of inducible heat shock protein 70.1 gene in tropically adapted Indian zebu cattle (Bos indicus) and riverine buffalo (Bubalus bubalis). Gene, **527(2)**: 606-615.

Srivastava, A. K. 2019. Conservation of Indigenous Cattle Breeds. Journal of Animal Research. **9**(1): 1-12

Stephan, W. 2015. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Molecular Ecology. **25(1)**: 79–88.

Sterken, R., Kiekens, R., Coppens, E., Vercauteren, I., Zabeau, M., Inzé, D., Flowers, J. and Vuylsteke, M. 2009. A population genomics study of the Arabidopsis core cell cycle genes shows the signature of natural selection. The Plant Cell. 21(10), pp.2987-2998.

Suekawa, Y., Aihara, H., Araki, M., Hosokawa, D., Mannen, H. and Sasazaki, S. 2010. Development of breed identification markers based on a bovine 50K SNP array, Meat Sci., **85**: 285–288.

Szmatoła, T., Gurgul, A., Ropka-Molik, K., Jasielczuk, I., Ząbek, T. and Bugno-Poniewierska, M. 2016. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. Livestock Science. **188**: pp.72-80.

Szpiech, Z. A., Novak, T. E., Bailey, N. P. and Stevison, L. S. 2020. Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. Cold Spring Harbor Laboratory.

Melo, T.P., Salinas Fortes, M.R., Hayes, B., Albuquerque, L.G., Carvalheiro, R. 2019. Across breed validation study confirms and identifies new loci associated with sexual precocity in Brahman and Nellore cattle. Journal of Animal Breed Genetics. **137**:139–154.

Tajima, F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. **123(3)**: 585–595.

Taye, M., Yoon, J., Dessie, T., Cho, S., Oh, S. J., Lee, H.-K. and Kim, H. 2017. Deciphering signature of selection affecting beef quality traits in Angus cattle. Genes & Genomics, **40(1)**: 63–75.

Tosser-Klopp G., Bardou P., Bouchez O., Cabau C., Crooijmans R. and Dong Y. 2014. Design and characterization of a 52K SNP chip for goats. PloS one. **9(1)**: e86227.

Visscher, P.M. and Goddard, M.E. 1995. Genetic parameters for milk yield, survival, workability, and type traits for Australian dairy cattle. Journal of Dairy Science. **78**: 205–220.

Vitti, J. J., Grossman, S. R. and Sabeti, P. C. 2013. Detecting Natural Selection in Genomic Data. Annual Review of Genetics, **47(1)**: 97–120.

Voight, B. F., Kudaravalli, S., Wen, X. and Pritchard, J. K. 2006. A Map of Recent Positive Selection in the Human Genome. PLoS Biology, **4(3)**: e72.

Wang, H.-D., Liu, L., Zhao, H.-R., Hou, Q.-F., Yan, J.-B., Shi, W.-L., Guo, Q.-N., Wang, L., Liao, S.-X. and Zhu, B.-F. 2017. Detection of fetal epigenetic biomarkers through genome-wide DNA methylation study for non-invasive prenatal diagnosis. Molecular Medicine Reports. **15(6)**: 3989–3998.

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D. and Taylor, J.F. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science. **92**: 3431-3436.

Wright, P. E. and Dyson, H. J. 2014. Intrinsically disordered proteins in cellular signalling and regulation. Nature Reviews Molecular Cell Biology. **16(1)**: 18–29.

Wu, Y., Fan, H., Wang, Y., Zhang, L., Gao, X., Chen, Y., Li, J., Ren, H. and Gao, H. 2014. Genome-Wide Association Studies Using Haplotypes and Individual SNPs in Simmental Cattle. PLoS ONE. **9(10)**: e109330.

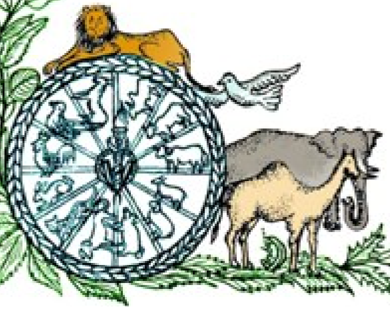
Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M. and Hulse-Kemp, A.M. 2015. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nature biotechnology. **33(5)**: 531-537.

Zhao, F., McParland, S., Kearney, F., Du, L. and Berry, D. P. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genetics Selection Evolution, 47(1).

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C. P., Sonstegard, T. S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J. A. and Salzberg, S. L. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biology, **10(4)**: R42.

* * *

Appendix



APPENDIX

GENERAL BUFFERS AND REAGENTS

1. **Phosphate Buffered Saline (PBS) pH 7.4**

Sodium chloride	10.0g
Potassium chloride	0.25g
Anhydrous disodium hydrogen phosphate	1.42g
Anhydrous potassium dihydrogen phosphate	0.25g
Distilled water	750ml

Dissolve and adjust to pH 7.4 with N/10 NaOH

Make the volume up to 1000 ml with distilled water

2. **Giemsa Stain**

Giemsa powder (Merck)	3.8 g
Methyl alcohol (Acetone free)	250 ml
Warm glycerine	250 ml

Dilute 1:20 in tap water and apply.

3. **Normal Saline Solution**

Sodium Chloride	0.9 g
Distilled Water up to	100 ml

SOLUTIONS FOR AGAROSE GEL ELECTROPHORESIS

1. **0.5M Ethylene Diamine Tetra Acetate (EDTA) pH 8.0**

EDTA	18.61 g
Distilled water	80 ml

Stir vigorously on a magnetic stirrer to mix. Adjust the pH to 8.0 and make the volume to 100 ml. Autoclave and store at 4^{o}C.

2. **Tris-Acetic acid-EDTA (TAE) Buffer (50X)**

Tris base	242 g
0.5M EDTA	100 ml
Acetic acid	57.1 ml
Distilled water up to	1000 ml

Store at 4°C.

3. **1X TAE Working Solution**

50X TAE buffer	20 ml
Distilled water	980 ml

4. **Ethidium bromide (10 mg/ml)**

Ethidium bromide	10 mg
Distilled water	1 ml

Store at 4^{0}C

5. **Gel Loading Dye (6X)**

Tris-HCl (pH 7.6)	10 mM
Bromophenol blue	0.03%
Xylene cyanol FF	0.03%
Glycerol	60%
EDTA	60 mM

REAGENTS AND SOLUTIONS FOR SDS-PAGE

1. **30% Acrylamide stock**

Acrylamide	29.2 g
N'-N bis-methylene acrylamide	0.8 g
Distilled water	50 ml

Dissolve to make a final volume of 100 ml. Filter and store at 4°C in dark bottles

2. **Resolving Gel Buffer (1.5M Tris- pH 8.8)**

Tris base	18.15 g
Distilled water	60 ml

Dissolve Tris and adjust the pH to 8.8 using 6N HCl.
Make up the volume to 100 ml with distilled water.

3. **Stacking Gel Buffer (0.5 M Tris-pH 6.8)**

Tris base	6.0g
Distilled water	60 ml

Dissolve Tris and adjust the pH to 6.8 using 6N HCl.
Make up the volume to 100 ml with distilled water.

4. **10% (w/v) Sodium Dodecyl Sulphate solution**

SDS	1.0 g
Distilled water	10 ml

5. **10% (w/v) Ammonium per sulphate**

APS	100 mg
Distilled water	1 ml

6. **10X SDS-PAGE Running Buffer (pH 8.3)**

Tris base	30.3g
Glycine	144.0 g
SDS	10.0 g
Distilled water up to	1000 ml

7. **2X loading dye**

Distilled water	3.55 ml
0.5 M Tris (pH 6.8)	1.25 ml
Glycerol	2.5 ml
10% (w/v) SDS	2.0 ml

0.5% (w/v) Bromophenol blue	0.2 ml
Total volume	9.5 ml

Store at room temperature

Add 50 µl β-Mercaptoethanol to 950 µl of sample buffer prior to use

8. **Coomassie Brilliant Blue (CBB) Staining solution**

CBB R- 250	250 mg
Methanol	50 ml
Glacial Acetic acid	10 ml
Distilled water up to	100 ml

9. **Destaining solution**

Methanol	50 ml
Glacial Acetic acid	10 ml
Distilled water	40 ml

REAGENTS FOR PURIFICATION OF RECOMBINANT PROTEIN UNDERNATIVE CONDITION

1. **Lysis Buffer (pH 8.0)**

NaH_2PO_4 (50 mM)	0.69 g
NaCl (500 mM)	2.92 g
Imidazole (20 mM)	0.136g
Distilled water	75 ml

Dissolve and adjust the pH to 8.0 using NaOH.

Make up the volume to 100 ml with distilled water.

2. **Wash Buffer (pH 8.0)**

NaH_2PO_4 (50 mM)	0.69 g
NaCl (500 mM)	2.92 g
Imidazole (40 mM)	0.272 g
Distilled water	75 ml

Dissolve and adjust the pH to 8.0 using NaOH.

Make up the volume to 100 ml with distilled water.

3. **Elution Buffer (pH 8.0)**

NaH_2PO_4 (50 mM)	0.345 g
NaCl (500 mM)	1.46 g
Imidazole (500 mM)	1.7 g
Distilled water	40 ml

Dissolve and adjust the pH to 8.0 using NaOH.

Make up the volume to 50 ml with distilled water.

REAGENTS FOR PURIFICATION OF RECOMBINANT PROTEIN WITH UREA LYSIS METHOD

1. **5X Tris-Phosphate Buffer (pH8.0)**

Tris base	608 mg
NaH_2PO_4	7.8 g

Add double distilled water upto 100 ml after adjusting the pH to 8.0.
Store at 4°C

2. **Base Solution**

Urea	16.8 g
5X Tris-Phosphate buffer (pH 8)	7 ml

Make up the voulume to 35 ml with distilled water.
Dissolve the contents.

3. **Lysis Buffer (pH 8.0)**

Base solution	5 ml
Imidazole (10 mM)	50 µl
β-mercaptoethanol (15mM)	5 µl
Triton-X (0.5%)	25 µl
Lysozyme	5 mg

Dissolve and adjust the pH to 8.0.

4. **Wash Buffer (pH 8.0)**

Base solution	15 ml
Imidazole (10 mM)	50 µl

Dissolve and adjust the pH to 8.0.

5. **Elution Buffer (pH 8.0)**

Base solution	5ml
Imidazole (200 mM)	1.36 g

Dissolve and adjust the pH to 8.0.

REAGENTS FOR WESTERN BLOTTING

1. **Transfer Buffer**

Tris base	3.03 g
Glycine	14.4 g
SDS	1.0 g
Methanol (20%)	200 ml

Distilled water to make up to 1000 ml

2. **Tris Buffered Saline (TBS) pH 7.5**

Tris.Cl (10 mM)	0.78 g
Sodium chloride (150 mM)	4.38 g
Distilled water	400 ml

Adjust pH to 7.5 with HCl and make the volume to 500 ml with distilled water.
Sterilize by autoclaving.

3. **TBS-Tween Buffer (TBST) pH 7.5**

Tris.Cl (20 mM)	1.57 g
Sodium chloride (500 mM)	14.61 g
Tween-20 (0.05% v/v)	250 µl
Distilled water	400 ml

Adjust pH to 7.5 with HCl and make the volume to 500 ml with distilled water.

4. **Blocking Buffer**

BSA (3% w/v)	1.5 g
TBS buffer (pH 7.5)	50 ml

5. **Substrate Buffer for HRPO conjugate (for development)**

PBS (pH 7.4)	10 ml
DAB tab	10 mg
30% H2O2	10 µl

REAGENTS AND SOLUTIONS FOR ELISA

1. **Coating Buffer (pH 9.6)**

Sodium carbonate ($\mathrm{Na2CO_3}$)	192 mg
Sodium bicarbonate ($\mathrm{NaHCO_3}$)	380 mg
Distilled water	90 ml

Dissolve and adjust pH to 9.6 with N/10 NaOH and make the volume up to 100 ml.Autoclave and store at 4^0C.

2. **10X Phosphate Buffered Saline (PBS), pH 7.4**

Sodium chloride	100 g
Potassium chloride	2.5 g
Anhydrous disodium hydrogen phosphate	14.2 g
Anhydrous potassium dihydrogen phosphate	2.5 g
Distilled water	750 ml

Dissolve and adjust to pH 7.4 with N/10 NaOH.Make the volume up to 1000 ml with distilled water.

3. **1X PBS Working Solution (pH 7.4)**

10X PBS stock	100 ml
Distilled water up to	1000 ml

Check the pH to 7.4. If not, adjust the pH to 7.4 either with N/10 NaOH or N/10HCl.

4. **Wash Buffer (PBS-T), pH 7.4**

1X PBS (pH 7.4)	100 ml
0.05% (v/v) Tween-20	50 µl

5. **Blocking buffer (pH 7.4)**

5% skim milk powder (w/v)	5.0 g
1X PBS (pH 7.4)	100 ml

6. **Substrate Buffer: Phosphate-Citrate Buffer (0.05M, pH 5.0)**

Solution A

Citric acid	2.1 g
Distilled water	100 ml

Solution B

Disodium hydrogen phosphate	3.56 g
Distilled water	100 ml

Mix 6.425 ml of solution A and 6.075 ml of solution B before use and make the volume up to 25 ml with distilled water. Add 10 mg OPD and 10 µl of H_2O_2 (30%, v/v) just before use.

Use within I h after preparation and keep in dark.

7. **Stopping Solution (1M H2SO4)**

Conc. H_2SO_4	5.6 ml
Distilled water up to	100 ml

Keep in amber coloured bottle and store at room temperature.

REAGENTS FOR IMMUNOCHROMATOGRAPHIC TEST

1. **Conjugation solutions**

A.	50mM KH_2PO_4	0.1360
	DEPC treated water up to	20 ml

Adjust pH of 15 ml to 7.5 and 5 ml in separate tube and adjust pH of 5ml to 9.0 using K_2CO_3

B. 5mM KH_2PO_4 (Dilute 50mM KH2PO4 10 times)Adjust pH to 7.5

2. **Blocking Solutions**

A. 1% PEG (20,000)- 0.01gm PEG in 50mM KH_2PO_4.Adjust pH to 9.0

B. 10% BSA- 2 gm BSA in 50 mM KH_2PO_4. Adjust pH to 7.5

3. **PA solution**

BSA	0.2gm
PEG	0.01gm
NaN3	0.02gm
Nacl	0.1755gm

Dissolve when need to use. Adjust pH to 8.2. All pH should be adjusted using K_2CO_3

4. **Sample buffer**

10 times diluted 1X PBS pH 8.5