ADDIS ABABA UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

Detection of Selection Signatures, Breed-Specific SNPs and Linkage Disequilibrium Analysis in Ethiopian Indigenous and European Dairy Cattle Breeds

MSc Thesis

By

Genet Dejene Dibaba

November, 2021

Addis Ababa, Ethiopia

Detection of Selection Signatures, Breed-Specific SNPs and Linkage DisequilibriumAnalysis in Ethiopian Indigenous Breeds and European Dairy Cattle Breeds

A Thesis Submitted to Institute of Biotechnology, Addis Ababa University in Partial Fulfillment of the Requirements for the Degree Master of Science in Biotechnology

By

Genet Dejene Dibaba

Addis Ababa, Ethiopia

Thesis Approval Sheet

We certify that Genet Dejene Dibaba MSc thesis research entitled "Detection of Selection Signature, Breed-Specific SNPs and Linkage disequilibrium Analysis in Ethiopian Indigenous Breeds and European Dairy Cattle Breeds" has been conducted under our direct supervision. Therefore, we kindly request the Institute of Biotechnology, Addis Ababa University to officially approve the thesis for open defense.

Supervisors:	Signature:	Date:
Tesfaye Sisay (PhD, Prof.)	_____	_____
Hailu Dadi (PhD)	_____	_____
Zewdu Edea (PhD)	_____	_____

Tesfaye Sisay (PhD, Prof.) _____ _____

Director, Institute of Biotechnology,
Addis Ababa University

Statement of Author

I, Genet Dejene, hereby declare that this thesis and its entirety is my own original work and no part of this has been previously presented or submitted for examination anywhere else. All assistance towards the synthesis of this thesis and entire references of others contained herein have been duly acknowledged.

Genet Dejene Dibaba

November, 2021

Acknowledgments

I express my gratitude and appreciation to Professor TESFAYE SISAY, my major advisor for his continuous guidance, invaluable suggestions and kindness. My sincere gratitude also goes to DR. HAILU DADI, my co-advisor, and encouragement starting from title selection, final proposal formulation and proposal implementation throughout this study. Further, I need to express my heartfelt gratitude to DR. ZEWDU EDEA co- advisor for his open interactions and his humble, his valuable advice, support, and assistance any time when I need him, fruitful discussions and continuous encouragement. In addition, I would like to extend my special thank for my friends who have directly and indirectly contributed to the successful accomplishment of the research work.

It gives me a great honor to thank ETHIOPIAN BIOTECHNOLOGY INSTITUTE (EBTi) for sponsoring my graduate study. My utmost appreciation also goes to CHUNGBUK NATIONAL UNIVERSITY which provided me the SNPs data used for this study.

Finally, I would like to express my gratitude to ADDIS ABABA UNIVERSITY, particularly the Institute of Biotechnology for accepting me to join the Master's Program, for furnishing a safe environment to get better knowledge in the field of Biotechnology and their financial support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CLR	Composite Likelihood Ratio Test
CNVs	Copy Number Variations
FAO	Food and Agricultural Organization
Fst	Fixation Index
GO	Gene Ontology
GWAS	Genome Wide Association Study
HP	Pooled Heterozygosity
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
NGS	Next Generation Sequencing

QTL	Quantitative Trait Loci
rEHH	relative Extended Haplotype Homozygosity
RFLP	Restricted Fragment Length Polymorphism
ROH	Runs of Homozygosity
SFS	Site Frequency Spectrum
SNPs	Single Nucleotide polymorphisms

Abstract

Ethiopia ranks first in Africa with the largest number of cattle populations adapted to diverse environments. Detection of selection signatures and assessment of linkage disequilibrium enable us to asses genetic diversity and genomic region under positive selections. However, in Ethiopian indigenous cattle there is few studies regarding their origin, divergency and genetic adaptability. This study investigated detection of selection signatures, breed-specific SNPs and linkage disequilibrium in Ethiopian cattle populations and European dairy breeds. A total 135 animals representing four Ethiopian indigenous cattle populations: Arsi (n = 29), Begait (n = 40), Boran (n = 40) and Sheko (n = 26) were genotyped with 80K SNP chip. Two European dairy breeds (Holstein, n = 60 and Jersey, n = 38) were used for comparison. The mean of minor allele frequency (MAF) 0.32 ± 0.12, 0.32 ± 0.12, 0.31 ± 0.13, 0.30 ± 0.13,0.19 ± 0.17 and 0.18 ± 0.17 for Arsi, Begait, Boran, Sheko, Holstein and Jersey, respectively. The common variant MAF (≥0.10 and 0.5) distribution across Ethiopian cattle populations and European cattle breeds were 89% and 57%, respectively. Ethiopian cattle specific SNPs were located in genes(*CSN2, ABCA7, PDE4B, ABCA1, JAK2, B4GALT4, FOXO3, GHR, ADCY8, ACACB*) associated with milk production traits, fertility and growth traits (*PGR, GHR, XKR4, ADCY5, POU2F1, IGF1, ABCC2, XKR4*) thermo-tolerant, coat color (*HSF1, HSPH1, HSPA4, KDR, RAD50, WNT1, KIT*), feed intake (*XKR4, ACCN1, ACAD11*) and fat thickness (*XKR4, IGFBP-3 and POU2F1*).The top 1% fixation index (*Fst*) values representing positive selection harbored candidate genes (*ABCG2, ABCA7, B4GALNT1, GHR, ITGAV*) involved in milk traits such as, milk protein, milk fat and mastitis, milk production and (*HSPH1, HSPA4, SOD1, MATR3, RAD50, KDR*) for tropical adaptation. The estimated observed heterozygosity (Ho) and expected heterozygosity (He) for Ethiopian cattle populations were found to be, 0.40 and 0.39 respectively. These values were 0.24 and 0.25 for European dairy cattle breeds respectively. Principal component analysis (PCA)clearly separated Sheko from

Ethiopian zebu populations with the value of 4.34%(PCA1) and 3.33% (PCA2). Similarly, PCA1 and PCA2 accounted for 65.97 % and 9.50 % of the variation and differentiated bos indicus from European taurine. Additionally, result of the phylogenetic tree analysis supporting PCA revealed that with the exception of Sheko, Ethiopian cattle population were closely clustered and two European dairy breeds (Holstein and Jersey) were located in a clade with Sheko. The overall mean of r^2 values were 0.22, 0.23, 0.23, 0.22, 0.24, 0.16, 0.16 in Arsi, Begait, Boran, Sheko, Holstein and Jersey, respectively. In broad, this study revealed that there are genomic regions under strong divergent selection harboring genes involved in milk production traits and in adaptation to tropical environment.

Keywords*: Adaptation, Linkage disequilibrium, Selection signature, SNPs*

CHAPTER ONE

1. INTRODUCTION

Ethiopia ranks first in Africa with the largest herd of cattle, reaching 70 million heads (CSA, 2021). Particularly in the horn of Africa, it has highest diversified cattle populations (Rege,1999). The hybrid genetic composition and evolution in different climatic andenvironmental conditions has shaped the genetic composition in possessing different adaptiveand productive characters. In Ethiopia cattle production plays a great role to sustain the livelihoods of most of the farming and pastoral communities. Indigenous cattle breeds are known for their better adaptation for low input systems, heat tolerance, disease resistance, local preferences, and cultural and aesthetic values. Their unique genetic background and adaptation to the low input system support their conservation and breed development. But, in most Ethiopian indigenous cattle populations the genetic diversity and genetic merits are not yet understood and exploited well (Shapiro *et al*., 2017). Few studies have been carried out using microsatellite, mitochondrial, or Y- chromosome markers on genetic diversity of Ethiopian cattle breeds (Dadi et al., 2008). Molecular markers for molecular characterization of cattle populations have been widely developed. Among them, single nucleotide polymorphisms (SNPs) are the most abundantmolecular markers in the genome with a variable distribution among species. The usefulness of SNP in the study of the diversity and structure of the population has been confirmed in several studies on livestock species (Zewdu *et al*., 2012; McKay *et al*., 2008; Williams *et al*., 2009).

Currently, 28 indigenous breeds of cattle have been recognized to exist in Ethiopia (Ethiopian Biodiversity Institute (EBI), 2016). As (CSA, 2015) reported 98.2% are indigenous tothe country in diverse ecosystems highland, dry mountain, lowlands, arid and forest and they are often said to

possess unique genetic traits. Among recognized indigenous breeds Arsi, Begait, Boran, Fogera, Horro and Ogaden are the main breeds in the country (Felleke,2010). Most cattle breeds have dual purpose for dairy and beef.

Currently genetic selection is the approach that opens the possibility to select animal with economically governing traits and help for the genetic improvement of livestock. To detect variation or polymorphism among individuals in the population for specific regions of DNA different molecular techniques has been used (Montaldo et al., 1998). Identification of candidate genes associated with economically important traits in selection signatures has become the best approach in Genomic wide association studies (Stranger et al., 2011). The variation between cattle populations is an important guide for promoting genetic resources of cattle populations andenables genetic improvement to meet the needs of production strategies and plans (Hu *et al.*, 2018).

Therefore, the current study estimated linkage disequilibrium, identified selections of signature and detected breed specific SNPs associated with milk production and ecological traits in selected Ethiopian cattle populations and European dairy breeds using Bovine GGP 80K Bead Chip genome-wide SNPs.

1.1. Objectives

1.1.1. General Objectives

- To evaluate genetic diversity and genomic region under preferential selection, in Ethiopian indigenous cattle populations and European dairy breeds using high density genetic markers.

1.1.1.1. Specific Objectives

- To detect breed-specific SNPs in Ethiopian indigenous cattle populations and European dairy breeds.
- To detect selection signatures in Ethiopian indigenous cattle populations and European dairy breeds.
- To estimate linkage disequilibrium in Ethiopian indigenous cattle populations and European dairy breeds.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Origin and Domestication of Cattle

Cattle are one of the most popular and widespread livestock species in the world. Due to genetic drift and natural selection, ancestral populations of domestic cattle evolved into distinct local populations under different environmental conditions. Around 8000-10000 BC, after the domestication of sheep and goats, cattle Taurine (*Bos taurus*) and Indicine (bos indicus) were domesticated respectively in the Middle East and the Indian subcontinent (Bollongino *et al*., 2012). Molecular-based studies also approve that world cattle populations, bos indicus and bos taurus were domesticated independently from aurochs subspecies (Bradley *et al*., 1996).

Bos Taurus - The domestication of bos taurus (humpless cattle) occurred ~ 10 000 years ago in the Near East (Gotherstrom *et al*., 2005). The first substantial evidence of cattle domestication anywhere in the world is the Neolithic culture prior to the pottery of the Taurus Mountains. Strong evidence for any plant or animal domestication site is genetic diversity. The place where plants or animals develop generally has a high degree of diversity among these species, the place where domesticated animals are introduced has less diversity. The greatest genetic diversity of livestock is found in the Taurus Mountains. Recent molecular characteristics from mitochondrial DNA (mtDNA) studies indicate that the Taurus was introduced to Europe and Africa and interbred with local wild animals (Beja-Pereira *et al*.,2006).

Bos Indicus - The mtDNA evidence from current research on domesticated zebu (humped cattle, bos indicus) shows that the two main lineages of bos indicus now exist in modern animals. The former species dominates Southeast Asia and South China, and is likely currently domesticated

in the Indus Valley region of Pakistan. Evidence of the transition from bos wild to bos *indicus* is in evidence in Harappan sites such as Mehrgahr about 7,000 years ago (Beja-Pereira *etal*.,2006).

2.1.1. Origin of African cattle

African cattle breeds are divided into two categories: - Bos taurus and bos indicus. Previous assessments maintained that African cattle were derived from three major introductions from Asia through the Nile Valley or via the Horn of Africa (Epstein, 1957). It was supposed that the Bos taurus, which occurred around 6000 BC was the first cattle breed introduced into Africa and Bos indicus the next breeds between 2750 and 2500 BC (Epstein, 1971). On the other hand, archaeological and molecular evidence supports the independent domestication of cattle (bos taurus) in Africa (Grigson, 1991; Bradley et al., 1996; Hanotte *et al*., 2002). However, until recently, the evidence for the domestication of bos taurus in Africa has been controversial (Decker *et al*., 2014). A recent genome-wide analysis revealed that African taurine was first domesticated in the Middle East and later crossed with Indian cattle.

The humped zebu (bos indicus) appears to have originated in Asia and spread to Africa mainly through the Horn of Africa and the east coast (Hanotte *et al*., 2002). The first zebu bulls introduced to Africa were crossed with taurine and produced sanga cattle, although the second wave and entrance resulted in the formation of zenga (sanga×zebu) (Epstein, 1971; Grigson, 1991; Rege, 1999). bos indicus is divided into zebu and zebu cross, which are phenotypically recognizable by the presence of significant breast-chest bulge (Rege, 1999). The position of the hunchback on the animal's back is used to classify the zebu proper and zebu crossbred types into cervico thoracic-humped and thoracic-humped stocks (Epstein, 1971). In hybrids of Zebu cattle with humps and breast humps, the humps are usually cervix and this type of cattle is called

sanga. Consequently, African cattle can be divided into four different groups, specifically Taurus, Indicus, Sanga and Sanga's type of cattle (Rege, 1999).

They were adapted to different agroecology and it is well known that bos indicus adapt to the subtropical regions of Africa and are more tolerant to various diseases (Muchenje *et al.*, 2008; Marufu et al., 2011). Promote effective management of indigenous genetic resources, because indigenous genetic resources may be more valuable in selection and breeding programs during periods of biological stress such as famine, drought or disease epidemics (FAO, 2010). In order to effectively manage these cattle breeds, a comprehensive understanding of their characteristics is required. These include the size and structure of the population, and the knowledge of differences within and between breeds (Boettcher *et al*., 2010; Groeneveld *et al*., 2010). Limited research has focused on the genetic characteristics of breeds. Emphasize the need to use these varieties as genetic resources for genetic identification.

Ethiopia is one of the countries which have diversified cattle populations. This is due to its diverse agro - ecology, topography, and proximity to Asia, which is the origin of most domestic animals in Africa (EBI, 2016). A breed is a homogenous group of livestock which are phenotypically unique from other groups or subpopulations of the same species (Halimanietal, 2010). There are around 28 indigenous breeds of cattle that have been recognized to exist in Ethiopia (EBI, 2016). These are grouped in to five major cattle types, large and small East African Zebu, Sanga, Zenga and Taurine types that are distributed to be found in all agro- ecological zones depending on their merit to a particular production system from arid tropical to afroalpine ecosystems (EBI, 2016).

There is no complete and up-to-date breed level population data for most of the breeds, and this makes determination of the status and trends difficult. Thus, characterization and identification of indigenous cattle breeds, including their unique traits, should be given high priority. That should be a precondition in designing conservation and sustainable utilization programs to help indigenous breeds to compete in changing production environments with limited production resources such as land, feed, labour and capital in the future (Reist-Marti *et al.,* 2004; Groeneveld *et al.*, 2010). Few studies were conducted in characterization of Ethiopian Indigenous cattle breeds. Recent molecular studies have characterized Ethiopian cattle breeds, as (Dadi et al.2008) Ethiopian cattle population is hybrid population depending microsatellite data analysis. Tarekegn *et al. (*2018) also reported that all Ethiopian cattle mtDNA sequences converged on one maternal lineage, which matches to African Bos taurus cattle; no zebu mtDNA haplotypes have been discovered in the Ethiopia population. Edea *et al. (*2014) conduct Genome-wide comparison of populations from different geographical regions and detected potential candidate genes associated with ecological adaptations and there is low genetic differentiation among Ethiopian cattle populations that own common historical origin and high gene flow.

2.2. Selection signatures

Detection of selection signatures is one of the most important approaches in modern genetics. It enables us in determining population divergence and differentiation. Also, it has become critical in biomedical sciences, in identifying genes related to disease resistance (Cagliani *et al.*, 2011), adaptation to climate (Rees and Harding, 2012) and altitude (Simonson *et al.*, 2010). Domestication selection has been carried out by humans in livestock species, to improve economic important traits, such as milk (Hayes *et al.*,2009) and meat (Kijas *et al.*, 2012).

production. Selection can be natural or artificial and it is effectiveness is depended on the time scale and methods used to detected selections.

Natural selection is a phenomenon driven by natural time, in which individuals with specific genotypes have different abilities to contribute to the gene pool of the next generation. (Templeton, 2006; Driscoll *et al.*, 2009). Fundamentally natural selection act in three ways, positive selection, negative selection and balancing selection. These methods of selection are a response to environmental pressure and acts differentially to change the allelic and genotypic frequencies (Harris and Meyer, 2006; Oleksyk *et al.*, 2010).

The difference between natural selection and artificial selection called selective breeding is a human-mediated process in which the gene pool of the next generation does not depend exclusively or necessarily on fitness components, but also on traits chosen by humans. Artificial selection can be classified as unconscious selection or methodical selection. The former occurs when there is no long-term objective, and this has been suggested as the cause of the early domestication process. The second occurs when a standard or objective drives the choice of parents for the next generation. Despite these differences and considering that the time frame in which these changes occur is often considerably different, the genetic consequences of natural and artificial selection are essentially the same (Avise and Ayala, 2009; Driscoll *et al.*, 2009; Gregory, 2009). Similarly, in livestock artificial selection has affected the development of specific traits of economic importance such as milk, meat and fertility, through genetic improvement (Randhawa *et al.*,2016).

Positive Selection – it is considered as positive selection when increment of phenotype increases the fitness of an organism and it becomes more prevalent in a population over time. Finally, the allele responsible for important phenotype would become more common at the population level

(Emlen *et al.,* 2005). When a new mutation has a selective advantage over other mutations and therefore increases its frequency in the population, positive selection occurs. (Kaplan *et al.*, 1989).

Negative selection - is the selective removal of deleterious alleles ,i.e., the detrimental variants that appear in the population tend to be removed (Charlesworth *et al.*, 1993). If the recombination rate in the region is limited or the population is highly inbred then negative selection reduces the variability around the deletion site (Charlesworth *et al.,* 1993, 1995;Andolfatto, 2001; Stephan, 2010). Too many low-frequency alleles are also observed in small and medium-sized populations (Charlesworth *et al.*, 1993, 1995), and the number of non- synonymous substitutions per non-synonymous locus is often less than the number of synonymous substitutions per site (Nei, 2005; Harris and Meyer, 2006). However, in regions with normal recombination rates, or when inbreeding is limited, no reduction in variability has been observed (Charlesworth et al., 1993, 1995; Stephan, 2010). Furthermore, the choice does not cause significant deviations in the spectrum (Charlesworth et al., 1993, 1995; Kim and Stephan, 2000; Andolfatto, 2001; Stephan, 2010).

Balancing selection - is when multiple alleles are retained in a population at an intermediate frequency, because of natural selection and when we expect that one allele will provide higher fitness than the other. In balancing selection, the fitness is depended on the allele frequencies of the alleles involved in phenotypic traits (Vitti *et al.,* 2013). If one allele starts to become more common, natural selection will favor individuals with the other allele. Other causes of balancing selection are frequency-dependent selection and selection in a fluctuating environment. Balancing selection maintains polymorphism in the population with respect to an allele of a trait

of interest and it favors the maintenance of polymorphism (Harris and Meyer, 2006; Oleksyk *et al.*, 2010).

Keeping the same alleles for a long time is called long-term balanced selection. In addition to maintaining polymorphisms at selected loci, it also tends to increase diversity in tightly linked neutral sites; if the region under selection has low recombination rates, then it generally also has longer coalescence times than other regions (Charlesworth, 2006).

2.2.1. Methods in detection of selection signatures

In detection of selection signature there are several methods based on the comparison of gene substitution patterns and divergence at the species level (Goldman and Yang, 1994; McDonald and Kreitman, 1991; Nielsen, 2005). Basically, approaches detecting selection signature are used to classify selective events that occurred in the distant past and that represent macroevolutionary patterns that arise as a result of divergent selection between the species (Kryazhimskiy and Plotkin, 2008; MacEachern *et al.,* 2009). Detection methods have been also used for detecting footprints of adaptive molecular evolution between the species based on nucleotide divergence have been used in selection signature (McDonald and Kreitman, 1991). Through time different methods have been developed as an advanced multilocus for comparing within-species nucleotide diversity at different DNA sites (Egea et al., 2008).

Later, selection signature studies are dependent on approaches to identify micro evolutionary selective events at the population level or within species. Traditional population genetics focuses on comparing a set of specific markers in a region with a neutral hypothesis through experiments or statistical models. Advances in SNP genotyping and high-throughput sequencing technologies have led to genome-wide scans of selective characteristics among populations within a species.

After the development of Genome-wide SNP array for cattle in 2008 (Matukumalli et al., 2009), there is a clear transition from microsatellite markers to dense SNP data. Similarly, various statistical methods have been developed for the detection of selection signatures using the DNA sequence or SNP genotype data in the livestock populations (Oleksyk *et al.,* 2010; Qanbari and Simianer, 2014).

The approach to detect selection signature can be classified into two groups: intra-population statistics and inter-populations statistics. Intra-population statistics detect for footprints of selection by comparing genomic data within populations. This type comprises three primary approaches based on the site frequency spectrum, linkage disequilibrium, and reduced local variability (Weigand and Leese, 2018). Inter-populations statistics mostly depend on the degree of differentiation due to locus-specific allele frequencies between the populations (Zhao *et al*., 2015). Single site and haplotype-based differentiation are categorized under these methods.

2.2.1. 1. Site frequency spectrum - based methods

Site frequency spectrum (SFS) is a method of assessments that rely on the distribution of allele frequencies in a population (Achaz, 2009; Ronen *et al.,* 2013). Selective sweeps produce an increased number of high- and low frequency variants and a decreased number of medium-frequency variants (Achaz, 2009). Different allele frequency-based neutrality tests include Tajima's D (Tajima, 1989), Fay and Wu's H statistic (Fay and Wu, 2000), and composite likelihood ratio test (Lindsay, 1988). Classical Tajima's D (Tajima, 1989) compares the difference between the average number of differences in nucleotides and the number of segregating sites estimated from polymorphism data (Carlson *et al.,* 2005). The D value will be zero under neutrality. In the case of positive selection, mutations will be rare that reduces heterozygosity and gives a negative value of D ($D < 0$). In contrast, balancing selection leads to

intermediate allele frequencies, and the D values will be positive (Tajima, 1989; Simonsen *et al.,* 1995). Fay and Wu's H statistic (Fay and Wu, 2000) is based on the frequency spectrum of ancestral and derived alleles, assuming that ancestral alleles are known.

The focus of this statistic is to detect the most recent positive selection, mainly for ancestral allel medium and high frequency alleles complementary to Tajima D, which can identify low and medium frequency alleles. (Cadzow *et al*., 2014). Another different approach from these methods is the Composite likelihood ratio test (CLR) statistic which assesses the skewness of thefrequency spectrum of alleles across the multiple loci and also incorporates recombination rate todifferentiate selection from demographic events (Nielsen, 2005; Vy and Kim, 2015).

The composite likelihood ratio test is based on hypothesis testing, which compares the neutral model of the genome window site spectrum with the selective scanning model (Chen *et al.,* 2018). The CLR test is extremely sensitive in detecting positive selection signals at multiple sitesin the population (Williamson *et al.,* 2007). Nonetheless, these SFS based methods were not suited for locating genome wide SNPs. Later the arrival of genomic technologies, haplotype phasing and imputation, more advanced methods based on linkage disequilibrium (LD) and haplotype homozygosity were developed.

2.2.1.2. Linkage disequilibrium-based methods

Linkage disequilibrium-based methods are the method particularly used to quantify variants under partial selective sweep. Sabetti *et al.* (2002) proposed an extended haplotype homozygosity (EHH) method based on linkage disequilibrium to detect selection signatures within a population. After determining the haplotype of the nucleus, the age of each nucleus is assessed by attenuation of LD as a function of distance (Sabeti *et al.,* 2002). EHH is theprobability that a pair of chromosomes carry homozygous core haplotypes. Then, relative

extended haplotype homozygosity (rEHH) is calculated to compare the EHH values of two core haplotypes. Core haplotypes with high rEHH values and high frequency in the population are said to be under positive selection (Sabeti *et al.,* 2002).

The method of relative extended haplotype homozygosity is used to identify regions that have recently been positively selected and do not require description of ancestral alleles. Voight *et al. (*2006) developed an integrated haplotype score by including the recombination distance into the statistics. After the development of the SNP chip, this is the most widely used method for detecting recently selected signatures. The integrated haplotype score is a measure of how abnormal haplotype is around the SNP compared to the entire genome (Voight *et al.,* 2006). Each SNP is treated as a core SNP and EHH values are considered for each core SNP according to their ancestral and derived allelic status using an out group.

This test shows points where ancestral and derived EHH values are below a certain threshold level. Next, integrate the EHH value into the map distance from the central allele to these SNPs. Compared to haplotypes related to ancestral alleles, their extremely negative values (iHS <2) indicate more widespread haplotypes in the background of derived alleles, while extreme positive values (iHS> 2) indicate that the ancestors are locus has swept the entire population (Weigand and Leese, 2018). Unlike the SFS method, it requires the haplotype phase, recombination map, genome location, and ancestry, and derived allele information for each core SNP. When the selected allele is at an intermediate frequency, this method is well suited for signature detection. Compared to relative extended haplotype homozygosity method, the iHS method is minimally affected by demographic factors and thus the probability of occurrence of false-positive results will be less (Voight *et al.,* 2006).

2.2.1.3. Approaches based on reduced local variability

The methods based on Reduced Local Variability basically focus on identifying genomic regions with reduced variation relative to the genome average. Runs of homozygosity (ROH) andpooled heterozygosity (HP) (McQuillan *et al.,* 2008, Rubin *et al.,* 2010) included under these approaches. Runs of homozygosity are contiguous lengths of homozygous genotypes that occur within an individual when two haplotypes share a recent common ancestor (Gibson *et al.,* 2006; Rebelato *et al.,* 2018). ROH is widely used to assess genomic inbreeding level, population structure, and demographic history of livestock populations (Curik *et al.,* 2014). Depending on the hitchhiking theory, selective scanning should have a homozygous loci spread, showing a higher degree of homozygosity than the average genome level (Almeida *et al.,* 2019). Therefore, ROH can be used to identify selection features, because individuals who have undergone the selection process will show homozygosity for long sequences around the target locus (Forutan *et al.,* 2018;Rebelato *et al.,* 2018; Xie *et al.,* 2019 year). On the other hand, heterozygosity pooling uses allele counts to calculate heterozygosity (Rubin *et al.,* 2010). This statistic estimates the deviation of expected local heterozygosity depression in chromosomal windows from the average heterozygosity of the genome.

2.2.1. 4. Haplotype based differentiation approaches

The haplotype-based differentiation methods comprise cross-population extended haplotype homozygosity (Sabeti *et al.,* 2007) and a haplotype-based extension of the FLK statistic (Fariello *et al.,* 2013). These methods use haplotype information in multiple populations and SNP ascertainment bias will be less. Cross population extended haplotype homozygosity is a haplotype-based differentiation method, introduced by Sabeti *et al*. (2007). To calculate Cross population extended haplotype homozygosity between populations A and B, first, iHH values

for each population are calculated separately by integrating the EHH of the entire sample in the population.

The homozygosity scores of the expanded haplotypes between populations are directional, and the positive and negative scores indicate that the selection occurred in populations A and B, respectively. Fariello *et al.* (2013) proposed a haplotype-based FLK method and it's statistic is based on both haplotype information and hierarchical structure of populations that result in a higher power for the detection of genomic regions under selection.

Unlike FST methods, it accounts for varying effective population sizes. The estimation of the population's kinship (F) matrix is similar to the FLK method, but instead of computing from allele frequencies, the statistics are determined from haplotype frequencies (Fariello *et al*., 2013).This method can be applied to an un phased SNP genotype data. The hierarchical population structure is estimated by using pairwise Reynolds' genetic distances (Reynolds *et al.,* 1983) between the populations. The pairwise Reynolds' distances between populations are calculated for each SNP and then averaged over the genome and converted into the Kinship matrix (Brito *etal.,* 2017).

2.2.1.5. Single site population differentiation methods

Single site population differentiation-based methods include FST (Wright, 1949) and FLK (Bonhomme *et al*., 2010). Fixation index (FST) (Wright, 1949) is based on the measure of differences in frequencies of alleles between populations (i.e., loci that are differentially fixed in different subpopulations). FST value ranges from 0 (no differentiation) to 1 (the fixed difference between populations). Highly differentiated allele frequency between the populations at any given locus (i.e., higher FST values) indicates positive selection, whereas low FST valuessuggest negative selection (Zhao *et al*., 2015). Wright's FST assumed infinite population sizes

and therefore, it may lead to the overestimation of FST for small sample sizes. But, a greater number of SNPs (> 500) increase the efficiency of detecting genetic differentiation even in case of small sample size (Willing *et al*., 2012).

Apart from Wright's FST, there were several other FST estimators proposed such as Nei's GST (Nei, 1973), Weir & Cockerham's FST (Weir and Cockerham, 1984), Hudson's FST (Hudson et al., 1992), Holsinger's FST (Holsinger, 2004), population specific FST (Weir and Hill, 2002), drift model-based FST (Nicholson et al., 2002) and Bayesian model-based FST (Gianola *et al.*, 2010). Each software uses different estimators, each with its advantages and disadvantages (Bhatia *et al*., 2013). An advantage of FST over other methods based on LD or SFS is that FST isSNP-specific and can detect the actual genetic variants under selection. Rather than analyzing each SNP separately, it is more advisable to scan for several consecutive SNPs with average FSTscore by use of genomic windows. When using hierarchically structured data sets, FST statistics can detect false positive/negative results (Fariello *et al*., 2014).

FLK is an extension of the original LK statistic (Lewontin and Krakauer, 1973) which compares observed and expected variances of FST estimated from data and through variance ratio test under neutrality respectively. FLK uses the phylogenetic estimation of the population kinship matrix, so it takes into account the dynamic changes of the effective population size over time and combines the hierarchical branching of the population (Bonhomme *et al*., 2010). The matrix F is a measure of the expected genetic drift of each population and the expected covariance between them. FLK is a powerful parameter test that can easily process massive genotype data sets to detect selection signatures between different populations. Compared with the FST method, FLK effectively reduces type 1 errors (false positives) in the process of selectingsignatures (Bonhomme *et al.*, 2010).

2.2.2. Selection signature studies in cattle

In livestock production detection of selection signatures is very important and currently it has been studied intensively. Genetic diversity research only analyzes the degree of differentiation between livestock breeds or populations, but the research on selection signatures goes further andfinds the real reason for this diversification. The detection of selective characteristics is important for characterizing livestock genetic resources and identifying genes that cause variation in important economic traits (Cesarani *et al.,* 2018). It is used to identify candidate genes related to ecological characteristics, such as adaptation to a specific environment (Yurchenko *et al.,* 2018). These studies help detect beneficial mutations that have selective advantages in specific populations or breeds.

The detection of selection characteristics is essential for a deeper understanding of the origin of the population and the genetic process that affects the phenotypic differentiation of cattle breeds. It can also better understand the progress of artificial selection and allow further genetic improvement of livestock populations. Understanding how choices play a role in specific populations can help us develop effective breeding programs to improve the important economic traits of these animals (Gurgul *et al.,* 2018).

Screening for traits can be used as a complementary method to genome-wide association studies (GWAS), associating selected candidate genes with phenotypes, which can then be used for genome selection (Chen *et al.,* 2016). The detection of selection characteristics generally takes a "top-down" approach, from genotype to phenotype, through which a statistical analysis of genomic SNP data is performed to identify the selection fingerprint / sign (Bomba *et al.,* 2015).

In GWAS, the entire genome of many individuals is scanned for genetic variation and the association between genotype and phenotype is assessed (Tam *et al*., 2019). However, the study of selected characteristics limits the analysis to only candidate genomic regions selected for specific traits of interest. It is based on the genetic parameters of the population and does not necessarily require phenotypic measurement (Zhao *et al*., 2015). Compared to classic GWAS, selection signature is better suited for studying traits that are very expensive, complicated, and sometimes impossible to study using GWAS methods. For example, tolerance to extreme temperatures or adaptation of animals to specific climates, resistance to diseases, response to various livestock husbandry strategies, advances in selection methods, etc. Sometimes, selection trait studies can replace GWAS to identify important SNPs related to genes selected for specific traits in breeds or populations (Maiorano *et al*., 2018). Igoshin *et al*. (2019) Attempted to use a comprehensive method of GWAS and selective trait scanning in Siberian cattle to identify candidate genes related to maintaining body temperature under acute cold stress.

2.2.3. Signature of positive selection for milk production traits in cattle

In the past, research on the characteristics of cattle selection was limited to specific markers (Wiener *et al*., 2003) or definite chromosomes (Hayes *et al*., 2008; Marques *et al*., 2008; Prasad et al., 2008). Several studies have been carried out on the detection of selection signatures relatedto milk production characteristics. Hayes *et al*. (2008), used the LD-based iHS method to detect the selection characteristics of Norwegian Red Bull BTA6 and observed several QTLs that affectmilk production traits. Similarly, Marques *et al. (*2008), used LD-based extended haplotype homozygosity (EHH) statistics to study the selection selection patterns of BTA14 in Angus and Holstein cattle populations. Prasad *et al*. (2008), used the EHH method to determine 8 areas of Holstein cattle (5 areas in BTA19 and 3 areas in BTA29) and 6 areas under the selection of

Angus cattle (in BTA19). 3 areas and 3 areas in BTA29). found that the candidate genes MRPS30 and FGF10 in BTA20 were associated with milk production, protein percentage, and mastitis resistance (Kadri *et al.*, 2015).

Compared with the Bos taurus breed, the number of studies on Bos indicus cattle is limited (Taye *et al.,* 2017; Maiorano *et al.,* 2018). Maorano *et al*. (2018) using iHS, FST and XPEHH methods showed that there are QTLs affecting milk and meat quality traits in dual-purpose Gir cattle populations. The NSG1 gene (associated with neuronal vesicle transport 1) identified in the FST method was previously reported to be in a selective state and associated with milk production traits, such as milk production, protein production, and fat composition (Lee *et al*., 2016).

In addition, several markers are being identified and widely accepted in a variety of economically important genes. Among the economically important traits, milk traits are the most important in the dairy industry. The association between DNA polymorphisms of multiple genes and milk production traits has been studied, including: prolactin (He *et al*., 2006); leptin (Liefers *et al*., 2005); diacylglycerol acyltransferase (DGAT1) (Grisart *et al.,* 2002; Thaller *et al*., 2003; stearoyl-CoA desaturase (Kgwatala *et al*., 2009; bovine leukocyte antigen (BoLA), DRB9 *et al*., Sharif9,. 2007 Growth hormone receptor gene, Bllott *et al*., 2003 Casein a s1 (CSN1S1); Kuss *et al.,* 2005 ATP binding cassette, subfamily G, member 2 (ABCG2) gene Cohen-Zinder *et al*.,2005 ; Olsen *et al*., 2007; Protease inhibitor gene (Khatib *et al*., 2005); Osteopontin gene (Lonard *et al*., 2005); Proliferation-activated receptor γ, co-activator gene (CoA) 1a (Weikard *et al.*, 2005), growth hormone (GH) gene (Zhou *et al*., 2005); signal transducer and activator of transcription (STAT) 1 (Brym *et al*., 2005; Cobanugu *et al* ., 2006) oxidized low-density lipoprotein receptor 1 (Khatib *et al*., 2006); cytochrome P450, subfamily XI B, peptide 1 (Kaupe *et al*., 2007); fatty acid synthase (Morris *et al*., 2007) ; the recru domain contains protein 15

(Pant *et al*., 2007) bovine Kcasein gene and CSN3 (Robitaill *et al*., 2007) thyroglobulin gene (Anton *et al*., 2008) caspase treatment; milk Globulin genes (Ganai *et al*., 2009); POU class 1 homeobox 1 (Huang *et al*., 2008); STAT5A (Khatib *et al*., 2008); and stearoyl CoA desaturase (Macciotta *et al*., 2008).

2.2.4. Signature of selection for adaptation to tropical environment

Several studies on selection signatures have focused on areas that may be related to adaptation to specific environmental conditions (Yurchenko *et al*., 2018; Igoshin *et al*., 2019; Weldenegodguad *et al*., 2019). Selected candidate genes, such as AQP5 (in BTA5), RAD50 (in BTA7) and RETREG1 (in BTA20), were found to be related to the response of Russian cattle breeds to cold / heat acclimatization (Yurchenko *et al*., 2018). To increase the power of detection of selection signatures (Yurchenko et al.,2018) combined several genome-wide statistics such as HapFLK, Tajima's D, π, and FST using DCMS. Multi-signal Decorrelated compounds compare the inter-pair correlations of various genome-wide univariate statistics, which will improve the signal-to-noise ratio during detection and selection signatures (Lotterhos et al., 2017). Igoshin *et al*., 2019, adopted a similar approach and they found a single candidate selection region on BTA15 harboring *MSANTD4* and*GRIA4* genes associated with cold-stress resistance phenotype. Apart from the nucleotide sequences and SNPs, several recent studies in cattle used Copy number variations (CNVs) for the detection of selection signatures (Yang et al., 2017; Zhang et al., 2020). Yang *et al*. (2017) conducted a genome-wide CNVs study in native Chinese cattlebreeds using a new statistic, *vi* (modified from *di*), and found 12 and 62 CNV regions underselection at the top 1 percent and top 5 percent of the *Vi* values. Similarly, Zhang *et al*. (2020)identified 11 CNV regions under selection involved in the adaptation to high altitudes using theFST approach.

2.3. Linkage Disequilibrium (LD)

Linkage disequilibrium is the "non-random association of alleles" in two or more linked and unlinked genes or loci in a population. Estimation of the LD coefficient between genetic markers can provide useful information to identify alternative evolutionary patterns of genomic variation within or between populations (Lewontin and Kojima, 1960). At present, association mapping based on linkage disequilibrium focuses on exploring the causal or mutational genetic polymorphisms of complex traits in plants, animals and humans (Easton *et al*., 2007; Remington *et al*., 2001; Valdar *et al*., 2006).

The linkage disequilibrium, different measurement methods were used. The most important and widely used statistics are Lewontin's D (Lewontin,1964) and the square of the genetic correlation coefficient r2 (Hill and Robertson, 1968). Measurement of r^2 is considered more robust for the use of biallelic markers and requires fewer samples than D` to obtain sufficient precision (Ardlie *et al*., 2002; Khatkar *et al*., 2008; Lipkin *et al*., 2009; Mustafa *et al*., 2018).

The level of LD around a specific allele is determined by different events in population history, including natural processes of genetic drift and hybridization of previously isolated populations (Reich *et al.*, 2001). Natural and artificial selection will also change the level of LD between specific alleles, helping to increase the frequency of certain alleles in the population gene pool, thus increasing the level of linkage disequilibrium in the genomic region affected by the selection process, thus increasing the correlation with other high frequency alleles, and pushing proximally connected alleles to higher frequencies (Frazer *et al*., 2007; Teo *et al.*, 2009). On the other hand, rapid population expansion will reduce LD. Over time, repeated recombination events tend to destroy large

alleles, thus reducing the overall level of LD (Hill and Robertson, 1968; Terwilliger *et al*., 1998) Years; Reich *et al*., 2001).

In population genetics studies LD has a wide range of applications like in the estimating of recombination rates (Li and Stephens,2003; Conrad *et al*.,2006), estimating effective population size (The Bovine HapMap Consortium 2009; Tenesaetal.,2007; Flury *et al*., 2010; Lu et al. 2012; Zhu *et al*. 2013), and inferring population history and demographic events (Reichetal.,2001; Freemanetal.,2006; McEvoy *et al*.,2011; Kim *et al*. 2009). Genome selection also use the connectivity of one or more SNPs with genes of interest (Meuwissen *et al*., 2001; Hayes *et al*., 2009; Prasad *et al*., 2008) and GWAS (Ardlie *et al*., 2002). Knowledge of the extent of LD in the genome the first step for determining the number of markers that were sufficient to obtain sound results in GWAS (Goldstein,2001; Carlson *et al*.,2004), QTL mapping bylinkage disequilibrium (McKayetal.,2007; Goddard and Hayes 2012; Edea, *et al*., 2015) and Genomic selection applications (Meuwissen *et al*.,2001; Khatkar et *al*.,2008; Sargolzaei *et al*.,2008; Qanbari *et al*.,2010).

2.3.1. Linkage disequilibrium study in cattle

In cattle population previous reports have been shown higher levels of LD and larger haplotype blocks in bovine populations, likely due to smaller effective population size and higher in breeding levels (Hayes, *et al*.,2003). McKay *et al*. (2007) suggested LD extended several Mega base pairs (Mb) using D' while r^2 indicated LD only up to 0.5Mb in different cattle breeds, and Khatkar *et al. (*2008) observed that the significant LD in Australian Holstein cattle extended up to 40kilo-base-pairs(kb) when estimated as r^2 and upto 8.2Mb when estimated as D'.

In general, most reports of LD in cattle have been based on a rather low number of markers, with as many as - 41,000 SNPs (Qanbari *et al*,.2010), and their results suggest that higher densities

will be required in the application of the SNPs for GWAS and Genomic selection (GS) studies. Based on a multi-species scenario and using 2670 markers (McKay *et al*., 2007), it is recommended to use 30-50,000 markers for future research, with a focus on QTL mapping. Degree of LD of the British dairy cattle population, Tenesa *et al*. (2003) genotyped 50 dairy cows and found that the LD among the same line group extended to approximately 10 cM. (Khatkar *et al.*, 2007) used 9195SNP to identify the Holstein cattle haplotype block. Their results showedthat 75,000-100,000 tagSNPs are needed for fine positioning research, of which approximately 250,000SNPs are used in the discovery phase. Finally, (Khatkar *et al*., 2008) established the total number of SNPs at 75,000 for low-power analysis and 300,000 markers for high-power genomic scanning of cattle.

Based on SNPs (IlluminaBovineSNP50BeadChip) derived from Western economically important beef and dairy breeds Edea *et al*. (2013) conduct genome-wide LD. The degree of LD affects the accuracy of genome association studies and genome prediction. Therefore, understanding the performance of different BeadChips in estimating the degree of LD helps to select the appropriate genotyping chip for genome-wide association studies (Lee *et al.*, 2014; Dadi *et al*., 2014) Edea (2015, 2013) surveyed genome-wide LD using three genotyping BeadChips (9,50, and 80 K) Ethiopian cattle populations and the Korean (Hanwoo) cattle breed.

2.4. Breed-specific Single-Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphism (SNP) is the most common type of genetic variation in the population. Each SNP characterizes a difference in a single DNA building block called a nucleotide. Most commonly, these variations are in the DNA between genes. They can be used as biomarkers. When SNPs appear in regulatory regions within or near genes, they may play a more direct role in economically important traits by affecting gene functions (Lewis *et al*., 2011). Numerous studies have shown the utility of single-nucleotide polymorphisms (SNP) markers for breed differentiation and individual assignment (Pariset *et al*., 2010; Kuehn *et al*., 2011; Lewis *et al*., 2011; Wilkinson *et al*., 2011; Hulsegge *et al*., 2013). Individual assignment uses genetic information to assign individuals to populations and determine the source of unknown individuals (Negrini *et al*., 2008). The molecular markers used to assign breed to individuals, as well as the allocation process used to assign a breed to individuals based on their genotypes at unique marker loci, assess the accuracy and efficiency of breed assignment to individuals (Zwane *et al*., 2016).

In cattle, phenotypic differences between breeds are much more defined. They are enhanced by strong artificial selection for different production goals, such as dairy products or beef. These phenotypic differences are driven by potential changes in genome structure, emphasizing the importance of species-specific genome inference. For this reason, species-specific reference genomes are a prerequisite (Czech, *et al*., 2018). Breed-specific reference genomes can improve the accuracy of SNP-based inferences, such as genome-wide association studies or SNP genotype imputation. In the future, when specific reference genomes for many varieties become available, they may become the basis for describing hybrid fluctuations and the dynamics of

changes in the Bos taurus genome caused by breeding programs (The Bovine HapMap Consortium, 2009).

2.4.1. Molecular Markers

For many years, biologists have tried to understand and explain the genetic code. Initially, there were few tools available to map features at genomic locations, although researchers had ideas on how to achieve this type of mapping (Haldane et al.,1948). Over the years, multiple types of markers have been developed that allow researchers to estimate the location of sequences that encode various phenotypes. These markers can be used as precursors for targeted sequencing to determine more precise locations and predict the phenotype of the offspring (Soller and Beckmann, 1983). There are several molecular tools in animal genetics, among which restriction fragment length polymorphisms (RFLPs), microsatellites, and single nucleotide polymorphisms (SNPs) are the most popular and widely used.

2.4.1.1. Restricting Fragment Length Polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) is one of the first types of markers of genetic sequence variation. The researchers initially digested the DNA with restriction enzymes and then used Southern blotting (Botstein *et al*.,1980) to identify the resulting target fragments. Subsequently, the RFLP method was combined with the polymerase chain reaction (PCR) to target specific DNA sequences (Saiki *et al*., 1985), simplifying the process.

Beckmann and Soller, (1983) proposed many uses of RFLP in plants and livestock, including group identification, quantitative trait locus (QTL) mapping, and MAS. Although RFLP iswidely distributed in the genome, each analysis fragment requires PCR or Southern blotting, and each polymorphism requires the use of specific restriction enzymes, so they are difficult to analyze on a large scale.

2.4.1.2. Microsatellites

Microsatellites, repeated short sequences found in tandem in the genome, were discovered in the early 1980s (Miesfeld *et al*., 1981; Spritz, 1981) and found to be more useful as genetic markers than RFLPs (Fredholm *et al*., 1993). Widely distributed within mammalian genomes (Toth *et al*., 2000), microsatellites were used to create linkage maps during the early 1990s for many species (Ellegren *et al*., 1994). Microsatellites tend to be highly polymorphic, with many alleles per locus (Fredholm *et al*., 1993), which reduces the number of families and/or loci that need to be assayed to obtain the same amount of distinctive information as bi-allelic markers can provide (Vignal *et al*., 2002). Although microsatellites can be found throughout the genome, they are lessabundant in the exons than in the non-coding regions of eukaryotes (Hancock, 1995) and thus arenot overly influential on many eukaryotic protein sequences. This limitation, as well as thehigher abundance of SNPs (Taillon-Miller *et al*., 1998), the difficulty to genotype many microsatellites across the genome simultaneously, the ease of statistical analysis of SNPs (Wakeley *et al.*, 2001), and the greater mutation rate of microsatellites, led to the replacement of microsatellites with SNPs as the marker of choice in the late 1990s and early 2000s.

2.4.1.3. Single Nucleotide Polymorphism (SNPs)

Single nucleotide polymorphisms (SNPs) are a bi-allelic type of marker, become popular because of many advantages. Compared with other genomic variants, SNP is by far the most abundant known in animals. Due to stability and high-throughput automated analysis, SNP may be a potential genetic marker and is of greater interest (Fries *et al*., 1990; Heaton *et al*., 2002). To complement the development of molecular markers on the basis of single or few loci, high put through genotyping via next generation sequencing (NGS) in the form of array or chip-based markers is more useful. Such markers could be used for a variety of purposes including genome-

wide association studies, population studies, bulk segregant analyses, quantitative trait loci(QTL) interval mapping, whole genome profiling, background screening (Kim *et al*., 2006; Wenzl *et al*., 2007; Gupta *et al*., 2008).

Although SNP discovery was much slower in production animal species than in humans, over one thousand SNPs had been discovered in pigs (Fahrenkrug *et al*., 2002) and chickens (Kim *et al*., 2002). Most of the major livestock species have used some form of reduced representation libraries (RRLs) to discover SNPs in different lines or breeds (Ramos *et al*., 2009). Genome sequencing efforts in production animals have led to the discovery of hundreds of thousands of additional SNPs (Wade *et al*., 2009).

The availability of tens of thousands to millions of SNPs per species has led to the development of SNP chips that are capable of assaying large numbers of SNPs per animal simultaneously. Illumina, Inc. (San Diego, CA) currently offers SNP chips that can assay approximately the following maximum numbers of SNPs for each species: 2.5 million in human; 700,000 in cattle; 60,000 in swine; and 50,000 in sheep (www.illumina.com). The ability to genotype so many SNPs at once has drastically reduced the cost per individual per genotype to less than half a cent for the entire process from DNA isolation to genotyping with a SNP chip.

The first-generation low-density magnetic bead chip Bovine 3K magnetic bead chip was launched in 2010 to increase the adoption of genomic testing (Illumina Inc. 2011a; Wiggans *et al.*, 2011). In addition, Illumina's Bovine 50K SNP high-density microbead chip is on the market.Unlike Bovine 3K, the Bovine 50K SNP (Infinium) chip can provide more than 50,000 informative SNPs that are evenly distributed throughout the bovine genome. The rapid detection of the Bovine 50K chip is evidenced by the number of new individuals tested (Illumina Inc. 2011b; Wiggans *et al*., 2011). Currently High-density SNP assays, such as the 80k BeadChip

(GeneSeek Genomic Profiler; GeneSeek, Lincoln, NE, USA) are available with large numbers of SNPs from which the most informative SNPs can be selected for breed assignment (Edea, *et al.,* 2015).

Several studies have been conducted in comparison of different Bechchips for breed assignment andin study of genetic diversity. BovineSNP50BeadChip are mainly derived from, Bos taurus breeds; use of these SNPs may lead to biased estimation of diversity indices in Bos indicus cattle breeds (Edea *et al.*, 2013). Recently, a higher SNP density 80K BeadChip (GeneSeek Genomic Profiler HD BeadChip) has been released, mainly of the Bos indicus variety. It is highly used in African Bos indicus cattle breeds in detection of more informative SNPs in comparison to others SNPs chips (Lee *et al*., 2014; Dadi *et al*., 2014; Edea *et al.*, 2015). Boichar *et al. (*2012). Compered 9, 50k and 80k BeadChips and reported that 80 K (GGP-80 K) contained approximately 80,000 SNP derived from bos indicus (www.geneseek.com). Edea et al. (2015) also compered 9, 50 and 80k SNP BeadChip and detected high autosomal SNPs in in bos indicus.

CHAPTER THREE

3. MATERIALS AND METHODS

3.1. Study breeds, samples collection and DNA extraction

DNA samples were collected from four Ethiopian indigenous cattle populations (n = 135)inhibited in different agro-ecologies: Arsi (n = 29), Begait (n = 40), Borana (n = 40) and Sheko(n = 26). Unrelated female and male animals were sampled based on available pedigree records or cattle owners' information. Nasal samples were collected using Performagene LIVESTOCK's nasal swab DNA collection kit and DNA were extracted from nasal samples according to the manufacturer's recommendations (DNA Genotek Inc.,2012). Two European dairy breeds (Holstein, n = 60) and Jersey (n = 38) were used for comparison.

3.2. Study breed and their descriptions

Arsi breed: Arsi breed is descended from the recent introductions of zebu into Africa from West Asia, and probably developed from a group of small shorthorn Abyssinian Zebu by the highland Oromo people (DAGRIS, 2006). Arsi cattle are mainly found in the central highlands of Ethiopia especially in Arsi, Shewa and Bale administrative regions. They are small, short and compact. Red, with a black muzzle, is the predominant color although many animals are black, light grey or white with black spots. It is classified in to zebu cattle type (Tesfaye *et al*., 1994).

Begait: This breed alternatively called Barca, is believed to have originated from Sudan and low-lands of Eritrea (DAGRIS, 2015). Begait cattle breed is one of the major breeds found in Ethiopia, used primarily for milk and meat production, and is usually white with black spots (Mason,1996). Begait cattle are phenotypically relatively large in size with a well-developed udder, small and stumpy horns in both male and females, long teats, a higher milk yielde and

aggressive nature. The common coat colors are grey, black and white. In terms of susceptibility, they are very vulnerable to food shortage. (Zerabruk *et al*., 2007; Mekonnen *et al*., 2010).

Boran: The breed is originally descended from the first introduction of Zebu into Africa from West Asia. Its presence was established first in the semi-arid and arid pastoral Boran plateau of southern Ethiopia. Pastoral movements and migrations led to spread of the Ethiopian Boran to the eastern rangelands in Ethiopia as well as into northern Kenya and south-western Somalia. The main location of Ethiopian Boran is the southern rangelands of Ethiopia, around Liben, Mega and Arero plains with the Borana pastoralists and bordering area of northern Kenya (DAGRIS,2007).

Sheko: Sheko breeds is believed to be the last remnant of the original humpless shorthorn (bos taurus) cattle in eastern Africa. It is the only known taurine type breed in Ethiopia. At present some of the Sheko manifest small humps inherited from zebu through introgression. The breed isnow considered endangered through gradual interbreeding with local zebu and sanga (Rege, 1999). The breed inhabits the humid parts of south-western Ethiopia with the Sheko tribe around Bench Maji zone and it is believed to have some level of trypanotolerance (DAGRIS, 2007).

Holstein-Friesian - Is a breed of large dairy cattle and it is native to Friesland and northern Holland. Its main features are its large size and spotted markings in black and white, sharply marked rather than blended. It is thought that these cattle have been selected for around 2,000 years for milk consistency. They have long been widely spread throughout continental Europe's more fertile lowlands, where they are highly valued for their milk-producing ability. The Holstein-Friesians in the United States out number all other milk breeds and produce nine-tenths of the milk supply (https://www.britannica.com/animal/Holstein-Friesian).

Jersey - The Jersey breed originated on the Island of Jersey, a small British island in the English Channel off the coast of France. Jersey is one of the oldest dairy breeds, having been reported by authorities as being purebred for nearly six centuries. The breed was known in England as early as 1771 and was regarded very favorably because of its milk and butterfat production. At that early date, the cattle of Jersey island were commonly referred to as Alderney cattle although the cattle of this island were later referred to only as Jerseys (Briggs and Hilton, 1980).

3.3. Genotyping and quality control

All selected Ethiopian indigenous cattle were genotyped using the GGP-80k Bead chip (GeneSeek Genomic Profiler). Autosomal SNP markers were screened for a call rate of ≥ 90%, aminor allele frequency (MAF) of > 0.01, and a sample call rate of ≥ 90%. After the above qualitymanagement parameters applied, the autosomal SNP markers obtained were used for analysis of linkage disequilibrium and selection signatures analysis. For MAF analysis, SNPs samples were filtered for a call rate ≥ 90% leaving 66432, 65460, 66811, 60679**,**67475 and 67462 for, Arsi, Begait, Borana, Sheko, Holstein and Jersey cattle, respectively. Quality control was implemented in PLINK (Purcell *et al*., 2007).

3.4. Data analysis

Minor allele frequency (MAF): within -breed genetic variation was estimated using PLINK version v1.09 (Purcell *et al.,* 2007). When it possesses an allele that is present only in specific breed, an SNP was stated to be breed specific (Ramos *et al*., 2011). Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 were used for functional analysis and analyses of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (Huang *et al.* 2008). A principal component analysis (PCA) and Manhattan Plots was carried out using R 4.5 studio.

Selection signatures analysis: Inter-population genetic differentiation (*Fst*) (Weir & Cockerham 1984) was calculated to identify selection signatures between the following breed pairs: Arsi-Holstein, Arsi-Jersey, Boran-Holstein, Boran-Jersey, Begait-Holstein, Begait-Jersey, Sheko-Holstein, and Sheko-Jersey using PLINK (Purcell *et al.*, 2007). The fixation index (*Fst*), is derived from the equation: $Fst = \frac{Ht-Hs}{Ht}$, Where, Ht = total genetic diversity, Hs= genetic diversity within sub-populations. The top1% *Fst* values were considered as an outlier loci or SNPs under positive selection. The Bovine UMD.3.1 genome assembly was used to annotate candidate genes associated with the outlier loci highly differentiated SNPs. The Database for Annotation, Visualization and Integrated Discovery (DAVID) variation 6.8 was used for functional analysis of the candidate genes (Huang *et al*., 2008). Quantitative traits loci (QTL) regions that overlapped with the identified candidate regions were also detected from cattle QTLdb (QTLdb/BT/index).

Linkage disequilibrium (LD): LD for adjacent SNPs were estimated using, distance of two loci from parent (D') and correlation coefficient of the frequencies (r2) (Hayes and Goddard, 2010) by employing SNP and variation suite v.8.5.0 (Golden Helix, Inc.; www.goldenhelix.com).These parameters vary between 0 and 1. D' and r2 were calculated as follow:

$r^2=\frac{D}{PA1PA2PB1PB2}$ r^2 = Correlation coefficient of the frequencies

D = P11p22 – p12p21 if D>0, $\frac{D}{Dmin}$ if D<0 D'= Distance of two loci from parent

$|D'| = \frac{D}{Dmax}$

Where, A and Bare two genes and each gene has only two alleles

A allele = A1 and A2, B allele = B1 and B2

Where, as

p11 = probability of A1B1 haplotype combination

p12 = probability of A1B2 haplotype combination

p21 = probability of A2B1 haplotype combination

p22 = probability of A2B2 haplotype combination

CHAPTER FOUR

4. RESULTS

4.1. Minor Allele Frequency

In this study minor allele frequency (MAF) was calculated for Ethiopian cattle populations and European dairy cattle and presented in Table 1. The mean MAF was 0.32 ± 0.12, 0.32 ± 0.12,0.31 ± 0.13, 0.30 ± 0.13, 0.19 ± 0.17 and 0.18 ± 0.17 for, Arsi, Begait, Borana, Sheko, Holstein and Jersey cattle, respectively. The overall mean MAF (0.31 ± 0.125) in Ethiopian cattle population was found to be greater than the value estimated in European dairy breeds (0.18 ± 0.17). Among Ethiopian cattle populations the lowest mean MAF was observed in Sheko.

Table 1. Mean (MAF) values in Ethiopian cattle populations and European dairy cattle breeds

Breed/Population	No	Mean ± SD
Arsi	29	0.32 ± 0.12
Begait	40	0.32 ± 0.12
Boran	40	0.31 ± 0.13
Sheko	26	0.30 ± 0.13
Over all	**135**	**0.31 ± 0.125**
Holstein	60	0.19 ± 0.17
Jersey	38	0.18 ± 0.17
Over all	**98**	**0.18 ± 0.17**

The distribution of MAF proportion across each chromosome was also detected as shown in Table 2. Across the Ethiopian cattle populations, common variant MAFs (≥0.10 and 0.5) were counted for 0.89% of the total SNPs. In European dairy breeds lower proportion of common variant SNPs 0.57% was observed. Intermediate variant MAFs (≥ 0.05 and 0.10) were observed in about 0.043% and 0.062% of markers in Ethiopian cattle populations and European dairy breeds, respectively. Rare Variant MAFs (≥ 0 and 0.05) detected with 0.018% in Ethiopian cattlepopulations and 0.064% in European dairy breeds. The proportion of fixed (0) MAFs accounted 0.0052% in Ethiopian cattle populations and 0.28% in Europeans dairy breeds.

Table 2. Minor allele frequency distribution of 80K Bead Chip SNP in Ethiopian cattle populations and European Cattle breeds

Breed/ Population	No.	Fixed (0)		Rare (>0 and 0.05)		Intermediate ≥0.05 and 0.10		Common ≥0.10 and 0.5	
		SNP	Pro	SNP	Pro	SNP	Pro	SNP	Pro
Arsi	29	282	0.0042	986	0.015	2481	0.037	60805	0.91
Begait	40	350	0.0053	1077	0.016	2693	0.041	61339	0.94
Boran	40	328	0.0049	1384	0.020	2806	0.042	60929	0.91
Sheko	26	438	0.0066	1441	0.021	3309	0.054	53715	0.82
Overall	**135**	**349**	**0.0052**	**1222**	**0.018**	**2822**	**0.043**	**59197**	**0.89**
Holstein	60	19962	0.29	4194	0.062	3521	0.052	40238	0.59
Jersey	38	19237	0.28	4550	0.067	4924	0.072	38108	0.56
Overall	**98**	**9799**	**0.28**	**4372**	**0.064**	**4222**	**0.062**	**39173**	**0.57**

SNP, Single nucleotide polymorphism; Pro, Proportion.

MAFs average was detected across each chromosome for Ethiopian cattle populations (Arsi, Begait, Borana and Sheko) and European dairy breeds (Holstein and Jersey). In Arsi cattle the average MAFs was 0.31% on BTA 5, 3 and 2 and on the rest of chromosome the MAFdistribution was 0.32. Average MAFs distribution accounted 0.32% across each chromosome for Begait. For Boran cattle populations on BTA 2(0.03%) and on BTA 5 it accounted 0.30% and 0.31% for the rest of chromosome. In Sheko breeds on BTA 1, 2 and 9 accounted 0.29% and on other 0.30% were detected. In European cattle breeds for Holstein on BTA 1 to 17 and 18 to 29 accounted 0.18% and 0.19% respectively. In Jersey Average MAFs distribution on BTA 1,2,3,4,5 and 8(0.16%), 9, 6,7 and 8 (0.17%) and 11 to 23; 24 to 29(0.18%) were accounted (Figure 1).

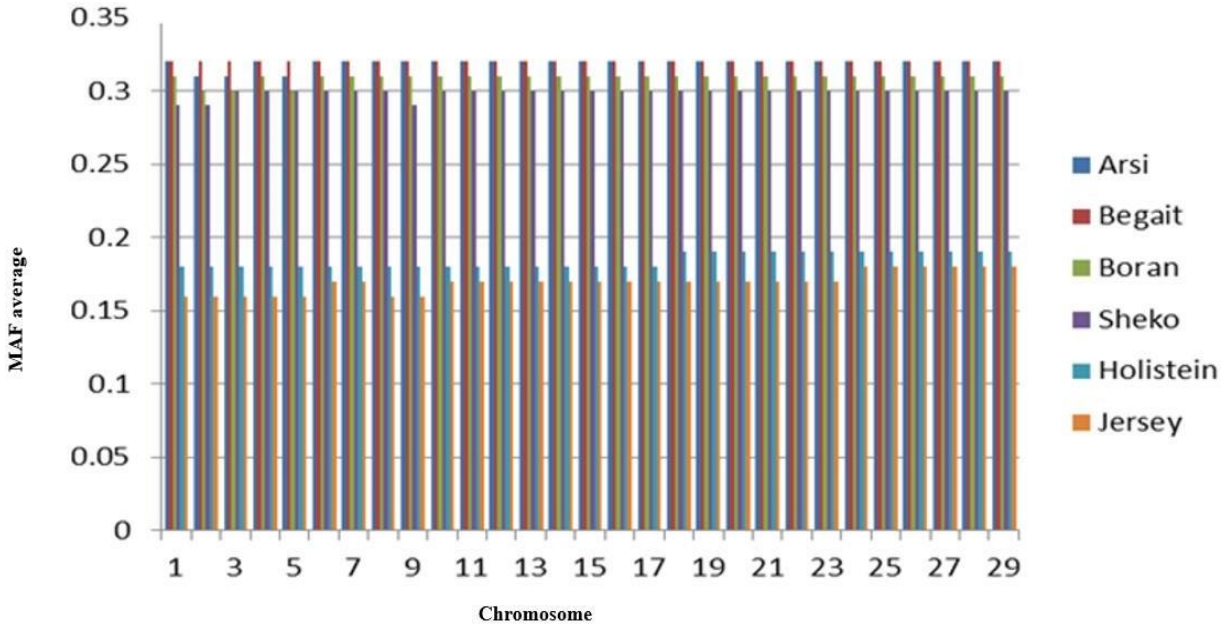


Figure 1. MAF distributions of Ethiopian cattle populations and European cattle breeds across chromosome.

4.1.2. Breed specific Single Nucleotide Polymorphisms

Breed specific SNPs among Ethiopian cattle populations and European dairy cattle breeds were detected. In Ethiopian cattle populations the total number of breed specific SNPs were 12, 16, 8 and 23 in Arsi, Begait, Borana and Sheko respectively. Among Ethiopian cattle populations the highest number of SNPs with 0.071% MAF average and 23 SNPs were observed in Sheko and whereas, in Boran lower number of SNPs with 0.051% MAF average and 8SNPs were detected. For European cattle breeds 2188 and 1958 were detected in Holstein and Jersey respectively as shown in Table 3.

Table 3. Breed-specific SNPs detected in Ethiopian cattle populations

Breed/Population	SNPs	MAF		
		Minimum	Maximum	Average
Arsi	12	0.017	0.068	0.033
Begait	16	0.013	0.15	0.051
Borana	8	0.013	0.16	0.062
Sheko	23	0.019	0.19	0.071

Commonly breed specific were also detected in Ethiopian cattle populations in comparison to European cattle breeds. The common SNPs for Arsi-European, Begait-European, Boran- European and Sheko-European were 16802, 16563, 16874, 15306 respectively. For Boran- European the highest number of common SNPs (0.34%) were detected and the lower number of common SNPs (0.29%) were accounted in Sheko. For Ethiopian-European 0.9% common NPs were detected. Similarly, for all Ethiopian common breed specific SNPs were also detected. A number of Ethiopian breed specific were 16387 with 0.34% Table 4.

Table 4. Breed - specific SNPs for Ethiopian cattle populations - European cattle breeds

Breed/Population	No. SNPs	MAF		
		Minimum	Maximum	Average
Arsi-European	16802	0.02	0.5	0.34
Begait-European	16563	0.01	0.5	0.34
Boran-European	16874	0.01	0.5	0.34
Sheko-European	16404	0.02	0.5	0.34
Ethiopian-European common SNPs	148	0.86	0.5	0.9
Ethiopian breed specific	16387	0.02	0.5	0.34

4.1.2.1. Annotation and functional analysis of Ethiopian breed specific cattle

Commonly specific SNPs in Ethiopian cattle 16387 and 1958 in European dairy breeds were annotated to the UMD3.1 genome reference respectively. From the total of annotated gene 67% Intergenic, 14% Intronic, 5% downstream, 9% upstream and 5% non-synonymouswere detected in Ethiopian cattle populations Figure 2.

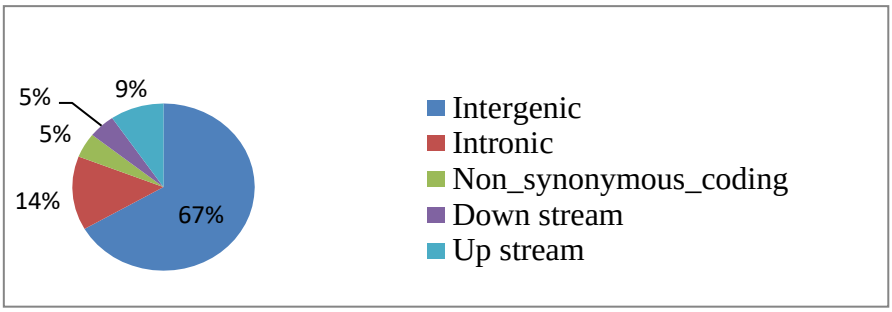


Figure 2. Annotated candidate gene in Ethiopian cattle populations

The SNPs detected as breed specific were associated with *ADCY5, ACAD11, ABCA12, ACTR3, POU2F1, IGFBP-3, WNT1, IGF1, A2M, ABCG2, KIT, KDR, CSN2, ABCA7, HSPA4, PDE4B, RAD50, ABCA1, JAK2, B4GALT4, FOXO3, GHR, PCCA, HSPH1, AHCY, POFUT1, ADCY8, HSF1, XKR4, PGR, ACACB, ACAD10, ACTN4, ACCN1, ACAA2, ABCC23* genes. From these genes *CSN2, ABCA7, PDE4B, ABCA1, JAK2, B4GALT4, FOXO3, GHR, ADCY8, ACACB, ACAA2, POFUT1, A2M* genes were associated with milk production traits like milk proteins, milk fat percentage, milk yield and other traits such as udder growth and mastitis resistance. Other potential gene associated with fertility and growth were *PGR, GHR, ACAD10, XKR4, ADCY5, POU2F1, IGF1, ABCC2*. Similarly, some genes were associated with adaptation traits *(ACC2, HSF1, HSPH1, HSPA4, KDR, RAD50)* and coat color *(WNT1, KIT, AHCY*) feed intake *(XKR4, ACCN1, ACAD11)* fat thickness *(IGFBP-3* and *POU2F1)* and birth weight (*ABCA12)* Table 5

Table 5. Annotated genes corresponding to Ethiopian cattle specific SNPs

Chr	Position	rs position	Gene	Function	Reference
1	138055411	rs136601562	*ACAD11*	Feed intake	Karisa *et al*., 2013
1	68382869	rs132888556	*ADCY5*	Fertility	Cai *et al*.,2019
2	103642042	rs136747972	*ABCA12*	Skin keratinization and calf birth weight	Charlier *et al*. 2008; Cole *et al.* 2014
2	65910313	rs137791631	*ACTR3*	Cytoskeleton organization	Mokhber *et al.*, 20118
3	1355306	rs133857868	*POU2F1*	Growth and fatness	Montarelo *et al*., 2014
4	76714543	rs133888844	*IGFBP-3*	Fat thickness	Venoroni *et al*., 2010
5	101342559	rs135159652	*A2M*	Preventing diseases that reduce milk yield	AC Freitas *et al.*, 2016
5	31016805	rs136527734	*WNT1*	Coat color	Lim *et al*., 2016
5	66654472	rs137398475	*IGF1*	Growth and feed efficiency	Stick *et al*., 1998
6	38014096	rs137490422	*ABCG2*	Milk protein	Yue *et al*. 2010
6	71791595	rs133971615	*KIT*	Coat color	Stella *et al.*, 2010
6	72271723	rs135309411	*KDR*	Thermo- tolerance	Ghadikolaci, *et al*.,2018
6	87186813	rs133265582	*CSN2*	Milk fat, protein and yield	Miluchova *et al.*, 2014
7	23150926	rs135764402	*RAD50*	Thermo- tolerance	Slipa *et al.*, 2021
7	45174539	rs134499400	*ABCA7*	Lactation	Mani *et al*., 2010

Table 5 Continued

7	46282289	rs137506643	*HSPA4*	Thermo- tolerance	Andrey *et al.,* 2018
7	79647636	rs135371193	*PDE4B*	Milk Production Traits	Kim *et al.,* 2021
8	39686168	rs135334215	*JAK2*	Milk fat percentage and mastitis	Khan *et al.,* 2019
8	76229925	rs132722340	*B4GALT4*	Milk production and Mastitis	Ogorevc *et al*.,2009
8	96333614	rs132638399	*ABCA1*	Milk fat metabolism	Jiang *et al*., 2019
9	42101159	rs134260716	*FOXO3*	Milk protein and fat	Gao *et al.,* 2017
10	32121397	rs136610400	*GHR*	Growth and milk	Xiaolong *et al.*,2018
12	29894900	rs133917282	*HSPH1*	Thermo- tolerance	Ben-Jemaa *et al*., 2020
12	81051206	rs134266870	*PCCA*	Neurotransmitter concentration	Chen *et al.,* 2020
13	62251016	rs133563880	*POFUT1*	Mammary growth, mastit defense and milk coagulation	Gutierrez *et al.,* 2014; Dettori *et al*., 2020
13	64269864	rs134537985	*AHCY*	Coat color	Nazir - Ghalilolae e*t al.,* 2018
14	11050557	rs132747781	*ADCY8*	Milk fat percentage, milk yield	Moradian *et al.,* 2019
14	1818986	rs135497883	*HSF1*	Heat tolerance	Baena *et al.,* 2018
14	24578344	rs133527588	*XKR4*	Growth, Feed intake	Perry *et al*., 2012
15	8128589	rs133643863	*PGR*	Fertility traits	Ying *et al*., 2000

Table 5 Continued

17	57975621	rs134943011	*ACAD10*	Fertility	Mota *et al*., 2017
17	66200226	rs137005387	*ACACB*	Reverse cholesterol transport and milk yield and composition	Wang *et al*., 2017; Han *et al*., 2017
18	48669914	rs134870838	*ACTN4*	Johne's disease	Neibergs *et al*., 2010
19	17081103	rs134892343	*ACCN1*	calf size, body weight (mature, yearling), residual feed intake	Lim *et al.,* 2016
24	49941216	rs134317084	*ACAA2*	Milk yield	Miltiadou *et al.,* 2017
26	20672654	rs137381680	*ABCC2*	production	Cheruiyot *et al*., 2018

The annotated genes were involved in different pathways, cAMP signaling path way, Rap1 signaling path way, calcium signaling pathway, Ras signaling pathway, Pathways in cancer, *GnRH* signaling pathway, thyroid hormone signaling pathway and *HIF-1* signaling pathway, ABC transporters, Adrenergic signaling in cardiomyocytes, Prolactin signaling pathway, Vasopressin regulated water absorption, and Estrogen signaling pathway (Table 6). ABC transporter family genes (ABCA1, ABCA7) were known to affect milk traits involved in this pathway. Another gene associated with milk traits (ADY8) was also involved in GnRH signaling pathway. AKT3 was involved in the estrogen signaling pathway. JAK2 and ITGAV genes associated with milk traits and mastitis, KIT and KDR genes associated with coat color

and GHR responsible for growth and milk production involved in PI3K – Akt signaling pathway were detected. The potential gene associated with milk production trait (CSN2) was involved in prolactin signaling pathway.

Table 6. KEGG path way analysis of Ethiopian cattle specific SNPs corresponding annotated genes

KEGG pathway	Count	P- value	Gene
Adrenergic signaling in cardiomyocytes	26	3.9E-14	*AKT2AKT3, ATP1A1, ATP1B1, ATP1B3, ATP2B1, BCL2, ACTC1, ATF2, AD, CY1, ADCY5, ADCY8, ADCY9, ADRA1A, ADRA1B, AGTR1, CACNA2D1, CACNA2D4, CACNB2, CACNB4, CACNG2, CACNG4CACNA1D, CAMK2B, CAMK2D*
cGMP-PKG signaling pathway	21	1.7E-8	*AKT2, AKT3, ATP1A1, ATP1B1, ATP1B3, ATP2B1, ATP2A2, BAD, ATF2, ADCY1ADCY5, ADCY8, ADCY9, ADRA1A, ADRA1B, ADRA2A, ADRB3*
Calcium signaling pathway	21	3.1E-7	*ATP2B1, ATP2A2, ADORA2A, ADORA2B, ADCY1, ADCY8, ADCY9, ADRA1A, ADRA1B, ADRB3AGTR1, AVPR1A, BDKRB2, CACNA1A, CACNA1B, CACNA1D, CACNA1E, CACNA1H, CAMK2A, CAMK2B, CAMK2D*
cAMP signaling pathway	18	4.0E-5	*ADCYAP1R1, AKT2, AKT3, ATP1A1, ATP1B1, ATP1B3, ATP2B1, BAD, ADORA2A, ADCY1ADCY5, ADCY8, ADCY9, BDNF, CACNA1D, CAMK2A, CAMK2B,*
ABC transporters	8	1.5E-4	*ABCA1, ABCA12, ABCA7, ABCB11, ABCC2, ABCC3, ABCC5, ABCG2*

Table 6 Continued

HIF-1 signaling pathway	8	1.8E-2	*AKT2, AKT3, BCL2, ANGPT1, ANGPT2, CAMK2A, CAMK2D*
Estrogen signaling pathway	7	5.8E-2	*AKT2, AKT3, ATF2, ADCY1, ADCY5, ADCY8, ADCY9*
PI3K – Akt signaling pathway	67	2.9E-10	*GNG2, GNGT1, HRAS, JAK1, JAK2, KITLG, KIT,MDM2, GHR, HGF, IKBKB, IGF1R, IGF1, INSR, IBSP, ITGAV, IL74, KDR, LAMA1, MTOR*
Prolactin signaling pathway	16	9.9E-4	*FOS, HRAS, JAK2, CSN2, ESR1, ESR2, FOXO3, GALT, GSK3B, CGA, IRF1, LHCGR, MAPK1, MAPK10, MAPK12, MAPK9*
Vasopressin regulated water absorption	40	8.0E2	*ARHGDIA, ADCY9, AQP3 AQP*

KEGG, Kyoto Encyclopedia of Genes and Genomes

4.1.3. Genetic Diversity and Population Structure

4.1.3.1. Population structure

Principal component analysis (PCA) was performed for the four Ethiopian cattle populations separately and for all breeds. The first and second principal components analysis accounted for 4.34% and 3.33% of the total variation and separated the four Ethiopian population according their geographical distribution. Sheko cattle were distantly separated from the remaining Zebu populations, whereas Boran and Arsi cattle closely clustered Figure 3. Figure 4 presents the relationship between Ethiopian cattle populations and European cattle breeds PCA1 and PCA2 accounted for 9.50% and 65.97% of the total variation respectively. PCA1 separated Ethiopian cattle populations from European cattle breeds whereas, PCA2 separates the two taurine breeds.

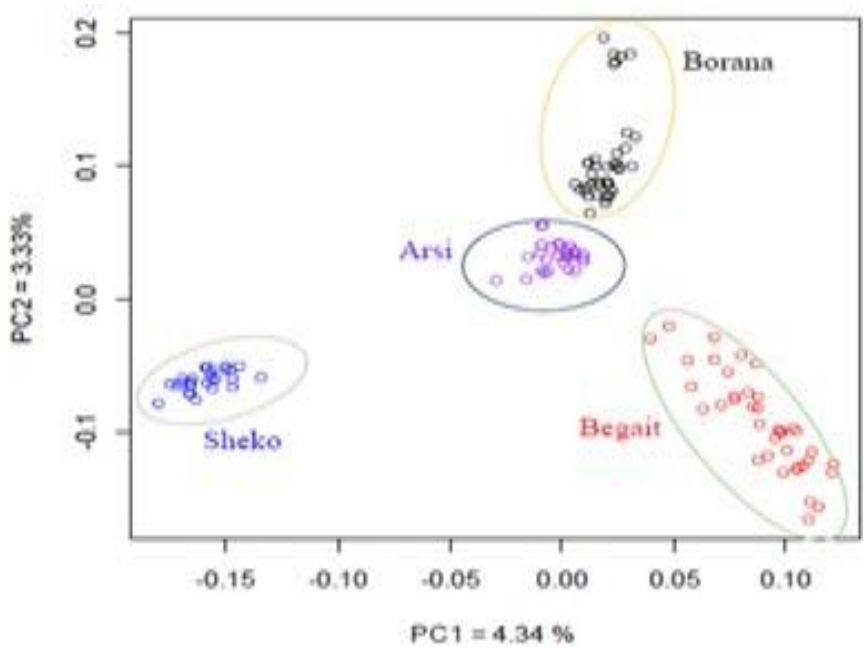


Figure 3. Ethiopian cattle populations clustering based on principal component analysis.

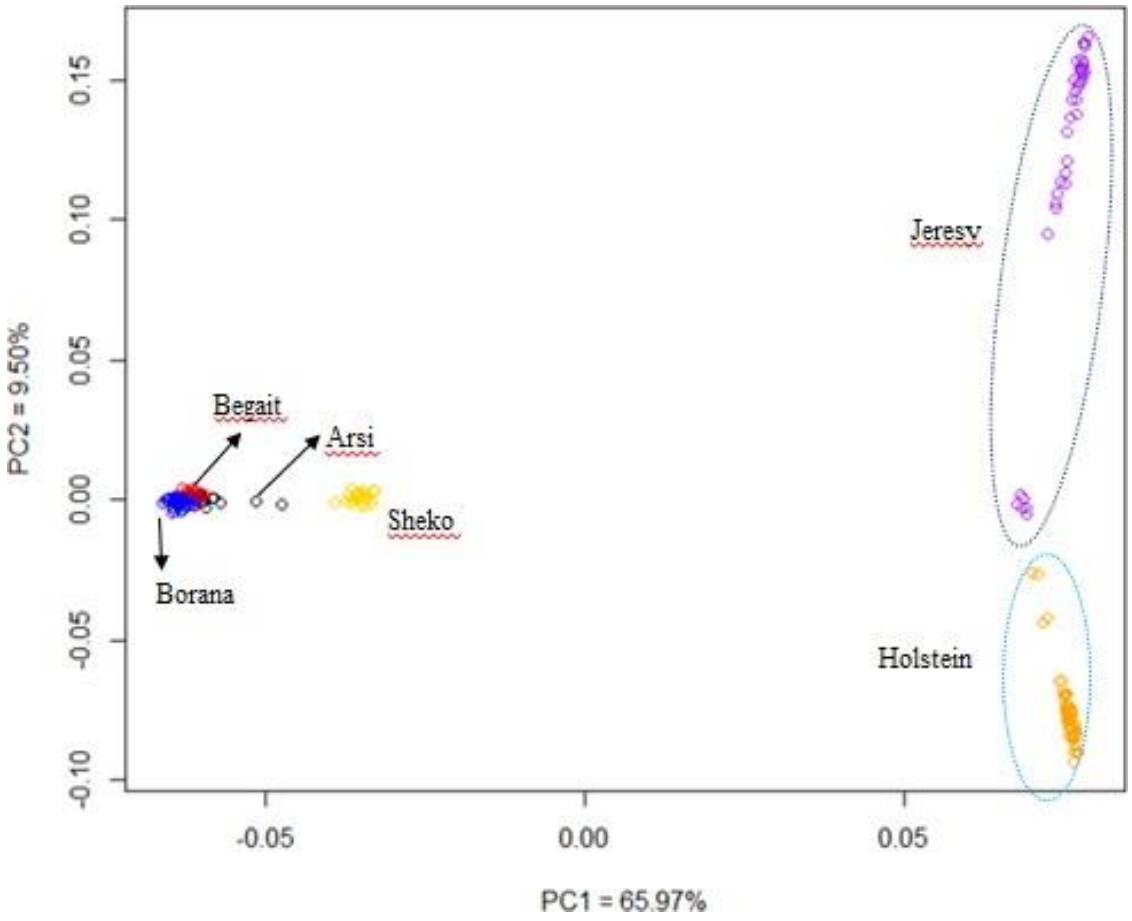


Figure 4. Ethiopian cattle populations and European dairy breeds clustering based on principal component analysis.

4.1.3.2. Genetic diversity

Observed heterozygosity (H_o) and expected heterozygosity (H_E) were used to calculate thegenetic diversity index (Table 7). The H_o ranged from 0.40 in Ethiopian Zebu to 0.23 in Jersey. Similarly genetic diversity (H_E) varied from 0.24 in Jersey to 0.40 in Ethiopian Zebu.

Table 7. Observed heterozygosity (H_o) and expected heterozygosity (H_E) in Ethiopian cattle populations and European cattle breeds

Breed/population	Observed hetrozygosity (Ho)	Expected heterozygosity (He)
Arsi	0.40	0.40
Begait	0.40	0.40
Borana	0.40	0.40
Sheko	0.40	0.39
Holstein	0.26	0.26
Jersey	0.23	0.24

4.1.3.3. Phylogenetic tree analysis

Phylogenetic tree was constructed using a neighbor-joining tree among four Ethiopian cattle populations and two European cattle breeds. According to the result of phylogenetic tree Ethiopian cattle populations were closely clustered together while Sheko the three Ethiopian cattle populations were closely clustered (Figure 5) as supported by PCA. With the exceptions ofSheko breeds other Ethiopian cattle populations (Arsi, Begait and Boran) were clustered in one branch of the clade (Figure 6). Based on UPGMA, the study cattle were distinctly segregated according to their breed origin (*Bos taurus* and Bos indicus). Among Ethiopian cattle, Sheko wasthe first to be separated, followed by Begait. The two Ethiopian Zebu cattle populations (Arsi and Borana) were closely clustered.

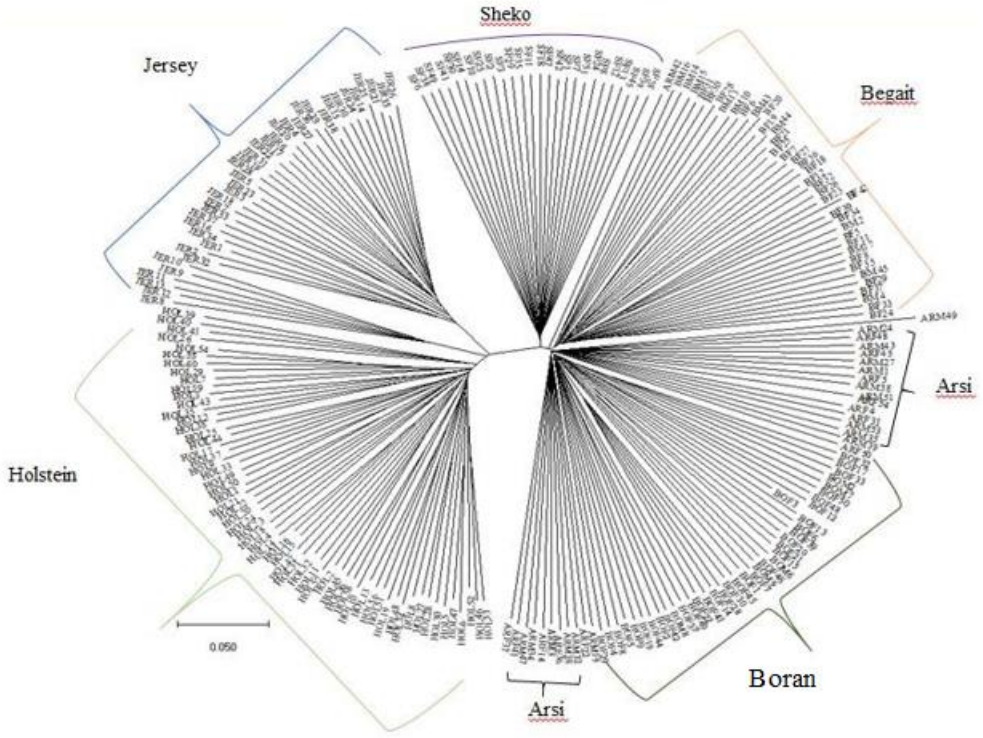


Figure 5. Genetic relationships among 6 cattle breeds constructed using a neighbor-joining tree

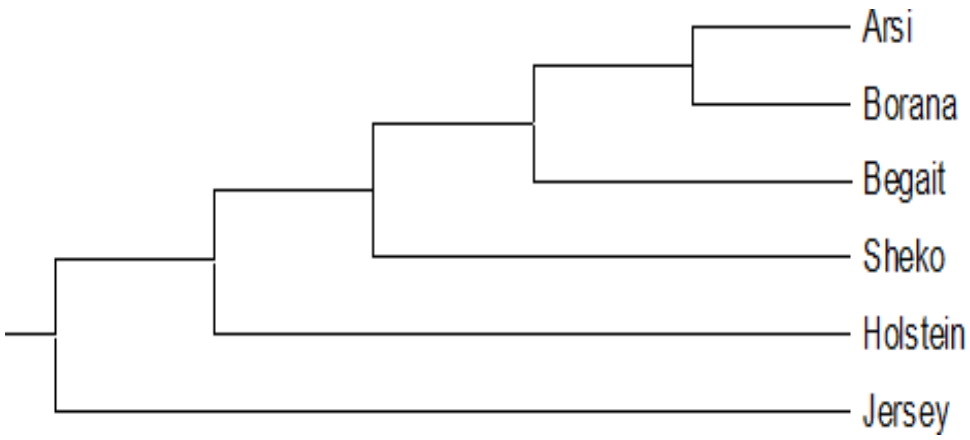


Figure 6. Genetic relationships among 6 cattle breeds constructed using a neighbor-joining tree

4.2. Analysis of linkage disequilibrium among pair wise SNPs

The overall mean average of r^2 and D' were detected in Ethiopian cattle populations and European cattle breeds (Table 8). The average mean of r^2 in Ethiopia cattle populations were higher than in European dairy cattle breeds. The average linkage disequilibrium LD (r^2) ranged from 0.16 in dairy breeds to 0.24. Similarly, the overall mean of D' varied from 0.42 in Holstein and Jersey to 0.69 in Sheko Ethiopian cattle. In all cattle populations, the average of linkage disequilibrium (LD) values declined with the increased physical distance. At shorter physical distances (0-1kb) European cattle breeds had higher r^2 values (0.73 - 0.74) Ethiopian cattle (0.22- 0.24). At 100 - 200 kb, Ethiopian cattle populations had higher r^2 (0,14 -0.16) than European dairy breeds (0.07 – 0.08) (Figure 7). The D' values followed the same trend that estimated values decreased with increase in distance between SNP markers in cattle populations (Figure 8).

Table 8. *r^2* and D' in Ethiopian cattle populations and European cattle breeds

Breed/ population	No. SNPs	*r^2* (Mean, SD)	D' (Mean, SD)
Arsi	66211	0.22 ± 0.25	0.66 ± 0.34
Begait	66211	0.23 ± 0.25	0.66 ± 0.34
Boran	66211	0.22 ± 0.25	0.67 ± 0.34
Sheko	66211	0.24 ± 0.25	0.69 ± 0.34
Holstein	66211	0.16 ± 0.27	0.42 ± 0.45
Jersey	66211	0.16 ± 0.28	0.43 ± 0.45

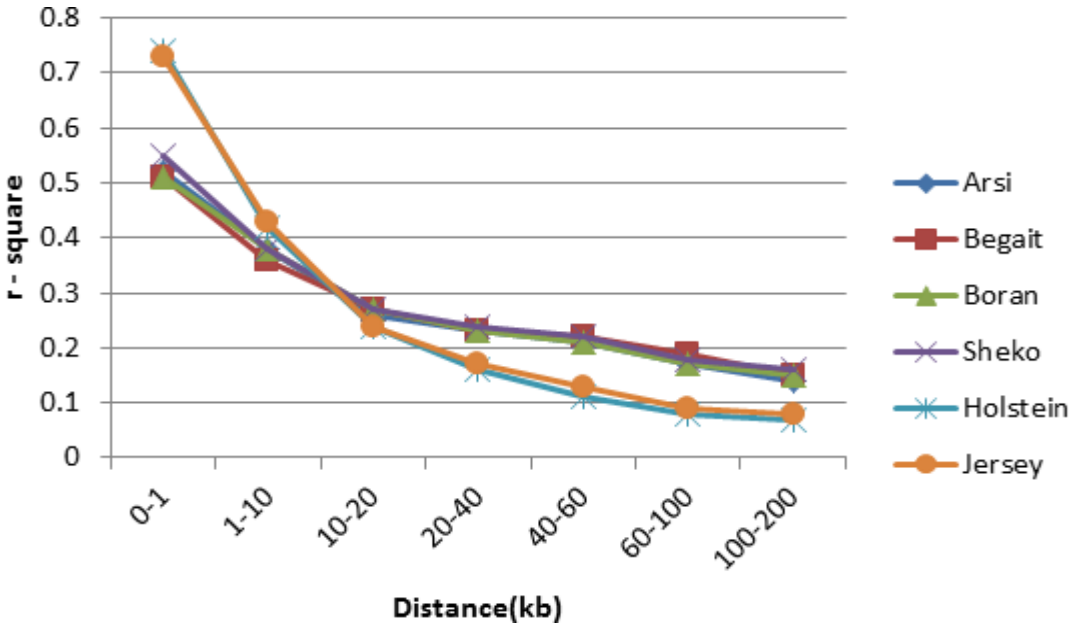


Figure 7. Linkage disequilibrium (r^2) between pairwise SNPs separated by different distances in Ethiopian cattle populations and European dairy cattle breeds

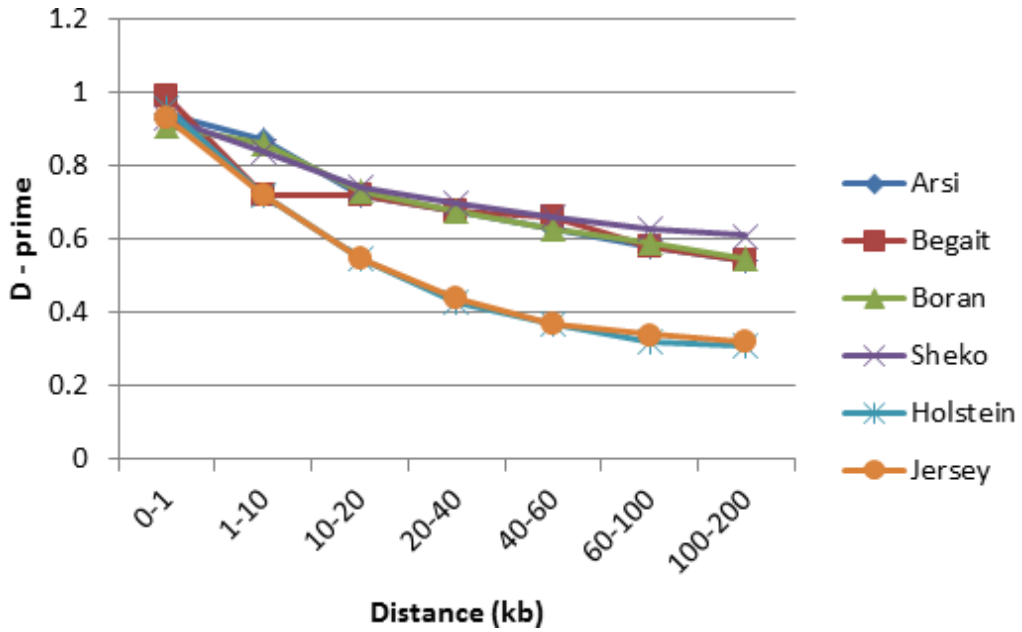


Figure 8. Linkage disequilibrium (D') between pairwise SNPs separated by different distances in Ethiopian cattle populations and European dairy cattle breeds

4. 3. Selection signature analysis

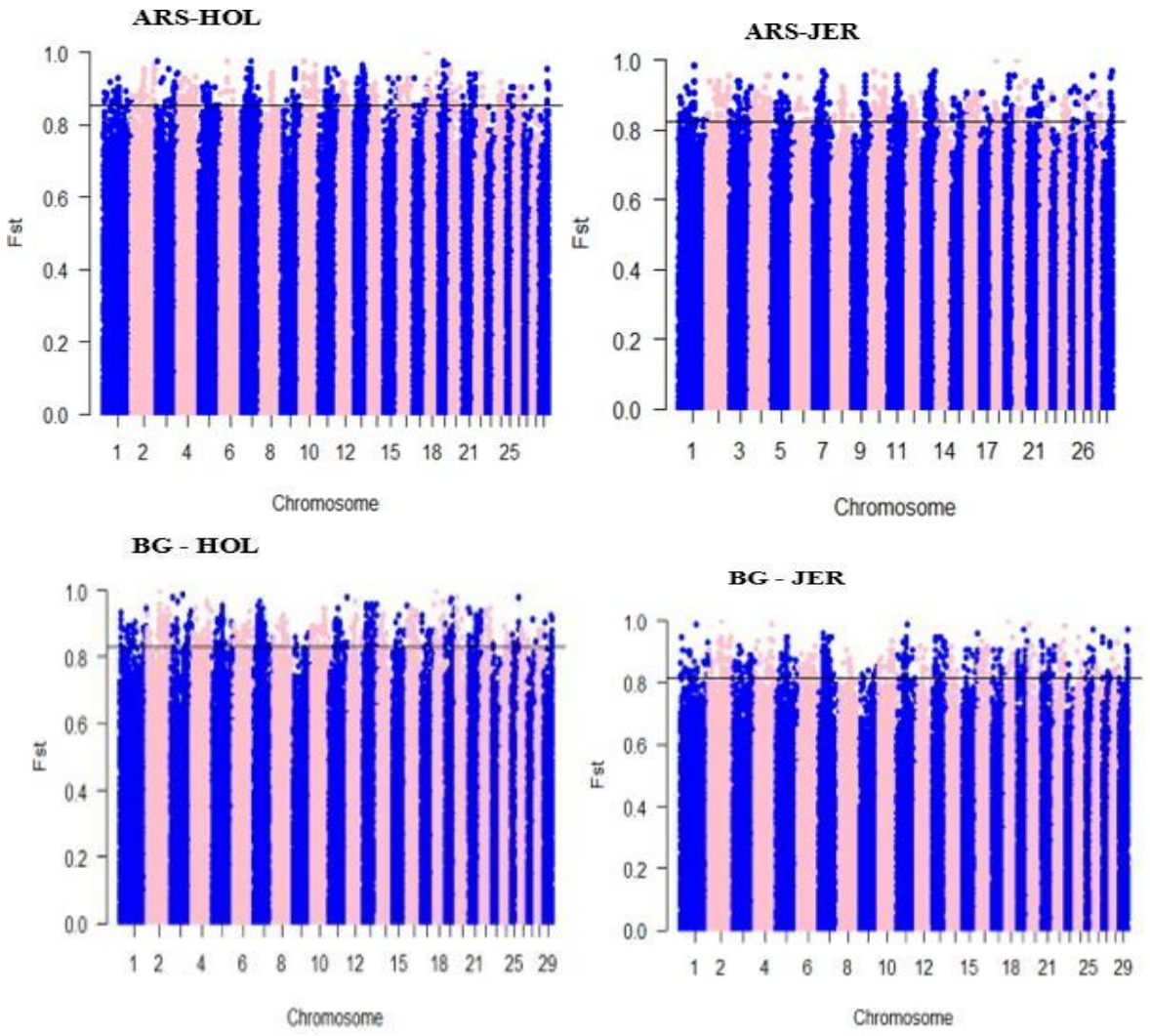
4.3.1. Genetic differentiation

The fixation index difference used to estimate genetic distance between Ethiopian cattle population -European dairy cattle breed and for within Ethiopian cattle populations were analysed (Table 9) Arsi- Begait, Arsi- Boran, Arsi- Sheko, Begait-Boran, Begait -Sheko, Bora-Sheko accounted 0.021, 0.011, 0.032, 0.028, 0.053, 0.044 respectively. The highest genetic distance was detected among pair - breeds for Begait - Sheko (0.053%). Boran - Sheko were the second to be distant by 0.044%. Although, in Boran - Begait pair breeds the lower genetic distance (0.011%) was detected. Generally, the higher genetic distance was detected within Ethiopian cattle populations specifically for Sheko breeds in comparison to other breeds. In Ethiopian cattle populations and European cattle breeds Sheko were exhibited lower distancethan all breeds with 0.24% and 0.23% with Holstein and Jersey pair breeds respectively.

Table 9. Genetic distance between Ethiopian cattle populations and European cattle breeds- based fixation index (Fst)

Breed	Arsi	Begait	Boran	Sheko	Holstein	Jersey
Arsi	0					
Begait	0.021	0				
Boran	0.011	0.028	0			
Sheko	0.032	0.053	0.044	0		
Holstein	0.28	0.27	0.28	0.24	0	
Jersey	0.27	0.27	0.27	0.23	0.22	0

Top 1% was considered in comparison of genetic divergence and differences between Ethiopian cattle populations and European cattle breeds. The genetic divergence between Aris- Holstein, Arsi-Jersey, Begait- Holstein, Begait-Jersey, Boran- Holstein, Boran-Jersey, Sheko-Holstein and Sheko - Jersey were detected across each chromosome as show in Manhattan plots (Figure 9). The difference was shown in all populations across all chromosomes. Arsi, Boran and Begait with Holstein and Jersey were highly diverged, whereas Sheko was lower in divergence in comparison to others.



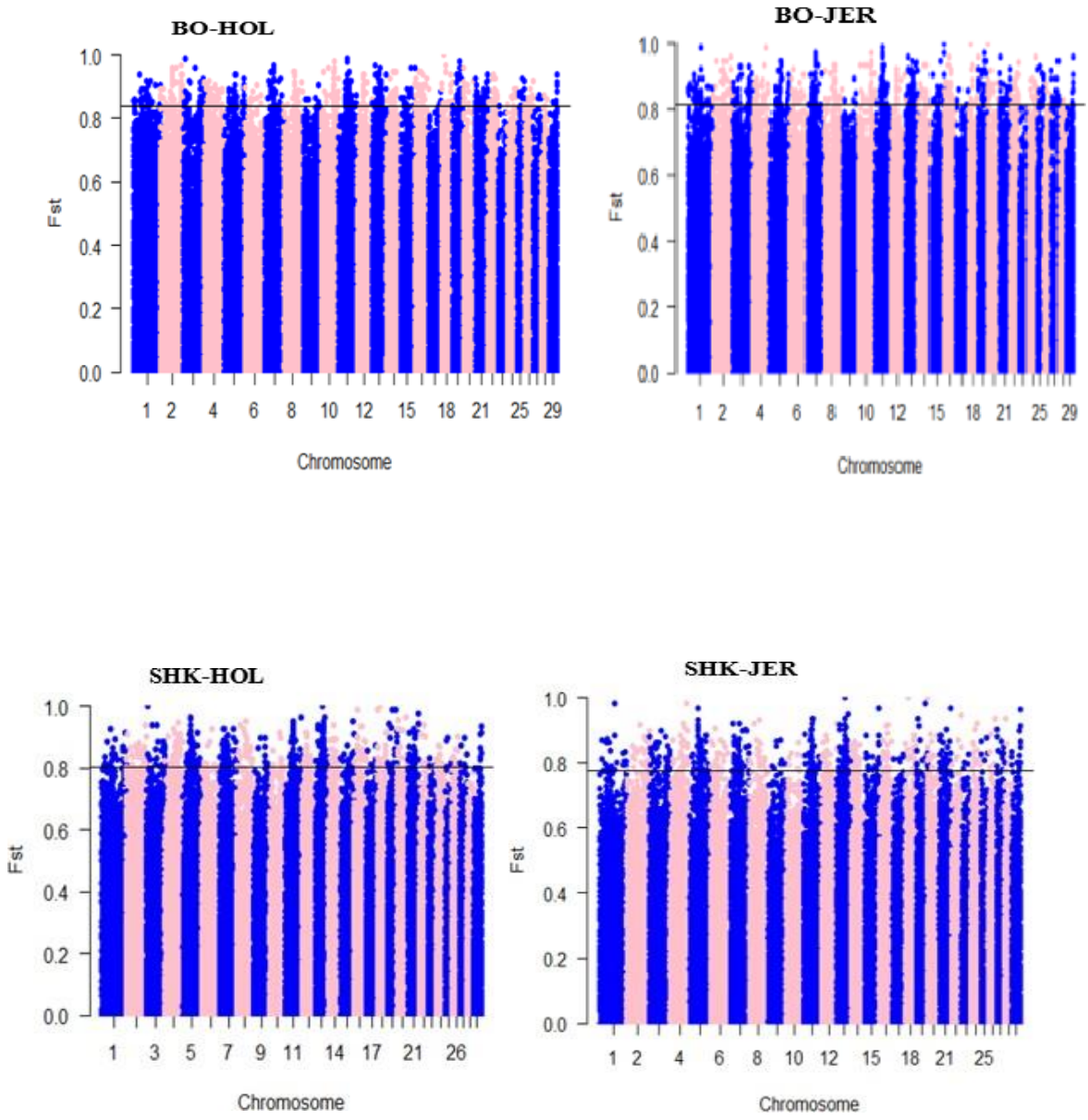


Figure 9. Manhattan Plots analysis between Ethiopian cattle populations and European cattle breeds divergence Arsi-Holstein (ARS-HOL), Arsi-Jersey (ARS-JER), Begait-Holstein (BG-HOL), Begait-Jersey (BG-JER), Boran-Holstein (BO-HOL), Boran-Jersey (BO-JER) and Sheko-Holstein (SHK-HOL), Sheko-Jersey (SHK-JER).

4.6.1. Annotation and functional analysis

The SNPs with top1% Fst value was considered to annotate gene associated with milk candidate gene and responsible genes for adaptation to tropical environment in Ethiopian cattle populations and European cattle breeds. Single nucleotide polymorphisms (SNPs) with top1% FSt value were 674, 674, 657, 666, 674,673, 674 and 674 in Arsi, Begait, Borana, Sheko, Holstein and Jersey were observed respectively. Common SNPs in Ethiopian -European pair wise were 450, 415, 508, 354 for Arsi-European, Begait-European, Boran-European, Sheko-European respectively.Boran has the highest number of common SNPs (508) with European breeds in comparison to other breeds. The lowest common SNPs were detected for sheko-European (354). After annotation 231, 201, 261 and 176 genes were detected for Arsi-European, Begait-European, Boran-European and Sheko-European respectively and 31 common genes were detected Table 10.

Table 10. SNPs and annotated genes in bracket with top 1 % Fst value for Ethiopian - European cattle breeds obtained from pair wise analysis.

Breeds	Arsi	Begait	Boran	Sheko
Holstein	674(331)	674(339)	674(303)	666(306)
Jersey	674(315)	673(329)	674(294)	657(315)
Common	450(231)	415(201)	508(261)	354(176)
Ethiopian-European common SNPs and genes			147 (31)	

Commonly selected SNPs were gnomically annotated to the UMD3.1 reference sequence to detect genes corresponding to those genes (Figure 10). Except for Begait - Holstein higher value of intergenic distribution was detected in all pair-wise cattle populations.

Figure 10. Distribution proportion of top 1 % annotated SNPs across genomic

4.6.2. Functional analysis of candidate genes under positive selection

Database for Annotation, Visualization and Integrated Discovery (DAVID) 6.8 versions were used in annotation of genes corresponding to selected SNPs. Functional enrichment analysis for retrieved genes were also analysed by DAVID (https://david.ncifcrf.gov/tools.jsp). Additionally, to consider the biological function and traits that are known to be affected by each annotated gene a comprehensive literature search was conducted. The biological mechanism, molecular function and signal path ways (KEGG) pathway were considered to describe significantly enriched Gene Ontology (GO) and functional processes. An adjusted Benjamini-corrected p- value = 0.05 was used as threshold value. DAVID gene annotation tools were used to identify several candidate genes linked to important biological functions and pathways in selected cattle populations. We were focused on responsible gene in milk production traits and candidate gene responsible for adaptation to tropical environment.

4.6.2.1. Candidate Genes associated with Milk Production Traits

Among annotated genes, *ABCG2, ABCA7, B4GALNT1 GHR and ITGAV* were the potential candidate genes for milk production traits and important in milk proteins, milk fat, for udder health, mastits and others (Table 12). Two genes *(B4GALNT1* and *GHR)* were under positive selection in all pair wise breeds analysis *(*Arsi-Holstein, Arsi-Jersey, Begait-Holstein, Begait- Jersey, Boran-Holstein, Boran-Jersey, Sheko-Holstein and Sheko-Jersey*).* *B4GALNT1* and *GHR* are genes responsible for milk production and mastitis disease resistance. Surprisingly an *ABCG2* gene was identified as potential candidate gene only for Sheko-Holstein pair and responsible in milk protein production. Similarly, *ITGAV* gene functioning in milk production

was detected in Arsi-Holstein pair breeds. For those genes quantitative trait loci (QTL) were also detected from animal QTL data base (https://www.animalgenome.org/cgi bin/QTLdb/index).

Table 11. Milk candidate genes under strong positive selection

BTA	Position	Genes	Functions	Reference
2	9701199	*ITGAV*	Milk production	Ogorenc *et al.*,2009
5	29667672	*B4GALNT1*	Milk production and Mastitis	Ogorevc *et al.*,2009
6	10664798	*ABCG2*	Milk protein	Yue *et al.*, 2010
7	45174539	ABCA7	Milk traits	Mani *et al.*, 2010
20	32121397	*GHR*	Udder health and milk fat	Xiaolong *et al.*,2018

4.6.2.1.1. Genotype variability for milk candidate genes

We found highly differentiated genotype between Ethiopian and European cattle breeds. In European cattle fixed genotype AA and GG were detected whereas Ethiopian cattle have more heterozygous AG. In detection of alleles variability alleles, A and G were also fixed in European cattle populations Table 12.

Table 12. Genotype variability and allele variability for milk candidate genes

Chr	Gene	SNP position	Genotype	Breeds					
				Arsi	Begait	Boran	Sheko	Holstein	Jersey
2	*ITGAV*	9701199	AA	0	0.21	0.08	0.19	1	1
			AG	0.41	0.35	0.51	0.54	0	0
			GG	0.59	0.29	0.41	0.31	0	0
			A	0.59	0.46	0.45	0.65	1	1
			G	0.41	0.54	0.55	0.35	0	0
7	ABCA7	45174539	AA	0	0.19	0.21	0.05	1	1
			AG	0.25	0.36	0.44	0.3	0	0
			GG	0.75	0.44	0.36	0.65	0	0
			A	0.13	0.38	0.38	0.21	1	1
			G	0.87	0.62	0.58	0.79	0	0
12	B3GALT1	29667672	AA	0.72	0.21	0.21	0.29	0	0
			AG	0.28	0.5	0.56	0.46	0	0
			GG	0	0.29	0.24	0.25	1	1
			A	0.89	0.89	0.77	0.73	0	0
			G	0.11	0.11	0.23	0.27	1	1
20	*GHR*	32121397	AA	0	0.053	0	0	1	1
			AG	0.21	0.24	0.13	0.46	0	0
			GG	0.79	0.71	0.87	0.54	0	0
			A	0.84	0.47	0.86	0.75	1	1
			G	0.16	0.53	0.14	0.25	0	0
26	ACTG2	10664798	AA	0.47	0.21	0.54	0.16	0	0
			AG	0.19	0.51	0.38	0.42	0	0
			GG	0.34	0.28	0.08	0.42	1	1
			A	0.72	0.74	0.73	0.71	1	1
			G	0.28	0.26	0.27	0.29	0	0

4.6.3.2. Pathway Analysis candidate genes associated within in milk production traits

In Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways four significant path ways related to milk production traits were detected. From the detected pathways (*cAMP*, *Rap1, PI3K-AKT* and *Wnt*) were the pathways with the highest number of differentially expressed genes. Among gene involved in pathway *ADCY3, ADCY7, GNAS, MC2R, PDE3B, RRAS* and *SLC9A1* genes were found in cAM signal pathway, *CSF1R, EFNA2 GHR, ITGAV*, IL2RA, *KDR, PPP2R5E, PPP2R2B, THBS2* were detected in PI3K-AKT signaling path way and GNAS,*ADCY3, ADCY7, CSF1R, DOCK4, EFNA2, KDR, RRAS in (Rap1), BAMBI, FBXW11, CSNK1A1, FZD7*, and *TCF7L1* were found in Wint signaling pathway (Table 14).

Table 13. KEGG Pathway of genes under positive selection for milk production traits

KEGG path way	Count	P- value	Function
cAMP	8	8.7E-2	Milk protein
Rap1	7	6.9E-2	Milk protein
PI3K-AKT	9	8.5 E-2	Glucose absorption and milk protein
Wnt signaling path way	5	8.5E -2	Regulating mammary growth

4.6.3.3. Gene ontology of candidate gene

A number of genes involved in different biological process were detected as shown in Table 15. A group of genes (*BARD1, ARIH1, KLHL20, KLHL24, KLHL29, KLHL3, GPR182, STAC3, ADCY7*, IRAK*3*, Rab*19, RAB22A, RAB3B, DOcK4, RRAS)* involved in protein ubiquitination biological process, intracellular signal transduction biological process and in smallGTPase mediated Signal transduction.

Table 14. Enrichment Analysis

GO terms	Count	P-value	Genes
Protein ubiquitination	6	9.0E-2	*BARD1, ARIH1, KLHL20, KLHL24, KLHL29, KLHL3*
Intracellular signal transduction	7	2.0E1	*GPR182, STAC3, ADCY7, IRAK3*
Small GTPase mediated signal transduction	5	3.2E-1	*Rab19, RAB22A, RAB3B, DOcK4, RRAS*

4 .6.4.1. Candidate gene for adaptation to tropical environment

Six potential candidate genes associated (*HSPA4, HSPH1, SOD1, MATR3, RAD50,* and *KDR*) for adaptation to tropical environment found under positive selection were detected. Among candidate genes, the *MATR3* functioning as thermo-tolerance was detected in all pair of Ethiopian-European pair breeds. The *SOD1* gene was detected in all pair except in Sheko pair European cattle breeds and act as thermo-tolerant. The *HSPA4* gene was found under positive selection for Arsi-Holstein, Arsi-Jersey, Begait-Holstein, Boran -Jersey pair breeds. The *HSPH1* gene was detected in Arsi-Holstein, Sheko-Holstein and Sheko-Jersey. *RAD50* was another gene found under positive selection for Arsi-Jersey, Begait-Holstein, Begait-Jersey, Sheko-Holstein and Sheko-Jersey pair breeds. Additionally, KDR gene responsible for coat color was also detected in Arsi-European pair breeds (Table 15).

Table 15. Candidate gene for adaptation to tropical environment

BTA	Position	Gene	Function	Reference
1	3113041	*SOD1*	Thermo- tolerance	Zeng *et al.*, 2018
6	72271723	*KDR*	Coat color	Ghadikolaci, *et al*.,2018
7	23070376	*RAD50*	Thermo- tolerance	Slipa *et al*., 2021
7	52234050	*MATR3*	Thermo- tolerance	Akakabe *et al*., 2013
7	46282289	*HSPA4*	Thermo- tolerance	Andrey *et al*., 2018
12	29,894,900	*HSPH1*	Thermo- tolerance	Howard *et al.* 2014

4.6.4.1.1. Genotype variability in candidate genes for tropical adaptation

Significant difference of genotype variability and allele frequency were detected between Ethiopian cattle populations and European cattle breeds as shown in Table 16. In European cattle breeds fixed genotype AA and GG were detected whereas Ethiopian cattle populations have more heterozygous AG and AC. In detection of allele variability allele, A, C and G were also fixed in European cattle breeds.

Table 16. Genotype variability and allele frequency for adaptation candidate genes

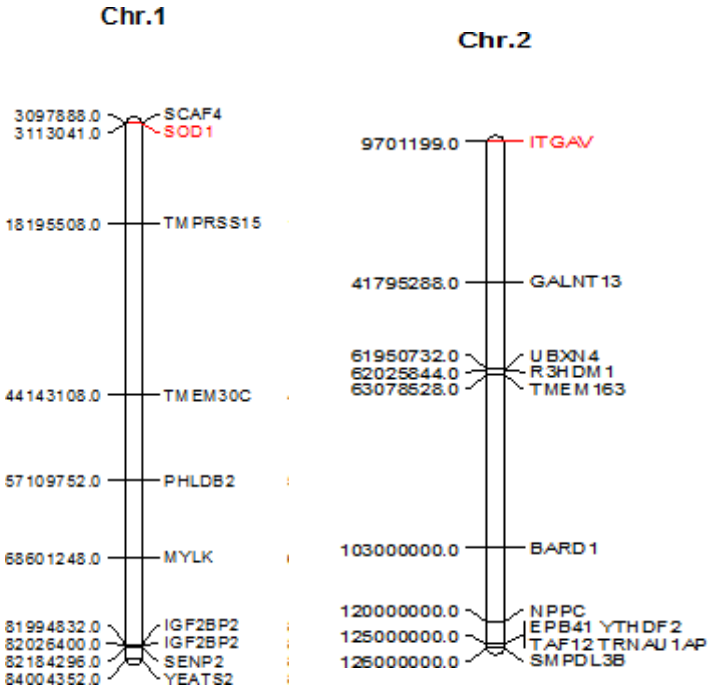
Chr	Gene	SNP position	Genotype	Breeds					
				Arsi	Begait	Boran	Sheko	Holstein	Jersey
1	*SOD1*	3113041	AA	0	0.026	0	0.35	0.97	1
			AC	0.24	0	0.31	0.54	0.03	0
			CC	0.79	0.92	0.69	0.11	0	0
			A	0.12	0.03	0.15	0.38	0.98	1
			C	0.88	0.97	0.85	0.62	0.02	0
6	*KDR*	72271723	AA	0	0.08	0.05	0.55	1	1
			AC	0.38	0.42	0.33	0.053	0	0
			CC	0.62	0.5	0.62	0.397	0	0
			A	0.19	0.16	0.23	0.34	1	1
			C	0.81	0.84	0.77	0.66	0	0
7	*RAD50*	23070376	AA	0.069	0	0.05	0.18	1	1
			AC	0.41	0.34	0.51	0	0	0
			CC	0.56	0.66	0.46	0.82	0	0
			A	0.28	0.17	0.31	0.18	1	1
			C	0.72	0.83	0.69	0.82	0	0
			AA	0	0.52	0	0	1	1
			AG	0.21	0.24	0.13	0.46	0	0
7	*MATR3*	52234050	GG	0.79	0.24	0.87	0.54	0	0
			A	0.83	0.87	0.58	0.38	1	1
			G	0.17	0.13	0.42	0.62	0	0
7	*HSPA4*	46282289	AA	0.66	0.79	0.31	0.32	0	0
			AG	0.31	0.16	0.54	0.5	0	0
			GG	0.034	0.053	0.15	0.27	1	1
			A	0.1	0.17	0.06	0.23	0	0
			G	0.9	0.83	0.94	0.27	1	1
12	*HSPH1*	29894900	AA	0.29	0.21	0.23	0	1	1
			AG	0.28	0.49	0.36	0.31	0	0
			GG	0.43	0.29	0.41	0.69	0	0
			A	0.21	0.46	0.41	0.15	1	1
			G	0.79	0.54	0.59	0.85	0	0

4.6.5.2. Pathway analysis for candidate gene related to adaptation to tropical environment

Five pathways Rap1, Camp, PI3K-AKT, wint signaling pathway and endocytosis were detected Table 17. Among candidate genes KDR gene was involved in Pap1 signaling path way.

Table 17. Pathway analysis of candidate genes related to adaptation to tropical environment

KEGG pathway	P- value	Genes
Rap1 signaling path way	2.2E-2	*GNAS, ADCY3, ADCY7, CSF1R, DOCK4, KDR, RRAS*
Camp signaling pathway	4.5E-2	*GNAS, ADCYE, ADCY7, CSF1R, MC2R, PDE3B, SLCA1*
PI3K-AKT signaling path way	8.5E-2	*GNAS, ADCY3, ADCY7, CSF1R, DOCK4, EFNA2, KDR, RRAS*
Wint signaling pathway	9.7E-2	*BAMBI, FBXW11, CSNK1A1, FZD7, TCF7L1*
Endocytosis	8.9E-2	*ASAP3, RAB22A, RAB7A, IL2RA, KIF5A PIP5K1A, ZFYVE27*


Chr.1
3097888.0
3113041.0
SCAF4
SOD1
18195508.0
TMPRSS15
44143108.0
TMEM30C
57109752.0
PHLDB2
68601248.0
MYLK
81994832.0
82026400.0
82184296.0
84004352.0
IGF2BP2
IGF2BP2
SENP2
YEATS2
Chr.2
9701199.0
ITGAV
41795288.0
GALNT13
61950732.0
62025844.0
63078528.0
UBXN4
R3HDM1
TMEM163
103000000.0
BARD1
120000000.0
125000000.0
126000000.0
NPPC
EPB41 YTHDF2
TAF12 TRNAU1AP
SMPDL3B
Chr.3
23445708.0
NOTCH2
80912304.0
JAK1
93706864.0
SLC1A7
105000000.0
106000000.0
110000000.0
118000000.0
119000000.0
121000000.0
RIMKLA
NFYC
TRAPPC3
MLPH
HDAC4 HDAC4
HDAC4
C3H2orf54 D2HGDH
Chr.5
47534832.0
47823696.0
CAND1
IRAK3
56181416.0
56232192.0
56241480.0
56247944.0
56264696.0
56275160.0
B4GALNT1
KIF5A
KIF5A
KIF5A
DCTN2
DCTN2
65524940.0
LOC510487
89040344.0
GYS2
98206392.0
BCL2L14

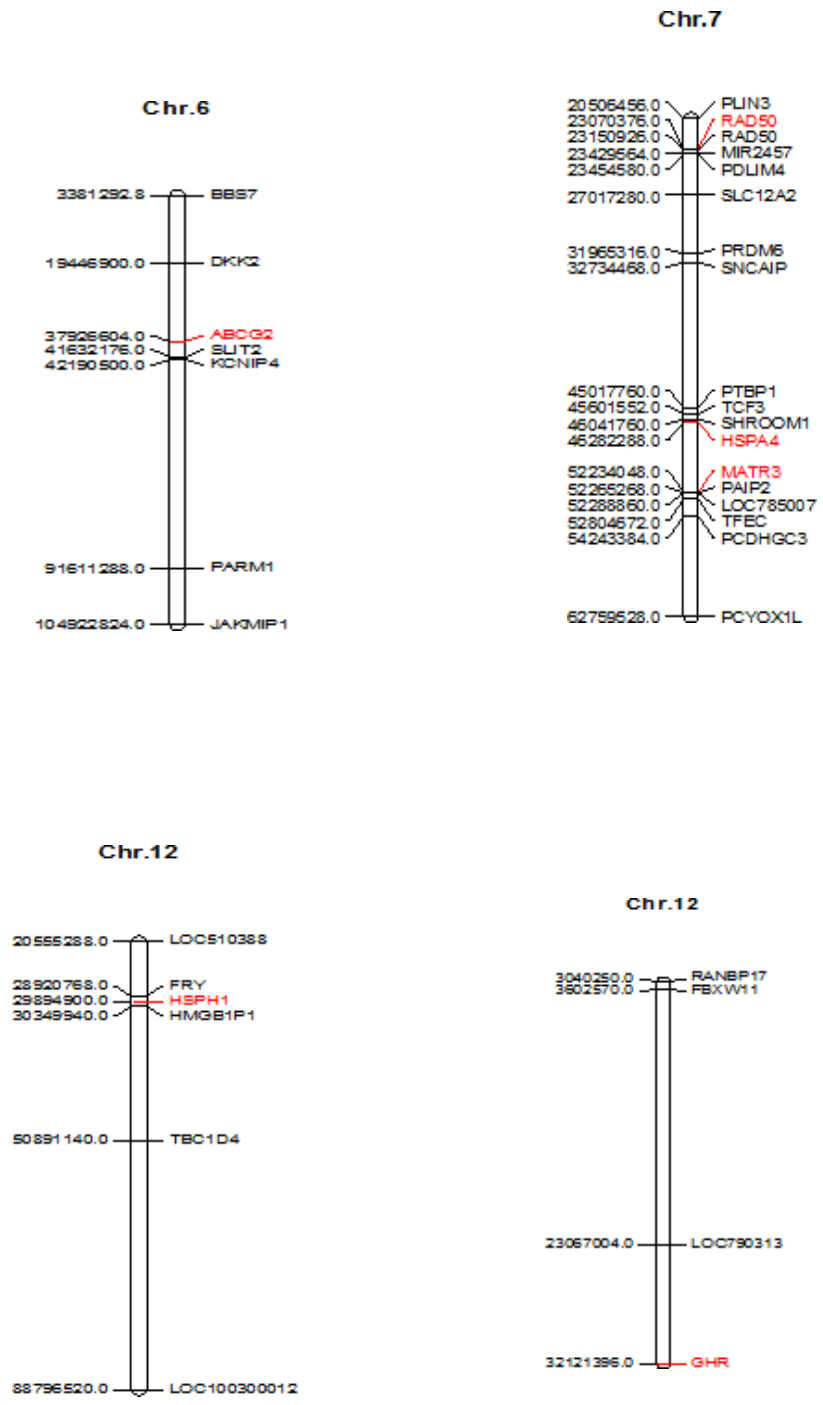


Figure 11. Functionally analysed candidate genes associated with milk trait and tropical adaptation physical maps showed the chromosomal location (Chr,1, 2,3,5,6,7 and 12) related SNPs on genome

CHAPTER FIVE

5. DISCUSSION

5.1. Minor allele frequency and within breed genetic diversity

Genetic variation within or among breeds is commonly explained in terms of allele frequency. The overall MAF in the present study for Ethiopian cattle populations 0.32 ± 0.12, 0.32 ± 0.12,0.31 ± 0.13, 0.30 ± 0.13, 0.19 for, Arsi, Begait, Borana, Sheko, which is higher than the previous study for Boran (0.21± 0.150) and Arsi (0.22 ± 0.148) (Edea, *et al.,* 2012). Also, Kim, *et al.*,Edea et al. (2015) and Mustafa, *et al*. (2017) reported a lowest MAF mean using 8K, 9k, 50K and 80k SNPs bead chip in Ethiopian cattle populations. The value for Holstein cattle was comparable with the previous report (McKay, *et al*., 2008). The differences between previous studies and the present results might be due to differences in sample size, chip bias and/or marker density. The 80k BeadChip was mainly from indicus, and indicus breeds had higher MAF values while taurus (Holstein, Jersey and Sheko) had lower values. In our current study, in Ethiopian cattle populations common variant accounted 89% of the total SNPs and higher than 74% reported for Ethiopian cattle (Edea, *et al.,* 2012). Thisdifference could be attributed to the SNP chip bias. Rare variant MAFs were observed in about 2% in the Ethiopian cattle populations and higher than previous study 0.09% (Edea, *et al*., 2012). The differences between previous studies and the present results might be due to differences in sample size, chip bias and/or marker density. The 80k Beadchip was mainly derived from indicus, and as expected Ethiopian zebu had higher MAF values than *Bos taurus* breeds (Holstein and Jersey).

5.1.1. Annotation of breed-specific SNPs

Breed-specific reference genomes can enhance the accuracy of SNP driven inferences such as Genome-wide Association Studies or SNP genotype imputation (Czech et al., 2018). From annotated breed specific SNPs, several potential genes associated with adaptive and productive traits were detected in Ethiopian cattle populations. For instance, *A2M* gene was reported as a potential gene associated with diseases resistance andit plays important roles in host immunity, particularly in animals with an inflammatory process such as mastitis (Wang *et al*., 2012). *ABCA1* was among a gene reported as a candidate gene associated with milk fat in Chinese Holstein dairy cows (Jiang *et al*., 2018). As a member of the superfamily of, *ABCA1* transports lipids and cholesterol across cell membranes.

ACACB, a member of ABC transporters family considered as a promising candidate gene potentially affecting milk composition traits in Chinese Holstein cattle (Han *et al.,* 2017). *ABCG2* was another gene affecting milk production traits (Yue et al., 2010). It is also responsiblefor the secretion of important substrates into the milk of cows (Jonker *et al*., 2005) and transportation of exotoxin and cytostatic drugs across the plasma membrane (Litman *et al.,* 2000). Similarly, ABCA*7* gene was also reported as potential gene expressed between lactation and non-lactating stages and in association with regulatory genes in bovine (Mani *et al*., 2010).

The *ACAA2* enzyme catalyzes the last step in mitochondrial fatty acid β-oxidation, thus playing a central role in the supply of energy for the animal (Bartlett and Eaton, 2004). Due to the chromosomal location of the ovine *ACAA2* gene in relation to the QTL described by Gutierrez-Gil *et al*. (2009) and its functional role in lipid metabolism, it was regarded as a putative functional and positional candidate gene that may affect milk yield and composition. The *ACAT2* gene, has been also identified as it associated with production and fertility traits (milk protein content, productive life, and conception and pregnancy rates) in Holstein cattle (Cochran *et al.*, 2013), whereas SNP within the swine *ACAT2* gene were suggested to influence the metabolic functions of the corresponding enzyme and thus may affect growth performance (Sodhi *et al.*, 2014).

The *POFUT1* gene also reported as potential gene involved in mammary development and differentiation and mastitis defense (Gutierrez *et al.*, 2014). The *ADCY8* gene was similarly reported that it is associated with important traits such as milk, reproduction and production traits in whole genome detection of recent selection signatures in Sarabi taurine breed (Moradian *et al.*, 2019). PDE4B gene was correspondingly reported as a gene underlying in genetic and molecular mechanisms for milk production phenotypes in the Korean Holstein population (Kim *et al*., 2021). The *FOXO3* was also another gene identified as potential gene associated with milk protein and fat traits in dairy cattle (Gao *et al.*, 2017). The *CSN2* gene was also reported as candidate gene in milk production traits such as milk yield, fat yield, and protein yield with in different cattle population (Miluchova et al., 2014). Khan *et al*. (2019) showed as *JAK2* gene associated with milk fat percentage and mastitis related traits in Holstein cattle.

In this study some of Ethiopian cattle-specific SNPs were located within potential genes associated to heat tolerance. Among those gene *HSPH1* gene was reported as candidate genes in heat resistance in cattle (Ben-Jemaa *et al*., 2020). The *HSF1* was also another gene identified as

related to adaptability in a subtropical climate (Baena *et al.*, 2018). Heat shock protein *(HSPB1)*, which is expressed in many tissues including muscle for protecting tissues from physiological stress, is known to enhance muscle development in bovine (Kim *et al.*,2018*). WNTI, KIT* and *AHCY* genes identified as candidate gene that affects coat color in cattle (Lim *et al.*, 2016, Stella *et al.*, 2010, Nazari - Ghalilolae *et al*.,2018).

Many studies showed that *XKR4* gene was a potential gene affecting feed intake and growth phenotypes in cattle (Lindholm-Perry, *et al.*, 2012, Bolormaa *et al*., 2011; Utsunomiya *et al.*, 2013,). Similarly, *XKR4* was reported as associated with rump fat thickness (Bolormaa *et al.*, 2011). The *XKR4 g*ene was related with weight and carcass traits in cattle (Porto Neto *et al*., 2012; Utsunomiya *et al.*, 2013).

Correspondingly, *PGR* was reported as potential gene effecting fertility traits in cattle (Khatib *et al.*, 2008a). Progesterone is hormone required for pregnancy initiation, implantation, and embryo development, which are mediated by the progesterone receptor (*PGR*) gene in mammals (McNeill *et al*., 2006). Certainly, several studies have revealed that high or low concentrations of progesterone are associated with low embryonic survival (Morris and Diskin, 2008). Also, ACAD*10* gene was also reported as candidate genes strongly associated with female fertility in beef cattle (Mota *et al.*, 2017). Likewise, *ADCY5* gene was stated as candidate genes for cow fertility (Cai *et al.*,2019).

The *IGF* genes are involved in the regulation of growth and cellular anabolism and have already been reported to be associated with feed efficiency (Bishop *et al*., 1989; Stick *et al.*, 1998). Also, *ACAD11* gene reported as candidate gene for feed efficiency in beef cattle (Karisa *et al.*, 2013). Mokhber *et al., (*2018) revealed *ACTR3* gene as a candidate gene involved in cytoskeleton

organization. *POU2F1* was identified by Montarelo *et al. (*2014), as one of the transcription factors that regulates the genes associated with the traits of growth and fat deposition in IberianX Landrace pigs, whereas, *IGFBP3* was reported as a candidate gene to influence fat thickness in Canchim beef cattle (Venoroni *et al.,* 2010). The PCCA gene was identified as gene associated with neurotransmitter concentration in cattle (Chen *et al.,* 2020). Neibergs *et al. (*2010) identified *ACTN4* gene as new putative candidate genes for Johne's disease in cattle.

In European cattle breeds a few genes corresponding specific SNPs were detected. Among them MIR34 gene reported involved in apoptosis, cell proliferation, migration, and differentiation. MiR-34a is one of several miRNAs that induce apoptosis in porcine granulosal cells (Pan et al., 2014), and its expression increases during the follicular-luteal transition in sheep (Bride et al., 2012). The *SLC26A10* was another gene identified as potential gene in meat tenderness across Merino, Border Leicester and Terminal sire types (Knight *et al*., 2012). *TCEB3* was also revealedas candidate gene in embryo development (Li *et al*., 2012).

5.1.2. Genetic differentiation and diversity

Heterozygosity measures the amount of genetic variation within a population and it indicates how much the variation exists in the population and how the variation is distributed across the alleles of analyzed markers (Nietlisbach *et al*., 2016). The observed heterozygosity (Ho) is the proportion of heterozygous individuals in population and expected heterozygosity (He) is the probability of an individual being heterozygous in any locus. The observed heterozygosity (Ho) and expected heterozygosity (He) was estimated in Ethiopian cattle and European cattle populations. In the present study the mean value of observed heterozygosity in Ethiopian cattle populations was higher than those found in previous investigation (Zerabruk et al., 2011; Edea, *et al*., 2012).

5.1.3. Principal Component Analysis (PCA)

Principal component analysis (PCA) was used to detect genetic relationship within Ethiopian cattle populations and between Ethiopian and European cattle breeds. The variation visualized onthe first and second principal components was 4.34% and 3.33% of the total variation within Ethiopian cattle populations. They grouped by their geographical back ground. The Ethiopian cattle populations is between the taurine and indicine breeds which is in agreement with their known lineage. PCA1 separated Sheko from the other breeds. It attributes that Sheko is from taurine and it is distant from the rest indicus (Arsi, Begait and Boran). PCA 2 separated Begait breed from Arsi and Boran breeds, depend on their geographical area and their origin. Similarly, on PCA 2 Arsi and Boran breeds were closer, depend on their geographical location there maybe the gene flow and they may be recently separated. The relationship between Ethiopian cattle populations and European cattle breeds were also estimated depending on their geographical location, PCA1 and PCA2 accounted 65.97% and 9.50% respectively. PCA1 separated Ethiopiancattle populations from European cattle breeds. It attributes those European cattle (Holstein and Jersey) were taurine and Ethiopian cattle (Arsi, Begait and Boran) were bos indicus and Sheko was the admixture of taurine and indicus. PCA2 separated Holstein and Jersey based on their area of origion.

5.1.4. Phylogenetic tree

Phylogenetic tree was constructed using a neighbor-joining tree among four Ethiopian cattle populations and two European cattle breeds. The current investigation presented that, with the exception of Sheko breed, Ethiopian cattle populations were closely clustered. Phylogenetic analysis revealed that Holstein and Jersey were located in a clade with Sheko, which indicated that these breeds share common ancestry. Other Ethiopian cattle Arsi, Begait and Boran were clustered in one branch of the clade. It illustrates the separation among bos taurus (Holstein, Jersey and Sheko) and bos indicus (Arsi, Begait and Boran). Among Ethiopian cattle, Sheko was the first to split, followed by Begait. Interestingly, the two Ethiopian cattle populations (Arsi and Borana) were closely clustered.

With the exception of Sheko, Ethiopian cattle populations have common historical origins and are thought to be founded as a consequence of previous admixture between *bos indicus* and *bos taurus* cattle manifesting a hybrid nature (Dadi *et al*., 2008, Edead *et al.*,2013). Hassan et al. (2007) also reported sheko breed consistently had the largest distance with all other breeds studied based on RAPD markers for genetic variably in Ethiopian indigenous cattle populations. Sheko is possibly associated with their common taurine background (Epstein and Masson.,1984).

5.2. Selection signature

The Fst statistics is the most widely used test to detect loci with outstanding genetic differentiation between populations (Myles et al.,2008). In the present study the top 1% Fst values were considered to present genomic regions under positive selection.

5.2.1. Candidate genes related to milk production traits

In dairy industry selecting super animal to produce high-yield and high-quality milk is the direct economic goals. Thus, understanding the genetic architecture and detecting milk-related candidate gene is beneficial in improvement of dairy breed and milk production simultaneously. In the present study the potential candidate genes found under positive selection were *ABCG2, B4GALNT1, GHR and ITGAV,* in milk production traits for commonly fixed SNPs in Ethiopian cattle populations and European dairy cattle breeds.

ABCG2, has been identified as a significance gene in regulation of bovine lactation. Many individual studies on different dairy cattle population proposed *ABCG2* gene as potential gene in milk production traits (Wayne & McIntyre 2002; Khatkar *et al*., 2004; Yue *et al*., 2010). It is also responsible for the secretion of important substrates into the milk of cows (Jonker *et al.*,2005) and transportation of exotoxin and cytostatic drugs across the plasma membrane (Litman*et al*., 2000). According to Cohen *et al*. (2004), *ABCG2* has also an important function in mammary gland differentiation and branching of the mammary epithelial ductal system. Additionally, *ABCG2* known as breast cancer resistance protein has been detected in alveolar epithelial cells in the mammary gland where it is strongly expressed during lactation (Cohen-Zinder *et al.*, 2005).

Several SNPs were also identified in *ABCG2* gene, of which the most important in milk productiontrait in cows (Cohen-Zinder *et al*., 2005; Olsen *et al*., 2005). It has been associated with milk fat yield and percentage and mammary epithelial cell proliferation (Schennink *et al*.,2009; Olsen *etal*., 2007 Jerry *et al*., 2012). Ali *et al*. (2013) reported one SNP in intron 7 of bovine *ABCG2*gene with significant association on milk protein yield and protein percentage . This study presented that *ABCG2*gene was found under positive selection in Sheko-Jersey pairwise breeds as potential candidate gene in milk production traits. Additionally, thisgene was associated with functional terms (ATP- binding) important for Epithelial cellproliferation in mammary gland. Similarly, high iHS values were found in the region near*ABCG2* in which mutations cause differences in milk yield and composition (Zinder, *et al*.,2005).

The *GHR* gene is also potential gene in milk production traits and it has been reported in many individual studies. Blott *et al*. (2003) and Viatal *et al. (*2006) reported *GHR* as a potential gene in production of milk traits such as milk yield, milk composition and milk fat. Additionally, Flori et al. (2009) identified 13 highly significant regions subjected to strong and recent positive selection by smoothing Fst values over each chromosome to explore the role of *GHR* in milk production in cattle.

Larkin *et al. (*2012) recently reconstructed the haplotypes of two influential sires of the contemporary Holstein- Friesian population with SNPs that have been subjected to artificial selection for milk production and identify *ITGAV* gene. In addition, *ITGAV* gene was detected based on the database of cattle candidate genes as genetic markers for milk production and mastitis developed (Ogorevc *et al*., 2009).

Similarly, *ITGAV* was positively identified as cardiomyopathy-associated genes in Holstein and although the increased frequency of alleles andthe increase in milk yield are beneficial for animal production (Agerholm *et al*. 1993).

5.2.2. Pathway analysis of genes associated with milk production traits

In the present investigation the candidate gene were associated with four significant *KEGG* pathways (*cAMP, Rap1, PI3K-AKT* and *Wnt* signaling path way). *cAMP, Rap1* and *PI3K-AKT* are the pathways with the highest number of differentially expressed genes, which have potential roles in mediating methionine-induced milk protein synthesis in bovine mammary epithelial cells (Hou *et al.,* 2019). According to Hou *et al*. (2019) *PI3K-AKT* signaling pathways is significantly enriched for unregulated differential expressed genes by methionine stimulation.

5.2.4. Candidate genes for adaptation to tropical environment

Heat stress is the most common cause of oxidative stress, as it causes mitochondrial oxidative stress and cell malfunction, which leads to cell death and damage cell survival in stressful situations necessitates rapid response mechanisms and, as a result, effective resumption of cell functioning when stress has been alleviated. When cells are exposed to heat stress, molecules are produced that are ready to mediate cell death and survival signals, as well as assist the cell's tolerance and recovery from damage (Zeng *et al*., 2018). Candidate genes for adaptation to tropical environment were selected using commonly fixed SNPs in both Ethiopian cattle populations and European cattle breeds. *HSPH1, SOD1, MATR3, KDR* were among the potential candidate genes for adaptation to tropical environment.

The *HSPH1* genes has been identified as candidate genes in adaptation to harsh environmental in cattle (Yurchenko *et, al.*,2018). Additionally, this gene was involved in physiological adaptations required to cope with environmental stressors (Ben-jemaa, *et al*.,2020). Howard et al. (2014) also

reported *HSPH1* candidate gene for adaptation to tropical environment in cattle. The study cattle breeds considered in the present study display arrange coat colors. Among candidate genes *KDR* gene has been reported as a gene associated with the reddening of coat color pattern in cattle (Hanna *et al*., 2014). Additionally, this gene has been identified as a putative candidate for coat color in the cattle study and more likely contributing to wattle development (Nazari - Ghadikolaci *et al.*, 2018). *RAD50* was identified as novel candidate gene for environmental adaptation and acclimation (Yurchenko *et al*.,2018).

5.6. Linkage Disequilibrium (LD) analysis

Linkage disequilibrium (LD) was evaluated between two SNPs using and D' and r^2. R^2 is the proportion of recombinant type and D' represents historical recombination via allelic association (Mustafa *et al*., 2018). The strength of the linkage can be characterized by the level of linkage disequilibrium (LD). D' and r^2 these measurements have a range of 0 to 1. D' = 1 indicates the absence of recombination between the two loci due to the presence of one of the polymorphisms, whereas D' < 1 represents the presence of historical recombination between the loci.

The overall average mean of r^2 in Ethiopian cattle populations (Arsi, Begait, Boran and sheko) were 0.22 ± 0.25, 0.23 ± 0.25, 0.22 ± 0.25, 0.24 ± 0.25 respectively. The higher mean of r^2 were detected in Ethiopian cattle populations in comparison to previous study using three genotyping Bead Chips (9, 50, and 80 K) (Edea, *et al*., 2015). The average mean of D' was 0.66 ± 0.34, 0.66 ± 0.34, 0.67 ± 0.34, 0.69 ± 0.34 in Arsi, Begait, Boran, Sheko, respectively and this may be due to the array and sample size used. For inter-SNP marker distances of 20-40 kb, the average estimated LD (r^2) values for Ethiopian cattle populations were higher than previous study in Ethiopian indigenous cattle

populations (Edea, et al., 2014). In European breeds the average estimated LD(r^2) for inter- SNP marker distance of 20- 40kb were lower than the finding (De roos *et al*., 2008). This could be attribute to difference in population history and chip effect.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

The patterns of genetic variation are commonly used for the consequences of domestication, breed formation, population structure, and selection. This study provides a brief overview of genetic variability, selection signatures, linkage disequilibrium and breed-specific SNPs in Ethiopian and European dairy cattle breeds. The observed high polymorphism of SNPs on the Bovines GGP-80K Bead Chip in Ethiopian cattle populations suggests that this genotyping platform could be used for further genetic analysis (genomic selection, genome-wide association studies) of African zebu and hybrid populations. Ethiopian cattle populations harbor high within-breed genetic variability which can be explored through appropriate breed improvement strategies (i.e., selection). The breed- specific SNPs identified and can be for breed assessment, product discrimination, and to characterize functional variants. The candidate genes identified in the present study will support selection of animals for improved milk production traits and adaptation to tropical environments. This selected SNPs associated to potential genes and they should be included in genetic selection and it help to identify causative genes and go further for whole genome selection (GWAS). Furthermore, it will provide information that could be used to monitor, conserve and manage the rich genetic diversity of our indigenous cattle populations and will help as guidelines in developing future animal breeding and strategies.

7. REFERNCES

Akakabe, Y., Koide, M., Kitamura, Y., Matsuo, K., Ueyama, T., Matoba, S., Yamada, H., Miyata, K., Oike, Y. and Ikeda, K. (2013). Ecscr regulates insulin sensitivity and predisposition to obesity by modulating endothelial cell functions. *Nat. comm.***4** :1-1.

Abu-Elheiga L., Oh W., Kordari P. & Wakil S.J. (2003). Acetyl-CoA carboxylase 2 mutant mice are protected against obesity and diabetes induced by high-fat/high-carbohydrate diets. *Acad. Sci. USA* .**100**: 10207–12.

Ageitos JM, Vallejo JA, Poza M. (2006). Fluorescein thiocarbamoyl-kappa-casein assay for the specific testing of milk-clotting proteases. *J. Dairy Sci*. **89**:3770-3777.

Alim, M.A., Xie, Y., Fan, Y., Wu, X., Sun, D., Zhang, Y., Zhang, S., Zhang, Y., Zhang, Q. and Liu L. (2013). Genetic effects of *ABCG2* polymorphism on milk production traits in the Chinese Holstein cattle. *J.Appl.Anim.Res.**41***:333-338.

Anton I, Kovacs K, Fesus L. (2008). Effect of DGAT1 and TG gene polymorphisms on intramuscular fat on milk production traits in different cattle breeds in Ilungary. *Acta. Vet.* **56**:181-186.

Ardlie, K.G., Kruglyak, L. and Seielstad, M. (2002). Patterns of linkage disequilibrium in the Hum. Genom. *Nat. Genet.***3**:299-309.

Baena, M.M., Tizioto, P.C., Meirelles, S.L.C. and Regitano, L.C.D.A. (2018). HSF1 and HSPA6 as functional candidate genes associated with heat tolerance in Angus cattle. *Brazilian J.Anim.Sci*. **47**.

Beckmann, J.S. and Soller, M. (1983). Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor. Appl. Genet.* **67**:35-43.

Beja-Pereira, A., Caramelli, D., Lalueza-Fox, C., Vernesi, C., Ferrand, N., Casoli, A., Goyache, F., Royo, L.J., Conti, S., Lari, M. and Martini, A. (2006). The origin of European cattle: evidence from modern and ancient DNA. *Proc. Nation. Acad. Sci.* **103**:8113-8118.

Ben-Jemaa, S., Mastrangelo, S., Lee, S.H., Lee, J.H. and Boussaha, M. (2020). Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle. *Sci. Rep*. **10**:1-14.

Bittante, G.; Contiero, B.; Cecchinato, A. (2013). Prolonged observation and modelling of milk coagulation, curd firming, and syneresis. *J. Int. Dairy.* **29**:115–123.

Blott S, Kim J-J, Moisio S, Schmidt-Ku¨ ntzel A, Cornet A. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *J. Genet*. **163**: 253–266.

Blott, S.C., Williams, J.L. and Haley, C.S. (1998). Genetic relationships among European cattle breeds. *Anim.Genet*. **29**:273-282.

Boettcher, P.J., Tixier-Boichard, M., Toro, M.A., Simianer, H., Eding, H., Gandini, G., Joost, S. Garcia, D., Colli, L.I.C.I.A., Ajmone-Marsan, P.A.O.L.O. and Globaldiv Consortium (2010). Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Anim.Genet.* **41**:64-77.

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

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. *Mol. Bio. and Evol.* **29**:2101-2104.

Bolormaa S, Porto Neto LR, Zhang YD, Bunch RJ, Harrison BE, Goddard ME, Barendse W. (2011). A genome-wide association study of meat and carcass traits in Australian cattle. *J Anim. Sci.* **89**:297–309.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Hum. Genet*.**32**:31

Bovine HapMap Consortium, Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse,W.,Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M. and Lien, S. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Sci*.**324**:528-532.

Bradley, D.G., MacHugh, D.E., Cunningham, P. and Loftus, R.T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Sci*.**93**:5131-5135.

Brym P, Kami_nski S, Ru_s_c A. (2004). New SSCP polymorphism within bovine STAT5A gene and its associations with milk performance traits in Black-and-White and Jersey cattle. *J. Appl. Genet*. **45**:445- 452.

Buono, K.D.; Robinson, G.W.; Martin, C.; Shi, S.; Stanley, P.; Tanigaki, K.; Honjo, T.; Hennighausen, L. (2006**).** The canonical Notch/RBP-J signaling pathway controls the balance of cell lineages in mammary epithelium during pregnancy. *Dev. Biol*.**293**:565–580.

Cai, Z., Guldbrandtsen, B., Lund, M.S. and Sahana, G. (2019). Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. *BMC. Genom*. **20**:1-9.

Central Statistical Agency, (2020/21). Report on livestock and livestock characteristics.

Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C. (2008). Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet*. **40**:449–454.

Chen, Q., Qu, K., Ma, Z., Zhan, J., Zhang, F., Shen, J., Ning, Q., Jia, P., Zhang, J., Chen, N. and Chen, H. (2020). Genome-Wide Association Study Identifies Genomic Loci Associated with Neurotransmitter Concentration in Cattle. *Front. Genet*. **11**:139.

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. *Front. Genet*. **9**:607.

Cobanoglu O, Zaitoun I, Chang YM. (2006). Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. *J. Dairy Sci*. **89**:4433 - 4437.

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

Cohen, M, Seroussi, E, Band, MR, Lewin, HA, Drackley, JK, Larkin, DM, Everts-van der Wind, A, Heon-Lee, J, Loor, JJ, Shani, M. (2004). SPP1 is a candidate gene for the QTL affecting milk protein concentration on BTA6 in Israeli Holstein. J. Anim. Genet. ISAG. F015. Tokyo, Japan.

Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loor, A. (2005). Evertsvan der Wind, J. H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. (2005). Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition inHolstein cattle. *Genomics. Res*. **15**:936–944.

Czech, B., Frąszczak, M., Mielczarek, M. and Szyda, J. (2018). Identification and annotation of breed-specific single nucleotide polymorphisms in Bos taurus genomes. *Plos.***13**:198-419.

Dadi Hailu, Kim, J. J., Yoon, D., Kim, K. S. (2012). Evaluation of single nucleotide polymorphisms (SNPs) genotyped by the Illumina Bovine SNP50K in cattle focusing on Hanwoo breed. *Asian-Austr. Anim. sci.* **25**:28.

Dadi Hailu, Tibbo, M., Takahashi, Y., Nomura, K., Hanada, H., Amano, T. (2008). Microsatellite analysis reveals high genetic diversity but low genetic structure in Ethiopian indigenous cattle populations. *Anim.Genet.* **39**: 425-431.

Dadousis, C.; Bi_ani, S.; Cipolat-Gotet, C.; Nicolazzi, E.L.; Rossoni, A.; Santus, E.; Bittante, G.; Cecchinato, A. (2016). Genome-wide association of coagulation properties, curd firmness modeling, protein percentage, and acidity in milk from Brown Swiss cows. J. *Dairy. Sci*. **2 99**:3654–3666.

Dadousis, C.; Pegolo, S.; Rosa, G.J.M.; Gianola, D.; Bittante, G.; Cecchinato, A. (2017). Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle. *J. Dairy.Sci.* **100**:1223- 1231.

DAGRIS, (2007). Domestic Animal Genetic Resources Information System (edited by S. Kempo, Y. Mamo, B. Astrat and Tadele Dessie). *ILRI*. http://dagris.ilri.cgiar.org

De Roos, A.P.W., Hayes, B.J., Spelman, R.J. and Goddard, M.E. (2008). Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genet.* **179**: 1503-1512.

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J. W., Molina Alcalá, A., Sonstegard, T. S., Taylor, J. F. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PloS.Genet*.**10**: 1004-254.

Deng, Y., Liu, T., Xie, Y., Wei, Y., Xie, Z., Shi, Y., Deng, X. (2020). High Genetic Diversity and Low Differentiation in Michelia shiluensis, an Endangered Magnolia Species in South China. *Fore*. **11**: 469.

Dettori, M.L., Pazzola, M., Petretto, E. and Vacca, G.M. (2020). Association Analysis between SPP1, POFUT1 and PRLR Gene Variation and Milk Yield, Composition andCoagulation Traits in Sarda Sheep. J.*Anim*. Sci. **10**:1216.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R. and Wareham, N., 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nat.* **447**:1087-1093.

Edea Zewdu, Bhuiyan, M.S.A., Dessie Tadelle, Rothschild, M.F., Dadi Hailu. and Kim, K.S. (2015). Genome-wide genetic diversity, population structure, and admixture analysis in African and Asian cattle breeds. *Anim*.Sci. **9**:218-226.

Edea Zewdu, Dadi Hailu, Kim, S. W., Park, J. H., Shin, G. H., Dessie, T., Kim, K. S. (2014). Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in E thiopia. *J. Anim. Genet*. **131**:358-366.

Edea Zewdu, Dadi Hailu, Kim, S., Dessie Tadele, Lee, T., Kim, H., Kim, K. (2013). Genetic diversity, population structure, and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front.Genet.***4**: 1-9.

Edea Zewdu, Dadi Hailu, Kim, S.W., Dessie, T. and Kim, K.S. (2012). Comparison of SNP variation and distribution in indigenous Ethiopian and Korean Cattle (Hanwoo) populations. *Geno.Info.***10**:200-205.

Edea Zewdu, Jeoung, Y. H., Shin, S. S., Ku, J., Seo, S., Kim, I. H., Kim, K. S. (2018). Genome-wide association study of carcass weight in commercial Hanwoo cattle.*J.Anim.Sci*. **31**:321-327.

Edea Zewdu, Jung, K. S., Shin, S., and Yoo, S. (2020). Signatures of positive selection underlying beef production traits in Korean cattle breeds. *J.Anim.Sci.Tech.* **62**:293-305.

Ellegren, H., Moore, S., Robinson, N., Byrne, K., Ward, W. and Sheldon, B.C.(1997). Microsatellite evolution--a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Mol.Bio and Evol*.**14**:854-860.

Epstein, H. (1957). The sanga cattle of East Africa. *Agri. J.* **22**:149-164.

Epstein, H. (1971). The origin of the domestic animals of Africa. Africana publishing corporation.

Ethiopian Biodiversity Institute (2016). Ethiopian National Strategy and Plan of Action for Conservation.

Fang, M., Fu, W., Jiang, D., Zhang, Q., Sun, D., Ding, X. and Liu, J. (2014). A multiple-SNP approach for genome-wide association study of milk production traits in Chinese Holstein cattle. *PloS one*, **9**: 99544.

Farrell HM, Jimenez-Flores R, Bleck GT. (2004). Nomenclature of the proteins of cows' milk sixth revision. *J. Dairy. Sci*. **87**:1641 - 1674.

Felleke, G., Woldearegay, M. and Haile, G. (2010). Inventory of Dairy Policy–Ethiopia. *Target* Business Consultants Plc, Netherlands Development Organization (SNV), Addis Ababa, Ethiopia.

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.L. and Gautier, M., (2009). The genome response to artificial selection: a case study in dairy cattle. *PloS.***4**:6595.

Fraser, D.J., Hansen, M.M., ØSTERGAARD, S., Tessier, N., Legault, M. and Bernatchez, L. (2007). Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Mol. Eco.***16**:3866-3889.

Freitas, A.C., De Camargo, G.M.F., Aspilcueta-Borquis, R.R., Stafuzza, N.B., Venturini, G.C., Tanamati, F., Hurtado-Lugo, N.A., Barros, C.C. and Tonhati, H. (2016). Polymorphism in the A2M gene associated with high-quality milk in Murrah buffaloes (Bubalus bubalis). *Genet. Mol. Res.*, ***15***:1-7.

Ganai NA, Bovenhuis H, van Arendonk JA. (2009). Novel polymorphisms in the bovine beta-lactoglobulin gene and their effects on beta-lactoglobulin protein concentration in milk. *J. Anim. Genet*. **40**:127-133.

Gao, Y., Jiang, J., Yang, S., Hou, Y., Liu, G.E., Zhang, S., Zhang, Q. and Sun, D. (2017). CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *BMC. Genom. **18***:1-12.

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J.G., Boichard, D., Lathrop, G.M. and Gut, I.G. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genet.* **177**:1059-1070.

Götherström, A., Anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D.G. and Ellegren, H. (2005). Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Bio. Sci.* **272**:2345-2351.

Grigson, C. (1991). An African origin for African cattle some archaeological evidence. *Afr. Arch. Rev.* **9**:119-144.

Grisart B, Coppieters W, Farnir F. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genom. Res.***12**: 222 - 231.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E.J.A.G. and Weigend, S. (2010). Genetic diversityin farm animals–a review. *Anim.Genet*. **41**:6-31.

Gutièrrez-Gil, B.; Jose Arranz, J.; Pong-Wong, R.; García-Gámez, E.; Kijas, J.; Wiener, P. (2014**)**. Application of selection mapping to identify genomic regions associated with dairy production in sheep. *Plos.One* .**9**:94-623.

Han, B., Liang, W., Liu, L., Li, Y. and Sun, D. (2018). Genetic association of the ACACB gene with milk yield and composition traits in dairy cattle. *Anim. Genet,.**49:***169-177.

Hancock, J.M. (1995). The contribution of slippage-like processes to genome evolution. *J.Mol. Evol.* **41**:1038-1047.

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. O. (2003). African pastoralism: Genetic imprints of origins and migrations. *Sci.* **296**: 336-339.

Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W. and Rege, J.E.O.(2002). African pastoralism: genetic imprints of origins and migrations. *Sci*.**296**:336-339.

Hassen, F., Bekele, E., Ayalew, W. and Dessie, T. (2007). Genetic variability of five indigenous Ethiopian cattle breeds using RAPD markers. *Afri. J.Biotech.* **6**:19.

Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res*. **13**:635–643.

He F, Sun DX, Yu Y. (2006). Association between SNPs within prolactin gene and milk performance traits in Holstein dairy cattle. *J. Anim.Sci*.**19**:1384 - 1389.

Hill, W.G. and Robertson, A., 1968. Linkage disequilibrium in finite populations. *Theor. and Appl. Genet* .**38**:226-231.

Hou, X., Jiang, M., Zhou, J., Song, S., Zhao, F. and Lin, Y. (2020). Examination of methionine stimulation of gene expression in dairy cow mammary epithelial cells using RNA-sequencing. *J.Dairy. Res.* **87**:226-231.

Hu, Z.L., Park, C.A. and Reecy, J.M. (2019). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nuc. Acid. Res*. **47**:701-710.

Huang DW, Sherman BT, Lempicki RA. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc*. **4**:44–57.

Huang W, Maltecca C, Khatib H. (2008). A proline-to-histidine mutation in POU1F1 is associated with production traits in dairy cattle. *J. Anim. Genet*. **39**:554 - 557.

Huang, W. and Marth, G. (2008). EagleView: a genome assembly viewer for next-generation sequencing technologies. *Genom. Res.* **18**:1538-1543.

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

Jerry Wei ,Pauline F Geale, Paul A Sheehy, Peter Williamson (2012). The impact of ABCG2 on bovine mammary epithelial cell proliferation. *J. Dairy.Sci.* **90**:1029-1038.

Jiang J., Gao Y., Hou Y., Li W., Zhang S., Zhang Q. & Sun D. (2016). Whole-genome resequencing of Holstein bulls for indel discovery and identification of genes associated with milk composition traits in dairy cattle. *Plos.One* .**11**: 168-946.

Jiang, P., Zhao, Z., Li, X., Wang, M., Xia, L., Cao, Y., Yang, R. and Fang, X. (2020). RNA interference mediated knockdown of ATP binding cassette subfamily A member 1 decreases the triglyceride content of bovine mammary epithelial cells. *J. Zool.* ***52***: 239.

Jonker JW, Merino G, Musters S, van Herwaarden AE, Bolscher E, Wagenaar E, Mesman E, Dale TC, Schinkel AH. (2005). The breast cancer resistance protein BCRP (ABCG2) concentrates drugs and carcinogenic xenotoxins into milk. *Nat. Med*. **11**:127-129.

Karisa, B.K., Thomson, J., Wang, Z., Stothard, P., Moore, S.S. and Plastow, G.S. (2013). Candidate genes and single nucleotide polymorphisms associated with variation in residual feed intake in beef cattle. *J. Anim. Sci.* ***91***:3502 - 3513.

Kaupe B, Brandt H, Prinzenberg EM. (2007). Joint analysis of the influence of *CYP11B1* and *DGAT1* genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *J. Anim Sci*. **85**:11 - 21.

Kgwatalala PM, Ibeagha-Awemu EM, Hayes JF. (2009). Stearoyl- CoA desaturase 1 3'UTR SNPs and their influence on milk fatty acid composition of Canadian Holstein cows. *J Anim. Genet*.**126**:394 - 403.

Khan, M.Z., Wang, D., Liu, L., Usman, T., Wen, H., Zhang, R., Liu, S., Shi, L., Mi, S., Xiao, W. and Yu, Y (2019). Significant genetic effects of JAK2 and DGAT1 mutations on milk fat content and mastitis resistance in Holsteins. *J. Dairy.Res.* ***86***:388 - 393.

Khatib H, Heifetz E, Dekkers JC. (2005). Association of the protease inhibitor gene with production traits in Holstein dairy cattle. *J. Dairy. Sci*. **88**:1208 - 1213.

Khatib H, Leonard SD, Schutzkus V. (2006). Association of the *OLR1* gene with milk composition in Holstein dairy cattle. *J. Dairy. Sci*.**89**:1753 - 1760.

Khatib H, Monson RL, Schutzkus V. (2008). Mutations in the STAT5A gene are associated with embryonic survival and milk composition in cattle. *J. Dairy. Sci*. **91**:784 - 793.

Khatkar MS, Thomson PC, Tammen I, Raadsma HW. (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* **36**:163 - 190.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A., Barris, W., Schnabel, R.D., Taylor, J.F. and Raadsma, H.W.(2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC. Genom*.**9**:1-18.

Kim, J.E., Ko, A.R., Hyun, H.W., Min, S.J. and Kang, T.C. (2018). P2RX7-MAPK1/2-SP1 axis inhibits MTOR independent HSPB1-mediated astroglial autophagy. *Cell death & diseas.* **9**:1-16.

Kim, S., Lim, B., Cho, J., Lee, S., Dang, C.G., Jeon, J.H., Kim, J.M. and Lee, J. (2021). Genome-Wide Identification of Candidate Genes for Milk Production Traits in Korean Holstein Cattle. J. *Anim*. Sci. ***11***:1392.

Kuehn, L.A., Keele, J.W., Bennett, G.L., McDaneld, T.G., Smith, T.P.L., Snelling, W.M., Sonstegard, T.S. and Thallman, R.M. (2011). Predicting breed composition using breed

frequencies of 50,000 markers from the US Meat Animal Research Center 2,000 Bull Project. *J.Anim.sci*. **89**:1742-1750.

Kuss AW, Gogol J, Bartenschlager H, et al. Polymorphic AP-1 binding site in bovine CSN1S1 shows quantitative differences in protein binding associated with milk protein expression. Leonard S, Khatib H, Schutzkus V. (2005). Effects of the osteopontin gene variants on milk production traits in dairy cattle. *J. Dairy.Sci*. **88**:4083 - 4086.

Larkin, D.M., Daetwyler, H.D., Hernandez, A.G., Wright, C.L., Hetrick, L.A., Boucek, L., Bachman, S.L., Band, M.R., Akraiko, T.V., Cohen-Zinder, M. and Thimmapuram, J. (2012). Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Sci.* **109**:7693-7698.

Lee, H.J., Kim, J., Lee, T., Son, J.K., Yoon, H.B., Baek, K.S., Jeong, J.Y., Cho, Y.M., Lee, K.T., Yang, B.C. and Lim, H.J. (2014). Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genom. Bio and Evol.***6**:1366-1374.

Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P. and Drineas, P. (2011). Tracing cattle breeds with principal components analysis ancestry informative SNPs. *PloS.* **6**:18007.

Lewontin, R.C. and Kojima, K.I. (1960). The evolutionary dynamics of complex polymorphisms. *Evol*.458-472.

Liefers SC, Veerkamp RF, Te Pas MF. (2005). Genetics and physiology of leptin in periparturient dairy cows. *Domestic. Anim Endoc.* **29**:227 - 238.

Lim, D., Strucken, E.M., Choi, B.H., Chai, H.H., Cho, Y.M., Jang, G.W., Kim, T.H., Gondro, C. and Lee, S.H. (2016). Genomic footprints in selected and unselected beef cattle breeds in Korea. *PloS.* **11**:151:324.

Lindholm-Perry AK, Kuehn LA, Smith TP, Ferrell CL, Jenkins TG, Freetly HC, Snelling WM. (2012). A region on BTA14 that includes the positional candidate genes *LYPLA1, XKR4* and *TMEM68* is associated with feed intake and growth phenotypes in cattle. *J. Anim. Genet.* **43**:216-219.

Lipkin, E., Straus, K., Stein, R.T., Bagnato, A., Schiavini, F., Fontanesi, L., Russo, V., Medugorac, I., Foerster, M., Solkner, J. and Dolezal, M. (2009). Extensive long-range and nonsyntenic linkage disequilibrium in livestock populations: deconstruction of a conundrum. *Genet.***181**:691-699.

Litman T, Brangi M, Hudson E, Fetsch P, Abati A, Ross DD, Miyake K, Resau JH, Bates SE. (2000). The multidrug- resistant phenotype associated with overexpression of the new ABC half-transporter, MXR (*ABCG2*). *J. Cell.Sci.* **113**:2011.

Loriol, C.; Dupuy, F.; Rampal, R.; Dlugosz, M.A.; Haltiwanger, R.S.; Maftah, A.; Germot, A. (2006**).** Molecular evolution of protein O-fucosyltransferase genes and splice variants. *Gly. Bio*.**16**, 736-747.

Macciotta NPP, Mele M, Conte G. (2008). Association between a polymorphism at the stearoyl CoA desaturase locus and milk production traits in Italian Holsteins. *J. Dairy. Sci*. **91**: 3184 - 3189.

Marufu, M.C., Qokweni, L., Chimonyo, M. and Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. *Ticks and tick-borne diseas*.**2**:172-177.

McBride D, Carre´ W, Sontakke SD, Hogg CO, Law A, Donadeu FX, Clinton M. (2012). Identification of miRNAs associated with the follicular-luteal transition in the ruminant ovary. *Repro*. **144**:221-233.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Neto, E.D., Gill, C.A., Gao, C. and Mannen, H. (2008). An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genet* .**9**:1-9.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Neto, E.D., Gill, C.A., Gao, C. and Mannen, H. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC. Genet.* **8**:1-12.

Miglior, F., B. L. Muir, and B. J. Van Doormaal. (2005). Selection indices in Holstein cattle of various countries. *J. Dairy. Sci*. **88**:1255-1263.

Miluchová, M., Gábor, M. and Trakovická, A. (2014). Analysis of beta-casein gene (CSN2) polymorphism in different breeds of cattle. *Sci.Pap. Anim. Sci and Biotech*. ***47***:56-59.

Mizuhashi K, Chaya T, Kanamoto T, Omori Y, Furukawa T. (2015). a trans membrane protein, is required for bone mineralization and spermatogenesis in mice. *PLoS. One*. **10**: 133- 704.

Mokhber, M., Moradi-Shahrbabak, M., Sadeghi, M., Moradi-Shahrbabak, H., Stella, A., Nicolzzi, E., Rahmaninia, J. and Williams, J.L. (2018). A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. *BMC. Genom.**19***:1-9.

Montaldo, H.H. and Meza-Herrera, C.A., 1998. Use of molecular markers and major genes in the genetic improvement of livestock. *J.Biotech.* **1**:15-16.

Moradian, H., Koshkoiyeh, A.E., Mohammadabadi, M. and Fozi, M.A. (2020). Whole genome detection of recent selection signatures in Sarabi cattle: a unique Iranian taurine breed. *Genes & Genom.* ***42***:203-215.

Moreno-Sanchez N, Rueda J, Carabaño MJ, Reverter A, Mcwilliam S, González C. (2010). Skeletal muscle specific genes networks in cattle. *Funct. Inte.Genom.* **10**:609–18.

Morris CA, Cullen NG, Glass BC. (2007). Fatty acid synthase effects on bovine adipose fat and milk fat. *Mamm. Genom*. **18**:64 -74.

Mota, R.R., Guimarães, S.E.F., Fortes, M.R.S., Hayes, B., Silva, F.F., Verardo, L.L., Kelly, M.J., de Campos, C.F., Guimarães, J.D., Wenceslau, R.R. and Penitente-Filho, J.M. (2017). Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle. *J. Anim. Breed and Genet*. ***134***:484-492.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G. and Strydom, P.E. (2008). Meat quality of Nguni, Bonsmara and Aberdeen Angus steers raised on natural pasture in the Eastern Cape, South Africa. *Meat. Sci.* **79**:.20-28.

Mustafa, H., Ahmad, N., Heather, H.J., Eui-Soo, K., Khan, W.A., Ajmal, A., Javed, K., Pasha, T.N., Ali, A., Kim, J.J. and Sonstegard, T.S. (2018). Whole genome study of linkage disequilibrium in Sahiwal cattle. *Anim. Sci.***48**:353-360.

Myles, S., Tang, K., Somel, M.E.H.M.E.T., Green, R.E., Kelso, J. and Stoneking, M. (2008). Identification and analysis of genomic regions with large between-population differentiation in humans. *Hum. Genet.***72**:99-110.

Navani A, Zhou M, McDonald J. (2008). Serum biomarker profiling by solid-phase extraction with particle embedded micro tips and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid. Commun. Spectrom*. **22**: 997-1008.

Nazari-Ghadikolaei, A., Mehrabani-Yeganeh, H., Miarei-Aashtiani, S.R., Staiger, E.A., Rashidi, A. and Huson, H.J (2018). Genome-wide association studies identify candidate genes for coat color and mohair traits in the Iranian Markhoz goat. *Front. Genet*. ***9***:105.

Negrini, R., Nicoloso, L., Crepaldi, P., Milanesi, E., Marino, R., Perini, D., Pariset, L., Dunner, S., Levéziel, H., Williams, J.L. and Marsan, P.A. (2008). Traceability of four European protected geographic indication (PGI) beef products using single nucleotide polymorphisms (SNP) and Bayesian statistics. *Meat.Sci*.**80**(4).1212-1217.

Neibergs, H.L., Settles, M.L., Whitlock, R.H. and Taylor, J.F. (2010). GSEA-SNP identifies genes associated with Johne's disease in cattle. *Mamm. Genom*. **21**:419-425.

Nicolini, P., Amorín, R., Han, Y. and Peñagaricano, F. (2018). Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. *BMC. Genet*. ***19***:1-8.

Ogorevc J, Kunej T, Razpet A, Dovc P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. *J.Anim .Genet*. **40**:832-851

Olsen HG, Lien S, Gautier M, Nilsen H, Roseth A, Berg PR, Sundsaasen KK, Svendsen M, Meuwissen THE. (2000). Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. *J. Genet*. **169**:275.

Olsen HG, Nilsen H, Hayes B, Berg PR, Svendsen M, Lien S, Meuwissen T. (2007): Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BMC. Genet*. **8**:32-41.

Pant SD, Schenkel FS, Leyva-Baca I. (2007). Identification of single nucleotide polymorphisms in bovine CARD15 and their associations with health and production traits in Canadian Holsteins. *BMC. Genom.* **8**:421.

Pazzola, M.; Dettori, M.L.; Cipolat-Gotet, C.; Cecchinato, A.; Bittante, G.; Vacca, G.M (2014). Phenotypic factors acting coagulation properties of milk from Sarda ewes. *J. Dairy. Sci*. **97** :7247–7257.

Pérez-Montarelo D, Madsen O, Alves E, Rodríguez MC, Folch JM, Noguera JL. (2014). Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis. *Physiol. Genom*. *46*: 195-206.

Purcell Neale B, Todd-Brown K. (2007). PLINK a toolset for whole-genome association and population-based linkage analysis. *Anim.J.Genet*. **81**:559-562.

Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. and Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Anim. Genet*.**41**(4).346-356.

Rahmatalla, S.A., Müller, U., Strucken, E.M., Reissmann, M. and Brockmann, G.A. (2011). The F279Y polymorphism of the GHR gene and its relation to milk production and somatic cell score in German Holstein dairy cattle. *J. Appl. Genet*, ***52***:459-465.

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

Rege, J. (1999). The state of African cattle genetic resources I. Classification framework and identification of threatened and extinct breeds. *Anim.Genet.Reso. Info.* **25**:1-25.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R. and Lander, E.S. (2001). Linkage disequilibrium in the human genome. *Nat*. **411**:199-204.

Reist-Marti, S.B. (2004). Analysis of methods for efficient biodiversity conservation with focus on African cattle breeds (Doctoral dissertation, ETH Zurich).

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M. and Buckler, E.S. (2001). Structure of linkage

disequilibrium and phenotypic associations in the maize genome. *Nation. Acad. Sci.* **98**:11479-11484.

Robitaille G, Britten M, Methot S. (2007). Polymorphisms within the 50-flanking region of bovine K-casein gene (CSN3) and milk production-related traits. *Milch.Wissen.Schaft*.**62**: 243-245.

Ron, M., Cohen-Zinder, M., Peter, C., Weller, J.I. and Erhardt, G. (2006). A polymorphism in ABCG2 in Bos indicus and Bos taurus cattle breeds. *J. Dairy. Sci*.***89***:4921-4923.

Rupp, R., Hernandez, A. and Mallard, B.A. (2007). Association of bovine leukocyte antigen (BoLA) DRB3. 2 with immune response, mastitis, and production and type traits in Canadian Holsteins. *J. Dairy. Sci.* ***90***:1029-1038.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Sci.***230**:1350-1354.

Schennink A, Bovenhuis H, Léon-Kloosterziel KM, van Arendonk JAM, Visker MHPW. (2009): Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. *J.Anim. Genet*. **40**:909-916.

Shapiro, B.I., Gebru, G., Desta, S., Negassa, A., Nigussie, K., Aboset, G. and Mechale, H. (2017). Ethiopia livestock sector analysis: A 15-year livestock sector strategy.

Sharif, S., Mallard, B.A., Wilkie, B.N., Sargeant, J.M., Scott, H.M., Dekkers, J.C.M. and Leslie, K.E. (1998). Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadiandairy cattle. *J. Anim. Genet*. **29**:185-193.

Singh U, Kumar A, Beniwal BK. (2008). Estimates of genetic parameters for economic traits in Ongole cattle. *Indian. Vet J*. **85**:167-169.

Stanley, P. (2007). Regulation of Notch signaling by glycosylation. *Curr. Opin. Struc. Bio*. **17**:530-535.

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

Tarekegn, G.M., Ji, X.Y., Bai, X., Liu, B., Zhang, W., Birungi, J., Djikeng, A. and Tesfaye, K. (2018). Variations in mitochondrial cytochrome b region among Ethiopian indigenous cattle populations assert Bos taurus maternal origin and historical dynamics. *Anim. Sci.* **31**:1393.

Tenesa, A., Knott, S.A., Ward, D., Smith, D., Williams, J.L. and Visscher, P.M.(2003). Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *J. Anim. Sci.* **81**:617-623.

Teo, Y.Y., Fry, A.E., Bhattacharya, K., Small, K.S., Kwiatkowski, D.P. and Clark, T.G. (2009). Genome-wide comparisons of variation in linkage disequilibrium. *Genom. res.***19**(10).1849-1860.

Terwilliger, J.D. and Weiss, K.M. (1998). Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr. Opin. Biotech.***9**:578-594.

Thaller G, Kra¨mer W, Winter A. (2003). Effects of DGAT1 variants on milk production traits in German cattle breeds. *J Anim. Sci*. **81**:1911-1918.

Tóth, G., Gáspári, Z. and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genom.Res*, **10**:967-981.

Tu F, Pan ZX, Yao Y, Liu HL, Liu SR, Xie Z, Li QF. (2014). MiR-34a targets the inhibin beta B gene, promoting granulosa cell apoptosis in the porcine ovary. *Gene.t Mol. Res*. **13**:2504-2512.

Udler, M.S., Tyrer, J. and Easton, D.F. (2010). Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. *Genet. Epidem.* **34**:463-468.

Vacca, G.M.; Pazzola, M.; Dettori, M.L.; Pira, E.; Malchiodi, F.; Cipolat-Gotet, C.; Cecchinato, A.; Bittante, G. (2015). Modeling of coagulation, curd firming, and syneresis of milk from Sarda ewes. *J. Dairy. Sci*. **98**:2245-2259.

Valdar, W., Solberg, L.C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W.O., Taylor, M.S., Rawlins, J.N.P., Mott, R. and Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38**:879-887.

Veneroni, G.B., Meirelles, S.L., Grossi, D.A., Gasparin, G., Ibelli, A.M.G., Tizioto, P.C., Oliveira, H.N.D., de ALENCAR, M.M. and REGITANO, L.D.A. (2010). Prospecting candidate SNPs for backfat in Canchim beef cattle. Embrapa Pecuária Sudeste-Artigo emperiódico indexado *(ALICE)*.

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

Viitala S, Szyda J, Blott S, Schulman N, Lidauer M. (2006). The Role of the Bovine Growth Hormone Receptor and Prolactin Receptor Genes in Milk, Fat and Protein Production in Finnish Ayrshire Dairy Cattle. *J. Genet*.**173**:2151-2164.

Wakil S.J. & Abu-Elheiga L.A. (2009). Fatty acid metabolism: target for metabolic syndrome. J. *Lipid. Res.* **50**:138-43.

Wang, X., Huang, J., Zhao, L., Wang, C., Ju, Z., Li, Q., Qi, C., Zhang, Y., Zhang, Z., Zhang, W. and Hou, M. (2012). The exon 29 c. 3535A> T in the alpha-2-macroglobulin gene causing aberrant splice variants is associated with mastitis in dairy cattle. *Immu. Genet*. **64**:807-816.

Wayne ML, McIntyre LM. (2002). Combining mapping and arraying: an approach to candidate gene identification. *Nat.Acad. Sci.* **99**:14903.

Weikard R, Ku¨hn C, Goldammer T. (2005). The bovine PPARGC1A gene: molecular characterization and association of a SNP with variation of milk fat synthesis. *Physiol. Genom.* **21**: 1-13.

Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evl.* **38**:1358-1370.

Weller, J. I. (2001). Quantitative Trait Loci Analysis in Animals. CABI Publishing, London, UK.Wilkinson, S., Wiener, P., Archibald, A.L., Law, A., Schnabel, R.D., McKay, S.D., Taylor, J.F. & Ogden, R., (2011). Evaluation of approaches for identifying population informative markers from high density SNP chip. *BMC. Genet*. **12**:1-14.

Wilkinson, J.M., Lee, M.R., Rivero, M.J. and Chamberlain, A.T. (2020). Some challenges and opportunities for grazing dairy cows on temperate pastures. *Grass and Forage Sci*.**75**:1-17.

Xiaolong Wang., Christine Wurmser., Hubert Pausch, Simone Jung, Friedrich Reinhardt, Jens Tetens, Georg Thaller, Ruedi Fries (2012). Identification and Dissection of Four Major QTL Affecting Milk Fat Content in the German Holstein-Friesian Population. *Plos.one*. **10**:1371

Ying, C., Y. C. Yang, W. F. Hong, W. T. Cheng, and W. L. Hsu. (2000). Progesterone receptor gene expression in preimplantation pig embryos. *Eur. J. Endocr*. **143**:697-703.

Yue W, Fang X, Zhang C, Pang Y, Xu H, Gu C, Shao R, Lei C, Chen H. (2010). Two novel SNPs of the ABCG2 gene and its associations with milk traits in Chinese Holsteins. *Mol. Biol. Rep*. **38**:2927-2932.

Yurchenko, A.A., Daetwyler, H.D., Yudin, N., Schnabel, R.D., Vander Jagt, C.J., Soloshenko, V., Lhasaranov, B., Popov, R., Taylor, J.F. and Larkin, D.M. (2018). Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Sci. Repor.**8***:1-16.

Yuri T Utsunomiya, Adriana S do Carmo, Roberto Carvalheiro, Haroldo HR Neves, Márcia C Matos, Ludmilla B Zavarez, Ana M Pérez O'Brien, Johann Sölkner, John C McEwan, John B Cole5, Curtis P Van Tassell, Flávio S Schenkel, Marcos VGB da Silva, Laercio R Porto Neto, Tad S Sonstegard6 and José F Garcia (2013). Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC. Genet*.**14**:52.

Zhou GL, Liu HG, Liu C (2005). Association of genetic polymorphism in GH gene with milk production traits in Beijing Holstein cows. *J. Bio.sci*. **30**:595-598.

Zwane, A.A., Maiwashe, A., Makgahlela, M.L., Choudhury, A., Taylor, J.F. and van Marle-Köster, E. (2016). Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds. *J.Anim.Sci,* **46:**302-312.